

2007

Choice among stimuli in equivalence classes

Christina Alligood
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Alligood, Christina, "Choice among stimuli in equivalence classes" (2007). *Graduate Theses, Dissertations, and Problem Reports*. 2590.

<https://researchrepository.wvu.edu/etd/2590>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

CHOICE AMONG STIMULI IN EQUIVALENCE CLASSES

Christina Alligood

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Psychology

Approved by

Philip N. Chase, Ph.D., Chair
Michael Perone, Ph.D.
B. Kent Parker, Ph.D.
Barry Edelstein, Ph.D.
A. Diane Williams, MSW

Department of Psychology

Morgantown, West Virginia

2007

Keywords: Stimulus relations, Stimulus equivalence, Equivalence classes
Copyright 2007 Christina A. Alligood

ABSTRACT

Choice Among Stimuli in Equivalence Classes

Christina A. Alligood

Stability in responding to stimuli within equivalence classes has implications for the maintenance of classes and the ease at which they can be reformed. One way of investigating stability is to examine accuracy and speed of responding as a function of nodal number, or the number of nodes between stimuli in a class. Previous research suggests that subjects respond more quickly and more accurately on relations involving fewer nodes (Fields, Adams, & Verhave, 1989 [May]; Fields, Adams, Verhave, & Newman, 1990; Fields, Adams, & Verhave, 1993; Fields, Landon-Jiminez, Buffington, & Adams, 1995; Spencer & Chase, 1996). A second way to investigate stability is to compare accuracy and speed of responding as a function of types of relations: trials that test baseline, symmetry, transitivity, and combined symmetry and transitivity relations. Research has shown that subjects typically respond to baseline and symmetric relations faster than transitive and combined relations in tests for emergent relations. In the current research, tests were conducted after stable responding in accordance with equivalence relations had been established. Within-class preference tests were used to assess the effects of nodality and relation types on stability. A within-class preference test consists of match-to-sample trials with three or more class-consistent comparisons that occur after confirmation of class formation. In the first experiment, subjects more frequently chose comparisons related to the sample via fewer nodes than those related via more nodes. In addition, subjects chose comparisons related to the sample via symmetry as often or more often than those related via trained baseline relations. Subjects also chose both symmetry and baseline more often than transitive and combined relations. Experiments 2 and 3 investigated the possibility that effects observed in the first experiment were due to the order of training and testing. The results of Experiment 2 were consistent with the results of Experiment 1. The order of testing in Experiment 3 revealed some differences. Performance on the nodal tests was more variable. In addition, only one subject demonstrated highly accurate and stable performance on tests for equivalence. In Experiments 2 and 3, the third comparison sometimes appeared to serve as a contextual stimulus for choosing between the other two comparison stimuli. Experiment 4 evaluated effects of a class-specific reinforcer arrangement during training on responding during post-class-formation within-class preference tests. The class-specific reinforcer arrangement increased stability on nodal-test responding. Relational test results were consistent with the previous three experiments. Results are discussed in terms of theoretical implications for the substitutability of stimuli in equivalence classes, and for application to education, particularly in learning languages and other complex curricula involving stimulus classes.

ACKNOWLEDGEMENTS

This research was partially funded by grants from the Office of Academic Affairs and the Psychology Department Alumni Fund. Thanks are due to my lab mate Harold Lobo for technical assistance and to my research assistant Adam Leroy for running many of the training and testing sessions. I would also like to thank my dissertation committee, Mike Perone, Kent Parker, Barry Edelstein, and Diane Williams, for their time and input.

My adviser and committee chair, Dr. Philip N. Chase, has been everything a graduate student could ask of a mentor and more. Phil is an incredible teacher, consistently applying behavior analytic principles to the benefit of his students. He sets a superb example of service and professionalism. He is warm, accessible, and always ready with sound guidance and a supportive pat on the back. Phil has been a patient cheerleader during my dissertation process, always ready to give detailed feedback and to prompt me when I needed prompting. Phil, thank you. It has been a pleasure to be your student.

I am fortunate enough to have both amazing friends and a wonderful family, and I appreciate their support of my education. Thank you, Vennessa, Elizabethann, and Misty, for sharing the graduate school experience with me and for becoming like family. Robin, thank you for continuing to cheer me on and for reminding me that there is life outside of academe. Christopher, thank you for your constant love, encouragement, and understanding. It means so much to know that my goals are important to you. And finally, I am grateful to my family for their never ending support of all of my endeavors. Mom, Kevin, Eric, and Kaitlyn, thanks for helping me persevere – I'm finally finished!

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
Table of contents	iv
Introduction.....	1
Experiment 1	9
Experiment 2	22
Experiment 3	27
Experiment 4	32
General Discussion.....	37
References	44
Appendices	48
Tables	53
Figure Captions	62
Figures.....	65

Choice Among Stimuli in Equivalence Classes

Behavior analysts have been interested in stimulus equivalence as a means of studying behavior that occurs in the presence of classes of arbitrary stimuli at least since Keller and Schoenfeld's classic 1950 text. Keller and Schoenfeld (1950) mentioned several early stimulus equivalence experiments (Cofer & Foley, 1942; Razran, 1939; Riess, 1940; Riess, 1946) in their discussion of concepts and mediated generalization (p. 160-161). Sidman and Tailby (1982) sparked renewed interest in this area. In a typical stimulus equivalence experiment, subjects receive direct training in match-to-sample (MTS) tasks. In MTS, one sample stimulus (e.g., A1) and two or more comparison stimuli (e.g., B1, B2, and B3) are presented. Subjects are directly trained through the reinforcement of their selection of a particular comparison stimulus (e.g., B1) in the presence of each sample stimulus. In this way, training establishes baseline relations between pairs of stimuli (e.g., A1B1). Following baseline training, subjects complete test trials in which the stimuli are rearranged to test for emergent (untrained) relations. Having been trained to match B1 to A1 and C1 to B1, a subject who matches A1 to B1 and B1 to C1 demonstrates symmetry. After the same training, matching A1 to C1 would demonstrate transitivity, and matching C1 to A1 would demonstrate combined transitivity and symmetry (hereafter referred to as a combined relation). Provided that the subject also demonstrates reflexivity by matching A1 to A1, B1 to B1, and C1 to C1, an equivalence class containing the stimuli A1, B1, and C1 is inferred.

The stability of relations among stimuli in equivalence classes also has been of interest to behavior-analytic researchers (e.g., Pilgrim & Galizio, 1995; Galizio, Stewart, & Pilgrim, 2004; Fields, Adams, Verhave, & Newman, 1990; Fields, Adams, & Verhave, 1993; Fields, Landon-Jiminez, Buffington, & Adams, 1995; Fields & Verhave, 1987; Spencer & Chase, 1996). The

stability of stimulus classes has implications for how easily classes are maintained, expanded, reestablished, and reformed with other stimuli. This interest is fundamental to the application of the concept of stimulus classes to understanding how complex environment-behavior relations, such as language, are learned. If language is said to involve classes of arbitrary stimuli, like synonyms and parts of speech, then these stimulus classes must be maintained in some contexts and also be flexible enough to change in other contexts. When the classes are stable, they may occur under conditions of disruption, long periods of time without use, and conditions of stress. When the relations are more variable, they may be more easily combined and recombined with other relations under conditions of adaptation.

The stability of responding to stimuli in equivalence classes may also prove important for a basic understanding of equivalence classes. Research has yet to provide a description of the necessary and sufficient conditions for establishing equivalence classes. For example, attempts to demonstrate equivalence with nonhuman animals have been largely unsuccessful (for a possible exception see Kastak, Schusterman, & Kastak, 2001). This lack of success may be due to a poor understanding of conditions necessary to produce stable equivalence classes.

Although stability has been investigated in a number of ways, as reported below, a unifying issue has been to determine the conditions that produce stable responding to stimuli in equivalence classes across manipulations. For example, investigators have examined variables that affect whether stimulus classes are maintained over a retention interval (Spradlin, Saunders, & Saunders, 1992).

For both applied and theoretical reasons, researchers have used investigations of the stability of responding to stimuli in equivalence classes to examine the issue of stimulus substitutability within equivalence classes (Fields et al., 1993; Fields et al., 1995). It has been

commonly held that stimuli in an equivalence class are substitutable for one another. For example, Green and Saunders (1998) stated:

Broadly defined, stimulus equivalence is synonymous with stimulus substitutability.

When a stimulus that controls a response can be replaced with another stimulus without altering the probability that the response will occur, the inference can be made that the two stimuli are the same, in some sense, to the organism. (p. 230)

Partial evidence for substitutability comes from finding that responding to stimuli in a class is stable. The more that responding is variable, however, the less substitutable the stimuli may be. Given these applied and theoretical reasons, a number of studies of stability have been undertaken.

Research on the Stability of Equivalence Relations

Investigations of stability of responding to stimuli in equivalence classes have used several different measures and methods to demonstrate stability in responding to stimuli in equivalence classes, including accuracy, speed, response generalization, and choices among comparisons within a stimulus class (Pilgrim & Galizio, 1995; Galizio et al., 2004; Fields et al., 1990; Fields et al., 1993; Fields et al., 1995; Fields & Verhave, 1987; Spencer & Chase, 1996). In general, accuracy and response speed have been used to measure differential responding during testing for equivalence classes, while response transfer and within-class choices have been used to measure differential responding following the demonstration of class formation.

Findings regardless of method have suggested that stability in responding to stimuli in equivalence classes is related to nodal number and type of relation tested (e.g., Fields et al., 1990; Fields et al., 1993; Fields et al., 1995; Fields & Verhave, 1987; Spencer & Chase, 1996). In a linear training sequence (e.g., A to B, B to C, C to D, and so on), a node is a step between

two related stimuli. Thus, if A goes with B and B goes with C, then A goes with C. In this instance, B is a node between A and C. The AC relation is a one-node relation. Similarly, if A goes with B, B goes with C, and C goes with D, then A goes with D. The AD relation is a two-node relation. Type of relation refers to the different relations tested to demonstrate equivalence (e.g., baseline, symmetry, transitivity, combined).

Fields et al. (1990) measured nodal differences in accuracy during tests for equivalence. Following two-choice linear training involving A, B, C, and D stimuli (three-syllable nonsense words), tests for symmetric, transitive, and combined relations were presented. Subjects responded with greater accuracy on tests for relations involving one node (e.g., AC relations) than on tests for relations involving two nodes (e.g., AD). Spencer and Chase (1996) measured both accuracy and response speed during tests for equivalence following three-choice linear training involving A, B, C, D, and E stimuli. Subjects responded more accurately and more quickly on tests for relations involving fewer nodes than on tests for relations involving more nodes. For example, subjects would respond more quickly, and would respond accurately more often, on a test of one-node transitivity (AC) than on a test of three-node transitivity (AE). In addition, Spencer and Chase also found differences in accuracy and speed of responding to the different types of relations tested. For example, subjects would respond more quickly, and would respond accurately more often, on a test of symmetry than on a test of transitivity. There were no significant accuracy or speed differences between responding to transitive relations and combined relations. In addition, Spencer and Chase found that when differences in accuracy diminished with repeated testing, differences in speed remained. Thus, the authors suggested that speed might be a more sensitive measure of nodal number effects than accuracy alone.

Other studies have involved measures of differential responding subsequent to successful performances on tests for equivalence. Fields et al. (1995) used two-choice linear training that involved A, B, C, D, and E stimuli (three-letter nonsense syllables). As in previous studies, subjects were more likely to respond accurately on tests for relations involving fewer nodes during tests for equivalence. Following successful equivalence test performances, subjects were trained to emit different responses in the presence of the A and E stimuli from each of the two classes. For example, the response emitted in the presence of the A1 stimulus would be different from the response emitted in the presence of the E1 stimulus. A2 and E2 stimuli would also each have a unique response. Subsequent tests measured the responses emitted in the presence of the B, C, and D stimuli from each class. Between-class errors were rare. That is, on 96% of test trials presenting a B, C, or D stimulus, subjects performed the response trained to the A or E stimulus from the same class as the sample. For example, in the presence of the B1 stimulus, subjects typically performed either the response trained to the A1 stimulus or the response trained to the E1 stimulus, and not the responses trained to the A2 or E2 stimuli. The particular response emitted, though, was related to nodal number such that the response trained to the E stimulus was more likely to be emitted on trials where the D stimulus was presented than on trials where the B stimulus was presented. Similarly, the response trained to the A stimulus was more likely to be emitted on trials where the B stimulus was presented than on trials where the D stimulus was presented.

Potential Methodological Explanations of Findings

Some explanations for the variability seen in experiments involving nodal number and relation type have focused on methodological issues. One such explanation was originally suggested by Spradlin and colleagues (e.g., Saunders, Saunders, Williams, & Spradlin, 1993;

Spradlin & Saunders, 1986) and was elaborated by Saunders and Green (1999). Saunders and Green's analysis begins with the assumption that

for performances to meet criteria for acquisition of the trained baseline relations as well as criteria for positive outcomes on all tests for stimulus equivalence, each stimulus must be discriminated from every other stimulus in the experiment (Saunders & Green, 1999, p. 120).

As noted by Sidman (1986), successful performance on the training provided in a typical stimulus equivalence experiment requires discrimination of each sample stimulus from every other sample stimulus presented across trials, discrimination of each sample from the comparisons presented within trials, and discrimination of each comparison stimulus from other comparisons presented within the same trial. Typical training does not require discrimination of comparisons presented on a given trial from other comparisons presented on different trials.

These between-trial discriminations are, however, required for successful performance on tests of equivalence because the comparison stimuli become the samples during combined and symmetry trials. Because the tests require simple discriminations that may not necessarily have been established during training, subjects may perform differently on trials testing for equivalence including stimuli that have been involved in all the necessary discriminations from those that have not (Saunders & Green, 1999).

This problem of differential performance has been addressed by distinguishing between training sequences. For example, the linear training sequence exacerbates the problem of not requiring all the simple discriminations during training. In linear training, the comparisons in the first stage of training become the samples for the second stage, and so on throughout the training series. Thus, the first stimuli in the series are never presented as comparisons and the last stimuli

in the series are never presented as samples. For example, in an AB-BC-CD-DE-EF-FG training series, the A stimuli would only be presented as samples, and the G stimuli would only be presented as comparisons during training. Saunders and Green (1999) hypothesized that the differences in speed reported by Spencer and Chase (1996) were artifacts of the differential acquisition of these discriminations caused by using a linear training structure. This is a plausible explanation for differences in speed on tests for equivalence in Spencer and Chase because they included both the relations involving the A stimuli and the G stimuli in their analyses and they did not include controls for the simple discriminations discussed by Saunders and Green (1999). Hypothetically, Saunders and Green's explanation would not hold, however, in a study that tested substitutability after obtaining highly accurate performance and stable response speeds on tests. Their original assumption is that all necessary simple discriminations must be acquired for performance on tests of equivalence to be consistently high. If such consistently high test performances were obtained, then one could conclude that all the necessary discriminations had been learned to a similar degree. In addition, this explanation does not seem to apply to the differences found among types of relations tested.

Another methodological explanation for variability related to nodal number was described by Imam (2001). Imam pointed out that linear training is typically conducted in a cumulative manner, such that in an AB-BC-CD training sequence, when the CD relation is trained AB and BC training trials are also included. Thus, when the subject finishes the sequence, many more AB trials will have been completed than CD trials. Imam posited that this difference in the number of training trials for different baseline relations may account for subsequent differences in relation to nodal number. When Imam controlled for differential practice, however, statistically different responding related to the number of nodes was still

obtained across all tests for one subject and some tests for the second subject. Although Imam interpreted a lack of negative slope in most of these tests as an absence of nodality effects, the linear trend evidenced by a negative slope is just one of many possible effects of nodal number. Some authors have reported linear nodality effects, but others have reported differential responding across nodes without specifying trends (e.g., linear, quadratic etc.). Because Imam (2001) obtained differential responding related to nodal number even when the number of training trials per conditional discrimination was controlled, his results may be added to the conclusion that nodal number affects stability of responding to stimuli in an equivalence class. In addition, Imam found speed differences in responding to different types of relations. Subjects responded more quickly on tests of baseline relations than on tests of symmetric relations, and more quickly on tests of symmetry than on tests of combined relations. Imam proposed that differences in the number of test trials of each relation type might have contributed to the speed differences.

Based on Saunders and Green's (1999) account, it seems important to implement a high standard of stability on initial equivalence tests before proceeding to tests for substitutability. In addition, based on Imam's (2001) arguments, it seems important to administer an equal number of training and testing trials within each subject for each relation trained and tested. It also seems important to separate the initial equivalence tests from tests of stability and substitutability so that stable class-consistent responding can be established before these tests are conducted. Given the concerns noted by Saunders and Green (1999) and Imam (2001), it appears that additional research is needed to isolate the variables responsible for producing stability in responding to the stimuli in an equivalence class.

Statement of the Problem

Previous research has measured differences in speed and accuracy of responding during tests for equivalence, and differences in response transfer and within-class preference following successful equivalence performances. The necessary and sufficient conditions for stability related to structural training variables, and to nodal number in particular, are still unclear. The present study investigated stability through post-class-formation tests that examined within-class preference (Fields, Adams, & Verhave, 1989 [May]). Within-class preference measures stability in responding to members of an equivalence class by presenting comparison stimuli that each belong to the same class as the sample. This type of test also provides a measure of substitutability within equivalence classes, separates initial tests from those of substitutability, and allows these tests after initial tests have shown highly accurate and stable rates of responding. If class members were perfectly substitutable for one another, subjects would be expected to choose each comparison equally often when presented with tests of within-class preference. The first experiment presented here used within-class preference tests to address whether unambiguous differences in preference for comparison stimuli could be obtained after an equivalence class had been formed.

Experiment 1

Method

Subjects

Five female undergraduate students attending West Virginia University and completing courses in psychology served as subjects in Experiment 1. Subjects were randomly selected from a pool of students recruited through the use of a recruitment form. Individuals completing the form were selected and contacted by the experimenter to schedule an appointment for an initial

session. Each subject read and signed a Consent and Information Form (Appendix A) before beginning the study.

Subjects received paper slips valid for extra credit in their psychology course. They also received a cash payment based on their performance. This payment was dependent on the number of points earned by the subject during the experiment. Each point earned had a monetary value of \$0.05. In addition, subjects received \$1.00 cash for each session attended upon completion of all scheduled sessions.

Apparatus and Setting

A specialized application programmed in Microsoft® Visual Basic® 6.0 presented the experimental training and testing tasks. An IBM®-compatible Pentium®-class computer was used to run the program. A 35.5-cm color monitor with a screen resolution of 800 x 600 pixels presented the MTS tasks and displayed points earned. A two-button wheel mouse was the input device, with all actions controlled by the left button. Sessions were conducted individually for each subject in a sound-attenuated room measuring approximately 180-cm x 180-cm. The room was furnished with a large desk, a chair, and a computer. A 30-cm x 30-cm wooden door, located to the right of the desk, was used to pass materials between the experimenter and the subject. A 117-cm x 50.8-cm one-way mirror, situated next to the wooden door, allowed the experimenter to observe the behavior of subject. The experimenter could also view the subject's responding on a monitor in the control room connected to the subject's computer. Throughout training and testing sessions, white noise played through headphones masked extraneous noises.

Stimuli

Eighteen Chinese characters were used as stimuli (see Figure 1). A notation including a capital letter and a number (e.g., A1) identifies each stimulus throughout the manuscript. Letters

designate sets of comparison stimuli (e.g., B1, B2, and B3 are the B set of comparison stimuli) for training and most tests (see below for exceptions). Numbers designate the stimulus classes that may emerge during testing (e.g., A1 and B1 are members of class 1). Subjects did not have access to this system of notation.

Matching to Sample

Each experimental session consisted of one or more blocks of matching-to-sample (MTS) trials. Blocks ranged from 36-90 MTS trials depending on the stage of the experiment. Two-minute intervals were programmed between blocks. At the beginning of each trial, a red square appeared on the computer screen, just above the vertical center and at the horizontal center. Three blue squares were aligned horizontally below the red square, and a point counter was located at the bottom-right corner of the screen. The sample stimulus was presented in the red square after a random delay from the beginning of the trial ranging from zero to three seconds (Carlin, Wirth, & Chase, 1998). If the computer's mouse was clicked before the sample appeared, the delay was reset to five seconds. After the sample appeared, a click of the mouse when the pointer was located on the stimulus produced three comparison stimuli, one in each of the blue squares. A mouse-click on one of these comparison stimuli was recorded as the subject's selection for the trial. During pretraining and most stages of baseline training, clicking the comparison stimulus designated as correct resulted in a 1-s tone, a display of either the word "Correct!" in green letters and one point added to the counter. Clicking a comparison stimulus other than the one designated as correct resulted in a different 1-s tone and a display of either the word "Wrong!" in red letters. During the last stage of baseline training and on test trials, the program continued to record subjects' choices to determine their pay at the end of each experimental phase, but differential consequences did not occur. Instead, an empty screen was

displayed for one second. No consequences were programmed for further responding by the subject during this 1-s interval in any experimental phase. During testing, subjects earned a point for each test trial regardless of the class consistency of their responses, but points were not displayed onscreen.

Each sample and correct comparison combination presented in a trial constitutes a trial type (e.g., A1B1). The trial types were presented randomly in each phase, with the restrictions that a single trial type was presented consecutively no more than two times, the comparison stimulus designated as correct was presented in the same location on no more than two consecutive trials, and the trial types for each set of comparison stimuli were presented equally often within a block of trials.

Procedure

All subjects received pretraining, baseline conditional discrimination training, and equivalence-class testing. Subjects who met class consistency and stability criteria for the equivalence-class testing then received nodal testing and relational testing. The dependent variables during these tests were the percent of test trials in which each comparison stimulus was chosen and the speed of subjects' responses upon presentation of the comparison stimuli. Of experimental interest in nodal testing was whether differences in percent of responses allocated to each comparison were related to nodal number. The interest in the relational testing was in whether differences were related to the type of relation (baseline, symmetry, transitivity, or combined).

Pretraining. One pretraining block was conducted. Before beginning the block, subjects read and signed the Consent and Information Form. The computer screen initially displayed a welcome message that included instructions (Appendix B). The subject then began a pretraining

block including 26 MTS trials. The task was to match an upper-case English letter to its lower-case equivalent (e.g., matching “A” to “a”). Consequences for correct and incorrect selections were as described above. The letter used as a sample for each trial was determined randomly, with no letter presented more than once. Subjects whose accuracy during pretraining was below 90% were dismissed from the study and received an extra-credit slip for their participation.

Baseline training and equivalence-class testing. Each remaining subject received a minimum of 14 baseline-training blocks. The accuracy criterion for baseline training blocks was the completion of two consecutive blocks with 90% or higher accuracy. The computer screen again presented the welcome message displayed prior to pretraining. Each subject was trained with five sets of three conditional discriminations among the arbitrary stimuli via MTS trials. The conditional discriminations trained were A1B1, A2B2, A3B3; B1C1, B2C2, B3C3; C1D1, C2D2, C3D3; D1E1, D2E2, D3E3; and E1F1, E2F2, E3F3. The discriminations were trained one set at a time, beginning with the AB discriminations. Once subjects met the accuracy criterion for this set, subsequent sets were trained following a linear training procedure (Green & Saunders, 1998) presented according to the order shown in Table 1.

Training blocks included between 36 and 90 trials, depending on the training stage. Stages AB through EF-1 involved cumulative training in that a new set of discriminations was trained and previously trained discriminations were also presented. In the last stage, general training, all sets of discriminations had been presented an equal number of times overall. By the end of training all sets of discriminations were presented an equal number of times (see Table 2). Differential consequences were not presented during the general training blocks. The accuracy criteria for advancement from a training stage to the next experimental stage required each subject to meet the less stringent of two requirements. The first requirement was that the subject

respond correctly on at least 90% of the trials in each of two consecutive blocks, and the second was that the subject respond incorrectly on no more than one trial in each of two consecutive blocks.

Equivalence-class testing blocks tested for reflexivity, symmetry, transitivity, and combined relations and were alternated with training blocks according to the order shown in Table 1. Trials for each type of equivalence-class testing block are listed in Table 3. Test trials did not include differential consequences. Stability was assessed after four blocks of each test type were presented. The stability criterion for advancing from a testing stage to the next experimental stage required that the difference between the average accuracy for the first two blocks and the average accuracy for the second two blocks was not greater than 10% of the average accuracy across all four blocks for each test type. If the stability criterion was not met after administration of four blocks of each test type, then one additional block of each test type was administered and stability was assessed using the last four blocks of each test type. This procedure continued until the stability criterion was met or until 10 blocks of each test type had been administered, whichever occurred first. If responding on the last phase of equivalence-class testing met the stability criterion and if the average accuracy across the last four blocks of this testing was greater than 70%, subjects advanced to nodal number testing.

Nodal testing. Table 4 lists trial types for the nodal test phase. Trials in this phase presented comparison stimuli from the same stimulus class as the sample and are referred to as either forward or backward testing according to the order in which they were presented in the linear training progression. For example, a trial presenting stimulus A1 as the sample and stimuli D1, E1, and F1 as the comparisons is referred to as forward testing, while a trial presenting stimulus F1 as the sample and stimuli A1, B1, and C1 as the comparisons is referred

to as backward testing. All nodal test trials were presented within a single block. As in previous testing blocks, differential consequences were not included. Stability was assessed after four testing blocks were administered. Because all three comparisons in nodal test trials were class consistent, stability was assessed using speed rather than accuracy. For this purpose, speed was calculated by dividing one by the time between the click on the sample stimulus and the click on a comparison stimulus. The stability criterion for speeds was the same as the stability criterion described above for accuracy. When the criterion was reached, subjects advanced to relational testing.

Relational testing. Table 5 lists trial types for the relational test phase. Trials in this phase presented comparison stimuli representing three different relations to the sample stimulus. For example, one trial type presented stimulus C1 as the sample and stimuli B1, D1, and E1 as the comparisons. In this case, selecting comparison B1 would represent a symmetric relation to the sample, comparison D1 would represent a baseline relation to the sample, and comparison E1 would represent a transitive relation to the sample. All relational test trials were presented within a single block, and test trials did not include differential consequences. Stability was assessed after administration of four testing blocks, with the stability criterion as described above. Each subject completed the experiment when the stability criterion was met or when 10 blocks were administered.

Results

All five subjects in Experiment 1 met accuracy criteria at each stage of training. Figure 2 shows that the subjects also met accuracy criteria for equivalence testing before moving on to nodal and relational testing. Each bar represents the mean percent correct responses across the last four testing blocks for one of four trial types. Labels on the X axis denote the four trial types,

with “Symm” denoting symmetry trials, “Trans” denoting transitivity trials, “Comb” denoting combined trials, and “Ref” denoting reflexivity trials. Four of the five subjects demonstrated 90% or greater accuracy on all four trial types. Equivalence performances were stable as determined by the four-test stability criterion described above.

All subjects met speed stability criteria for nodal and relational testing. In previous studies speed data have been analyzed in relation to accuracy on tests of emergent relations. Because all choices on nodal and relational tests were class consistent and therefore accuracy was not relevant, speed data were used only to assess stability of responding on these tests.

Figure 3 shows the percent of responses allocated to each comparison on the nodal tests for all five subjects. Each group of three bars represents a trial type, with each bar showing the percent of total responses allocated to a particular comparison. On nodal testing trials, each comparison was related to the sample via a different number of nodes. The number below each bar denotes the number of nodes in the relation between a particular comparison and the sample. Subsequent figures showing nodal test results will follow the same format.

For 19 of the 20 possible comparisons of trial types on the nodal tests subjects chose most often the comparison related to the sample via the fewest nodes. On one set of comparisons for one subject (011), the majority of responses were allocated to the comparison related to the sample via three nodes. In addition, for most subjects, the relation between the number of nodes and the proportion of responses allocated was a linear function. For two subjects (009 and 010), choices on all trial types were linear in order of the number of nodes in the relation. For subject 006, the pattern of choices on trials comparing relations with two, three, and four nodes and on trials comparing relations with one, two, and three nodes also was linear (37%, 33%, and 30%; and 71%, 17%, and 12%, respectively). For subject 007, the pattern of choices was linear on

trials comparing relations with two, three, and four nodes (71%, 21%, and 8%), on trials comparing relations with one, three, and four nodes (96%, 4%, and 0%), and on trials comparing relations with one, two, and four nodes (92%, 4%, and 4%). For Subject 011, the pattern of choices was linear on trials comparing relations with one, two, and three nodes (59%, 33%, and 8% respectively).

The effects of nodal number were also shown in the relational test trials that did not include baseline relations. For the sake of comparisons of nodal number, symmetric relations have zero nodes. Figure 4 shows the percent of responses allocated to each comparison on relational test trial types that did not include baseline relations. Each group of bars again represents a particular trial type, with each individual bar showing the percent of total responses allocated to a particular comparison. The three comparisons in each trial of the relational tests were each related to the sample via a different relation. The letters and numbers below each bar denote the relation between a particular comparison and the sample. An “S” denotes a symmetric relation, a “T” denotes a transitive relation, and a “C” denotes a combined relation. Numbers following these letters indicate the number of nodes between that comparison and the sample. Subsequent figures showing relational test results will follow the same format.

For 23 of the 25 possible comparisons of nodal number on the relational tests subjects chose most often the comparison related to the sample via the fewest nodes. Subject 007 chose the comparison related to the sample via the fewest nodes most often on three of the five trial types. On trial types comparing symmetric to one-node combined and one-node transitive relations, and combined one-node to combined three-node and combined four-node relations she selected another stimulus most often. Subject 006 chose the comparison related to the sample via the fewest nodes exclusively on trials comparing symmetric, three-node combined, and four-

node combined relations and on trials comparing symmetric, two-node transitive, and three-node transitive relations. This subject's pattern of responding was slightly more variable on other trial type comparisons. Subjects 009 and 011 chose the comparison related to the sample via the fewest nodes most often on all trial types. For subject 010, the comparison related to the sample via the fewest nodes was chosen exclusively on four of the five trial types. The patterns of responding to nodal number on relational test trials were typically linear, but some response patterns formed other functions.

Figure 5 shows the percent of responses allocated to each comparison on relational test trial types that included baseline relations. Subjects allocated the majority of responses to comparisons related to the sample via baseline or symmetric relations on all trial types. Preference for baseline or symmetric relations varied across subjects. On trials comparing baseline, symmetric, and one-node transitive relations, two subjects (009 and 010) chose the comparison related to the sample via the baseline relation most often, two subjects (006 and 007) chose the comparison related to the sample via the symmetric relation most often, and one subject (011) chose the comparisons related to the sample via the baseline and symmetric relations equally often. On trials comparing baseline, symmetric, and one-node combined relations, two subjects (010 and 011) chose the comparison related to the sample via the baseline relation most often, two subjects (006 and 007) chose the comparison related to the sample via the symmetric relation most often, and one (009) subject chose the comparisons related to the sample via the baseline and symmetric relations equally often.

Discussion

In Experiment 1, undergraduate psychology students completed three-choice linear training involving A, B, C, D, E, and F stimuli (Chinese characters), which controlled for the

number of trials of each relation trained. Subjects then completed testing for the emergence of three six-member classes. After passing the equivalence tests at a high level of accuracy, subjects completed nodal testing, in which all comparisons were members of the same equivalence class as the sample. The relation of the comparisons to the sample varied in terms of nodal number. On these nodal tests, all five subjects were more likely to choose the comparison related to the sample by the fewest nodes (see Figure 3). These data suggest that choice among stimuli in an equivalence class was influenced by the number of nodes: stimuli separated from the sample by more nodes were less likely to be chosen than stimuli separated from the sample by fewer nodes.

After completing this testing, subjects moved on to relational testing, in which all comparisons were again members of the same equivalence class as the sample. The comparisons in these test trials each bore a different type of relation to the sample. Extending an investigation of Fields et al. (1989 [May]), some relational testing trials also presented comparisons related to the sample by directly trained baseline relations. On these relational tests, all five subjects were more likely to choose the comparison related to the sample by a baseline (directly trained) or symmetrical relation than those related by transitive or combined relations (see Figures 4 and 5).

The relational tests also allowed further comparisons of nodal number because in addition to varying relation types, the comparisons' relation to the sample also varied in terms of nodal number. For example, symmetry and baseline relations involve zero nodes, while transitivity and combined relations involve one or more nodes. Nodal number also affected performance during relational testing as all five subjects chose comparisons related to the sample by transitive or combined relations involving fewer nodes more often than those involving more nodes (see Figure 4).

In trials including comparisons related to the sample by baseline and symmetric relations, all five subjects chose the comparison related by the symmetric relation at least as often as the one related by the directly trained baseline relation. This result is particularly interesting because it might not be expected given that demonstrations of symmetric relations had not been reinforced in the experiment, but demonstrations of the baseline relation had. Given that baseline and symmetric relations have zero nodes and transitive and combined relations have one or more nodes, these results suggest that whether the relation is trained or emergent is less important to the substitutability of the stimuli in an equivalence class than the number of nodes between stimuli.

The results of this study are consistent with data reported by Pilgrim and Galizio (1990). In this study, an initial equivalence class was established after training of arbitrary conditional discriminations. When the baseline relations were changed through further training, three of four subjects responded consistently with the new baseline relations on symmetry tests, but responded consistently with the original baseline relations on transitivity tests. The authors noted that this result seemed to be “inconsistent with the functional substitutability of stimuli that defines equivalence classes” (p. 223). The same might be said of the data in the present study.

The training and testing procedures in Experiment 1 were designed to eliminate several previously posed explanations for the kind of differential responding seen here. These differences in response allocation could not have arisen from unequal numbers of training trials among the conditional discriminations as suggested by Imam (2001), because the number of training trials per conditional discrimination was equated. In addition, class-consistent performances on equivalence tests administered prior to nodal testing show that all necessary simple discriminations were acquired. Thus differential response allocation cannot be attributed

to failure to acquire certain simple discriminations (Saunders & Green, 1999). Finally, the stable response speeds obtained in nodal testing suggest that the differences in response allocation seen here are not artifacts of differential acquisition caused by linear training, as suggested by Saunders and Green (1999).

These results are consistent with previous results showing that accuracy and speed of responding varies in connection with the type of relation between the sample and comparison. Further experiments are needed, however, to provide a more thorough analysis of this effect. For example, in Experiment 1, all subjects were exposed to the same testing order. Therefore, the observed variability may have been an artifact of this particular order. Because equivalence tests were administered before other tests in this case, it is possible that the equivalence tests themselves provided a history sufficient to produce this result. Also, the nodal testing may have influenced responding on the relational tests. Both nodal tests and relational tests presented comparison fields in which all comparisons were in the same previously established class as the sample. These two types of tests may therefore have appeared very similar to subjects. Following nodal testing, in which they chose the comparison whose relation to the sample involved the fewest nodes, subjects may have simply continued this pattern of responding in relational testing. Therefore, further research controlling for test order is needed.

Experiment 2 was designed to evaluate whether results similar to Experiment 1 would be obtained when the order of relational and nodal testing were reversed. In addition, some subjects in Experiment 2 were exposed to training with the purported reinforcer (“Correct”), while some subjects were exposed to training with three different purported reinforcers (business logos) that were associated with discount coupons from local businesses through instructions for these subjects. This arrangement allowed comparison between the use of the word “correct” and the

business logos as purported reinforcers. This was done because if business logos function similarly to correct feedback backed up by money, then subsequent experiments could use business logos as purported reinforcers.

Experiment 2

In this experiment, subjects completed relational testing immediately following equivalence testing. Following relational testing, subjects completed nodal testing. By administering relational testing before nodal testing, this experiment examined whether differences in responding to different stimulus arrangements found in Experiment 1 might have been an artifact of the particular testing sequence. This experiment used business logos (the logos matched discount coupons delivered as rewards for participation) as consequences during training for some subjects (045, 047, and 050) instead of the “Correct” and “Wrong” messages used in Experiment 1. This was done to verify whether there were any systematic differences in training results or subsequent test results due to this difference in consequences.

Method

Subjects

Five female undergraduate students (different than the subjects in Experiment 1) attending West Virginia University and completing courses in psychology served as subjects in Experiment 2. Subjects were randomly selected from a pool of students recruited through the use of a recruitment form. Individuals completing the form were selected and contacted by the experimenter to schedule an appointment for an initial session. Each subject read and signed a Consent and Information Form before beginning the study.

Subjects received paper slips valid for extra credit in their psychology course. They also received either cash (Subjects 014 and 015) or coupons for use at local businesses (Subjects 045,

047, and 050) based on their performance. Cash or coupon reimbursement based on performance was dependent on the number of points earned by the subject during the experiment. Each point earned had a monetary value of \$0.05. In addition, subjects received \$1.00 cash for each session attended upon completion of all scheduled sessions. Subjects who received coupon reimbursement experienced the presentation of a business logo or an “X” rather than the presentation of “Correct” or “Wrong” messages. Points were earned for correct responses independent of the presentation of these consequences.

Apparatus/Setting

The apparatus and setting, as well as the match-to-sample program, were the same as in Experiment 1.

Procedure

As in Experiment 1, all subjects received pretraining and baseline conditional discrimination training and equivalence testing. Subjects who met the accuracy criterion for baseline training and the stability criterion for equivalence testing then received relational testing and nodal testing.

Pretraining. Pretraining was the same as in Experiment 1.

Baseline training. Baseline training was the same as in Experiment 1.

Equivalence-class testing. Following baseline training, equivalence-class testing blocks were administered the same as in Experiment 1.

Relational testing. Following equivalence-class testing, relational testing was administered as in Experiment 1.

Nodal testing. Following relational testing, nodal testing was administered as in Experiment 1.

Results

All subjects in Experiment 2 met accuracy criteria for training regardless of whether they received the “Correct!” and “Wrong” (subjects 014 and 015) or the business logo and red “X” (subjects 045, 047, and 050) consequences. Subjects also met the accuracy criterion for equivalence testing before advancing to relational and nodal tests. Figure 6 shows the mean percent correct across the last four testing blocks for each subject. Four of the five subjects scored over 90% correct on all four trial types. Equivalence performances were stable as determined by the four-test stability criterion.

Figure 7 shows the percent of responses allocated to each comparison on nodal testing trials for all five subjects in Experiment 2. For 14 of the 20 possible comparisons of trial types on the nodal tests, subjects chose most often the comparison related to the sample via the fewest nodes. Exceptions were the “234”, “134”, and “124” trial types for subjects 047 and 050. In these cases, subject 047 chose the comparisons related to the sample via the fewest nodes and via the most nodes equally often. Subject 050 most often chose the comparison related to the sample via the most nodes.

Figure 8 shows the percent of responses allocated to each comparison on relational test trials that did not include baseline relations. For 22 of the 25 possible comparisons of nodal number on the relational tests, subjects chose most often the comparison related to the sample via the fewest nodes. Subjects 006 and 007 chose the comparison related to the sample via the fewest nodes most often on all trial types. Subject 045 selected the comparison related to the sample via the fewest nodes most often on four of the five trial types, selecting the comparison related to the sample via a one-node transitive relation most often on trials comparing symmetric, on-node transitive, and one-node combined relations. Subjects 047 and 050 also chose the

comparison related to the sample via the fewest nodes most often on four of the five trial types, selecting the comparison related to the sample via a four-node combined relation most often on trials comparing two-node, three-node, and four-node combined relations. The patterns of responding to nodal number on these relational tests were typically linear, but some response patterns indicated other functions.

Figure 9 shows the percent of responses allocated to each comparison on relational test trials that included baseline relations. On 8 of the 10 possible comparisons on these tests, subjects chose most often the comparison related to the sample via a baseline relation. The exceptions were the “BSC” trial type for subjects 014 and 015, on which the comparison related to the sample via symmetry was chosen most often.

Discussion

In this experiment, five female undergraduate subjects completed relational and nodal testing in the opposite order from subjects in Experiment 1 to assess possible test order effects. On nodal tests, four of the five subjects in this experiment exhibited the linear and/or u-shaped patterns of responding also seen in Experiment 1. The results of relational tests that did not include baseline relations were also consistent with Experiment 1. The consistency observed between the results of Experiment 1 and Experiment 2 suggests that the order of testing was not responsible for the results seen in the Experiment 1. In addition, the absence of a systematic acquisition effect of the business logos used as reinforcers for some subjects suggests that the logos and discount coupons can be used instead of points and money as reinforcing consequences.

Another interesting effect was observed in Experiment 2. In relational tests that included baseline relations, four of the five subjects chose the comparison related to the sample via a

baseline relation most often when the third comparison in the trial was related to the sample via a transitive relation. Conversely, they chose a comparison related to the sample via a relation other than baseline most often when the third comparison in the trial was related to the sample via a combined relation. It is possible that the relation of the third comparison to the sample served as a contextual stimulus for choosing between the other two comparisons. In Experiment 1, it was noted that choices among comparison stimuli might be influenced by whether alternative choices were members of the same class or different classes. Relational test results from Experiment 2 suggest that this context may be even more fine grained in some cases. That is, when the available choices are all members of the same equivalence class, responding may be influenced by nodal number, as shown in Experiment 1. If nodal number does not differentiate between comparisons, as is the case when baseline and symmetric comparisons are both available, then other contextual variables play a role. For example, the combined relation is a combination of symmetric and transitive relation. Perhaps the presence of a stimulus involving the property of symmetry may be a context for selecting the stimulus related to the sample via a symmetric relation. Further, the transitive relation does not include a symmetric relation. Therefore, the presence of a stimulus involving the property of transitivity may have served as a context for choosing the stimulus related to the sample via relations other than symmetry, in this case the baseline relation. This result is consistent with the idea that substitutability of class members may depend on which class members are present in the selection environment.

Order of testing may have also played a role in these results. Apparently the order of nodal testing and relational testing is inconsequential, but equivalence tests, which were interspersed with training as in Experiment 1, may have influenced responding on both nodal and relational tests by providing a testing history sufficient to produce the results seen here.

Therefore, further investigation of possible order effects was needed. Toward this end, another experiment was conducted in which equivalence testing was performed after nodal and relational testing.

Experiment 3

In this experiment, the order of testing was again changed. This time the order was relational, then nodal, then equivalence testing to further examine the possibility that differences in responding to different stimulus arrangements might be an artifact of the testing sequence.

Method

Subjects

Five female undergraduate students (different than the subjects in Experiment 2) attending West Virginia University and completing courses in psychology served as subjects in Experiment 3. The subjects were randomly selected from a pool of students recruited through the use of a recruitment form. Individuals completing the form were selected and contacted by the experimenter to schedule an appointment for an initial session. Each subject read and signed a Consent and Information Form before beginning the study. Compensation was the same as in Experiment 2 with the exception that business logos and coupons were used with all subjects in Experiment 3.

Apparatus/Setting

The apparatus and setting, as well as the match-to-sample program, were the same as in Experiments 1 and 2.

Procedure

As in the previous experiments, all subjects received pretraining and baseline conditional discrimination training. Subjects who met the accuracy criterion for baseline training then received relational testing, nodal testing, and equivalence testing.

Pretraining. Pretraining was the same as in Experiments 1 and 2.

Baseline training. Baseline training was similar to Experiments 1 and 2. However, unlike the previous experiments, equivalence testing in Experiment 3 was not interspersed with training blocks (see Table 6). Thus, training blocks were administered continuously until the accuracy criterion was met for each block type.

Relational testing. Following baseline training, relational testing was administered as in Experiments 1 and 2.

Nodal testing. Following relational testing, nodal testing was administered as in Experiments 1 and 2.

Equivalence-class testing. Following nodal number testing, equivalence-class testing blocks were administered similarly to Experiments 1 and 2. Unlike the previous experiments, here equivalence-class testing blocks were not interspersed with baseline training blocks. Instead, the testing blocks were administered continuously at the end of the experiment until responding (a) met the stability criterion and average class consistency was over 80%, or (b) 10 blocks were administered.

Results

All subjects in Experiment 3 met accuracy criteria in training before proceeding to nodal and relational tests. Figure 10 shows the percent of responses allocated to each comparison on nodal testing trials for all five subjects in Experiment 3. Responding on these tests was quite variable. Subjects chose the comparison related to the sample via the fewest nodes most often on

only 5 of the 20 possible comparisons of nodal number. The comparison stimulus related to the sample via the most nodes was chosen most often on seven nodal number comparisons, while the comparison stimulus related to the sample via the median number of nodes was chosen most often on two nodal number comparisons. Choices were allocated equally to each of the three comparison stimuli on six nodal number comparisons.

Figure 11 shows the percent of responses allocated to each comparison on relational tests that did not include baseline relations. For 20 of the 25 possible comparisons of nodal number on the relational tests, subjects chose most often the comparison stimulus related to the sample via the fewest nodes. Subjects 020 and 106 chose most often the comparison stimulus related to the sample via the fewest nodes on all five trial types. Subject 017 chose most often the comparison stimulus related to the sample via the fewest nodes on three trial types. This subject chose the comparison stimuli related to the sample via symmetry and a four-node combined relation equally often on trials comparing symmetry, three-node combined, and four-node combined relations. She chose most often the comparison stimulus related to the sample via a four-node combined relation on trials comparing two-node, three-node, and four-node combined relations. Subject 021 chose most often the comparison related to the sample via the fewest nodes on four trial types, choosing most often the comparison related to the sample via a one-node transitive relation on trials comparing symmetric, one-node transitive, and one-node combined relations. Subject 048 also chose most often the comparison stimulus related to the sample via the fewest nodes on four trial types, choosing most often the comparison stimulus related to the sample via a four-node combined relation on trials comparing two-node, three-node, and four-node combined relations. The patterns of responding to nodal number on relational test trials in this

experiment were typically linear, but several response patterns formed u-shaped functions, and one response pattern formed an inverted u-shaped function.

Figure 12 shows the percent of responses allocated to each comparison on relational tests that included baseline relations. For 7 of the 10 possible combinations on these tests, subjects chose most often the comparison related to the sample via a baseline relation. The exceptions were the “BSC” trial type for subjects 020, 021, and 106, in which subjects chose most often the comparison related to the sample via symmetry.

Figure 13 shows the mean percent correct responses across the last four blocks of equivalence testing for subjects in Experiment 3. No equivalence data are available for subject 020 because she withdrew from participation before these data could be collected. Subject 048 withdrew before stability could be reached in the equivalence-testing phase, so the data for this subject are drawn from a single testing block. Only one subject in this experiment, subject 021, showed high accuracy on equivalence tests following nodal and relational tests. One other subject, 017, showed high accuracy on symmetry and reflexivity tests but not on transitivity and combined tests. Subjects 048 and 106 both showed high accuracy on reflexivity tests only.

Discussion

In Experiment 3, responding on nodal tests was more variable than in previous experiments. This suggests that the equivalence tests, which were administered before nodal tests in previous experiments but were administered later in Experiment 3, may have influenced responding on nodal tests. In the first two experiments, the equivalence tests had shown highly accurate and stable responding, demonstrating that all the necessary simple discriminations had been acquired (Saunders & Green, 1999). The results from Experiment 3 seem to confirm the importance of this requirement. Performance on the relational tests, however, showed little effect

of order of testing. On relational tests that did not include baseline relations, subjects in Experiment 3 responded in a similar manner to subjects in the previous two experiments. On relational tests that included baseline relations, subjects in this experiment chose either stimuli related to the sample via baseline or symmetry over transitivity and combination, but also often chose the comparison stimulus related to the sample via symmetry over the comparison stimulus related via baseline. As in Experiment 2, this was especially true when the third comparison stimulus was related via combined. Thus, it appears that the type of relation of the third comparison to the sample may have influenced the stability of responding on certain types of trials.

On tests of equivalence, administered following nodal and relational tests, subjects in Experiment 3 performed differently from subjects in the previous two experiments. Only one subject showed a high degree of accuracy on these tests. It is possible that the nodal and relational tests disrupted equivalence performances. This is unclear because equivalence tests were not administered before the nodal and relational tests. Thus it is possible that highly accurate equivalence performances would not have been demonstrated at that point either. The particular order of testing implemented in this experiment was designed to investigate effects of prior equivalence testing on nodal and relational test performances by withholding the equivalence tests until the end of the testing sequence. To test whether the nodal and relational tests disrupt equivalence performances, future experiments could present equivalence tests both before and after the other tests.

Experiments 2 and 3, which controlled for test order, produced relational-test findings similar to those of Experiment 1. Despite minor differences between Experiments 1 and 2, and the differences found in Experiment 3 on nodal tests, the most consistent results on all three

experiments suggest that the substitutability of stimuli in a class can be disrupted by the comparisons that are provided in a test. In all three experiments, the particular comparisons present in a trial seemed to override class membership as a controlling variable of response allocation. The circumstances under which this type of contextual control does and does not occur are still unclear. The next experiment was designed to examine whether it would occur in a circumstance that has been shown to be effective at increasing class-consistent responding. Specifically, Experiment 4 was designed to examine whether context would override class membership as a controlling variable when the baseline conditional discriminations were trained using a class-specific reinforcer arrangement.

In a typical matching-to-sample training arrangement, the same reinforcer is used for all training trials (e.g., points exchangeable for money). In a class-specific reinforcer arrangement (e.g., Litt & Schreibman, 1981; Estevez, Fuentes, Mari-Beffa, Gonzalez, & Alvarez, 2001), a distinct reinforcer is used for trials involving stimuli in each experimenter-defined class. For example, selecting B1 in the presence of A1 might be reinforced with R1, while selecting B2 in the presence of A2 might be reinforced with R2, and selecting B3 in the presence of A3 might be reinforced with R3. Class-specific reinforcement arrangements have been shown to increase stability in responding on match-to-sample trials, as evidenced by more efficient acquisition of the baseline relations that are prerequisites for equivalence-class formation. Like nodal number, class-specific reinforcement is a characteristic of the training procedures. The question in Experiment 4 was whether the contextual control seen in the first three experiments would override class membership even in classes produced through training with class-specific reinforcers.

Experiment 4

In this experiment, subjects were trained with class-specific reinforcers. Rather than delivering a randomly selected business logo or the word “correct” following each correct response during training, the computer program delivered business logos following correct responses that were specific to each of three experimenter-defined classes. Subsequently, subjects were exposed to equivalence, nodal, and relational testing. The aim of this experiment was to examine whether the stability of equivalence classes can be improved through the use of class-specific reinforcers.

Method

Subjects

Five female undergraduate students (different than the subjects in Experiments 1, 2, and 3) attending West Virginia University and completing courses in psychology served as subjects in Experiment 4. Subjects were randomly selected from a pool of students recruited through the use of a recruitment form. Individuals completing the form were selected and contacted by the experimenter to schedule an appointment for an initial session. Each subject read and signed a Consent and Information Form before beginning the study. Compensation was the same as in Experiment 3.

Apparatus/Setting

The apparatus and setting, as well as the match-to-sample program, were the same as in Experiments 1, 2, and 3.

Procedure

The order of training and testing were the same as in Experiment 1. The only difference was in the arrangement of reinforcing stimuli during training trials.

Pretraining. Pretraining was the same as in Experiments 1, 2, and 3.

Baseline training. Baseline training was similar to Experiments 1 and 2. However, instead of presenting a reinforcer randomly selected from the group of three chosen by the subject, the training program in this experiment produced a particular reinforcer from the group of three, also chosen by the subject, depending on the type of training trial. For example, correct selections of comparison B1 in the presence of sample A1 was always be followed by Reinforcer 1 (R1), whereas correct selections of comparison B2 in the presence of sample A2 was always be followed by Reinforcer 2 (R2), and so on.

Equivalence-class testing. Following baseline training, equivalence-class testing blocks were administered as shown in Table 1.

Nodal testing. Following equivalence-class testing, nodal number testing was administered as shown in Table 1.

Relational testing. Following nodal testing, relational testing was administered as in Table 1.

Results

All five subjects in Experiment 4 met accuracy criteria for training. Figure 14 shows the mean percent correct responses across the last four blocks of equivalence tests, on which subjects met the accuracy criterion before proceeding to nodal and relational testing. All five subjects demonstrated 90% or greater accuracy on all four trial types. Equivalence performances also were stable as determined by the four-test stability criterion.

Figure 15 shows the percent of responses allocated to each comparison on nodal testing trials. On 18 of the 20 possible nodal test combinations, subjects chose most often the comparison related to the sample via the fewest nodes. The exceptions were the “234” and “134” trial types for subject 030. The comparison related to the sample via the median number of nodes

was most often chosen for this subject on “234” trials, and each comparison was chosen equally often on “134” trials. Thus, subject 030 is the only subject in Experiment 4 who showed any variability of responding on nodal tests. Each of the other four subjects exclusively chose the comparison related to the sample via the fewest nodes on every trial in nodal testing.

Figure 16 shows the percent of responses allocated to each comparison on relational test trials that did not include baseline relations. For 23 of the 25 possible comparisons of nodal number on the relational tests, subjects chose most often the comparison related to the sample via the fewest nodes. Subjects 036, 043, and 049 chose this way on all five trial types. Subject 030 chose most often the comparison stimulus related to the sample via the fewest nodes on four of the five trial types, choosing most often the comparison stimulus related to the sample via a three-node combined relation on trials comparing two-node, three-node, and four-node combined relations. Subject 044 also chose most often the comparison stimulus related to the sample via the fewest nodes on four of the five trial types, choosing most often the comparison related to the sample via a four-node combined relation on trials comparing two-node, three-node, and four-node combined relations. The patterns of responding to nodal number on relational test trials were almost always linear, but one response pattern formed a u-shaped function and another formed an inverted u-shaped function.

Figure 17 shows the percent of responses allocated to each comparison on relational test trials that included baseline relations. For 7 of the 10 possible comparisons of relation on these tests, subjects chose most often the comparison related to the sample via a baseline relation. The exceptions were “BST” and “BSC” trials for subject 030, and “BSC” trials for subject 043. Subject 030 chose most often the comparison related to the sample via symmetry on both trial

types, and subject 043 chose the comparisons related to the sample via baseline and symmetry relations equally often.

Discussion

The arrangement of reinforcers in a class-specific manner seems to have increased stability on the nodal tests. This result may be viewed as consistent with the literature showing that class-specific reinforcers have increased the stability of responding. While every comparison on the nodal tests represented a class-consistent response, the response allocations of four of the five subjects in Experiment 4 represent extremely stable patterns.

Performance on relational tests that did not include baseline relations was consistent with performances on these tests in previous experiments. On relational tests that included baseline relations, the arrangement of class-specific reinforcers may have changed the way subjects responded to comparisons that were related to the sample by symmetry and baseline. Unlike Experiments 2 and 3, only one subject (043) chose the comparison related to the sample via baseline most often when the third comparison was related via transitivity, and chose the comparisons related to the sample via baseline and symmetric relations equally often when the third comparison was related via a combined relation. The class-specific reinforcer arrangement may have increased within-subject stability of responding in the sense that subjects seemed more likely to choose a particular stimulus relation (baseline or symmetric) regardless of whether the third choice represented a transitive relation or a combined relation. This would be consistent with the idea that class-specific reinforcers increase the stability of responding in general.

This result is also consistent with the literature on the Differential Outcome Effect (DOE), another term that has been used to describe class-specific reinforcer arrangements (see Goeters, Blakely, & Poling, 1992 for a review). Some authors investigating this effect have

found a facilitative effect of class-specific reinforcement arrangements on acquisition and retention and have hypothesized that these effects are related to remembering. The behavior of remembering may also be important to the increased stability seen with the use of class-specific reinforcers in the present study.

General Discussion

The four experiments in this study were designed to investigate choice among stimuli in equivalence classes. In this section, results will be discussed in the following order. First, expected results will be discussed, including implications for the flexibility of classes. Next, unexpected results will be discussed, in particular the allocation of responses to symmetric relations on the relational testing trials that included baseline relations. Finally, possible reasons for these results and limitations of the current experiments will be discussed in the context of suggesting further experiments.

On nodal tests and on relational tests that did not include a baseline relation, subjects most often chose the comparison related to the sample via the fewest nodes. This was expected based on previous research (Fields et al., 1990; Fields et al., 1993; Fields et al., 1995; Fields & Verhave, 1987; Spencer & Chase, 1996). Although comparisons related to the sample via the fewest nodes were chosen most often overall, the relation between response allocation and the number of nodes in the relation was not always linear. This too was expected given past research (e.g., Spencer & Chase, 1996; Imam, 2001). Moreover, these results aid arguments against some methodological explanations put forward regarding similar findings in past research. The present procedures controlled for number of presentations of training trials for each conditional discrimination, so differential responding could not have occurred due to inconsistencies in these numbers. In addition, tests for equivalence-class formation were completed prior to the within-

class nodal and relational tests on three of the four experiments, showing that all simple discriminations necessary for class-consistent responding had been established. Finally, stable response speeds were obtained on nodal and relational testing for each subject, suggesting that differential response allocation was not caused by differential acquisition arising from linear training.

The results of this experiment seem to support Fields and Moss's (in press) contention that all members of an equivalence class are substitutable for one another only within certain contexts, but not others. One relevant contextual variable is the particular comparisons provided to subjects. In contexts in which each of the comparisons is a member of a different class and only one comparison is a member of the same class as the sample, class membership appears to dictate response allocation among comparisons. This is the context in which equivalence-class testing occurred in this experiment, which is typical of equivalence research. Conversely, in contexts in which all comparisons are members of the same class as the sample, variables other than class membership appear to dictate response allocation. This is the context in which nodal and relational testing occurred in this experiment. One of the variables that may dictate responding in this context is the number of nodes between the sample and comparison, as seen in the present results. In other words, members of an equivalence class are equally substitutable for one another only in contexts in which they are contrasted with stimuli from other classes.

These results support the idea that, given a conditional stimulus and a choice between several stimuli that are members of the same class as that stimulus, adult humans tend to choose the class member that is most closely related to the conditional stimulus. By extension, members that are more distally related to each other may be more flexible (less stable). The finding that variables other than class membership affect response stability suggests one way in which

existing classes might be disrupted and new classes formed. For example, we might ask a student to select the best synonym for *talk* from among the choices *gab*, *jaw*, and *speak* given the following sentence: “The principal asked to _____ with the student in her office.” Given this context, the best answer may be the more formal synonym, *speak*. Consistent allocation of responses to this choice in such contexts may create a new class that includes formal synonyms for *talk*, such as *speak*, but excludes more colloquial synonyms such as *gab* and *jaw*. Skinner (1957, p. 91-102) discussed the differences between these synonyms as the difference between types of extension. *Speak* is a generic extension of *talk* whereas *gab* and *jaw* are either metaphoric or metonymic. The difference between generic, metaphoric, and metonymic extensions is the degree to which the stimuli in the extended class share controlling attributes. This too is related to the relative differences between stimuli in a class.

Unexpected results also emerged in the present study, beginning with relational testing in Experiment 1. Subjects often chose the comparison stimulus related to the sample via symmetry over the comparison stimulus related via baseline. This pattern of responding was not expected given the reinforcement history established by training in the experiments. Choosing the comparison stimulus related to the sample via baseline had been repeatedly reinforced in training whereas choosing the comparison stimulus related to the sample via symmetry had never been reinforced in the experiments. Interestingly, in Experiments 2 and 3, there seemed to be an association between the third comparison and whether subjects allocated the majority of responses to the comparison stimulus related to the sample via symmetry or baseline. In these cases, the third comparison seemed to serve as an even more specific context for choosing between the comparison stimuli related to the sample via baseline and symmetry. When the third comparison was related to the sample via transitivity, some subjects were more likely to choose

the comparison related to the sample via baseline. Conversely, when the third comparison was related to the sample via a combined relation, some subjects were more likely to choose the comparison related to the sample via symmetry. Thus, choosing among class members seems to have been dependent on what the third comparison stimulus was, even though the third comparison was almost never chosen. The fact that the combined relation includes a combination of the symmetric and transitive relations may provide a possible explanation for the function of the comparison stimulus related to the sample via a combined relation as a contextual stimulus for choosing the comparison related via symmetry.

This association, however, was not evident in Experiments 1 and 4. While the precise reasons for this are unclear, one possibility is that the order of testing had an effect on this pattern of responding. Experiments 1 and 4 both used the same order of testing, with equivalence tests presented first, followed by nodal and relational tests. Experiments 2 and 3 altered this order, with Experiment 2 presenting equivalence tests followed by relational and then nodal tests, and Experiment 3 presenting nodal and relational tests followed by equivalence tests. It may be that the combined effects of the equivalence tests and the nodal tests, which both preceded the relational tests in Experiments 1 and 4 but not in Experiments 2 and 3, decreased the likelihood of the third-comparison-dependent pattern of responding.

Another limitation of the current experiments is that only one training structure was investigated. The top panel of Figure 18 shows the structure of training that was used in the present study, in which each class member was directly related through training to only one other class member. All other class members were related via one to four nodes. This has been described as a linear training structure (Green & Saunders, 1998). Responding to classes of stimuli, however, can be trained so that the relations between each stimulus pair may be closer.

For example, consider the bottom panel of Figure 18, which shows a diagram of a situation in which a stimulus, A1, is related through training to the stimulus B1. The same original stimulus, A1, may then be related to multiple other stimuli (C1, D1, E1, and so on) through direct training. A1 would be related to every stimulus in the resulting class via a trained baseline relation. Each stimulus would be related to A1 via symmetry, and every stimulus would be related to all stimuli other than A1 via a one-node transitive relation. This has been described as sample-as-node training (or many-to-one) in contrast to the linear training structure used in the current experiment (Green & Saunders, 1998). Based on the present results, allocation of choices among stimuli in such a class might be expected to be more equally distributed across comparisons than that seen in the current experiments.

Future research could examine this possibility. For example, future work might compare the stability of classes in which the baseline relations are trained using a linear training structure like that used here, and other classes in which the baseline relations are trained using other training structures, such as comparison-as-node and sample-as-node training. After obtaining stable and accurate equivalence performance in each case, within-class tests like those used here could be used to investigate stability.

Investigating different training structures would also allow the separation of nodal number and training order because in linear training, nodal number differences are confounded with training order. For example, following linear training, in which the sample stimulus is A1 and the comparisons are C1, D1, and E1, most subjects allocated the majority of responses to C1. This allocation was consistent with control by nodal number, but this result would also be consistent with order of training as a controlling variable. Stimulus C1 would have been the first of the three comparison stimuli to be introduced in a linear training sequence. Experiments

involving the creation of large classes through many-to-one (sample-as-node) training, such as the bottom panel of Figure 18 would control for nodal number while testing whether the kinds of results seen here might be produced by training order. In such a training structure, there is one node between all stimuli in the class except the sample, but relations could be trained in different orders to test the order hypothesis. It may be that order of training is a more direct construct for describing the relations among stimuli than nodal number.

Another avenue for future research is an investigation using the training and testing order used in Experiments 1 and 4 with varied reinforcers such as business logos. In the present study, Experiment 1 examined performance with this training and testing order using the words “Correct” and “Wrong” as consequences, and Experiment 4 examined performance with the same training and testing order using class-specific reinforcers. An experiment using this same order and employing business logos in a non-class-specific arrangement would add to the information regarding differential performance related to the type of reinforcer and reinforcer arrangement used in training.

Finally, another limitation of the present study is that the equivalence classes had only six members each. Because of this, it was only possible to present two types of relational test trials with comparisons related to the sample via baseline, symmetry, and transitivity or a combined relation without including the “A” or “F” stimuli, which may have been chosen in such trials for other reasons (e.g., primacy or recency effects). Future experiments could investigate this phenomenon further by establishing larger equivalence classes so that a greater variety of tests like the ones performed here would be possible. For example, given an eight-member class with stimuli A-H, these types of test trials would be possible with the E and F stimuli as samples in addition to the C and D stimuli, which were samples in the tests of this type performed in the

present study. This would allow for comparisons of baseline, symmetric, and transitive relations and of baseline, symmetric and combined relations with the same stimulus as the sample. For example, trial types could include E:FDC (where E is the sample and F, D, and C are comparisons; these comparisons would represent baseline, symmetric, and combined relations) and E:FDG (baseline, symmetric, and transitive).

In summary, the present study replicated and extended previous work on responding to stimuli within equivalence classes. The results suggest that responding to stimuli in equivalence classes may be heavily context-dependent. Given the context of a sample stimulus that is a member of an equivalence class, stimuli within that class may be substitutable for another in that they would each be chosen over other, non-class-member stimuli. In other contexts, however, variables other than class membership may occasion choices between comparisons (Fields & Moss, in press). In the present case, the linear training structure produced relations with different numbers of nodes. In the contexts of the nodal and relational tests, in which the comparison stimuli were all members of the same class, choices may have been occasioned by these differences in nodal number. The present study raised additional questions to be investigated in future work. Such investigations will continue to add to our understanding of the nature of equivalence relations and their involvement in understanding complex responding such as verbal behavior.

References

- Carlin, L. A., Wirth, O., & Chase, P. N. (1998). Effects of sample response requirements on matching-to-sample performance with humans. *Experimental Analysis of Human Behavior Bulletin*, 16, 2-5.
- Cofer, C. N., & Foley, J. P. (1942). Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychological Review*, 49, 513-540.
- Estevez, A. F., Fuentes, L. J., Mari-Beffa, P., Gonzalez, C., & Alvarez, D. (2001). The differential outcome effect as a useful tool to improve conditional discrimination learning in children. *Learning and Motivation*, 32, 48-64.
- Fields, L., Adams, B. J., & Verhave, T. (1989, May). Relative strength of symmetrical, transitive, and equivalence relations after equivalence classes are fully formed. Paper presented in a symposium at the Annual Meeting of the Association for Behavior Analysis, Milwaukee, WI.
- Fields, L., Adams, B. J., & Verhave, T. (1993). The effects of equivalence class structure on test performances. *The Psychological Record*, 43, 697-712.
- 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., 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. & Moss, P. (in press). Stimulus relatedness in equivalence classes: Interaction of nodality and contingency. *European Journal of Behavior Analysis*.

- Fields, L. & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 48, 317-332.
- Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Typicality effects in contingency-shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behavior*, 82, 253-273.
- Goeters, S., Blakely, E., and Poling, A. (1992). The differential outcomes effect. *The Psychological Record*, 42, 389-411.
- Green, G. & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone, (Eds.) *Handbook of Research Methods in Human Operant Behavior*. New York: Plenum Press.
- 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.
- Kastak, C. R., Schusterman, R. J., and Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior*, 76, 131-158.
- Keller, F. S. & Schoenfeld, W. N. (1950). *Principles of Psychology: A systematic text in the science of behavior*. Acton, Massachusetts: Copley Publishing Group.
- Litt, M. D., & Schreibman, L. (1981). Stimulus-specific reinforcement in the acquisition of receptive labels by autistic children. *Analysis and Intervention in Developmental Disabilities*, 1, 171-186.
- Pilgrim, C. & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior*, 54, 213-224.
- Pilgrim, C. & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I:

- adults. *Journal of the Experimental Analysis of Behavior*, 63, 225-238.
- Razran, G. H. S. (1939). A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 90, 89-90.
- Riess, B. F. (1940). Semantic conditioning involving the galvanic skin reflex. *Journal of Experimental Psychology*, 26, 238-240.
- Riess, B. F. (1946). Genetic changes in semantic conditioning. *Journal of Experimental Psychology*, 36, 143-152.
- 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., & 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. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M.D. Zeiler (Eds.), *Analysis and Integration of behavioral units*. Hillsdale, NJ: Erlbaum.
- 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.
- Skinner, B. F. (1957). *Verbal Behavior*. Acton, Massachusetts: Copley Publishing Group.
- 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.

Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding Verbal Relations*. Reno, NV: Context Press.

Appendix A

CONSENT AND INFORMATION FORM**Examining Complexity Effects in Human Problem Solving Performances****Introduction**

I, _____, have been invited to participate in this research study, which has been explained to me by Christy Alligood. This research is being conducted by Christy Alligood, M.A. for professional purposes in the Department of Psychology at West Virginia University, under the supervision of Philip N. Chase, Ph.D.

Purposes of the Study

The purpose of this study is to learn more about the effects of different types of tasks on problem solving. I understand that the information collected from my participation in this study might be used in Christy Alligood's published research.

Description of Procedures

This study involves playing games on the computer. The games involve using a mouse to select Chinese characters displayed on the computer monitor. I will earn points that will be exchanged for discount coupons that may be used at local businesses. I will receive these coupons at the end of the study. I understand that the number of points I earn will depend on my performance on the games. I have been informed that this study will take approximately twelve hours for me to complete and that I will also receive payment based on my attendance at scheduled sessions, as well as extra credit.

I understand that because of the experimental protocol, it is important for me to come every day at my agreed-upon time. I understand that if I miss a session, I will be asked to come in for a make-up session within a week of the missed session. I also understand that if I miss two or more sessions, or I do not call in advance of missing a session, I may be dropped from the experiment. If I become ineligible to continue because of missed sessions, I understand that I will not receive extra credit or payment for attendance. Approximately 10 subjects are expected to participate in this study.

Submission date _____

Page 1 of 3

initials_____
date

Risks and Discomforts

There are no known or expected risks from participating in this study, except for the mild frustration associated with performance on the computer games.

Alternatives

I understand that I do not have to participate in this study and that I will not suffer any type of negative consequences if I decline to participate in the study.

Benefits

I understand that this study is not expected to be of direct benefit to me, but the knowledge gained may be of benefit to others. I will receive extra credit in my class for participation and money based on my attendance and performance, equaling approximately \$5 per session.

Contact Persons

For more information about this research, I can contact Christy Alligood, at cashford@mix.wvu.edu, or her supervisor, Dr. Philip N. Chase at 304/293-2001 ext. 626. For information regarding my rights as a research subject, I may contact the Office of Research Compliance at 304/293-7073.

Confidentiality

I understand that any information about me obtained as a result of my participation in this research will be kept as confidential as legally possible. I understand that my research records and test results, just like hospital records, may be subpoenaed by court order or may be inspected by the study sponsor or federal regulatory authorities (including the FDA if applicable) without my additional consent. In any publications that result from this research, neither my name nor any information from which I might be identified will be published without my consent.

Submission date _____

Page 2 of 3

initials

date

Voluntary Participation

Participation in this study is voluntary. I understand that I am free to withdraw my consent to participate in this study at any time and that such refusal to participate will not affect my student status at West Virginia University. Refusal to participate or withdrawal will involve no penalty to me. I have been given the opportunity to ask questions about the research, and I have received answers concerning areas I did not understand. In the event new information becomes available that may affect my willingness to continue to participate in the study, this information will be given to me so I may make an informed decision about my participation.

Upon signing this form, I will receive a copy.

I willingly consent to participate in this research.

Signature of Subject or Subject’s Legal Representative

Date

Time

Signature of Investigator or Co-Investigator

Date

Time

Submission date _____

Appendix B

Welcome Message

Welcome to the Emergent Behavior Laboratory!

As part of this study, you will be asked to play a game on the computer for points. Initially, a picture will be displayed at the top of the screen. If you click on this picture, three additional pictures will appear below the original. Click one of these pictures.

Feedback on your choices will include a brief computer-generated message, and the display of any points that you earn by a counter at the bottom right corner of the screen. Sometimes you will receive feedback on your choices, and other times you will not. However, your choices and points earned will always be recorded. Each point is worth \$0.05 (five cents) toward discount coupons for use at local businesses.

Please get comfortable before the session begins. When you are ready to start the session, click the button below. When the session is over, the computer will prompt you to call the experimenter by knocking on the small wooden door to your right.

Welcome Message

(Correct/Wrong Condition)

Welcome to the Emergent Behavior Laboratory!

As part of this study, you will be asked to play a game on the computer for points. Initially, a picture will be displayed at the top of the screen. If you click on this picture, three additional pictures will appear below the original. Click one of these pictures.

Feedback on your choices will include a brief computer-generated message, and the display of any points that you earn by a counter at the bottom right corner of the screen. Sometimes you will receive feedback on your choices, and other times you will not. However, your choices and points earned will always be recorded. Each point is worth \$0.05 (five cents).

Please get comfortable before the session begins. When you are ready to start the session, click the button below. When the session is over, the computer will prompt you to call the experimenter by knocking on the small wooden door to your right.

Table 1

Order of Experimental Phases for Experiments 1 and 4, including Training and Testing Blocks, Minimum Number of Blocks for Each Stage, Reinforcement Density, and Criterion for Advancement to the Next Stage.

Block Type	Minimum Number of Blocks	Reinforcement Density	Criterion for Advancement
Pretraining	1	100%	Accuracy
AB Training	2	100%	Accuracy
BC Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
CD Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
DE Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
EF Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
General Training	2	0%	Accuracy
Equivalence Testing	4 each relation	0%	Stability and Class Consistency
Relational Testing	4	0%	Speed Stability
Nodal Testing	4	0%	Speed Stability

Table 2

Number of Training Trials per Block for Each Set of Discriminations at Each Stage of Training

Training Stage	Number of trials of each set of discriminations per block					Total
	AB	BC	CD	DE	EF	
AB	36					36
BC	12	36				48
CD	9	12	36			57
DE	6	12	18	36		72
EF-1	3	6	9	18	36	72
EF-2	6	6	9	18	36	75
General	18	18	18	18	18	90
Total	90	90	90	90	90	

Table 3

Test Trials for Symmetry, Transitivity, Combined, and Reflexivity Relations During Equivalence

Testing.

Symmetry	Transitivity	Combined	Reflexivity
B1A1, B2A2, B3A3	A1C1, A2C2, A3C3	C1A1, C2A2, C3A3	A1A1, A2A2, A3A3
C1B1, C2B2, C3B3	B1D1, B2D2, B3D3	D1B1, D2B2, D3B3	B1B1, B2B2, B3B3
D1C1, D2C2, D3C3	C1E1, C2E2, C3E3	E1C1, E2C2, E3C3	C1C1, C2C2, C3C3
E1D1, E2D2, E3D3	D1F1, D2F2, D3F3	F1D1, F2D2, F3D3	D1D1, D2D2, D3D3
F1E1, F2E2, F3E3	A1F1, A2F2, A3F3	F1A1, F2A2, F3A3	E1E1, E2E2, E3E3
	B1F1, B2F2, B3F3	F1B2, F2B2, F3B3	F1F1, F2F2, F3F3
	C1F1, C2F2, C3F3	F1C1, F2C2, F3C3	
	A1D1, A2D2, A3D3	D1A1, D2A2, D3A3	
	A1E1, A2E2, A3E3	E1A1, E2A2, E3A3	
	B1E1, B2E2, B3E3	E1B1, E2B2, E3B3	

Note: The first stimulus listed for each trial represents the sample, and the second stimulus represents the comparison designated as accurate. For example, in a B1A1 symmetry trial,

stimulus B1 would be presented as the sample, with A1, A2, and A3 as comparison stimuli; A1 would be the class-consistent selection. Nodal number is listed for Transitive and Combined Equivalence Relations. In addition, baseline trials were included so that each block includes a total of 60 trials.

Table 4

Test Trials for Nodal Number Testing and Nodal numbers from Sample to Comparisons for Each Trial.

Number of Nodes from Sample to Comparisons	Conditional Discriminations	Trials per Block
1, 2, 3	A: CDE	3
1, 2, 3	B: DEF	3
3, 2, 1	E: ABC	3
3, 2, 1	F: BCD	3
1, 2, 4	A: CDF	3
4, 2, 1	F: ACD	3
1, 3, 4	A: CEF	3
4, 3, 1	F: ABD	3
2, 3, 4	A: DEF	3
4, 3, 2	F: ABC	3

Note: The first stimulus listed for each trial represents the sample, and the last three stimuli listed represent the comparisons, listed from the least to greatest number of nodes for forward testing and from the greatest to least number of nodes for backward testing.

Table 5

Trial Types and Relations Indicated By Each Comparison in Relational Testing.

Trial	Comparison	Relation	Trials per Block
B: AEF	A	Symmetry	3
	E	Transitivity (2 nodes)	
	F	Transitivity (3 nodes)	
C: BDF	B	Baseline	3
	D	Symmetry	
	F	Transitivity (2 nodes)	
C: BAE	B	Symmetry	3
	A	Combined (1 node)	
	E	Transitivity (1 node)	
D: ECA	E	Baseline	3
	C	Symmetry	
	A	Combined (2 nodes)	
F: EBA	E	Symmetry	3
	B	Combined (3 nodes)	
	A	Combined (4 nodes)	
F: ECB	E	Symmetry	3
	C	Combined (2 nodes)	
	B	Combined (3 nodes)	

Trial	Comparison	Relation	Trials per Block
F: CBA	C	Combined (2 nodes)	3
	B	Combined (3 nodes)	
	A	Combined (4 nodes)	

Note: For each trial type, the first stimulus listed represents the sample, and the last three stimuli listed represent the comparisons. Three trials of each type were presented, one for each stimulus class 1-3.

Table 6

Order of Experimental Phases for Experiment 3, including Training and Testing Blocks, Minimum Number of Blocks for Each Stage, Reinforcement Density, and Criterion for Advancement to the Next Stage

Block Type	Minimum Number of Blocks	Reinforcement Density	Criterion for Advancement
Pretraining	1	100%	Accuracy
AB Training	2	100%	Accuracy
BC Training	2	100%	Accuracy
CD Training	2	100%	Accuracy
DE Training	2	100%	Accuracy
EF Training	2	100%	Accuracy
General Training	2	0%	Accuracy
Relational Testing	4	0%	Speed Stability
Nodal Testing	4	0%	Speed Stability
Equivalence-Class Testing	4 each relation	0%	Stability and Class Consistency

Table 7

Order of Experimental Phases for Experiment 4, including Training and Testing Blocks, Minimum Number of Blocks for Each Stage, Reinforcement Density, and Criterion for Advancement to the Next Stage

Block Type	Minimum Number of Blocks	Reinforcement Density	Criterion for Advancement
Pretraining	1	100%	Accuracy
AB Training	2	100%	Accuracy
BC Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
CD Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
DE Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
EF Training	2	100%	Accuracy
Equivalence Testing	4 each relation	0%	Stability
General Training	2	0%	Accuracy
Equivalence Testing	4 each relation	0%	Stability and Class Consistency
Nodal Testing	4	0%	Speed Stability
Relational Testing	4	0%	Speed Stability

Figure Captions

Figure 1. The 18 Chinese characters used as experimental stimuli used in testing and training. A notation including a capital letter and a number (e.g., A1) identifies each stimulus. Letters designate sets of comparison stimuli (e.g., B1, B2, and B3 are the B set of comparison stimuli) for training and most tests (see below for exceptions). Numbers designate the stimulus classes that may emerge during testing (e.g., A1, B1, are members of class 1). Participants did not have access to this system of notation.

Figure 2. Percent correct responses on equivalence tests in Experiment 1. On the X axis, “Symm” denotes the bar representing percent correct on symmetry trials, “Trans” denotes this measure for transitivity trials, “Comb” for combined trials, and “Ref” for reflexivity trials.

Figure 3. Percent of responses allocated to each comparison on nodal testing trials in Experiment 1. Numbers on the X axis represent the number of nodes between sample and comparison for each comparison in a given trial. For example, “234” denotes a trial in which one comparison was related to the sample via two nodes, another was related via three nodes, and a third was related via four nodes. Each bar represents the percent of total responses allocated to a particular comparison on each trial type. For example, in the “234” trial type, the bar under “2” represents the percent of total responses on the “234” trial type that were allocated to the comparison related to the sample via two nodes.

Figure 4. Percent of responses allocated to each comparison on relational testing trials in Experiment 1 that did **not** include baseline relations. For this figure and Figure 5, letters on the X axis represent the type of relation of each comparison to the sample in a given trial, with S referring to symmetry, T referring to transitivity, C referring to combined, and B referring to baseline. Numbers refer to the number of nodes in a given relation. For example, “S C3 C4”

denotes a trial on which one comparison was related to the sample via symmetry, another was related via 3-node Combined, and a third was related via 4-node Combined.

Figure 5. Percent of responses allocated to each comparison on relational testing trials in Experiment 1 that included baseline relations. Same labels as Figure 4.

Figure 6. Percent correct responses on equivalence tests in Experiment 2. Same labels as Figure 2.

Figure 7. Percent of responses allocated to each comparison on nodal testing trials in Experiment 2. Same labels as Figure 3.

Figure 8. Percent of responses allocated to each comparison on relational testing trials in Experiment 2 that did **not** include baseline relations. Same labels as Figure 4.

Figure 9. Percent of responses allocated to each comparison on relational testing trials in Experiment 2 that included baseline relations. Same labels as Figure 5.

Figure 10. Percent of responses allocated to each comparison on nodal testing trials in Experiment 3. Same labels as Figure 3.

Figure 11. Percent of responses allocated to each comparison on relational testing trials in Experiment 3 that did **not** include baseline relations. Same labels as Figure 4.

Figure 12. Percent of responses allocated to each comparison on relational testing trials in Experiment 3 that included baseline relations. Same labels as Figure 5.

Figure 13. Percent correct responses on equivalence tests in Experiment 3. Same labels as Figure 2.

Figure 14. Percent correct responses on equivalence tests in Experiment 4. Same labels as Figure 2.

Figure 15. Percent of responses allocated to each comparison on nodal testing trials in Experiment 4. Same labels as Figure 3.

Figure 16. Percent of responses allocated to each comparison on relational testing trials in Experiment 4 that did not include baseline relations. Same labels as Figure 4.

Figure 17. Percent of responses allocated to each comparison on relational testing trials in Experiment 4 that included baseline relations. Same labels as Figure 5.

Figure 18. Depictions of the structure of classes in the present study (top panel) and a hypothetical class whose members are all related to each other by relations involving one or fewer nodes (bottom panel). In both panels, directly trained relations are depicted with solid arrows. In the bottom panel, some of the resulting emergent transitive relations are depicted with broken arrows. Note that each stimulus B1-F1 would also be related to each other stimulus B1-F1 via a one-node transitive relation with A1 as the node. Classes in the present study included six members, which were each directly related through training to only one other member. Other members were related via emergent relations involving zero to four nodes. Given a conditional (sample) stimulus, allocation of choices among other class members varied according to the number of nodes in nodal tests and in relational tests. Hypothetically, differential responding such as that seen here might be reduced in tests requiring a choice among class members that are more closely related, such as those in the bottom panel.

	1	2	3
A	鬥	齒	弋
B	自	豕	韋
C	隶	几	龠
D	龍	韭	鬼
E	及	齊	竹
F	鳥	豆	麻

Figure 1

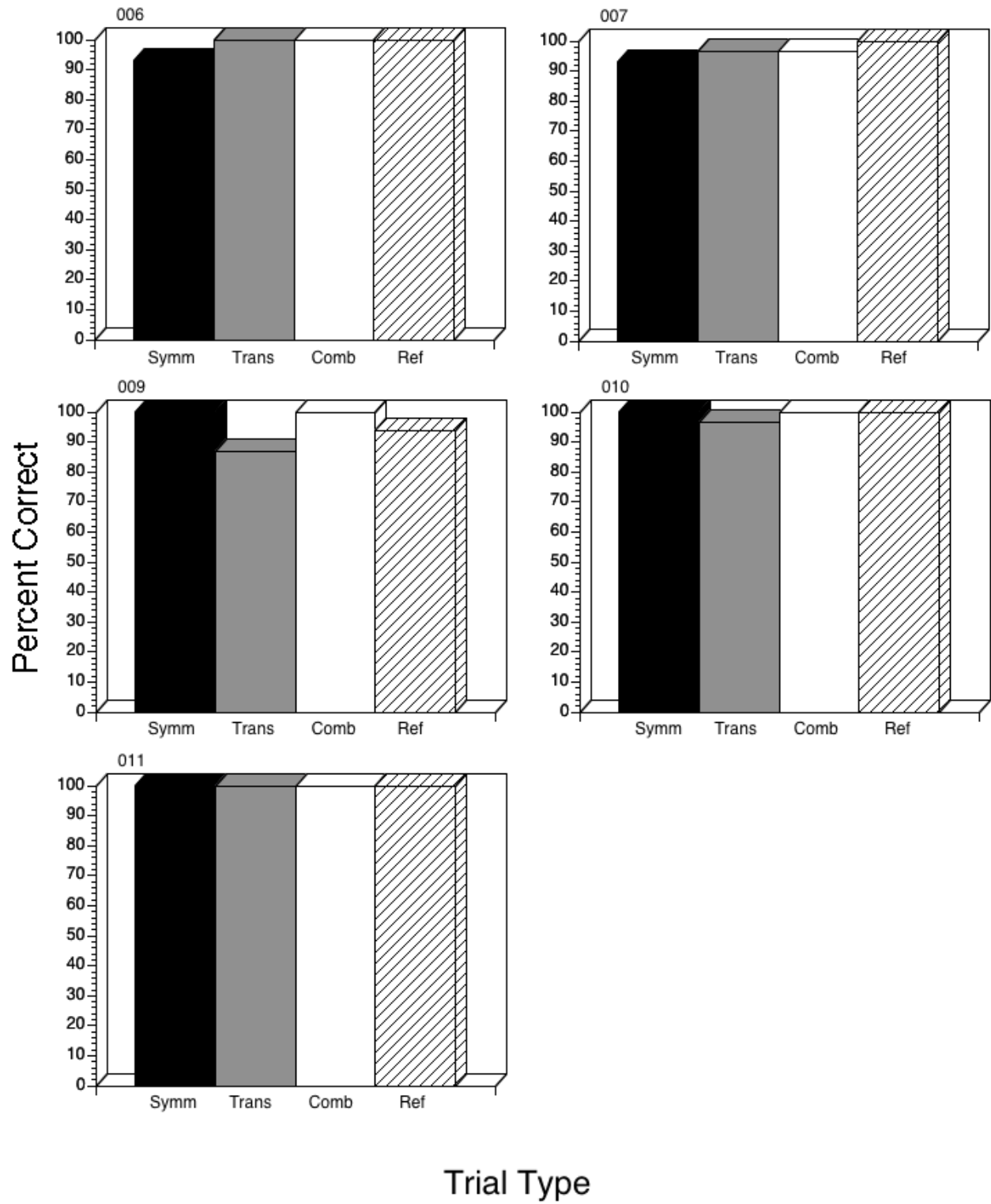


Figure 2

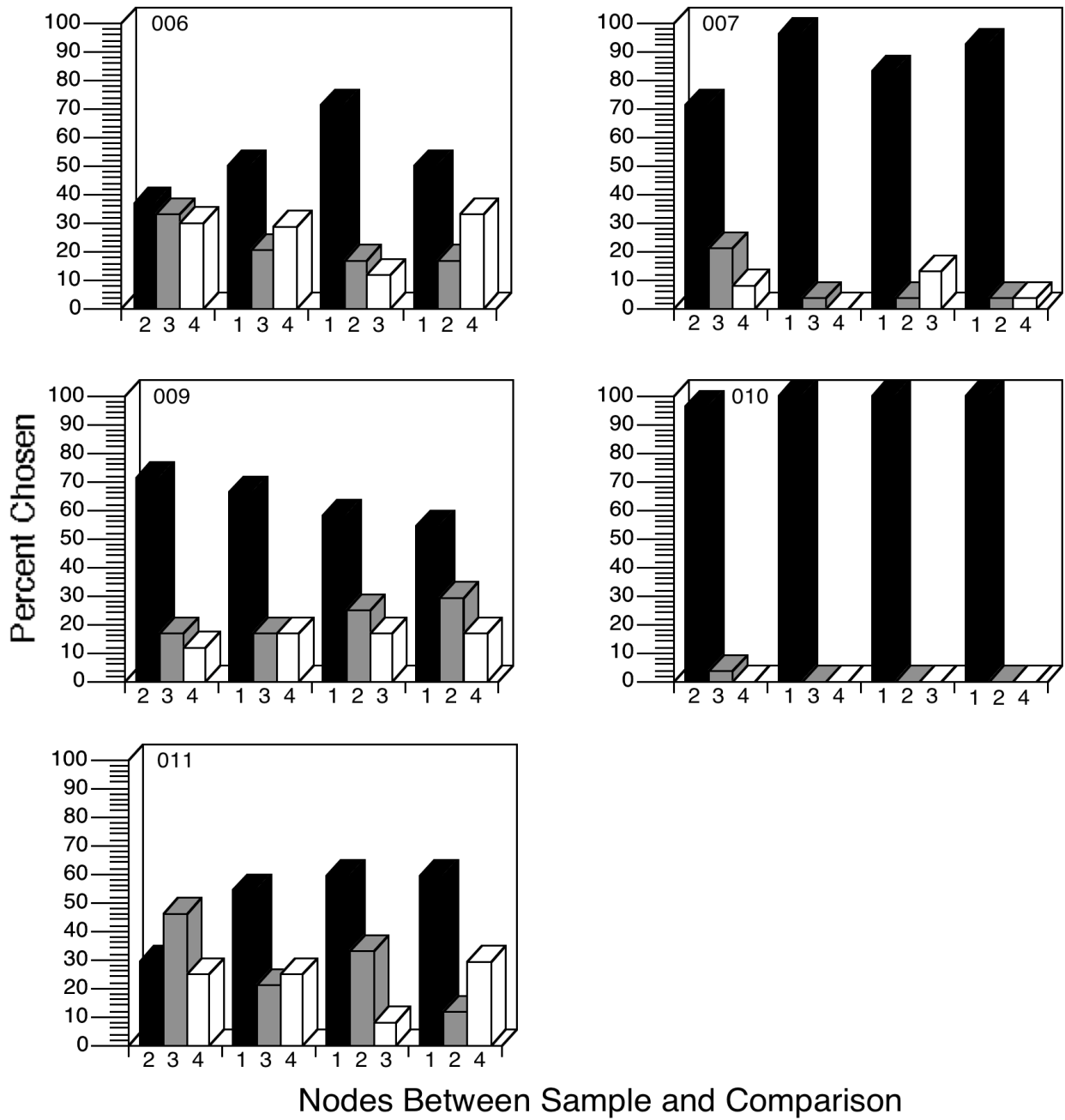
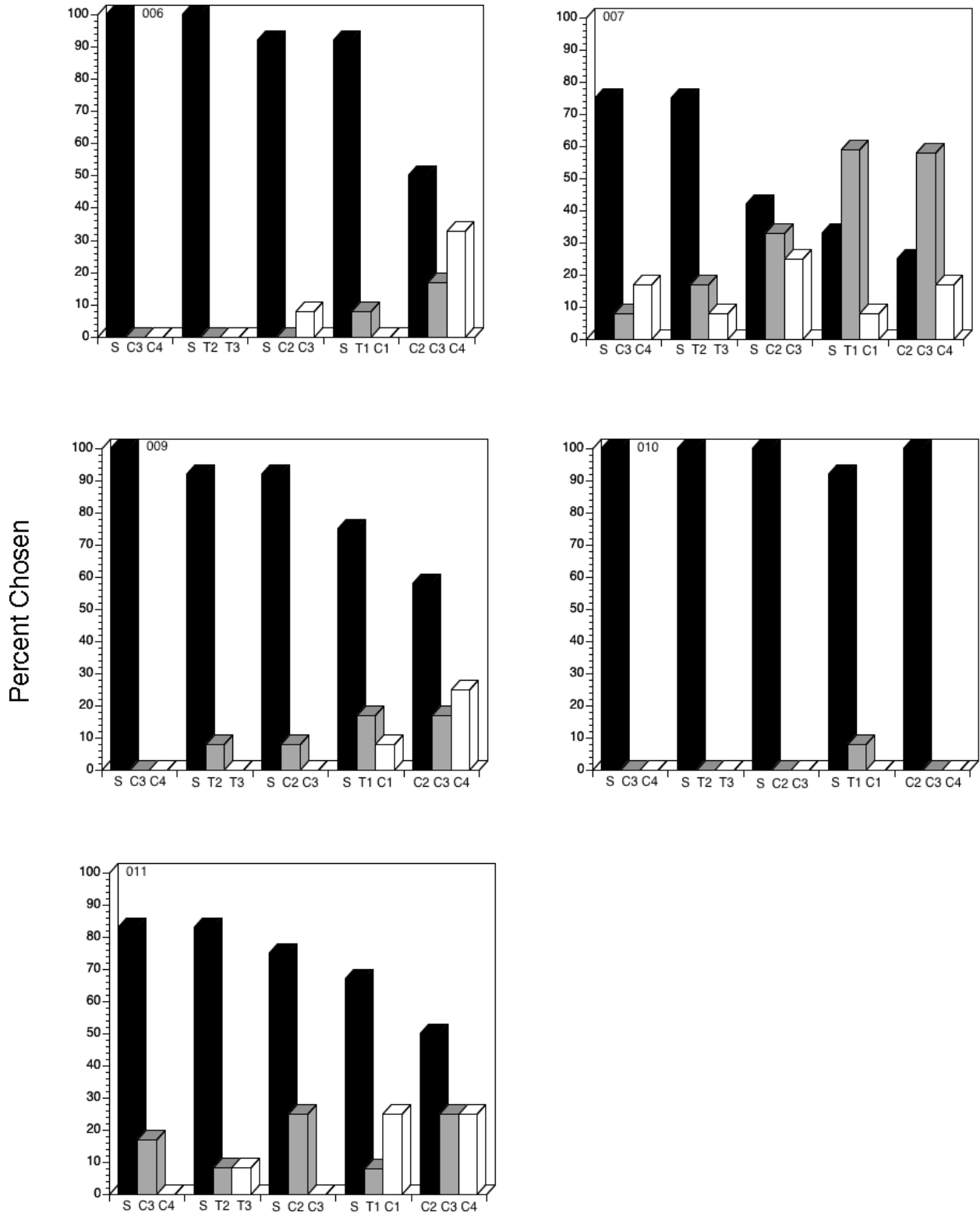


Figure 3



Relation Between Sample and Comparison

Figure 4

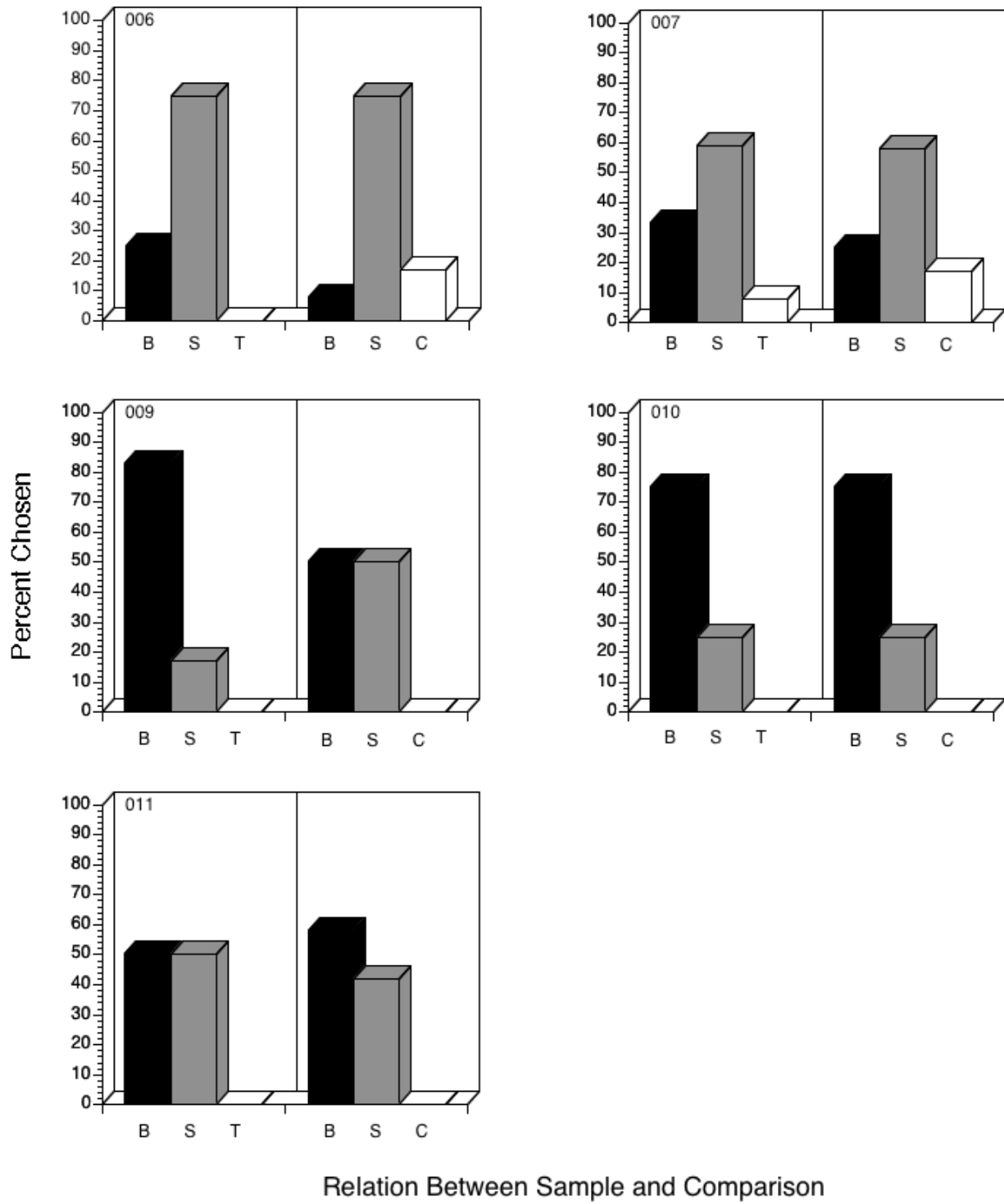


Figure 5

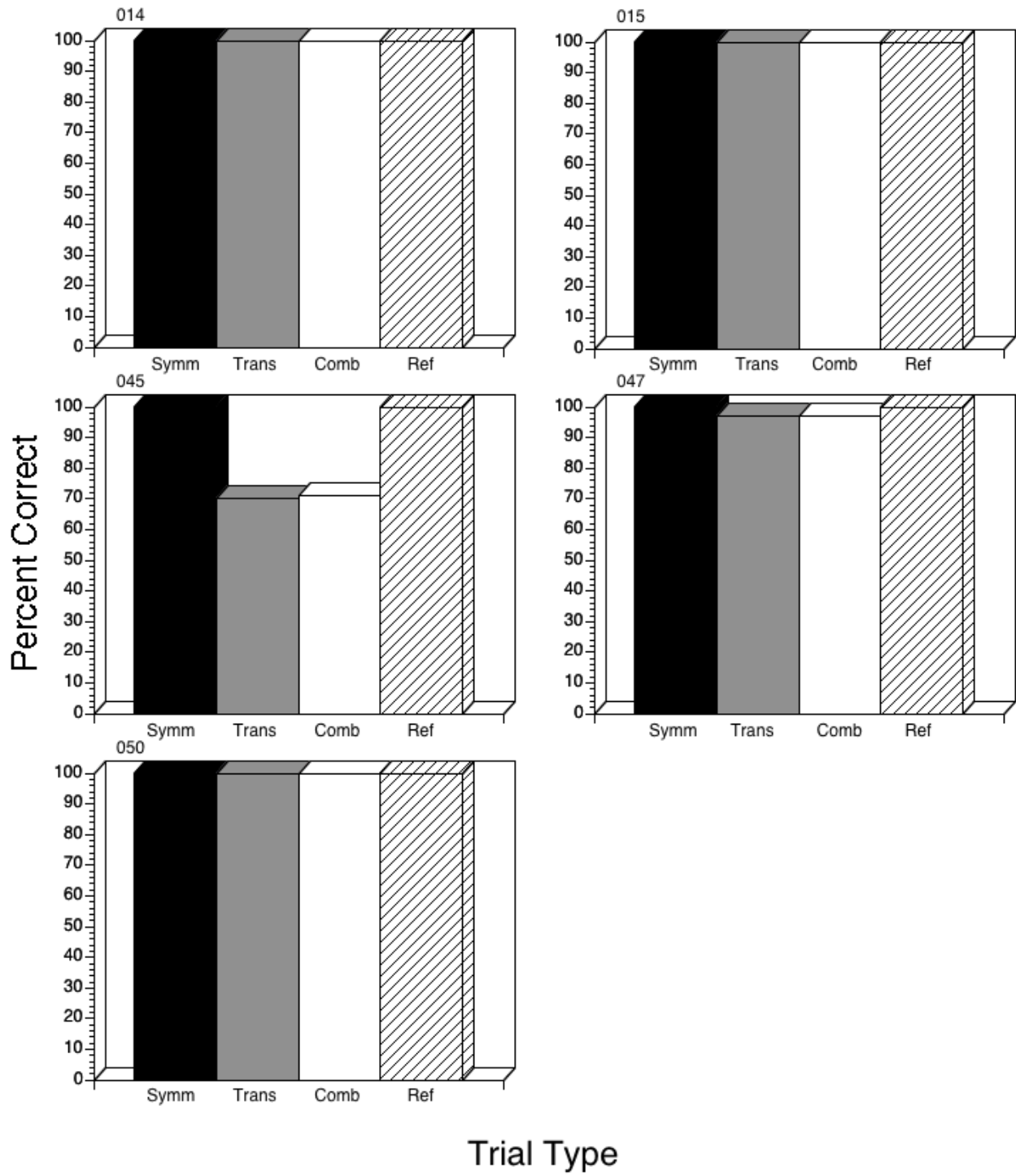
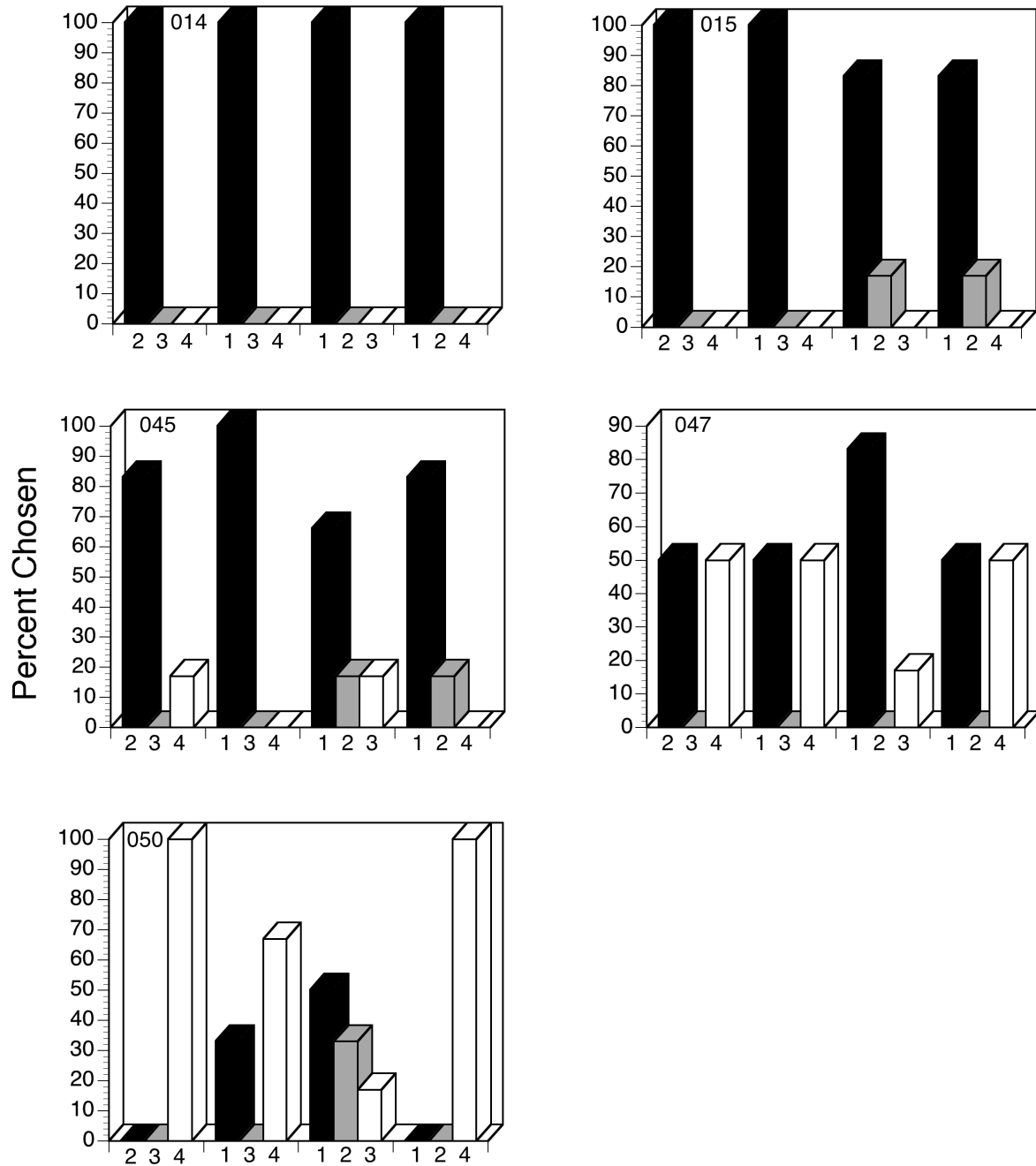
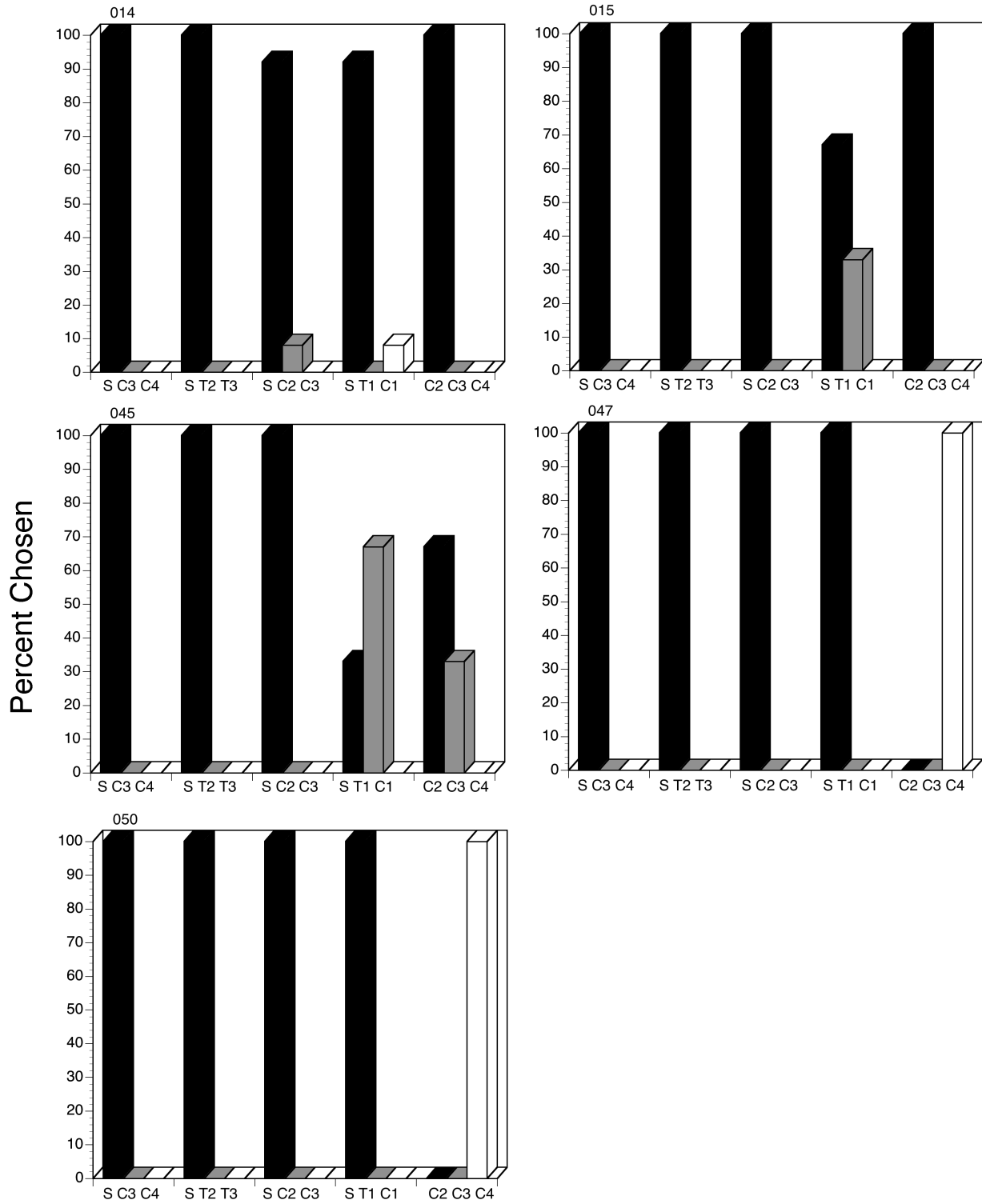


Figure 6



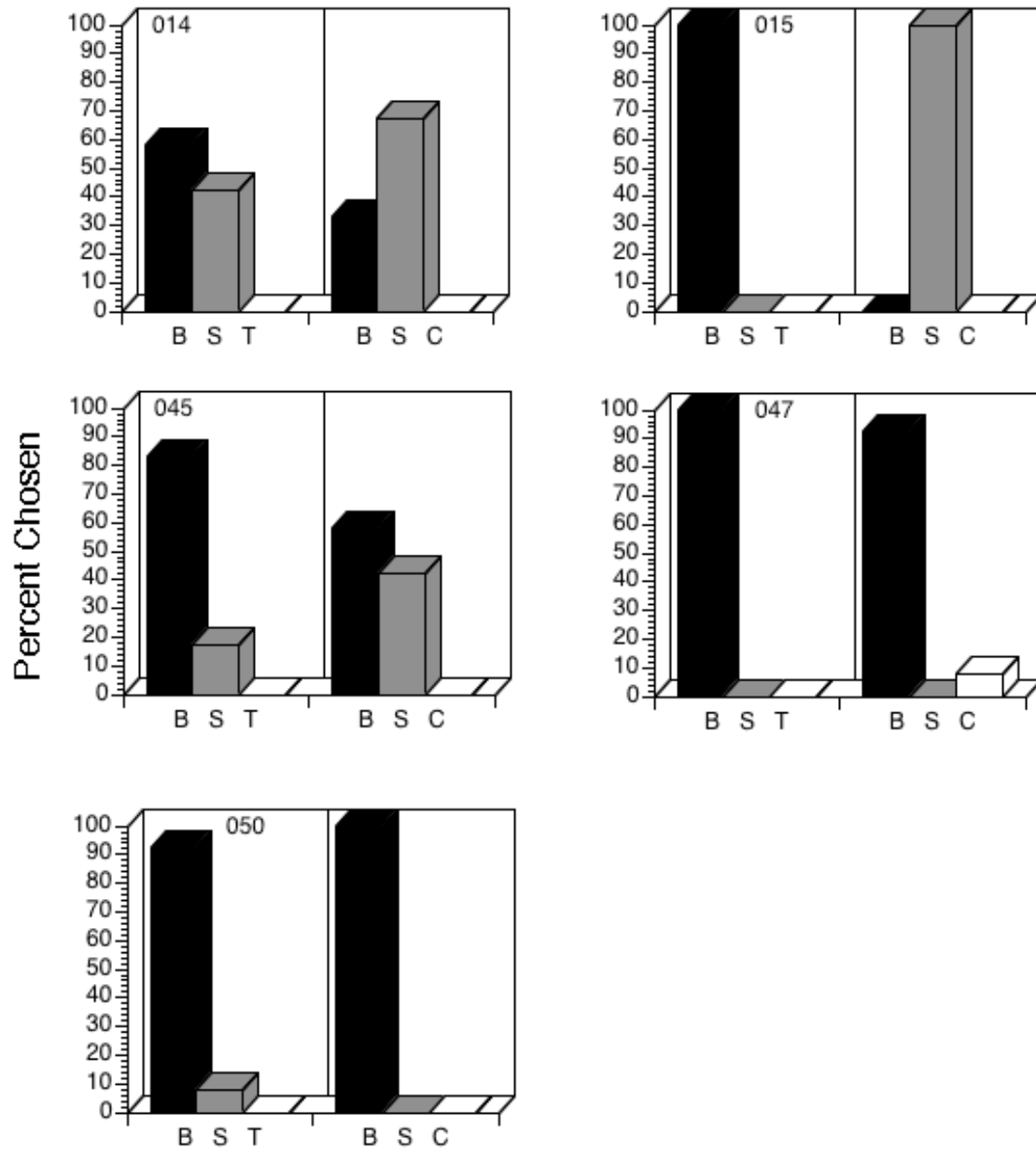
Nodes Between Sample and Comparison

Figure 7



Relation Between Sample and Comparison

Figure 8



Relation Between Sample and Comparison

Figure 9

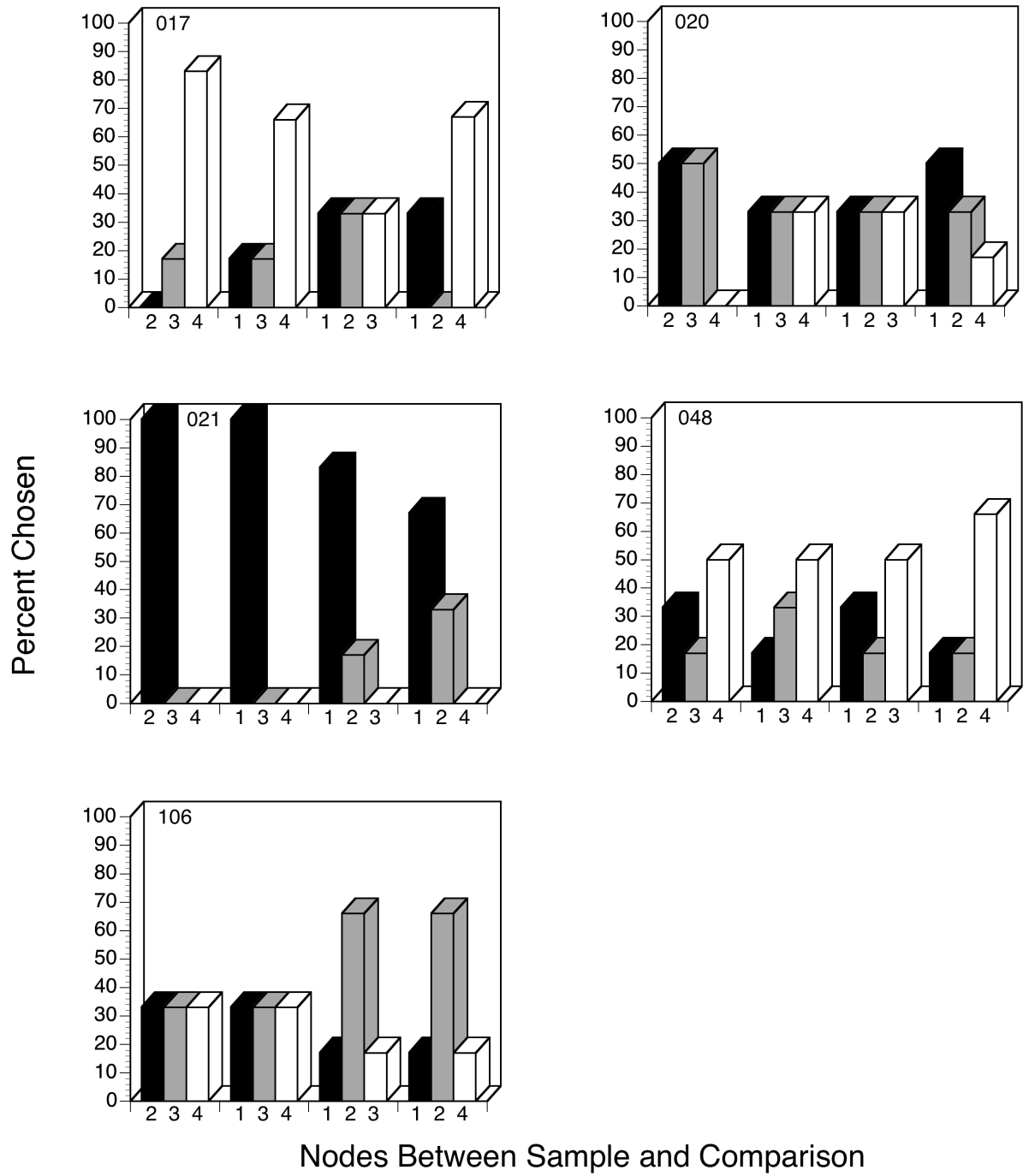
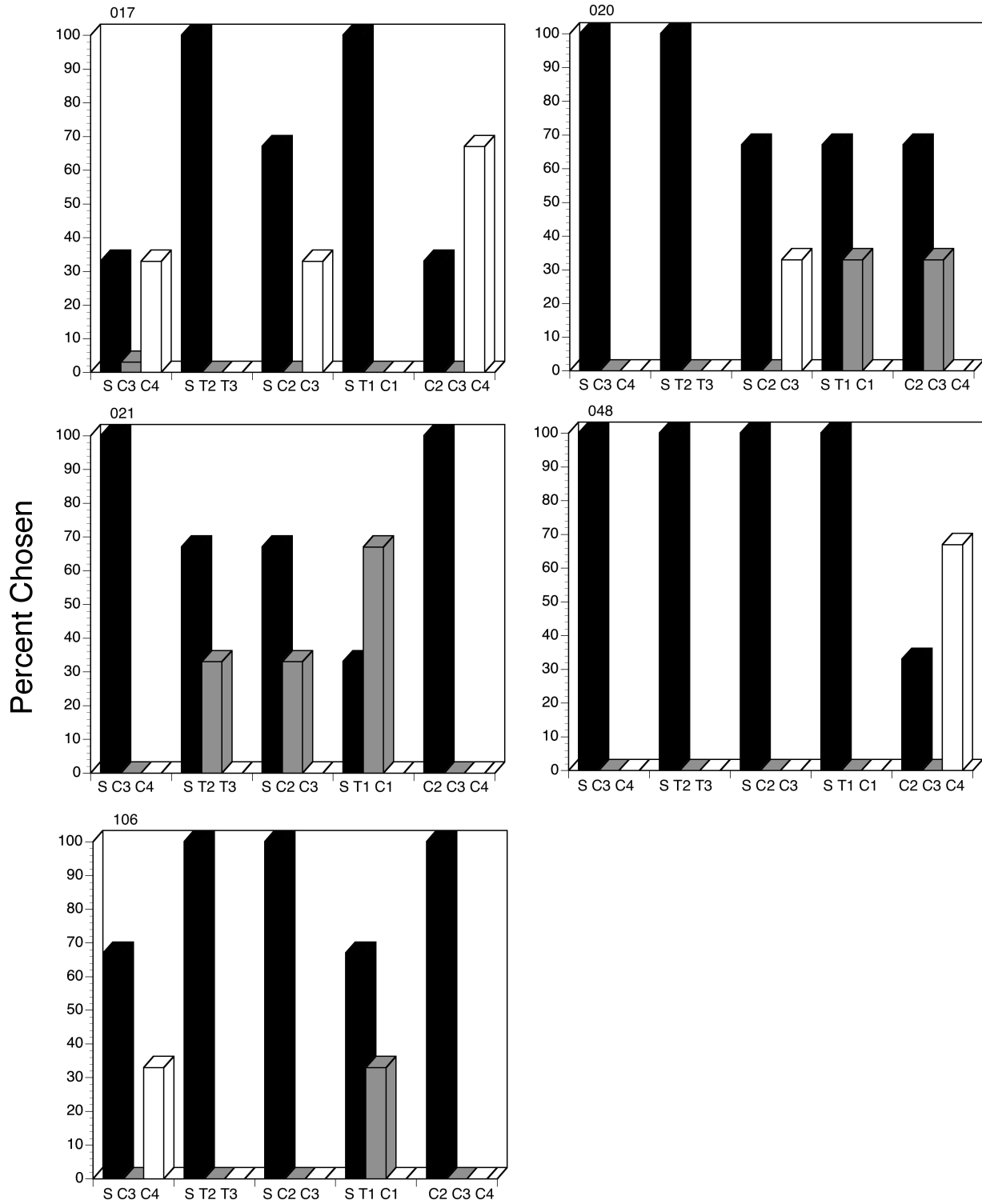
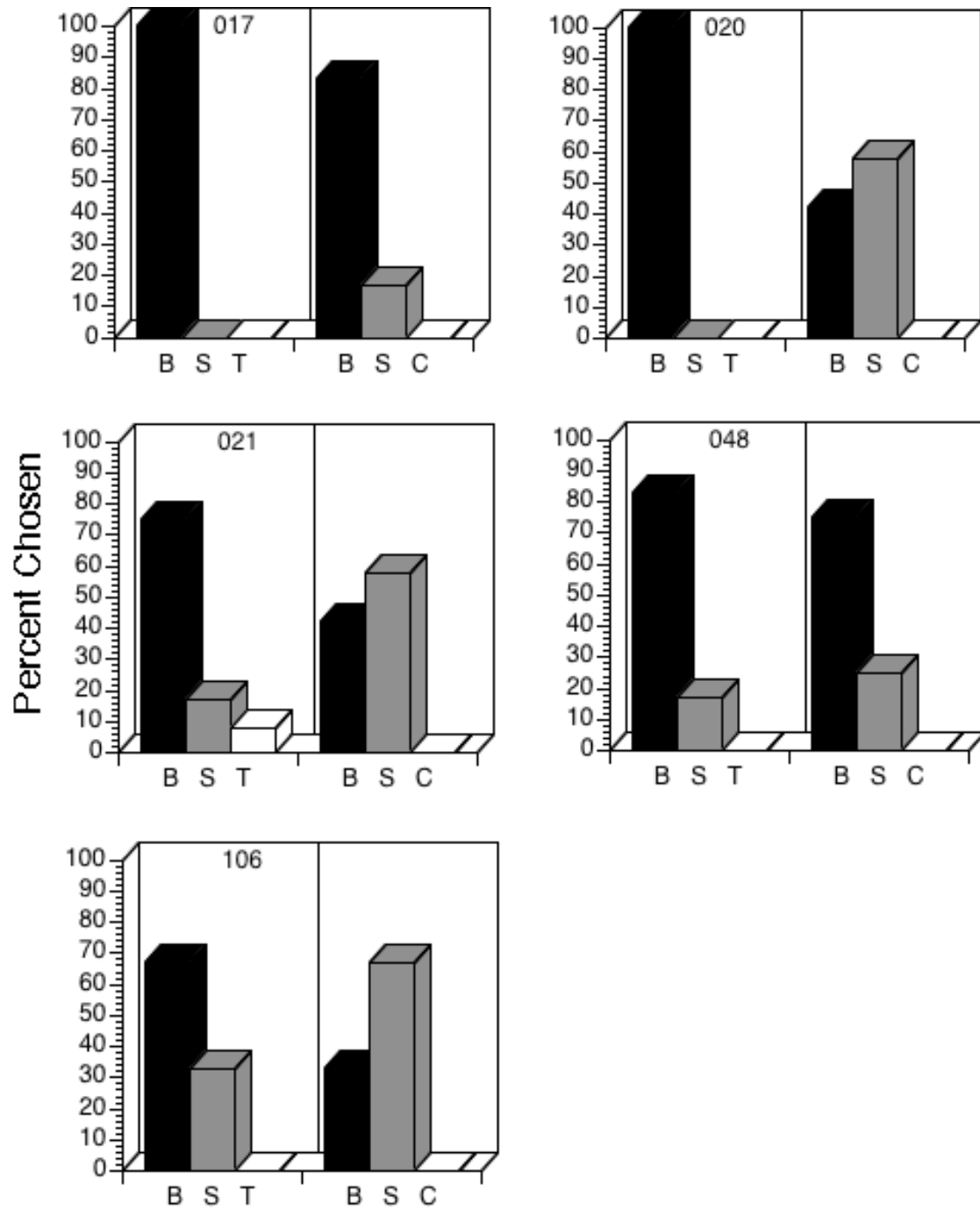


Figure 10



Relation Between Sample and Comparison

Figure 11



Relation Between Sample and Comparison

Figure 12

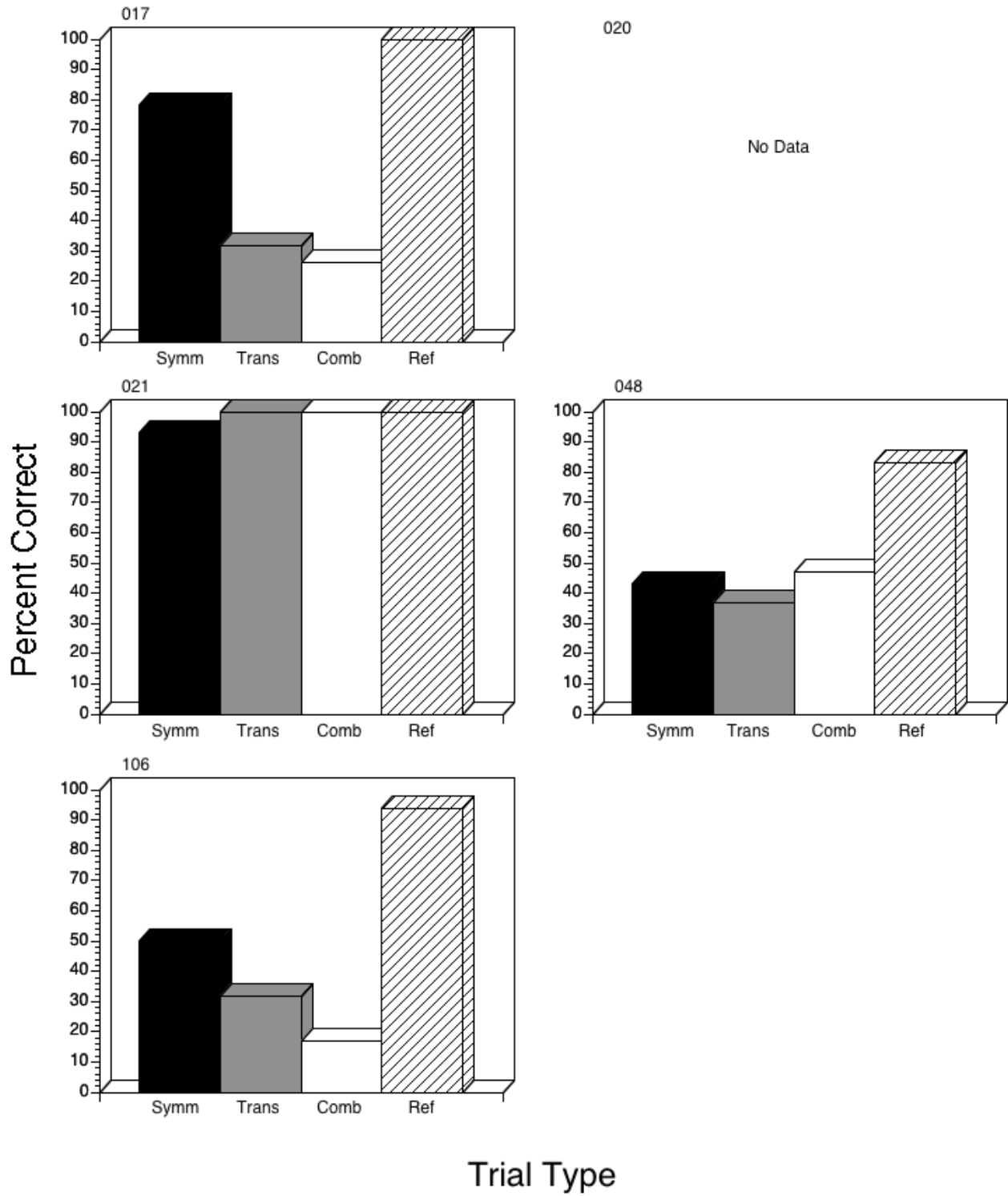


Figure 13

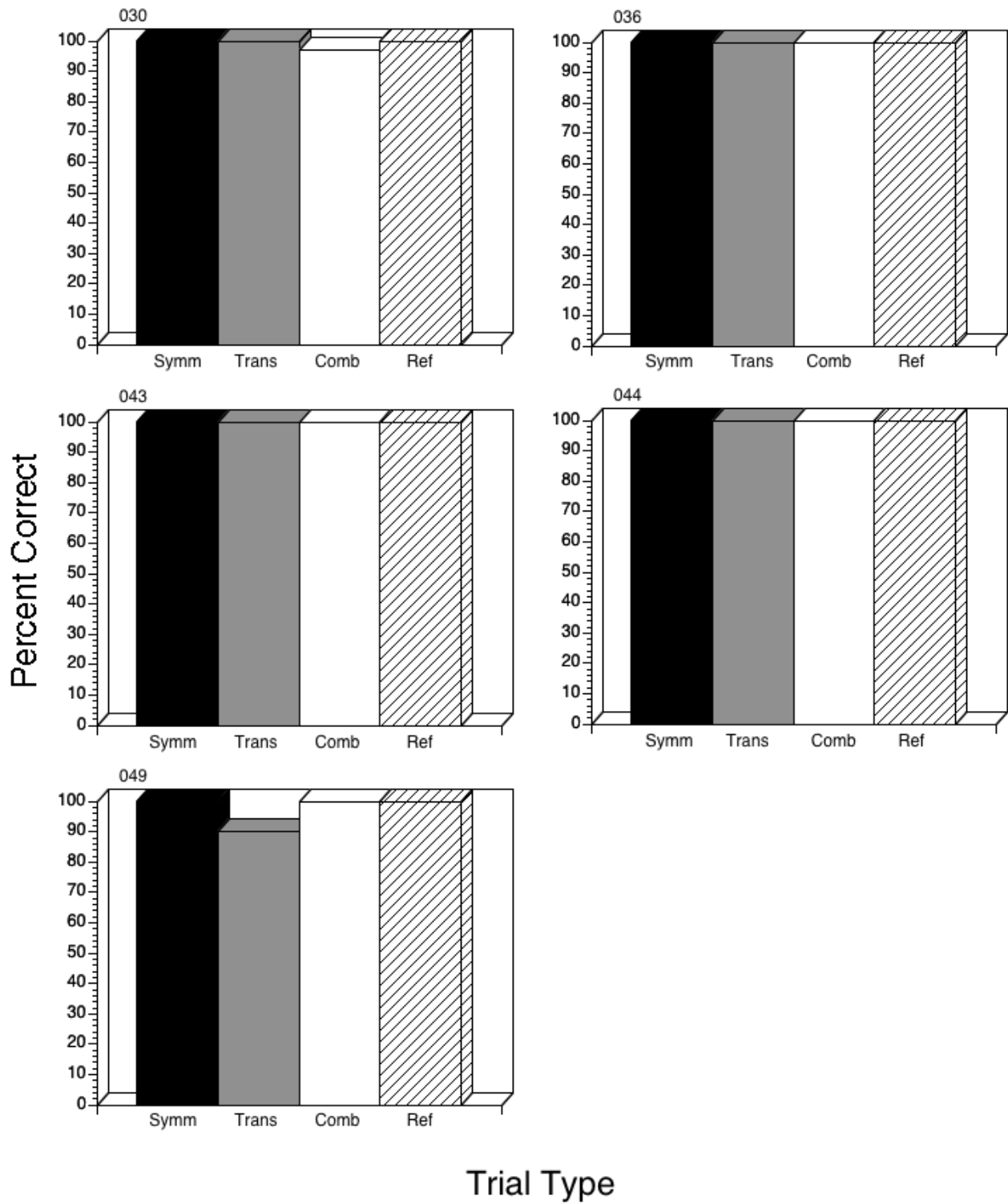


Figure 14

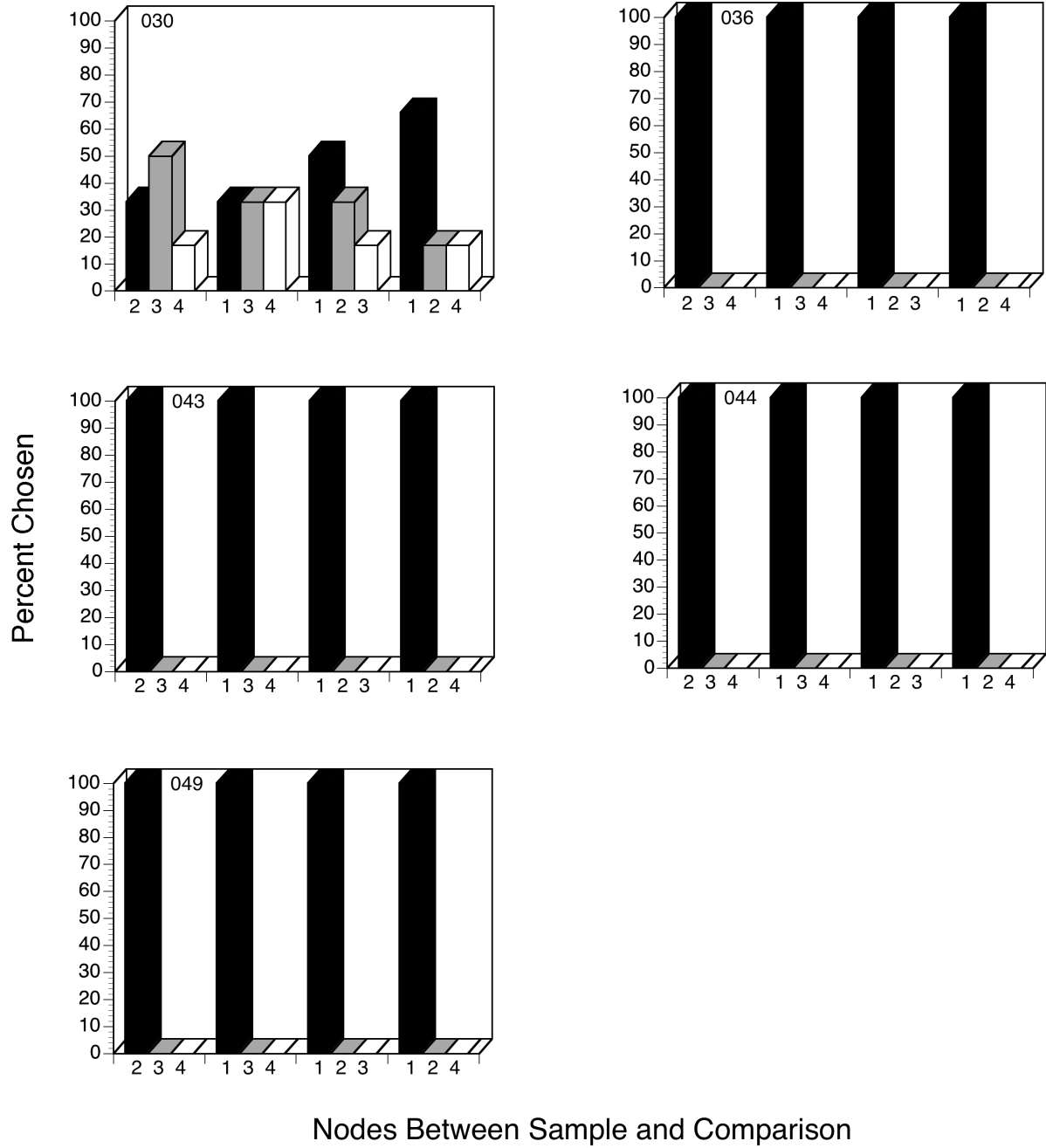
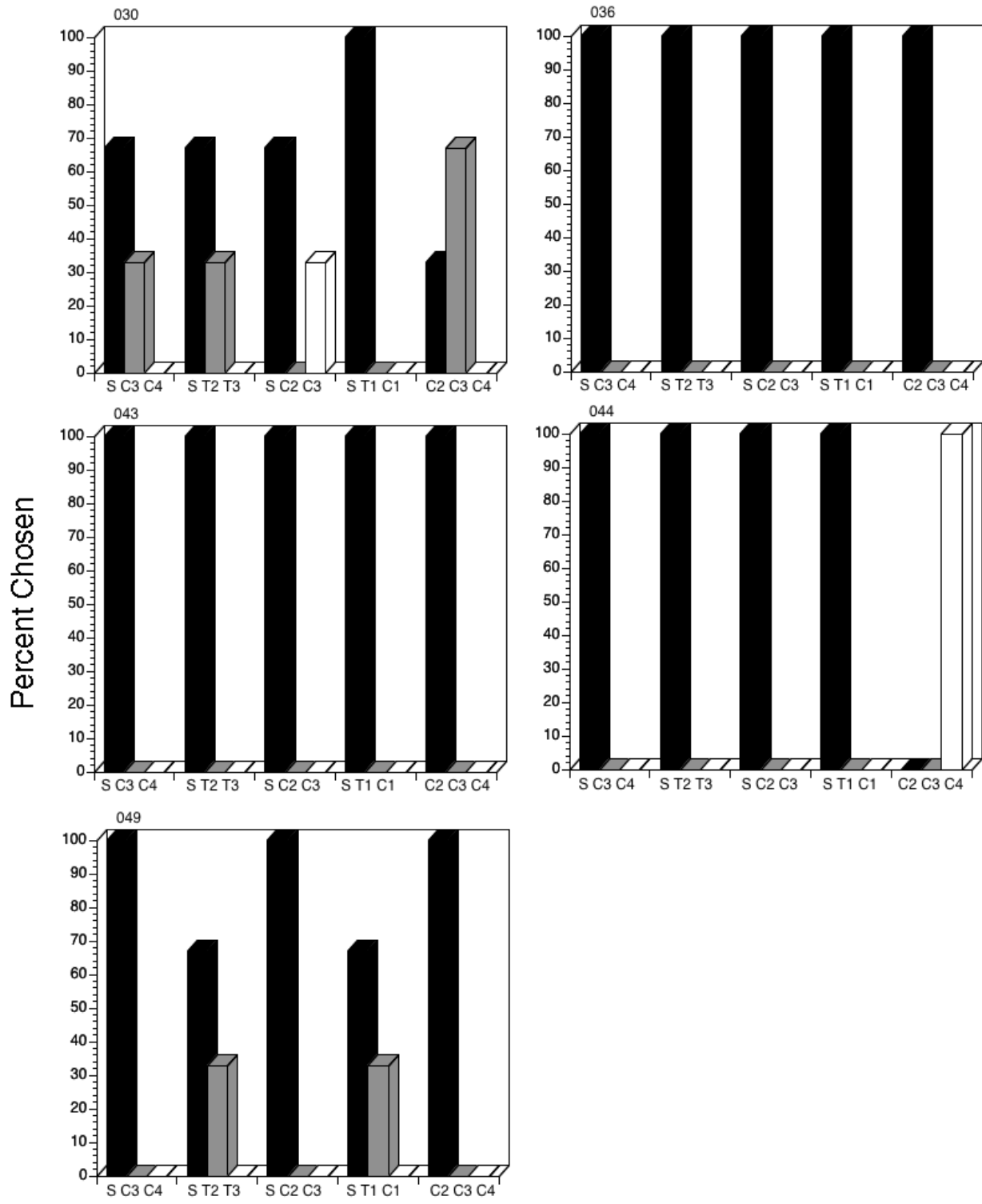
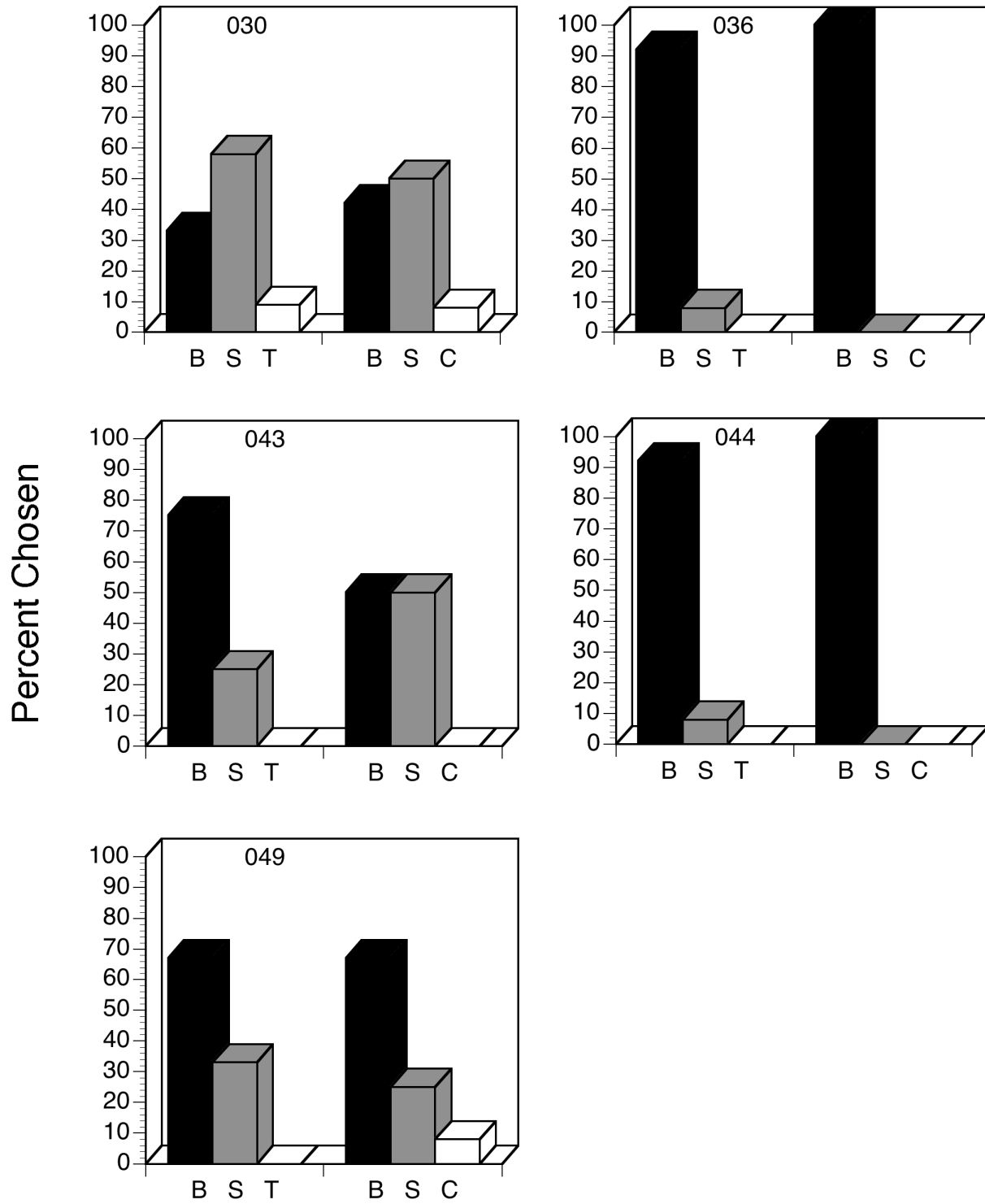


Figure 15



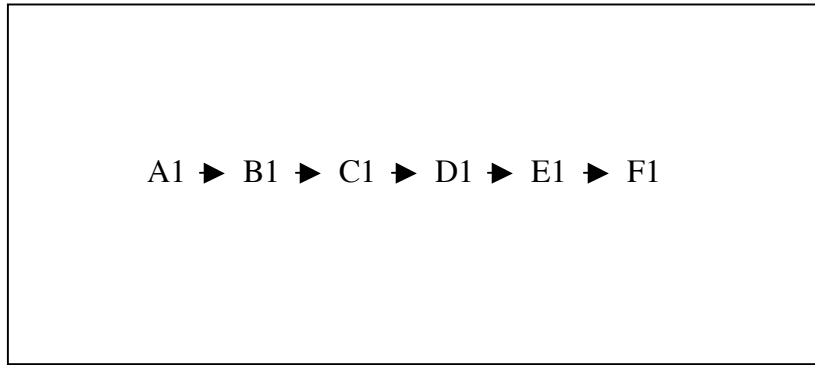
Relation Between Sample and Comparison

Figure 16

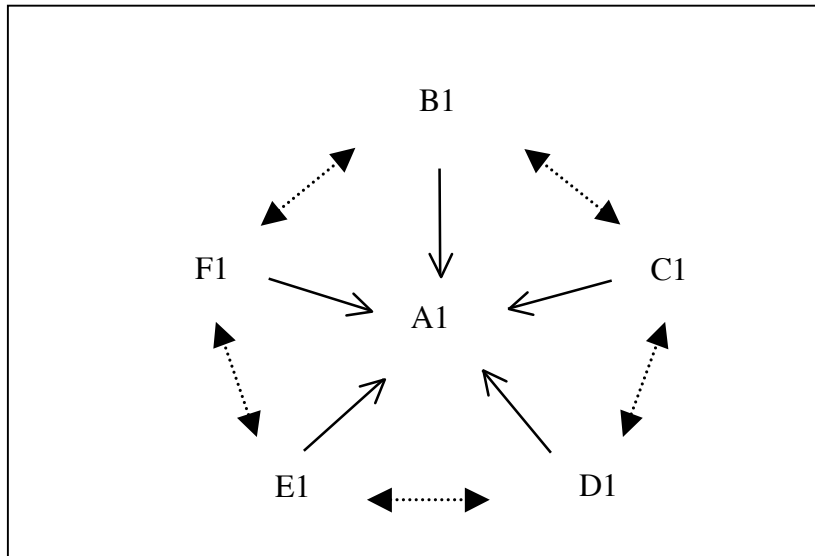


Relation Between Sample and Comparison

Figure 17



Linear Training



Sample-As-Node Training

Figure 18