EVALUATING EQUIVALENCE RELATIONS IN RATS USING AN OLFACTORY MATCHING-TO-SAMPLE PROCEDURE

L. Brooke Poerstel

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Master of Arts

Department of Psychology

University of North Carolina Wilmington

2007

Approved by

Advisory Committee

_____ _____

_____ _____
Chair

Accepted by

_____
Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

Equivalence classes can be characterized as groups of stimuli which control responding based on relations among members of the class, rather than absolute stimulus features such as shape, size, or color. Formation of equivalence classes often includes conditional discrimination training, which establishes contingency relations between physically dissimilar, arbitrary stimuli using a Match-To-Sample (MTS) paradigm. If, through the conditional discrimination training, the arbitrary stimuli become members of predetermined equivalence classes, then four untrained properties of equivalence classes (identity or reflexivity, symmetry, transitivity, and equivalency) should also emerge during test configurations. Unlike humans, evidence of responding characteristic of equivalence class formation in nonhuman animals is rarely, if ever, found. Regardless, classification of environmental stimuli based on abstract, relational features may be a fundamental aspect in learning and adaptation, as well as a possible indication of nonhuman symbolic behavior. The equivalence model also provides a parsimonious account of the often complex social and communicative behaviors observed in nonhuman animals in natural settings. Perhaps traditional laboratory equivalence procedures require modification such that these behaviors may be more readily observed in nonhuman subjects. The current paper evaluated whether rats could demonstrate (1) acquisition of conditional discriminations, both identity and arbitrary, within the training framework and (2) relational responding in the presence of novel testing configurations (emergence of generalized identity matching, symmetry, transitivity, and equivalence relations) through a modified MTS procedure using olfactory stimuli and class-specific reinforcers. Three experiments were conducted to evaluate generalized identity MTS (Experiment 1),

arbitrary MTS and emergent equivalence (Experiment 2), and training experience effects on emergent equivalence performance (extensive identity MTS pre-training versus no identity MTS pre-training, Experiment 3). Subjects, 11 Male HSD rats, were trained to retrieve reinforcers from cups of scented sand, which served as the stimuli throughout the experiment. Nine subjects were trained identity MTS discriminations during Experiment 1. Five of the nine demonstrated convincing evidence for generalized identity matching. The same five animals were then trained arbitrary MTS discriminations and given emergent equivalence tests during Experiment 2. Four of the five Experiment 2 subjects performed above chance during equivalence tests. Experiment 3 consisted of arbitrary MTS training with two naïve subjects that had no identity MTS experience. One Experiment 3 subject received an emergent equivalence test, but did not perform above chance levels. The results of the three experiments suggest that olfactory stimuli and class-specific reinforcers allow for transfer of responding during both generalized identity and emergent equivalence tests. The exact effect of pre-training on emergent equivalence performances is still unclear, as the results of Experiment 3 are currently inconclusive.

ACKNOWLEDGEMENTS

I sincerely thank Dr. Kate Bruce and Dr. Mark Galizio for providing me such a wonderful opportunity to learn about learning, about research, about myself, and about appreciating the influential people in your life. I cannot express how grateful I am to both of you for your guidance, your friendship and your patience and understanding. Truthfully, I am just honored to have spent time with you.

Many thanks to my committee members Dr. Ray Pitts and Dr. Kim Sawrey, who kindly endured two defenses, multiple presentations, and edited some lengthy documents. I appreciate all of the suggestions and input you have provided on this project and otherwise.

Although she has moved on, I also must thank Becky Rayburn-Reeves for all of her help and insight throughout the course of this project. I am so thankful that we shared the graduate school and laboratory experience together. I also want to thank Laura Bullard for her assistance and dedication to this project. Thanks as well to Kelly Weiland, who conducted much of Experiment 3, and to the lab members who are now carrying the torch.

And last, but absolutely not least, I have to thank my family and friends for tolerating countless hours of stimulus equivalence talk and still remaining supportive and understanding. I'm sure none of it made any sense, but you listened nonetheless. I certainly owe you all one.

DEDICATION

Although they can't read these symbols, I would like to dedicate this thesis to the rats that made it possible. Each animal taught me so much about animal behavior and learning and life in general.

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Categorization and Abstraction

Classification, as a behavior, can be characterized as differential responding occasioned by a given set of stimuli, which may or may not share physical or relational elements. Grouping stimuli encountered in an individuals' environment allows both human and nonhuman animals to efficiently manage the large numbers of external stimuli they regularly encounter (Wasserman, 1993). The ability to respond to environmental stimuli in such a fashion allows for allocation of energy and resources to other important survival processes, as these resources are not being consumed by the need to continuously interpret incoming stimuli (Schusterman & Kastak, 1993). Thus, 'sensitivity' to specific stimulus elements, and the resulting differential response to those stimuli which share certain elements, is potentially adaptive in nature.

With exposure to sufficient exemplars possessing a given element, category membership may generalize to novel stimuli that meet specifications for membership of that class. In other words, any novel stimulus which shares some element, whether physical or relational in nature, with an existing category of stimuli can be potentially incorporated through stimulus generalization. The novel stimulus becomes a member of said category when responding in its presence resembles behaviors occasioned by the other members of the category to which it was inducted ("transfer" of responding). The novel stimulus may be treated as a new member of the class without prior exposure (Vauclair, 2001), where responding comes under the control of the novel stimulus during the first experience. From ontogenic selection perspective, this type of responding may be crucial in order for an organism to adapt to fluctuating environmental conditions

(Schusterman & Kastak, 1993, 1998) and a wide variety of environmental stimuli. Responding to features of both the physical and relational aspects of stimuli may certainly confer an advantage during the lifetime of the organism. Schusterman and Kastak (1998) and Schusterman, Kastak, & Kastak (2003) suggest that such behaviors are potentially relevant for social and communicative situations in which nonhuman animals must be able to classify and identify individuals (e.g. kin, offspring, mates, competitors), recognize relationships among individuals (e.g. within dominance hierarchies, coalitions, and between mother and offspring), and respond appropriately to alarm and food calls. Behavior observed following equivalence class training and testing provides a parsimonious model for the complex social and communicative behaviors observed in natural settings (Schusterman et al., 2003). Thus, categorization of external stimuli may be an important aspect of learning for both humans and nonhumans.

Similar responses in the presence of a physically dissimilar set of stimuli often constitutes relational or functional categorization in which responding is controlled by common function or consequences. In this case, stimulus generalization and induction are dependent on the function of the novel stimulus, rather than its physical elements, and whether or not it matches the function of the existing category. As a result, stimuli within a given category often become substitutable for one another, and accordingly any one stimulus may be replaced with a fellow class member without observing any changes in responding. The common responses occasioned by all of the members of functionally-related classes are the elements that define what is typically referred to as "associative concepts" (Zentall, Galizio, & Critchfield, 2002), also termed equivalence classes. The formation of equivalence classes is often used to investigate the presence of "concepts"

or conceptual behavior in both humans and nonhumans within a behavioral framework. Conceptual behaviors are said to be derived from the functional relations between environment and behavior, in this case between stimuli within a given equivalence class. Various emergent relations (or transfer of relational responding) between members of the same class occur as a result of their shared functions. When physically dissimilar, arbitrary stimuli control the same responses based on shared function, and these relations emerge without explicit training (during tests of identity, symmetry, and transitivity), an equivalence class, or classes, are said to exist.

The Equivalence Relation

In 1982, Sidman and Tailby described how arbitrary, physically dissimilar stimuli can become members of an equivalence class and outlined specific procedures designed to evaluate stimulus equivalency. According to Sidman and Tailby, equivalence class membership requires that stimuli within a given class become replaceable and substitutable for one another and also function as conditional stimuli for members of that same class. In addition, Sidman and Tailby argued that these class members, if truly equivalent, must control responding such that each of the three properties of the equivalence relation, identity, symmetry, and transitivity, are observed. The importance of the reinforcement contingency (shared stimulus function) on equivalence class formation (e.g. Kastak, Schusterman, & Kastak, 2001; Meehan, 1999; Schusterman & Kastak, 1998; Sidman, Wynne, Maguire, & Barnes, 1989; Vaughan, 1988) was not explicitly included in the equivalence paradigm until Sidman (2000). This inclusion described various analytical units within the equivalence relation, including conditional and discriminative stimuli, as well as responses and reinforcers. In other words, the

entire equivalence relation not only includes stimulus replacement and substitutability and the three properties, but also the responses and reinforcers associated with each stimulus. Before the explicit inclusion of reinforcement contingencies into equivalence relations, the notion of shared consequences within classes, or amongst class members, was more commonly associated with functional classes (Sidman, 2000). Functional classes were defined as groups of arbitrary stimuli which share common functions, and this functional connectivity is the mechanism creating the classes and linking members together.

The inclusion of responses and reinforcement contingencies into the traditional definition of equivalence relations blurred any preexisting differences between equivalence and functional classes, as both forms of categories are controlled by the reinforcement contingency. However, although both functional and equivalence classes are derived from the reinforcement contingency, their characterization often differs along several important dimensions. The majority of these differences exist in procedural variations and the methods used to test class membership. Most notably, and unlike equivalence classes, establishment of functional classes is not typically created using match-to-sample procedures. As a result, it has been suggested that the stimuli within a functional class framework may not necessarily serve as conditional stimuli for one another, and are not necessarily interchangeable or substitutable because most investigations of functional class formation do not test for the three equivalence properties (Sidman, 2000), which typically exemplify equivalence class formation. Whether or not functional and equivalence classes are essentially interchangeable concepts remains to be established.

Again, the stimuli within an equivalence class do not share similar physical elements, but become members of an equivalence class and control common responses based on shared contingency relations (e.g. reinforcement). In a typical Sidman equivalence format, classes are established through match-to-sample training procedures and, as a result, the properties of identity, symmetry, and transitivity emerge untrained. The reinforcement contingency creates what we define as the equivalence class by linking the arbitrary stimuli through shared function so that the stimuli become equivalent for one another (Sidman, 2000). Stimulus equivalency is demonstrated when, following the training of certain discriminations, tests for emergent equivalence relations reveal responding indicative of the existence of the three equivalence properties.

Sidman and Tailby (1982) developed laboratory procedures which train the behavioral prerequisites for equivalence classes and test for equivalence class formation using the properties that define mathematical equivalence. Sidman, Rauzin, Lazar, Cunningham, Tailby, & Carrigan (1982), Sidman (2000), and Sidman & Tailby (1982) proposed that training of specific conditional discriminations using a match-to-sample paradigm links stimuli into equivalence classes based on common contingencies. When establishing an equivalence class, certain relations between class members are explicitly trained or reinforced. Since all equivalence class members are affected by the same variables, the influence of reinforcement history for the trained relations extends to novel stimuli and novel relations between class members (Sidman & Tailby, 1982). Traditionally, match-to-sample (MTS) procedures have been used to train a variety of conditional discriminations in order to test for equivalence class formation. In a MTS procedure, trials first include the presentation of a stimulus which subjects are required to

respond to in some fashion (e.g. key peck, nose press, lever press). This first stimulus is also known as the sample, or conditional, stimulus. Following responding to the sample stimulus, two or more discriminative or comparison stimuli appear - one correct (S+) and the other incorrect (S-). Conditional discriminations are defined when the functions (correct or incorrect) of comparison stimuli are conditional upon the sample stimulus so that responding to the comparison stimuli depends on the sample stimulus (Dube, McIlvane, & Green, 1992).

The conditional discriminations described by Sidman, et al. (1982), Sidman (2000), and Sidman & Tailby (1982) train various stimulus-stimulus relations and test for the three emergent relations characteristic of equivalence classes; identity, symmetry, and transitivity. Through MTS procedures, baseline discriminations are trained. Provided that these stimuli become members of the same equivalence class through conditional relation training, these three untrained equivalence properties or relations between class members should also emerge. If responding during baseline conditional discrimination training 'transfers' in the presence of emergent identity, symmetry, and transitivity test configurations, an equivalence relation is said to have formed between the stimuli (Dube et al., 1992). For this to be the case, responding must eventually come under relational stimulus control (and not absolute stimulus control) via the relation between the sample and comparison stimuli. If behavior is controlled by absolute properties of the stimuli, such as physical stimulus dimensions, the relational responding required for the emergent properties of equivalence were absent (Dube et al., 1992). The definition of equivalence therefore involves more than just accurate responding during baseline discriminations;

responding must also generalize during trails that include discriminations not previously experienced.

The first emergent property of equivalence classes is reflexivity (or identity): if A, then A, or for the conditional relation "r", ArA (Sidman & Tailby, 1982). When the emergent identity relation is tested following only arbitrary conditional discrimination training (e.g. ArB, ArC), as is common in human equivalence research, the term identity is typically applied. However, when the identity relation is trained through multiple exemplars, as is common in nonhuman equivalence research, it is known as generalized identity matching. Unlike arbitrary conditional discriminations, which train relations between physically dissimilar stimuli (and theoretically create derived stimulus relations, symmetry and transitivity), discriminations which set up the conditions necessary for identity relations train subjects to respond to sample and comparison stimuli which are physically similar. As a result, the identity relation has often been conceptualized as a separate phenomenon from the other equivalence relations, in the sense that responding is controlled by the physical dimensions of the stimuli (choose the comparison that matches the sample stimulus). Nevertheless, identity matching still requires that the relation between sample and comparison stimuli controls responding, since the reinforced choice depends on the sample identity, but this occurs in a non-arbitrary fashion.

In Sidman and Tailby (1982) first, and others that followed, MTS procedures were used to train baselines and test for emergent identity. A typical identity matching trial would consist of the presentation of a sample stimulus (A) to which the subject must respond (e.g. key peck, lever press). Upon responding to the sample, two or more comparison stimuli appear. Reinforcement is delivered following a response to the

comparison stimulus, which is identical to the sample stimulus (A), and not delivered following responding to the dissimilar comparison (B). For example, given the sample stimulus shape "star", responses to its identical comparison stimulus "star" are reinforced, and responses to the physically dissimilar comparison stimulus "circle" are not.

Using a multiple exemplar format, the trained discriminations (e.g. A to A and not B, or star to star and not circle) serve as a baseline by which novel discriminations may be tested for generalization of responding - generalized identity matching. Tests for generalization of identity matching consist of the presentation of novel sample and comparison stimuli, XrX (e.g. X and Y, or "arrow" and "diamond"). For example, based on the history of reinforcement for responses to the physically similar comparison during identity matching training, given an arrow shaped sample stimulus (X), subjects should choose the identical arrow comparison (X) stimulus, and not the diamond shaped comparison (Y). During these tests, using novel stimuli in both sample and comparison positions (X and Y), rather than familiar comparisons (A or B), controls for responding based on exclusion of the familiar, defined stimulus (A or B). Holding all else constant, if both sample and comparison stimuli are novel during generalized identity matching testing, behavior must come under control of the identity relation (between sample and comparison) in order to generalize matching performance to the new stimuli (Dube et al., 1992; Peña, Pitts, and Galizio, 2006).

The second emergent property of equivalence classes is the symmetry relation; if A, then B, so if B, then A, or for the conditional relation "r", ArB; BrA (Sidman & Tailby, 1982). Arbitrary, physically dissimilar, conditional MTS procedures are used to

train baseline discriminations and test for the emergent symmetrical relation. Using arbitrary conditional discriminations in a MTS format, direct matching training occurs for the relation "if A, then B", or ArB. Following training of the ArB discrimination, subjects are then tested for the emergent symmetrical relation "so if B, then A", or BrA. During arbitrary MTS, the functions of the comparison stimuli (S+ or S-) are still conditional upon the sample stimulus, however, the sample stimulus and correct (S+) comparison are now physically dissimilar, or arbitrary. The sample stimulus and correct comparison (S+) are designated as members of the same stimulus class, as denoted by subscript numerals. A conditional discrimination training trial consists of the presentation of a sample stimulus (A_1) to which the subject responds. Upon responding to the sample, two or more comparison stimuli appear (B_1 and B_2). A response to the comparison stimulus (B_1), which is different from (but within the same designated class as) the sample stimulus (A_1) will result in reinforcer delivery, whereas responding to the incorrect comparison from a different class (B_2) will not. For example, given the sample stimulus shape "star", subjects are trained to choose its designated class member "crescent", and not the comparison stimulus "cross" from the other class. Responses to the comparison stimulus "star" (S+) were reinforced while responses to the comparison stimulus "cross" will not be reinforced.

Symmetrical responding demonstrates the functional substitutability of sample and comparison stimulus positions (Sidman et al., 1982). Evidence that the trained conditional discriminations (e.g. A_1 to B_1 and not B_2, or star to crescent and not cross) are also symmetrical relations occurs when relations in the opposite direction (e.g. B_1 to A_1 and not A_2) are demonstrated without further training. For example, given a star shaped

sample stimulus (A_1), subjects are trained choose the designated class member shape "crescent" (B_1), and not the cross shape (B_2). If the trained conditional discrimination is also symmetrical, then, given "crescent" (B_1) as the sample, subjects should choose the star shape (A_1) and not the cross (A_2). Behavior must come under control of the relational properties of the stimuli within the designated class in order to demonstrate the symmetrical relation between arbitrary stimuli.

The third emergent property of equivalence classes is the transitive relation; if A, then B, if B, then C, so if A, then C, or for the conditional relation "r", ArB, BrC; ArC (Sidman & Tailby, 1982). The transitive relation requires training of two arbitrary conditional discriminations, ArB and BrC, where the comparison stimulus B in the first trained relation serves as the sample stimulus in the second trained relation. In this way, the stimulus B serves as the nodal stimulus, theoretically linking stimuli A and C into a transitive relation based on shared function. If the trained conditional discriminations are also transitive relations, emergent responding will occur for the untrained relation ArC via mutual associations with nodal stimulus B. Using arbitrary conditional discriminations in a MTS format, direct training occurs for the relation "if A, then B", or ArB, and "if B, then C", or BrC. Following training of the ArB relation, subjects are trained a second conditional relation, BrC, and then tested for the emergent transitive relation "so if A, then C", or ArC. The functions of the comparison stimuli (S+ or S-) are again conditional upon the sample stimulus and the sample stimulus and correct (S+) comparison are physically dissimilar, or arbitrary. The sample stimulus and correct comparison (S+) are members of the same designated stimulus class, denoted by subscript numerals. A conditional discrimination training trial consists of the presentation

of a sample stimulus (A_1) to which the subject responds. Upon responding to the sample, two or more comparison stimuli appear (B_1 and B_2). Reinforcement is delivered following a response to the comparison stimulus B_1, which is different from (but within the same class as) the sample stimulus A_1. Again, no reinforcement is delivered for responding to the incorrect comparison from a different class (B_2). For example, given the sample stimulus shape "star", subjects are trained to choose its fellow class member "crescent", and not the comparison stimulus "cross" from the other class. Responses to the comparison stimulus "crescent" were reinforced (S+) while responses to the comparison stimulus "cross" from a different class will not be reinforced. A second conditional discrimination is then trained between the sample stimulus B_1 and correct comparison C_1, in the presence of the incorrect comparison, C_2. For example, given the sample stimulus shape "crescent", responses to "square" were reinforced, and responses to "triangle" will not.

Transitive responding involves an untrained conditional relation between the sample stimulus from the first trained relation (A_1) and the comparison stimulus from the second trained relation (C_1). Evidence that the trained conditional discriminations (e.g. A_1 to B_1 and B_1 to C_1, or star to crescent and crescent to square) are also transitive relations occurs when a relation between the sample from the first trained relation and the comparison from the second trained relation (e.g. A_1 to C_1 and not C_2) is demonstrated without further training. For example, given a star shaped sample stimulus (A_1), reinforcement occurs for choosing the fellow class member crescent (B_1), and not the cross shape (B_2). Then, given the crescent (B_1) as the sample, reinforcement occurs for choosing the square shaped comparison stimulus (C_1), and not the triangle (C_2). If these

two trained relations are transitive, responding to the square comparison (C_1) will occur in the presence of the star sample stimulus (A_1) without explicit training. Subjects' behavior must come under control of the relation between the stimuli in each designated class in order to demonstrate the transitive relation during testing configurations. Finally, the relation between sample stimulus C and comparison stimulus A, which may require symmetrical and transitive relational responding, may also be evaluated following the conditional discriminations trained prior to transitivity testing (e.g. given C_1 as the sample, choose A_1 and not A_2, or given square, choose star and not cross).

Functional Classes versus Equivalence Classes

Failures to demonstrate strong evidence of equivalence in nonhumans using conditional discrimination training in match-to-sample procedures have led researchers to investigate equivalence from a different procedural angle, specifically, the training of functional classes through repeated reversals using simple discriminations (Dube & McIlvane, 1993; Kastak et al., 2001; Schusterman & Kastak, 1998; Vaughan, 1988).

Although a complete comparison of functional and equivalence classes is beyond the scope of the current discussion, and potentially not crucial, a brief example serves to clarify some basic procedural differences and address current research approaches. Kastak et al. (2001) first explicitly trained two functional classes with arbitrary visual stimuli using simple discrimination, repeated reversal procedures developed by Vaughan (1988). Functional classes were trained through repeated simple discriminations in which the contingencies for each class of stimuli initially alternated between S+ and S- across sessions. So, in a given session, all class 1 stimuli would be correct (S+) and all class 2 stimuli would be incorrect (S-). Following a contingency reversal, all responses to class 2

stimuli would now be reinforced, while responses to class 1 stimuli would not. Since contingency switches were not signaled to the subjects prior to session initiation, the first few exposures to stimuli from both classes during that session determined which class was correct or incorrect that day. The goal of repeated reversal training was this: following some number of contingency reversals and given that classes had formed, subjects should transfer responding to all members of the correct class (S+) for that session following exposure to the first few members of that class. In other words, based on the function (S+ or S-) of a given class member or members during that first exposure, all other class members should control the same function and subsequent responding to remaining class members should follow suit.

Through studies such as that of Kastak et al. (2001), the lines between functional and equivalence classes have been blurred, as shown by Sidman's (2000) reconsideration of equivalence classes and the procedures necessary to establish them. However, as Sidman (1998, 2000) notes, many studies lack a complete analysis of all three equivalence properties. Since the transfer of responding observed following functional class procedures is often not evaluated in an equivalence relation format, the distinction between functional and equivalence classes remains. Whether there exists important functional differences between functional and equivalence classes, above and beyond procedural differences, is not clear. A complete demonstration of the emergent equivalence properties following functional class training seems appropriate.

Class-specific Reinforcement

If MTS procedures occasionally, but not always or never, produce demonstrations of equivalence transfers, what are the important variables in those situations where it was

observed? What modifications to MTS will increase acquisition of these relations in nonhumans? To start, differentiating the reinforcement contingencies for each equivalence class may create more definitive separation of classes. Sidman (2000) specifically stated that the reinforcement contingency is what links each unit of the greater equivalence relation together. This contingency is the essential variable in the formation of equivalence classes and the common reinforcer, or shared consequence, creates the equivalence classes between groups of arbitrary stimuli. He argued that each element of the entire equivalence relation (conditional stimulus, discriminative stimulus, response, and reinforcer) is important for the formation of that relation and the equivalence class that arises through conditional discrimination training. So, if both classes share a common response and/or reinforcer, these must drop out of the entire relation before equivalence classes can be formed based on the conditional and discriminative stimuli alone. In humans, equivalence classes have been demonstrated without the use of differential responses and reinforcers (Harrison & Green, 1990). Yet, most nonhuman animals have been unable to show similar performance. Since each element of the relation is an important feature for class differentiation, Sidman suggested that using class-specific responses and/or reinforcers will greatly increase the likelihood that classes will form.

The use of differential, or class-specific, reinforcement to establish arbitrary classes was elegantly demonstrated with nonhumans in the same study by Kastak et al. (2001). Although training procedures differed from the conditional discrimination methodology outlined by Sidman (2000) and Sidman & Tailby (1982), the two California sea lion subjects (*Zalophus californianus*) only showed evidence of class formation after

functional class training with two types of fish reinforcers differentially associated with each set. At the beginning, the two fish reinforcers were randomized across both sets so that reinforcers were not specifically associated with either class. During non-specific reinforcement phases, the sea lions failed to show evidence of functional class formation during sessions in which contingencies were reversed. So at this point, responding did not transfer to members of a given class based on the function of the first stimulus. Only after reinforcers were later made class-specific did the sea lions exhibit behaviors indicative of functional classes. Presumably, each fish was differentially associated with one of the two classes and thus, became a class member as well. The class-specific reinforcers may have helped to distinguish the two classes from one another. Additionally, the sea lions later extended their functional class training to conditional discriminations characteristic of Sidman-style equivalence classes.

Equivalence in Nonhuman Animals

Several studies have demonstrated the formation of equivalence classes and conceptual behavior in humans (Sidman et al., 1982; Sidman & Tailby, 1982). However, there has been considerable debate about whether research conducted with nonhuman animals actually meets criteria for conceptual categorization behaviors (Dube & McIlvane, 1993), especially compared to similar studies using human participants. Some research suggests that nonhumans cannot form equivalence classes in the same manner as humans, and therefore do not appear to use symbols or conceptualize in the same manner either. In comparison, some studies have shown at least partial evidence for equivalence-like responding in nonhumans (D'Amato, Salmon, Loukas, & Tomie, 1985; Kastak et al., 2001; Schusterman & Kastak, 1993; Vaughan, 1988). Most of these investigations have

used procedures that often vary from those used with humans. As a result of the procedural variations across human and nonhuman research, much of the nonhuman equivalence data have been disregarded as evidence for conceptual behavior. And overall, most of the original questions comparing nonhumans and the use of symbols remain largely unanswered (e.g. How do nonhumans compare with humans in terms of conceptual behavior? Are nonhuman animals capable of demonstrating evidence of symbol use through the formation of equivalence classes?). It is argued here that perhaps investigations should focus on the important, relevant features of equivalence classes for different species rather than focusing on a direct human versus nonhuman comparisons. Thus, a more appropriate research question might ask: under what conditions, if any, are these classes formed for the particular species in question?

In 1982, Sidman et al. conducted pioneering research comparing conceptual behavior in nonhuman primates, specifically rhesus monkeys (*Macaca mulatta*), olive baboons (*Papio anubis*), and human children, using equivalence-class training and testing. The study consisted of five consecutive experiments, allowing the researchers to assess the current experimental phase and then troubleshoot during the subsequent phase. Throughout the study, match-to-sample procedures were used to train conditional discriminations and test for emergent stimulus relations. For the nonhuman subjects, training baseline discriminations and testing emergent relations occurred in a modified chamber using visual stimuli (e.g. horizontal and vertical line orientations and hues) presented on keys arranged in a square array. In this study, the sample, or conditional, stimuli were always presented in the center of the array, while comparisons were equally presented in any of the four keys positioned at each corner of the square array.

Reinforcers were delivered into one of two feeder trays, and side of delivery was counterbalanced. Initial training and habituation stages consisted of only sample and comparison keys, but in the final stages of apparatus training, subjects were also required to press a trial-initiation key, which produced the sample stimulus. Pressing the sample stimulus key darkened the trial-initiation key and produced two comparison keys, one correct (S+) and one incorrect (S-).

The subjects in first experimental phase were three rhesus monkeys and the goal was to establish whether subjects could show emergent symmetry after conditional discrimination training. Using lines and hues, the subjects were trained various conditional discriminations, both identity MTS and conditional MTS types. To familiarize subjects with the various stimuli, animals were first trained with identity MTS trials, where samples and correct comparison choices were identical stimuli. For example, 'line-line' identity MTS trial types (1A and 1B) consisted of either a horizontal line sample and a horizontal line correct comparison (S+) with a vertical line incorrect comparison stimulus (S-), or a vertical line sample and correct comparison (S+) with a horizontal line incorrect comparison (S-). Trial types 2A and 2B, or 'hue-hue' trials, presented either a green sample and green correct comparison (S+) with a red incorrect comparison (S-), or a red sample and correct comparison (S+) with a green incorrect comparison (S-). At this time it became apparent that certain subjects had difficulty learning the 'line-line' MTS task, so different sample schedules were introduced for each sample stimulus. One schedule required five responses to the sample before comparisons were presented (fixed ratio, FR5), while the other presented comparison stimuli after two seconds had elapsed with no response (differential reinforcement of low rates, DRL2").

Subjects were also trained conditional discriminations consisting of line sample stimuli and corresponding hue comparisons (trials 3A and 3B). Specifically, for the vertical line sample, the green comparison was correct (S+), and for the horizontal line sample, the red comparison was correct (S+). To establish whether the vertical line and the green hue, and the horizontal line and the red hue had become two separate classes, the final trial types (4A and 4B) consisted of untrained symmetry probes. Untrained symmetry probe trials consisted of the reverse configuration of trial types 3A and 3B, where green and red hues served as samples and vertical and horizontal lines as correct comparisons (S+), respectively. If subjects had formed classes of stimuli, performance during the untrained trials should have resembled performance during baseline training trials.

During the first experiment, subjects readily acquired and mastered the trained identity and conditional discriminations, with accuracies above 90%. Yet, during symmetry probes trials, the subjects' performance declined significantly from baseline measures, with average symmetry probe accuracies at 55% or lower. Although subjects were able to learn two individual baseline conditional discriminations, vertical-green and not red and horizontal-red and not green, they were unable to demonstrate the symmetrical relation, green-vertical and not horizontal and red-horizontal and not vertical, perhaps because of the unchanging function of sample and comparison stimuli during training trials. It may have been beneficial for the authors to vary the position of the stimuli, serving as both sample and comparison, during training sessions in order to promote transfer during symmetry probe trials.

The second phase of the Sidman et al. (1982) study used one subject (R44) from the preceding experiment. Since probe trials in Experiment 1 were conducted in

extinction for two subjects, these animals were considered untrained and therefore able to be tested further. Experiment 2 addressed the issue of whether the identity of the incorrect comparison affected performance during the subjects' unsuccessful symmetry probes from Experiment 1. In Experiment 1, the incorrect comparisons during identity and conditional discrimination training were always of the same stimulus modality as the sample stimulus. So, hue samples always had another hue as S-, and line samples always had the other line counterpart as S-. In other words, if 'vertical line' was the sample and the correct comparison (S+) during identity MTS, 'horizontal line' was always the incorrect comparison rather than both 'horizontal line' and corresponding class member 'red hue' equally serving as S-. Similarly, during conditional discrimination training, if 'vertical line' was the sample and 'green hue' was the correct comparison (S+), 'red hue' was always the incorrect comparison (S-) rather than both 'red hue' and corresponding class member 'horizontal line' equally serving as S-. Finally, unlike Experiment 1, the symmetry probe trials in this phase were not conducted in extinction, thus reinforcement was delivered following correct responses. The authors hypothesized that by training the substitutability of incorrect comparisons, the subject may be able to form classes more readily and prevent the use of exclusion via unchanging comparisons (S-), thereby increasing the likelihood of control by the sample stimulus.

Following baseline reestablishment and training trails with changing incorrect comparisons (S-), the remaining subject (R44) performed no better during reinforced symmetry probes than when probed in extinction during Experiment 1 ($\bar{x} \approx 40\%$). Varying the S- stimuli, probing with reinforcement, and retraining the baseline did not

appear to assist in any sort of class formation. Moreover, the reinforced probes appeared to disrupt the subject's baseline performance, rather than enhance categorization.

After Experiment 2, the authors questioned why the nonhuman primate subjects from Experiments 1 and 2 were unable to demonstrate symmetrical responding when past research using human subjects was successful (Sidman et al., 1982). They hypothesized that perhaps differences in methodology between the past human equivalence studies and the current nonhuman primate study were the reason for the opposing findings. Thus, Experiment 3 was designed to closely replicate the methodology used for Experiments 1 and 2 with six normally functioning (e.g. normal functioning for given age range, no learning or behavioral problems) human children. For the human subjects, comparison stimuli were presented on touch-sensitive keys using a projector, and arranged in a circular array of eight keys. Only four keys, which formed the four corners of a square, were illuminated with comparison stimuli. The ninth key, located in the center of the array, displayed the sample stimuli. All stimuli were identical to those used previously (e.g. hues and lines). The primary reinforcers were pennies, which were paired with a chime that served as a conditioned reinforcer or signal that the primary reinforcer was to follow. Symmetry probes were not reinforced.

Similar to the nonhuman subjects, two children had difficulty acquiring the line-line and line-hue discriminations, and sample schedules of responding were introduced. Subjects also learned and mastered the same baseline identity and conditional discriminations, just as in the earlier phases; however, they did not receive the additional S- variability training trials as in Experiment 2 involving subject R44. Symmetry probes were conducted after the baseline trials were mastered. Four out of six subjects

demonstrated significant transfer during the emergent symmetry probes ($\bar{x}$ ≈90%,100%, 85%, 95%) even without training for sample and comparison position substitutability, while the remaining two did not ($\bar{x}$ =17%, 56%). For the 4 subjects that were successful, it was concluded that the differences between the human and nonhuman data were probably not due to procedural variations across species. The remaining two human subjects received additional training with new baseline trials and were eventually successful in demonstrating emergent symmetry. The additional baseline trials addressed some differences between training and probe trials, wherein hue stimuli were present simultaneously during some baseline trials, allowing for direct comparison, but not during probe trials. The authors speculated that these trial differences may have required the nonhuman subjects to also learn the discriminations in different ways during training versus testing trials and that perhaps applying the new baseline trials would be beneficial.

In Experiment 4, using two of the same nonhuman primate subjects (R44 and R46), new baseline trials were added to the subjects' previous training that more closely resembled trial configurations during probes. To achieve this, an additional stimulus dimension, 'form', was added and consisted of X (vertical and green class) and + (horizontal and red class). Subjects had difficulty acquiring the 'form-line' discriminations, so a yes-no procedure with a limited hold and a correction procedure were used to facilitate learning. Following mastery with the yes-no procedure, the remaining discriminations were introduced, and the baseline performance declined. Although the subject received more probe-representative training trials and interventions designed to increase stimulus control, performance during symmetry probes did not improve ($\bar{x}$ ≈30% for R46, R44 similar).

The authors chose to replicate using a different, nonhuman primate species, the olive baboon. The baboons were selected because, after four experimental stages, it remained unclear whether differences in data resulted from differences across species or experience. Experiment 5 resembled that of 1 and 3, and was conducted using two baboons, Bab-Sim and Bab-Win. Apparatus specifications were similar to those used with the macaques, but sized for the larger baboons. Subjects acquired the baseline discriminations and were tested for symmetry once in extinction and twice with reinforcement for subject Bab-Sim, and only once with reinforcement for Bab-Win. Subjects did not transfer their accurate baseline performance during symmetry probes; means for probe trials were 46%, 44%, and 48% for Bab-Sim, and 49% for Bab-Win. As before, although they learned the individual conditional discriminations, no symmetrical relations were formed between the stimuli.

A potential issue with the Sidman et al. (1982) study was the use of identical stimuli for both the children and nonhuman primates, specifically the unchanging horizontal and vertical line comparisons (D'Amato et al., 1985). As previously noted, in an attempt to overcome initial difficulties with various sample stimuli and promote discrimination amongst those stimuli, procedures were modified such that both the human and nonhuman subjects were temporarily required to respond differentially to the sample stimuli (e.g. fixed ratio, FR5, for vertical line sample stimuli and differential reinforcement of low rates, DRL2s, for horizontal line sample stimuli) before any comparison stimuli were presented. Using identical stimuli allowed for direct comparison across species; however, because nonhumans often have difficulties discriminating these types of stimuli (e.g. during training phases of this study), it may have been biased

towards the human subjects. Using more species-relevant stimuli may have increased the likelihood of transfer in the nonhuman primates. Yet, because both humans and nonhumans had difficulty with these stimuli, it may be pertinent to just replace them altogether.

According to Sidman et al. (1982), symmetry and transitivity are both necessary properties of equivalence classes, thus only the human children showed any convincing evidence for equivalence class formation through their performance in symmetry probes. The macaques and baboons apparently learned many separate, individual conditional discriminations rather than forming classes of stimuli, perhaps because of the unchanging function of sample and comparison during training (e.g. no training using nodal stimuli). The authors state that, in order to overcome control by stimulus function (always a sample or comparison), the interchangeability of the sample and comparison functions may have to be explicitly trained, yet this important aspect was not manipulated in the 1982 study. Training additional conditional discriminations with nodal stimuli functioning as both samples and comparisons would have provided the needed exposure and set up an opportunity to test transitive relations as well. Also, the current study tested subjects with a limited number of stimuli. The use of multiple exemplar training may be necessary before the emergent symmetry and transitivity relations characteristic of equivalence classes can be shown in nonhumans. Finally, perhaps MTS procedures are not an effective means of establishing classes in nonhumans, and a different procedure is necessary (e.g. repeated reversal training).

Using conditional discrimination procedures similar to Sidman et al. (1982), D'Amato et al. (1985) evaluated equivalence tasks in monkeys (*Cebus apella*) and

pigeons (*Columbia livia*) across three different experiments. The first experiment trained and tested six monkeys in operant chambers containing five projectors; four were arranged to illuminate comparison stimuli in the corners of a square pattern and the fifth illuminated the sample stimulus in the center position. Visual stimuli consisted of various black and white shapes (e.g. triangle, circle, vertical line, dot, and plus) and one red key stimulus. Although a vertical line sample was used, no horizontal lines were included to control for difficulties in discrimination between these stimuli. Following a trial initiation response, sample stimuli were presented for approximately 1s, then disappeared with the presentation of comparison stimuli (0s delay procedure). Two subjects first received ArB (e.g. triangle to red key and not vertical line, and dot to vertical line and not red key) conditional discrimination training, while the remaining four had prior conditional MTS training and required no additional training sessions. All subjects reached high levels of accuracy on the final few baseline training sessions prior to BrA symmetry testing ($\bar{x}$ = 95%).

Symmetry testing was not conducted in extinction, thus only the first presentation of novel relations were analyzed as evidence for equivalence transfer. Reinforcement of testing trials is often used when testing nonhumans because probing in extinction typically causes disruptions in performance and emotional behaviors (Schusterman & Kastak, 1993). BrA (e.g. red key to triangle and not dot) symmetry testing reversed the sample and comparison function of the A and B stimuli. As a control, two different BrA conditions were tested; BrA+ and BrA-. The BrA+ condition was the reverse relation consistent with the trained ArB relation, as stated above. The BrA- condition swapped the correct comparisons so, if during ArB training subjects were reinforced for selecting the

red key in the presence of the triangle sample, the inconsistent BrA- relation now reinforced responses to the dot comparison in the presence of the red key sample. Five out of six subjects performed better on BrA+ than BrA- tests, although this difference was not statistically significant ($.05 < p < .10$, t-test), perhaps as a result of individual differences characteristic of between-subjects designs. Four of the subjects were then trained and tested for the BrC and CrB relations, respectively. Like the first task, CrB+ and CrB- testing formats were used during the second round of testing as well. Three of the four subjects performed better on CrB+ than CrB- test types, however, these results were not statistically significant. The authors concluded that, overall, subjects did not show evidence of symmetrical responding and may have been discriminating based on features of the procedures other than the relevant stimulus dimensions (e.g. configuration, position, or stimulus sequences). The weak demonstration of bidirectional symmetry in these monkeys, as well as most nonhuman animals, is common (D'Amato et al., 1985; Schusterman & Kastak, 1993) and is possibly the result of alternative sources of control (e.g. changing sample and comparison function during testing that is not explicitly trained during baseline).

In the second experiment, D'Amato et al. (1985) addressed the ability for the same four primate subjects to form unidirectional transitivity relations using baseline training of two conditional discriminations, which initially occurred during the first experiment. The authors hypothesized that the null findings in the Sidman et al. (1982) study were the result of the vertical/horizontal line stimuli confound, and therefore testing with more discriminable stimuli would promote emergent relations. ArC transitivity testing began immediately following Experiment 1 baselines, and again included baseline

consistent (T+) and inconsistent contingency (T-) formats. Consistent with the notion that unidirectional relations are more readily observed in nonhumans, all four subjects demonstrated strong evidence for emergent transitivity, and also performed better on the T+ format (T+ $\bar{x} = 91\%$ vs. T- $\bar{x} = 11\%$, $p < .001$, t-test). Following Experiment 2, D'Amato et al. concluded that the use of vertical and horizontal comparison stimuli in Sidman et al. (1982) may have prevented subjects from showing emergent transitivity, as demonstrated here.

Finally, a third experiment was conducted to evaluate transitivity in a non-primate species, specifically, the pigeon. Using operant chambers, three subjects were trained to peck keys which were illuminated using the same stimuli from Experiments 1 and 2. All subjects readily acquired the baseline discriminations for ArB and BrC relations. Transitivity tests were similar to those in the earlier experiments, including the use of T+ and T- formats. Unlike the results obtained with the primate subjects, there was no evidence of transitivity transfer in the pigeons, and only one subject performed better on T+ than T- formats. Although Experiment 3 was designed to replicate Experiment 2 as closely as possible, the authors noted that more testing across a wider variety of contexts would be necessary before jumping to species-related conclusions. Perhaps, as in Sidman et al. (1982), the stimuli used with pigeons were not optimal, or perhaps pigeons require multiple exemplar training as well.

Schusterman and Kastak (1993) continued the search for nonhuman symbolic behavior by testing two California sea lions on equivalence tasks using conditional discrimination procedures. The sea lions received baseline identity matching and conditional discriminations training for all three emergent properties of equivalence using

arbitrary visual cues mounted on a large, wooden display. The sample stimulus was presented through a box in the center section of the display, with comparisons located on either side. Sample stimuli were presented for a few seconds while the subjects observed from a holding station. Comparisons were then presented and the subjects were required to select the correct comparison using the nose. Correct responses resulted in the delivery of a fish reinforcer.

Training included multiple exemplars, or repeated MTS training with various different stimuli, to expand the breadth of training, increase the likelihood that emergent relations would occur, and allow for reinforcement during testing trials. Again, probing in extinction typically causes disruptions in performance and emotional behaviors, but reinforcing responses during test trials significantly reduces the number of trials that can be evaluated for emergent relations, since all subsequent trials are essentially trained. However, using multiple exemplars increases the number of valid testing trials and permits reinforcement during all test probes. In response to the failure to demonstrate symmetry in previous nonhuman studies (Dugdale & Lowe, 1990 and Tomonaga, Matsuzawa, Fujita, & Yamamoto, 1991: as cited in Schusterman & Kastak, 1993), the functions of sample and comparison stimuli were explicitly trained prior to testing for the symmetry relations to facilitate symmetrical responding. Generalized identity matching with the same two sea lions was established in an earlier study, and maintenance of these relations through novel stimulus MTS presentations occurred prior to training of new MTS relations. The older subject (Rocky) struggled during identity MTS maintenance and new training phases, thus all analyses were of the remaining subject's (Rio) performance.

Rio readily acquired the arbitrary conditional relations that were explicitly trained, specifically ArB and BrC, and showed many fewer errors during training of the second relation (ArC $\bar{x}$ = 36.2 $\pm$ 31.7 errors vs. BrC $\bar{x}$ = 17.8 $\pm$ 13.5 errors). Emergent relations were tested using the first six presentations of novel stimulus-stimulus relations during symmetry and transitivity. Relational responding was defined as accurate performances during these novel stimulus-stimulus relation trials. Using multiple exemplars, Rio was tested on two separate occasions for BrA symmetry, and although her performance overall was not significantly better than chance (8/12, $p > .10$, binomial test), she improved across the two testing sessions (3/6 vs. 5/6). Rio was also tested twice for CrB symmetry and performed significantly better than chance, correctly responding 10/12 times ($p < .05$). Transitivity tests for ArC were similarly promising, with Rio correctly responded to 11/12 possible trials ($p < .01$), as well as the symmetry-transitivity CrA relation, with correct responses for 10/12 possible trials ($p < .05$).

Rio's accurate performance during almost all of the symmetry and transitivity probe trials demonstrates behavior indicative of equivalence class formation. However, her initial failures during BrA symmetry tests and later successes with BrC tests, suggest that the multiple exemplar training was a crucial factor in her ability to respond in a relational fashion. As suggested by Sidman and colleagues (1982), the training of the "symmetry concept", or the substitutability of sample and comparison functions, through multiple examples appears to be necessary in order for most nonhuman animals to show emergent symmetry. Because transitivity is unidirectional and stimulus positions are stable (e.g. the sample is always the sample and the comparison is always the comparison), nonhumans more readily demonstrate this relation, as compared to

symmetry. As a result of the sea lion's relational transfer, Schusterman and Kastak conclude that, again, perhaps the cognitive abilities required to group stimuli in equivalence classes precedes language, and not the opposite. The authors also suggest that equivalence, from an evolutionary standpoint, is an adaptation necessary for efficient social and cognitive functioning.

Although there are a few notable exceptions (D' Amato et al., 1985; Schusterman & Kastak, 1993), research using basic conditional discrimination training to evaluate the three properties of equivalence classes, identity, symmetry, or transitivity, has largely been unsuccessful in demonstrating class formation in many different nonhuman species including rats and nonhuman primates (Iversen, 1997; Lipkins, Kop, & Matthijs, 1988; Sidman et al., 1982).

For example, Lipkins et al. (1988) extended the investigation of conditional discriminations and equivalence relations in pigeons. Subjects were trained and tested in operant chambers which contained three adjacent walls with two pecking keys per wall. Stimuli consisted of combinations of lines and hues. Trials began with the presentation of sample stimuli, and upon response to the sample, two comparisons appeared. Unlike the D'Amato et al. (1985) study, the current procedure used simultaneous presentations of sample and comparison stimuli (sample stimuli remained present, vs. absent in 0s delay procedure). A non-correction procedure was used, wherein incorrect responses resulted in a 8s time out followed by inter-trial interval initiation.

Nine subjects were trained and tested on several conditional discriminations in a variety of orders for two different classes (1 and 2). Some subjects were explicitly trained ArB and BrC, followed by testing for symmetrical BrA and CrB relations, as well as

transitive ArC and dual symmetrical-transitive (equivalence) CrA relations. Others were trained BrA and BrC relations, followed by testing transitive ArC and the dual equivalence relation CrA. Finally, some subjects were trained ArB and CrB relations and tested for symmetrical BrA and BrC, transitive ArC, and the dual equivalence relation CrA. Most importantly, these training procedures were designed to establish equivalence classes comprised of A, B, and C members.

Of the nine subjects in this study, only four ever met criterion during baseline training sessions. Of those four, none demonstrated significantly above chance (50%) performance accuracy during emergent symmetry and transitivity probes. In conclusion, and in contrast to the nonhuman primate transitivity data obtained by D'Amato et al. (1985), it appeared that the pigeons in the current study were not capable of forming equivalence classes through conditional discrimination training, as indicated by their chance-level performance during probe tests. Moreover, the majority of the subjects in the current study were not able to acquire baseline discriminations. External sources of control, such as spatial location, may have affected performances. In addition, no measure of identity matching was evaluated prior to training and testing for symmetry and transitivity. The authors note that this may be a necessary first step to establish the consistency of each stimulus across trials and sessions.

Equivalence and Rats

Compared to species such as pigeons (Meehan, 1999; Vaughan, 1988), nonhuman primates (Vauclair, 2001), and marine mammals (Kastak et al., 2001; Schusterman & Kastak, 1993), there is little evidence of equivalence class formation using conditional

discrimination training in rats, as well as functional classes. However, generalized identity matching has been investigated with rats.

For example, in two studies conducted by Iversen (1993, 1997), Long-Evans rats (*Rattus novegicus*) were trained using a visual, identity MTS procedure that was designed to evaluate the identity relation or Sidman's first property of equivalence; identity (ArA). Subjects were tested in an operant chamber containing three nose-poke keys which were illuminated with various colors. During Experiment 1 (1993) baseline training, the sample stimulus key was always located in the center position and subjects were required to press the center sample key for access to comparison keys. In these identity MTS procedures, the sample and one comparison key were identical in color, while the second comparison was a different color. The subjects required hundreds of these training trials before 90% criterion level performances were shown. Then, during initial testing phases, the location of the sample key was varied across the three possible stimulus keys. This sample position variation caused a significant decline in subject performance, from 90% during training to around 60% during testing. During a follow-up experiment, Iversen (1997) varied the position of the sample during training sessions, yet almost all subjects were still unable to transfer above chance (50%) performance during trials with the novel sample locations.

Many of the rats in this study were apparently responding based on individual stimulus-stimulus relations and stimulus positions. Through Iversen's (1993, 1997) studies, it became clear that position effects may be a significant issue in MTS procedures with rats, and possibly other nonhuman species. In addition, this study showed that control by stimulus position was completely overriding control by the visual

stimuli. Perhaps using different stimuli would have promoted control by the visual features of the stimuli rather than by their relative spatial positions.

In two experiments conducted by Eichenbaum (1998) and Dudchenko, Wood, and Eichenbaum (2000), olfactory stimuli were used to investigate memory and learning in the rat. A digging procedure was developed which required subjects to dig in cups containing a mixture of scent and play sand to obtain cereal reinforcers. The digging procedure used in these studies facilitated rapid odor discrimination learning and successful performance on several complex memory tasks (e.g. memory span tasks). Eichenbaum et al. attributed the observed performance to the exploitation of the rat's olfactory system and the natural foraging behaviors inherent in the digging procedure. No MTS procedures or evaluation of conceptual behavior was investigated; however, both studies demonstrated the utility of this olfactory procedure with rats.

Adaptation of Eichenbaum's (1998) olfactory digging procedure to a MTS procedure was first undertaken in a study by Peña et al. (2006). In contrast to the relatively slow and fragile acquisition of visual MTS tasks in the Iversen studies (1993, 1997), the MTS procedures using olfactory stimuli produced rapid learning, stable performance, and generalization to novel stimuli in rats. Using identity MTS procedures, cups of scented sand were delivered using a tray that was inserted into a modified operant chamber. Holes were drilled in the tray to hold the cups of stimuli. Stimuli were presented sequentially across sessions so that subjects began training with a few scents and following criterion level performances, were presented increasingly more scents in addition to the familiar scents already trained. Thus, sessions eventually included many familiar scents as well as any novel scents introduced on a given day. At each stage

subjects required significantly fewer training sessions to meet criterion; approximately 35 to 40 sessions were required for each subject before criterion for the first stimuli were met, followed by approximately 5 to 10 for the second stimuli, and 5 or less for all of the remaining stimuli. Some subjects reached stages where sessions included up to 30 or more stimuli and performance remained significantly above chance. Rapid acquisition of the identity MTS procedures and transfer to novel stimuli indicated that subjects' behavior had come under the control of the olfactory stimuli and not the position of the sample or comparisons. In this case, relations between the olfactory stimuli were more readily learned and discriminated than the visual stimuli, which resulted in one of the first convincing demonstrations of generalized identity matching using conditional discriminations in rats.

The data obtained from the Peña et al. (2006) study show the importance of evaluating the relative benefits of standardization inherent in using visual stimuli against the inevitable costs of using those stimuli. Sacrificing the complete automation and standardization typical in studies using visual stimuli may be necessary in order to assess the actual capabilities of rats with the use of olfactory stimuli. The study conducted by Peña et al. (2006) also provides evidence that nonhumans are capable of exhibiting at least reflexive relations using conditional MTS procedures and relevant stimulus dimensions.

Olfaction and Rats

The study conducted by Peña et al. (2006) suggests the value of studying conditional discriminations and MTS procedures with olfactory stimuli. Instead of replacing the conditional discrimination and MTS procedures with functional class

training, maybe a modification of stimulus dimensions could promote relational responding.

All animals, including humans, are sensory specialists. In other words, animals often primarily rely on one sensory modality to the exclusion of others (Slotnick, 2001). For example, many nonhuman species such as canines and rodents have evolved highly specific olfactory abilities. These species will tend to favor olfaction as a primary means of contacting their environment, often to the exclusion of the other sensory modalities; vision, audition, gustation, and proprioception. On the other hand, humans tend to favor vision as their primary means of sensing and perceiving the surrounding environment. As a result, much research with both human and nonhuman animals uses visual stimuli, which can be standardized and highly controlled. For most nonhumans (e.g. pigeons, primates), the biased use of visual stimuli is in line with their own sensory biases, and their efficient performances show this. However, for those animals that do not significantly rely on vision (e.g. rodents), this is a great disadvantage. In terms of the current equivalence discussion, the historical lack of equivalence demonstration in rats may be directly related to the use of suboptimal stimuli.

According to Slotnick (2001), procedures that use olfactory stimuli with rats may be more successful because the olfactory stimuli are more biologically relevant than visual stimuli. Since rats are essentially 'smelling' organisms, they are notoriously poor performers in tests using other stimulus modalities, such as vision. When tested with olfactory stimuli, rats exhibit performances that resemble the accuracy and efficiency of their primate counterparts. Although olfaction may be primitive in terms of structural

anatomy and appearance in evolutionary time, the olfactory abilities of rats appear to be anything but primitive.

The rodent olfactory system is structurally simple. Olfactory receptors located in the nasal passages have direct connections to the limbic system, hypothalamus, and the prefrontal cortex, without many filters in between (Slotnick, 2001). These connections have implications for effects of odors on basic biological functions, as well as memory, learning, cognition, and emotional behavior.

Various researchers have used olfactory stimuli with rodent subjects across a variety of contexts, including learning set studies, memory span tasks, associative learning, MTS, and neurobiological lesion experiments (Slotnick, 2001). For example, when olfactory stimuli were adapted to learning set studies, rats displayed learning set acquisition in a relatively small number of trials (Slotnick, Kufera & Silberberg, 1991; Slotnick, Hanford & Hodos, 2000: as cited in Slotnick, 2001). In the case of memory span tasks, rats have been shown to recall odors learned many days prior. And, as previously noted, rats are notoriously slow to acquire even identity MTS discriminations. However, using olfactory stimuli in a MTS format greatly increases rates of learning and acquisition of the MTS task (e.g. Peña et al., 2006).

If olfactory stimuli promote efficient learning and accuracy in rats, why are these stimuli not widely adopted? For the most part, difficulties with odor control and standardization of scents have plagued olfactory research. Methods of odor presentation that completely automate trials, standardize scent concentrations, and control for odor mixing are effective (e.g. odor puff chambers). Nonetheless, procedures such as the Eichenbaum (1998) digging procedure used in Peña et al. (2006), that are not as

automated or standardized, still facilitated rapid odor discrimination learning and successful performances.

Current Investigation

The present study sought to investigate the formation of equivalence classes in rats using olfactory stimuli. The use of olfactory stimuli in the current study was hypothesized to increase the rate of acquisition of the discriminations and facilitate class formation. Compared to the relatively slow acquisition of visual MTS procedures with rats observed in the Iversen (1993, 1997) studies, Peña et al. (2006) demonstrated rapid acquisition of olfactory MTS procedures using the digging procedure developed by Eichenbaum (1998). The combined use of odor stimuli, multiple exemplars, and a naturalistic digging response requirement was predicted to facilitate acquisition of the discriminations in a more effective and efficient manner than comparable studies using visual stimuli. The stimuli and the response requirement characteristic of the current study were used to maximize the probability of observing responding indicative of equivalence class formation.

In the current study we slightly modified the digging procedure by using two different types of pellet reinforcers rather than crushed rat chow (Eichenbaum, 1998) or cereal reinforcers (Dudchenko et al., 2000; Eichenbaum et al., 1998). The pellet reinforcers used in the current study were class-specific or, in other words, pellet reinforcers were differentially associated with a particular class of stimuli and never mixed across classes for an individual subject. The class-specific reinforcer assignment was included to maximize class formation and assess the value of differential outcomes on equivalence class formation, as proposed by Sidman (2000) and used by Kastak et al.

(2001) with sea lions and Meehan (1999) with pigeons. It is hypothesized that the use of class-specific reinforcers, both sugar and grain, will increase the speed of discrimination mastery (lower number of sessions to meet criterion).

Using the modified Eichenbaum (1998) olfactory digging procedure, the current study was divided into three experiments. The first experiment began as a systematic replication of Peña et al. (2006). Experiment 1 was designed to examine generalized identity MTS using a modification of the multiple exemplar training procedure used in Peña et al. (2006) and two different apparatus. Differences in apparatus were also hypothesized to influence the acquisition and mastery of MTS discriminations. Subjects that achieved statistically significant performance (see Method for explanation) during Experiment 1 were advanced to Experiment 2, arbitrary conditional discrimination training and emergent equivalence testing. Experiment 2 sought to clarify whether rats could, given the extensive history of identity matching achieved through Experiment 1, demonstrate relational responding and equivalence class formation through arbitrary conditional MTS training. Finally, to further test the hypothesis that generalized identity matching experience promotes accurate equivalence performance, a third experiment was conducted. Experiment 3 included the addition of two experimentally naïve subjects that received only arbitrary MTS training and equivalence testing without prior exposure with multiple exemplars of the identity relation.

METHOD

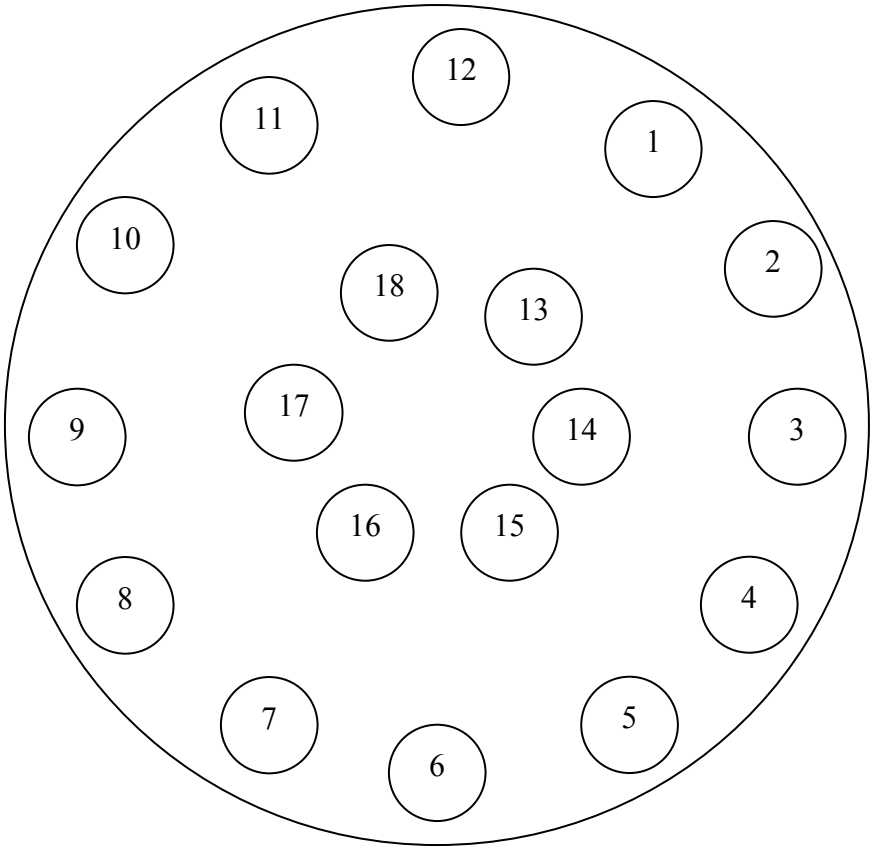
Subjects

The subjects for the current study were 11 male Holtzman (Sprague-Dawley) albino laboratory rats between 90 and 120 days old at onset of testing. Subjects were kept at approximately 85% of their free feeding weight through restricted food access, while water was provided *ad libitum*. Each subject and food ration was weighed daily to maintain stable weights. Food rations varied depending on the subject's weight, consumption, and observed performance during sessions, but ranged between 16g and 10g of chow per day. Subjects were fed following testing sessions, approximately 0.5 hours after the conclusion of the session. Subjects were individually housed and maintained on a 12:12 hr light-dark cycle.

Apparatus

Two different apparatus were used for data collection as a between-subjects manipulation during Experiment 1 and 2. The first apparatus, the open-field Odor Arena (see Figure 1) is a circular acrylic table equipped with a swivel mechanism which allows the Arena to be rotated. The Arena is 94cm in diameter and contains eighteen, 5cm drilled holes that will hold 2oz translucent condiment cups containing scented sand. The 18 hole positions are spaced approximately 13cm apart and numbered consecutively in two circular, clockwise arrays. Positions 1 through 12 are located on the outermost ring and numbers 13 through 18 are located on the inner ring (see Figure 1). Aluminum baffling approximately 30cm high surrounds the perimeter of the Arena, and prevents subjects from exiting the apparatus. During MTS, sample cups filled with scented sand were presented in the home cage of the subject, located on the table adjacent to the

a.



b.

Figure 1. a.) Odor Arena Apparatus.
b.) Odor Arena Diagram with numbered hole positions designed to hold comparison stimulus cups.

Arena. Comparison scents (cups) were pseudo-randomly arranged inside the apparatus in two of the possible 18 holes on a given trial, with the remaining 16 holes occupied by empty condiment cups.

The second apparatus, which was ultimately faded out of the study, is the Operant Match-To-Sample Chamber (see Figure 2). The Operant MTS Chamber is a modified operant chamber (28cm X 26cm X 30cm) with front, rear, and top Plexiglas® panels for ease of observation, two metal sides, and metal rods below. A 5cm horizontal section on the base of the front panel was removed in order to fit the plastic tray used for the match-to-sample (MTS) procedures. The plastic MTS tray consisted of three 5cm holes drilled in a triangular shape so that a single hole was positioned above two adjacent comparison holes (see Figure 2). Each 5cm hole of the MTS tray held a 2oz stimulus cup designated as either a sample or comparison. This triangular configuration allows access to only the sample when the tray is inserted approximately halfway into the chamber and full access to comparisons when completely inserted.

Stimuli

All plastic condiment cups and perforated lids (Fabri-kal, see Figure 3) were labeled according to odor and stored in scented sand containers to ensure separation of scents and saturation of lids with odors. Lids were perforated using a small sewing needle to puncture holes in a standard circular pattern with approximately equal numbers of holes for all lids. Lids were labeled to avoid confusion during sessions with multiple stimuli and to ensure proper storage following each session. To prevent scent marking, each subject had individual lids and cups for each session. Lids and cups were rotated so

a.

b.

Figure 2. a.) Operant Match-To-Sample Chamber with stimulus tray.
b.) Operant Match-To-Sample stimulus tray. The upper hole position always contained the sample stimulus, the lower two holes contained the two comparison choices, S+ and S-.

that they were not reused during consecutive sessions. The olfactory stimuli contained a specific mixture of household spices or liquid oil odorants when necessary (see Appendix A for entire odor list) and fine-grain, white play sand (The Home Depot). Sand scented with household spices was mixed at a ratio of 10 grams to 1000 grams, spice to sand, respectively. Sand scented with liquid oil odorants was mixed at a ratio of 24 drops of liquid from a standard eyedropper per 1000g of sand. Sand and liquid odorants were mixed thoroughly to guarantee even dispersion of the liquid. Liquid odorants were used when subjects required additional scents and only after all dry spice scents had been used during training. The specific 10g/1000g spice to sand ratio was selected based upon data from previous probe studies which indicated sufficient masking of pellet reinforcer odor at this concentration of scented sand (Peña et al., 2006). Subjects in these earlier probes showed no evidence of reinforcer detection when tested at this concentration of spice to sand. In the instance that a subject showed an aversion to certain scents, wherein responding to the particular stimulus ceased, these particular scent(s) were discarded from that subject's set and replaced with a different scent. The olfactory stimulus sets for each subject across all three experiments are shown in Table 1.

Cups were filled with scented sand to approximately 1cm below the rim, and perforated lids were placed on the cup rim. Reinforcers were placed into stimulus cups and depressed below the surface using metal tweezers. All experimenters wore latex examination gloves to transfer and remove cups and lids in all of the preparation and experimentation phases.

Figure 3. Perforated stimulus lids used in the Odor Arena Apparatus.

General Procedure

Testing was typically conducted five days per week, Monday through Friday, with all subjects participating in one session per day. All testing occurred in either the Odor Arena room or the Operant MTS Chamber room, which were both equipped with a 70dB white noise generator. All sessions were video recorded using a computer web camera and Windows MovieMaker software.

Initial Training and Habituation

All subjects were initially handled and food restricted to facilitate reinforcer consumption and establish responding during the tasks. Habituation to reinforcers, sugar and grain pellets, and the apparatuses followed. Early training sessions allowed subjects to habituate to the apparatus and materials and allowed shaping of responding to the sand, lids, and MTS procedure. The first exposures introduced the reinforcers, sand, and cups by placing pellets at the surface of unscented sand of the sample and one comparison stimulus. Once subjects consumed all reinforcers during 12 trials of each of five consecutive sessions, pellets were incrementally buried to deeper levels; beginning with surface placement, then half buried, then only one plane visible, and finally completely covered by sand. The final depth was approximately 1cm below the surface of the sand, which was designed to prevent any visual or olfactory detection of the pellet. Following consistent retrieval of buried pellets for 12 trials across five consecutive sessions, perforated lids were placed on the cup rim (for Odor Arena subjects). Lid training began with lids covering approximately half of the underlying cup of sand. Coverage was gradually increased until the lid completely covered the cup. Once lid removal became reliable, which varied across subjects, scented sand was introduced. At the same time, a

Table 1. Individual Subject Stimulus Class Assignments for all phases of the study.
Asterisks denote stimuli that were created using liquid odorants rather than dry, powdered spices. For all subjects except F6, Class 1 stimuli were paired with sugar pellets and Class 2 stimuli were paired with grain pellets. F6 received only sugar pellets for both classes throughout the study.

Subject	F6		F3		F5		H7	
Stimuli	1(sugar)	2(sugar)	1(sugar)	2(grain)	1(sugar)	2(sugar)	1(sugar)	2(grain)
A	Nutmeg	Dill	Clove	Rosemary	Nutmeg	Dill	Clove	Rosemary
B	Celery	Cinnamon	Onion	Sage	Coriander	Sage	Onion	Sage
C	Clove	Ginger	Sumac	Thyme	Turmeric	Garlic	Sumac	Thyme
D	Oregano	Onion	Bay	Paprika	Thyme	Clove	Bay	Paprika
E	Thyme	Coriander	Marjoram	Cumin	Paprika	Mustard	Marjoram	Cumin
F	Mustard	Sumac	Turmeric	Oregano	Marjoram	Sumac	Turmeric	Oregano
G	Cumin	Marjoram	Cinnamon	Nutmeg	Celery	Cumin	Cinnamon	Nutmeg
H	Garlic	Rosemary	Ginger	Mustard	Ginger	Oregano	Ginger	Mustard
I	Sage	Turmeric	Dill	Celery	Onion	Bay	Dill	Celery
J	Bay	Paprika	Coriander	Garlic	Rosemary	Cinnamon	Coriander	Garlic
K	Sassafras	Hickory	Fennel	Allspice			Fennel	Allspice
L	Worcestershire	Orange	Beet	Carob			Beet	Carob
M	Savory	Fennel	Lime	Tomato			Tomato	Lime
N	Carob	Allspice	Hickory	Savory			Raspberry	Savory
O	Beet	Caraway	Orange	Worcestershire			Orange	Worcestershire
P	Lime	Tomato	Caraway	Sassafras			Caraway	Sassafras
Q	Spinach	Grape*	Spinach	Peppermint*			Spinach	Peppermint*
R	Raspberry	Peppermint*	Raspberry	Grape*			Hickory	Grape*
S	Almond*	Maple*	Almond*	Maple*			Almond*	Maple*

Subject	H8		G13		G8		I30	
Stimuli	1(sugar)	2 (sugar)	1 (sugar)	2 (grain)	1 (sugar)	2 (grain)	1 (sugar)	2 (sugar)
A	Dill	Nutmeg	Clove	Rosemary	Mustard	Thyme	Sumac	Oregano
B	Celery	Cinnamon	Onion	Sage	Caraway	Orange	Mustard	Bay
C	Clove	Ginger	Sumac	Thyme	Savory	Oregano	Cinnamon	Clove
D	Oregano	Onion	Bay	Paprika	Marjoram	Garlic	Tomato	Raspberry
E	Garlic	Thyme	Marjoram	Cumin	Raspberry	Celery	Sage	Nutmeg
F	Rosemary	Mustard	Turmeric	Oregano	Beet	Cumin	Turmeric	Garlic
G	Savory	Sage	Cinnamon	Nutmeg	Rosemary	Turmeric	Fennel	Lime
H	Paprika	Turmeric	Ginger	Mustard	Sage	Spinach	Orange	Carob
I	Cumin	Bay	Dill	Celery	Paprika	Ginger	Rosemary	Celery
J	Coriander	Marjoram	Coriander	Garlic	Carob	Tomato	Thyme	Onion
K			Fennel	Allspice	Sassafras	Bay	Paprika	Cumin
L			Beet	Carob	Sumac	Cinnamon	Savory	Marjoram
M			Lime	Tomato	Nutmeg	Onion	Dill	Worcestershire
N			Hickory	Savory	Dill	Allspice	Coriander	Allspice
O			Orange	Worcestershire	Clove	Garlic	Caraway	Beet
P			Caraway	Sassafras	Lime	Hickory	Grape*	Sassafras
Q			Spinach	Peppermint*	Fennel	Almond*	Spinach	Ginger*
R			Raspberry	Grape*	Peppermint*	Grape*	Hickory	Peppermint*
S			Almond*	Maple*	Maple*	Worcestershire	Almond*	Maple*
T			Peach*	Chocolate*				
U			Lemon*	Vanilla*				
V			Root beer*	Coconut*				
W			Pineapple*	Strawberry*				
X			Banana*	Brandy*				
Y			Apple*	Rum*				

Subject	J6		I4		I29	
Stimuli	1 (sugar)	2 (grain)	1 (sugar)	2 (grain)	1 (sugar)	2 (grain)
A	Raspberry	Sage	Allspice	Orange	Spinach	Marjoram
B	Sumac	Fennel	Bay	Onion	Rosemary	Lime
C	Carob	Cumin	Beet	Oregano	Thyme	Garlic
D	Coriander	Nutmeg	Carob	Paprika	Tomato	Mustard
E	Celery	Marjoram	Celery	Raspberry	Sumac	Nutmeg
F	Cinnamon	Ginger	Cinnamon	Rosemary	Paprika	Fennel
G	Paprika	Onion	Clove	Sage	Savory	Dill
H	Dill	Lime	Coriander	Sassafras	Sassafras	Caraway
I	Tomato	Allspice	Cumin	Savory	Sage	Cumin
J	Caraway	Spinach	Caraway	Spinach	Turmeric	Coriander
K	Worcestershire	Beet	Dill	Sumac	Raspberry	Clove
L	Turmeric	Mustard	Fennel	Thyme	Worcestershire	Cinnamon
M	Savory	Thyme	Garlic	Tomato	Oregano	Celery
N	Garlic	Sassafras	Ginger	Turmeric	Onion	Carob
O	Rosemary	Bay	Hickory	Worcestershire	Orange	Beet
P	Oregano	Orange	Lime	Almond	Grape*	Bay
Q	Hickory	Almond*	Marjoram	Grape*	Maple*	Allspice
R	Peppermint*	Clove	Mustard	Peppermint*	Almond*	Hickory
S	Maple*	Grape*	Nutmeg	Maple*	Peppermint*	Ginger

second comparison cup was added, completing the MTS set-up for the rest of the experiment. Preliminary sessions consisting of only 12 trials were used until subjects were completing each of 12 trials within three minutes, responding to both the sample and the comparison stimuli. Complete sessions, consisting of 24 trials, were then implemented and remained the same length for the duration of the study.

Match-To-Sample Procedure

Typical matching-to-sample (MTS) procedures were used to train baseline discriminations and test during probe phases. The MTS procedure involved the presentation of a single sample stimulus followed by the simultaneous presentation of two, different comparison stimuli (S+, S-). Only responses to the sample and correct comparison stimuli (S+) resulted in reinforcement. Using pellet reinforcers, these MTS procedures trained subjects to match or choose the correct comparison stimulus (S+), and not the incorrect comparison stimulus (S-), based on the sample stimulus. In other words, whether the response to a given comparison stimulus resulted in reinforcement was conditional on the sample stimulus, and reinforcement of responses to the comparison stimuli differed across trials depending on the sample stimulus for that given trial. Two types of responses were scored during experimentation; correct responses and incorrect responses. In the Odor Arena, complete responses were defined as any displacement, either in whole or in part, of the perforated lid from the rim of the underlying cup using the snout, face, or front paws. Thus, this provided a clear response definition. In the Operant MTS Chamber, stimuli did not include the perforated lids. Responses were scored when subjects displaced the scented sand with the snout, face, or paws using a digging motion with the front paws. Digging motions were defined as the use of the front

paws to displace the sand from the stimulus cup in the direction towards the subjects' body. Simply placing one or both paws on the surface of the sand without digging did not meet the requirements for digging, and were not scored as responses. In both apparatus, a correction procedure was used so that subjects always ended trials with a response to S+, either as an initial correct response, or following an incorrect response to S-.

Trials began with the presentation of the stimulus cup containing the sample scent and reinforcer. For all training and testing phases in the Odor Arena, the sample stimulus cup was presented in the home cage, located adjacent to the Arena apparatus. Following the completion of a digging response and retrieval of the reinforcer from the sample stimulus, subjects were removed from the home cage and physically placed in the center of the Arena, allowing access to S+ and S- comparison cups. After displacing the lid of the correct stimulus (S+) on a given trial, subjects were then required to dig in the scented sand and retrieve reinforcers. Once the reinforcer was consumed and the subjects physically moved away from the correct stimulus (S+), the trial ended. Following trial completion, the subjects were physically removed from the apparatus and placed in the home cage for a variable inter-trial interval while experimenters prepared the next trial. For subjects trained and tested in the Operant MTS Chamber, the MTS tray was used to deliver sample and comparison cups during all training and testing phases. Subjects were required to displace the sand using a digging motion and retrieve reinforcers. Once reinforcers were consumed, the current trial ended and the MTS tray containing the stimuli cups was withdrawn from the chamber. An inter-trial interval of varying duration began and preparation for the next trial occurred. For all subjects, a variable inter-trial interval of approximately 15 to 30 seconds was required for the next trial set up. If a

response to S+ was not made within 3 minutes, the trial was ended (timed out) and scored as incorrect. If subjects timed out on three consecutive trials, the session was ended.

During all stages of MTS training, each stimulus occurred as a sample and as a correct and incorrect comparison an equal number of times (See Appendix B-G). Stimulus combinations (sample, S+, and S-) were balanced so that all stimuli were paired together equally often. For the Odor Arena, location of the stimuli was pseudo-randomly selected from the 18 possible hole locations for both comparison cups such that no single hole location contained correct or incorrect comparisons for more than two consecutive trials. The sequence of stimulus positions also varied across sessions. For the Operant MTS Chamber, the location of comparison stimuli was counterbalanced across the left and right positions. Neither the left nor the right position contained the correct or incorrect stimulus for more than two consecutive trials. Generally, each phase of the experiment consisted of ten versions of stimulus position configurations that were cycled through in a random order. In some instances, subjects remained at a given stage for a prolonged period, and so additional versions of the data collection sheets were created to prevent control by stimulus location resulting from extensive training with the same position sequences. These randomizing and counterbalancing techniques were used to control for any associations between stimuli and stimuli locations that may have formed.

A mastery criterion was designated at 90% or greater performance on two consecutive sessions. Performance was calculated by dividing the number of trials in which a correct response was made to S+ by the total number of trials possible in the session, typically 24. This proportion of correct trials to all possible trials per session was then multiplied by 100 to generate a percent correct measure. To meet this criterion,

subjects could score incorrectly on only two of 24 trials, or 91.67%. This criterion was used for the entire experiment for many subjects; however, some subjects required a criterion of 87.5% or greater for two consecutive sessions in order to progress through the experimental phases in a timely fashion. This criterion required three or fewer incorrect responses out of 24 trials.

For all three experiments, responses during testing trials with novel stimuli or novel stimulus relations were reinforced. Research with nonhumans has shown that using extinction during testing trials disrupts performance and elicits emotional behaviors (Peña et al., 2006; Schusterman & Kastak, 1993). Therefore, any subsequent presentation of those stimuli or relations were considered, and scored as, training trials rather than probe trials. Thus, only the first presentations of novel stimuli and novel stimulus combinations during test sessions were considered in the evaluation of emergent relations.

Special Procedures

Throughout the study it became necessary to develop additional measures and adapt procedures to ongoing changes in subject's behaviors, to unexpected questions, and to assess any performance issues that arose. For example, pellet detection probes were incorporated into all three experiments for all subjects to control for responding due to pellet odor detection. Such probes were called no-baited trials and were conducted such that the correct comparison stimulus (S+) was unbaited. Once subjects responded to the S+, either by displacing the lid in the Arena or by contacting the surface of the sand in the MTS Chamber, the reinforcer pellet was delivered into the stimulus cup with tweezers by the experimenter. These no-baited trails were conducted during baseline and testing sessions. To control for scoring biases, occasional sessions were conducted with a

second, blind rater present. Scoring records were compared for the principal and blind raters to determine inter-observer agreement. Additionally, by video recording experimental sessions, we are able to review sessions and conduct further inter-observer agreement tests in the future if needed. Video recording sessions also allowed for evaluation of experimenter cuing, as the videos included audio and visual information during the entire session of both the experimenter and the subject.

Another set of special procedures adapted during the study was intended to facilitate stimulus control and to advance subjects through training phases more rapidly. These instances, or 'therapy' procedures, were often used to correct behavioral problems or to encourage control by stimulus odor. For example, some subjects in the Arena underwent a fastened lid therapy wherein the perforated plastic lids were snapped tightly onto the stimulus cup below, thereby preventing access to the scented sand. Lid fastening was applied to the incorrect stimuli (S-) only and was designed to increase the response cost for displacing the S- lid if encountered before the S+ stimulus. Lid fastening was also intended to increase sampling behaviors to the S- stimuli such that subjects encountered the S-, but did not displace the lid. For example, in many cases, subjects in the Arena would respond to the first stimulus encountered, regardless of correctness. Lid fastening was intended to make this behavior more costly, and was often highly effective in increasing the occurrence of S- sampling behaviors upon removal of the intervention.

Unfortunately, the lid-fastening intervention made scoring of responses more difficult, since it prevented their displacement and therefore affected our response definition. In this case, the response definition for fastened S- lids was redefined as any contact with the lid with the paws, nose, or face. Typically, subjects would remove the

entire cup, with the fastened lid, from the hole for the first few trials or sessions with the lid fastening therapy in place. This behavior was clear and easy to score as a response. After some exposure to the fastened lids, subjects' responses became more cryptic, often using their nose or face to contact the lid. Because their behavior towards the fastened lids became so discreet, experimenter scoring was inevitably affected. So, in general, subjects' performances during lid fastening therapy were scored as being much more accurate than their performances without fastened lids, undoubtedly resulting from the difficulty in detecting responses to S-. As a result, any criterion-level performances achieved during lid fastening therapy were not considered for advancement.

Another form of therapy occurred in the MTS Chamber for similar purposes. Because the chamber only varied comparison stimulus positions on the left and right sides, subjects often showed side preferences. In an attempt to overcome these biases, a time out procedure was used and consisted of removing the MTS tray for 10 seconds if the preferred side was contacted first after the tray was inserted. Most subjects assigned to the chamber were eventually transferred to the Arena, and the remaining chamber subject (H7) continued to demonstrate a strong side preference regardless of the time out.

The third therapy procedure was used for some Arena subjects and was called the dummy procedure. Similar to the lid fastening therapy, the dummy procedure was designed to increase response costs for responding to the first stimulus encountered. In addition to the S+ and S- stimuli in the Arena, three 'dummy' stimuli of the same odor were added to increase sampling behaviors. Dummy odors were never the same as the S+ or S- odors being used and responses to dummy stimuli were never reinforced. This therapy was marginally effective and was instituted for only a few subjects. All three

therapies and control procedures (no-bait trials, inter-observer agreement) are noted and described in detail for each applicable subject.

Design and Statistical Analyses

The current study used a small-n probe procedure. The probe design consisted of within-subject repeated baseline training and probe tests. Data for the current study were quantified using percent correct, which was calculated by dividing the number of correct trials per session by the total number of trials per session (typically 24). Percent correct for pellet detection probe trials and emergent identity, symmetry, transitivity, and equivalence probe trials was scored in addition to baseline accuracy, and was also calculated by dividing the number of correct trials by the total number of probe trials in that session. Emergent identity, symmetry, transitivity, and equivalence data were analyzed using one-tailed binomial tests to determine statistical significance ($p<.05$).

EXPERIMENT 1

Experiment 1 attempted to systematically replicate the generalized identity matching observed in Peña et al. (2006) with changes in apparatus and different reinforcer conditions. Experiment 1 examined generalized identity MTS using a modification of the multiple exemplar training procedure used in Peña et al. (2006). Experiment 1 was also designed to be a platform for later experiments that would test for equivalence relations.

METHOD

Subjects

Nine experimentally naïve male H-SD rats were used during Experiment 1 (F6, F3, F5, H7, H8, G13, G8, I30, J6).

Apparatus

Both the Odor Arena and the Operant MTS Chamber were used during Experiment 1. The first four Experiment 1 subjects (Arena F6, F3, G13; Chamber H7, F5, H8) were randomly assigned to apparatus condition. The Chamber was used only initially and only one subject experienced the chamber for the entirety of training and testing, across both Experiment 1 and 2 (H7). All other subjects trained in the chamber were eventually transferred to the Arena (F5, H8) during Experiment 1. Three additional subjects were added during Experiment 1 (G8, I30, J6) and were all trained in the Odor Arena.

Stimuli

Stimuli used during Experiment 1 were the same as described in the General Method.

Procedure

Experiment 1 Training: Identity Matching-To-Sample Baseline

Stimuli used during identity MTS were presented and tested for generalization while still novel to the subjects, then added to the training set, trained to criterion, and then discarded from further identity training. A second set of stimuli were then introduced and tested for generalized matching, and then became the new training set. Following criterion performance, this second set of stimuli was dropped from training, and a third set was introduced for testing and training in the same fashion as the previous set. Pilot research conducted prior to the current study indicated that such procedures established strong baselines and generalized to novel stimuli during probes (Thomas, 2006) without the inclusion of previous baseline stimuli in subsequent testing sessions.

Table 1 lists the two stimulus classes for all subjects across all three experiments. The order of stimulus presentation for each subject during Experiment 1, identity MTS, is shown in Table 2. Identity MTS procedures trained subjects using physically identical sample and correct comparison (S+) stimuli. Subjects were required to match the sample stimulus to its identical comparison counterpart (S+), and not the dissimilar, incorrect comparison (S-). During identity MTS baseline training, subjects were required to remain in training with a given set of stimuli until overall performance on two consecutive sessions was equal to, or above, 90 percent correct. Exceptions are noted within tables.

Experiment 1 Training: First Baseline Stimulus Set, A_1 and A_2

All subjects began training with two different stimuli, one from each designated class based (see Table 2, Figure 4). The first two stimuli, labeled A_1 and A_2 respectively, served both as samples and correct (S+) and incorrect comparisons (S-), conditional upon the sample stimulus for that particular trial. For example, when **A_1** served as the sample stimulus, reinforcement (denoted by arrow) occurred following responses to the identical, correct comparison (**A_1+**), while no reinforcement occurred following responses to the incorrect comparison stimulus (**A_2-**). Similarly, when the sample stimulus was **A_2**, reinforcement occurred following responses to the identical, correct comparison (**A_2+**), and not for responses to the incorrect comparison (**A_1**).

Table 2. Experiment 1 Identity Matching-To-Sample Stimulus Presentation Order. Asterisks denote stimuli that were created using liquid odorants rather than dry, powdered spices. A superscript numeral one indicates the last stimulus set completed before identity MTS training concluded.

Subject	F6		F3		F5		H7	
Stimuli	1 (sugar)	2 (sugar)	1 (sugar)	2 (grain)	1 (sugar)	2 (sugar)	1 (sugar)	2 (grain)
A	Nutmeg	Dill	Clove	Rosemary	Nutmeg[1]	Dill[1]	Clove	Rosemary
B	Celery	Cinnamon	Onion	Sage	Coriander	Sage	Onion	Sage
C	Clove	Ginger	Sumac	Thyme	Turmeric	Garlic	Sumac	Thyme
D	Oregano	Onion	Bay	Paprika	Thyme	Clove	Bay	Paprika
E	Thyme	Coriander	Marjoram	Cumin	Paprika	Mustard	Marjoram	Cumin
F	Mustard	Sumac	Turmeric	Oregano	Marjoram	Sumac	Turmeric	Oregano
G	Cumin	Marjoram	Cinnamon	Nutmeg	Celery	Cumin	Cinnamon	Nutmeg
H	Garlic	Rosemary	Ginger	Mustard	Ginger	Oregano	Ginger	Mustard
I	Sage	Turmeric	Dill	Celery	Onion	Bay	Dill	Celery
J	Bay	Paprika	Coriander	Garlic	Rosemary	Cinnamon	Coriander	Garlic
K	Sassafras	Hickory	Fennel	Allspice			Fennel	Allspice
L	Worcestershire	Orange	Beet	Carob			Beet	Carob
M	Savory	Fennel	Lime	Tomato			Tomato	Lime
N	Carob	Allspice	Hickory	Savory			Raspberry	Savory
O	Beet	Caraway	Orange	Worcestershire			Orange	Worcestershire
P	Lime	Tomato	Caraway[1]	Sassafras[1]			Caraway[1]	Sassafras[1]
Q	Spinach	Grape*	Spinach	Peppermint*			Spinach	Peppermint*
R	Raspberry	Peppermint*	Raspberry	Grape*			Hickory	Grape*
S	Almond*	Maple*	Almond*	Maple*			Almond*	Maple*

Subject	H8		G13		G8		I30	
Stimuli	1 (sugar)	2 (sugar)	1 (sugar)	2 (grain)	1 (sugar)	2 (grain)	1 (sugar)	2 (sugar)
A	Dill	Nutmeg	Clove	Rosemary	Mustard	Thyme	Sumac	Oregano
B	Celery	Cinnamon	Onion	Sage	Caraway	Orange	Mustard[1]	Bay[1]
C	Clove	Ginger	Sumac	Thyme	Savory	Oregano	Cinnamon	Clove
D	Oregano[1]	Onion[1]	Bay	Paprika	Marjoram	Garlic	Tomato	Raspberry
E	Garlic	Thyme	Marjoram	Cumin	Raspberry	Celery	Sage	Nutmeg
F	Rosemary	Mustard	Turmeric	Oregano	Beet	Cumin	Turmeric	Garlic
G	Savory	Sage	Cinnamon	Nutmeg	Rosemary	Turmeric	Fennel	Lime
H	Paprika	Turmeric	Ginger	Mustard	Sage	Spinach	Orange	Carob
I	Cumin	Bay	Dill	Celery	Paprika	Ginger	Rosemary	Celery
J	Coriander	Marjoram	Coriander	Garlic	Carob	Tomato	Thyme	Onion
K			Fennel	Allspice	Sassafras	Bay	Paprika	Cumin
L			Beet	Carob	Sumac	Cinnamon	Savory	Marjoram
M			Lime	Tomato	Nutmeg	Onion	Dill	Worcestershire
N			Hickory	Savory	Dill	Allspice	Coriander	Allspice
O			Orange	Worcestershire	Clove	Garlic	Caraway	Beet
P			Caraway	Sassafras	Lime	Hickory	Grape*	Sassafras
Q			Spinach	Peppermint*	Fennel	Almond*	Spinach	Ginger
R			Raspberry	Grape*	Peppermint*	Grape*	Hickory	Peppermint*
S			Almond*	Maple*	Maple*	Worcestershire	Almond*	Maple*

	G13		Subject	J6	
	1 (sugar)	2 (grain)	Stimuli	1 (sugar)	2 (grain)
T	Peach*	Chocolate*	A	Raspberry	Sage
U	Lemon*	Vanilla*			
V	Root beer*	Coconut*	B	Sumac	Fennel
			C	Carob	Cumin
W	Pineapple*	Strawberry*	D	Coriander	Nutmeg
X	Banana*	Brandy*			
Y	Apple*	Rum*	E	Celery	Marjoram
			F	Cinnamon	Ginger (dropped)
			G	Paprika	Onion
			H	Dill	Lime
			I	Tomato	Allspice
			J	Caraway	Spinach
			K	Worcestershire	Beet
			L	Turmeric	Mustard
			M	Savory	Thyme
			N	Garlic	Sassafras
			O	Rosemary	Bay
			P	Oregano	Orange
			Q	Hickory	Almond*
			R	Peppermint*	Clove
			S	Maple*	Grape*

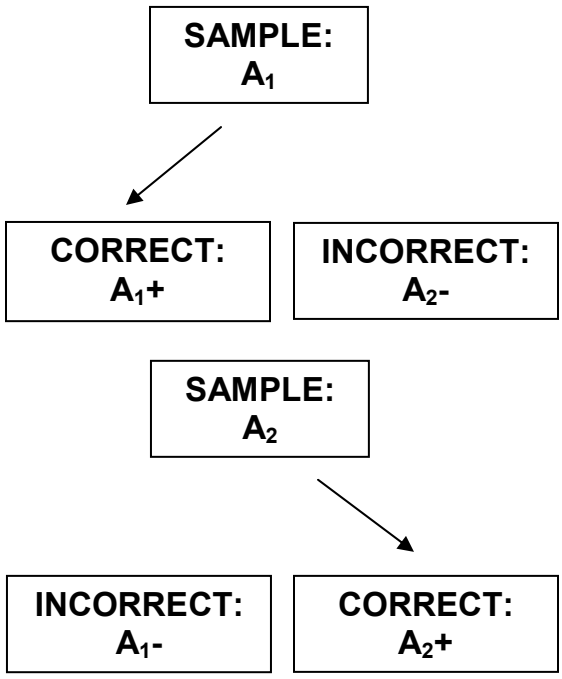


Figure 4. Identity Matching Training Configurations: A_1 & A_2

Additionally, **A_1** and **A_2** stimuli occurred as samples and correct and incorrect comparisons within and across sessions with equal frequency. Comparison stimulus (S+ and S-) locations were randomized such that any two of the possible 18 hole positions contained either the S+ or S- stimulus on each trial. Subjects continued training at this beginning, "A" stimuli, stage until criterion-level performance, or two consecutive sessions with overall performance of 90 percent or greater, was achieved.

Experiment 1 Probes: Generalized Identity Matching-To-Sample Testing

Initial trials during the first session following advancement to any new stimulus set were scored slightly differently from the rest of the sessions during training. All responses to sample and correct comparison stimuli were reinforced during probe trials. Therefore, only the first occurrence of each novel stimulus as a sample within this new set of stimuli was considered a completely novel trial. These Novel Probe trials were scored in addition to the rest of the familiar trials within that session. Performance on Novel Probe trials was critical in order to establish whether generalized identity matching had emerged from identity matching training. For example, subjects were trained to match the sample stimulus **A_1** to the identical, correct comparison (**A_1+**) when stimulus **A_2** is the incorrect comparison (**A_2-**), as well as match sample stimulus **A_2** to the identical, correct comparison (**A_2+**) when A_1 is the incorrect comparison (see Figure 5). If subjects learned to identity match, responding to the correct novel comparison (X_1+ or X_2+) should occur in the presence of the identical novel sample stimulus during probe trials. Novel stimuli are denoted as X_1 or X_2, and arrows indicate correct responses (see Figure 5). Trials which include novel stimulus configurations, or combinations of sample and comparison stimuli which have never been presented in a given configuration, were

also scored separately and accurate performances were interpreted as further evidence of generalization of the identity relation concept. These trial types were labeled Novel Combination Probes.

After criterion, two consecutive sessions with 90% or greater performance accuracy, was met for the first trained stimulus pair labeled "A" (**A_1** and **A_2**), a second set of six novel stimuli, denoted "B, C, and D", (specifically B_1, B_2, C_1, C_2, D_1, D_2), were introduced for testing (see Table 2). The first set of two stimuli (A_1 and A_2) were removed and no longer used for identity matching training. Novel stimuli B_1, B_2, C_1, C_2, D_1, and D_2 were first used for probing for the identity relation (see Figure 6). For example, if the identity relation generalized from training to these novel stimuli, subjects should have chosen the B_1 comparison when B_1 was also the sample stimulus, D_2 when D_2 was the sample stimulus, and C_1 when C_1 was the sample, and so on. Since reinforcement occurred for all trials, novel stimuli became training stimuli during the first session in which they were presented. So, the current set of stimuli (B_1, B_2, C_1, C_2, D_1, and D_2) became training stimuli after generalized matching was assessed during their initial presentations.

Experiment 1 Training, continued: Identity Matching-To-Sample Training

Training continued in much the same fashion as before, using the "A" stimuli, but with a greater number of stimuli and associated constraints. First, any one of the six stimuli from this second set served as the sample stimulus during a given trial, as depicted in the top box of Figure 6. Second, the same stimulus also served as a correct comparison (S+), as depicted by the arrows and lower left and right hand boxes. So, if **B_1**

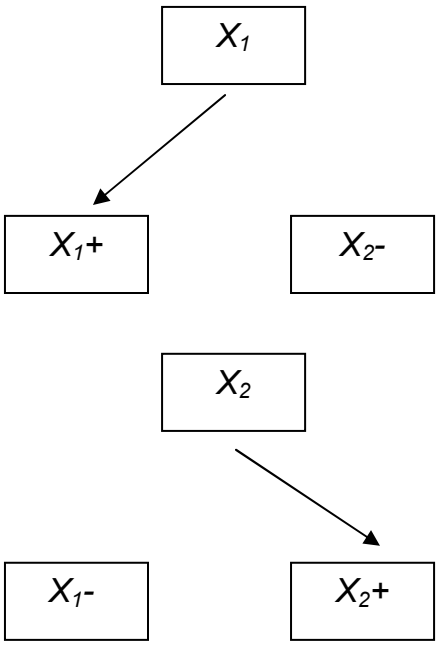


Figure 5. Generalized Identity Matching Transfer Probe Configuration

was the sample, **B_1** was also the S+, or if **C_2** was the sample, **C_2** was also the S+, and so forth. Third, any one of the three stimuli from the opposing designated class served as the incorrect comparison (S-), depending on the sample and correct comparison. So, class 1 stimuli would serve as the S- comparison if the sample and S+ stimuli were from class 2, and class 2 stimuli would serve as the S- comparison if the sample and S+ stimuli were from class 1. So, if **D_1** was the sample and S+, then either **B_2**, **C_2**, or **D_2** served as the S-, and so forth.

Again, all stimuli occurred equally as samples and as correct and incorrect comparisons. Stimuli from class one (B_1, C_1, and D_1) were equally paired with stimuli from class two (B_2, C_2, and D_2) and comparison locations (hole position for Arena or left and right for MTS Chamber) were randomized and counterbalanced, respectively. Subjects remained at the B, C, and D training stage until criterion performance was met once again.

Following criterion performance, the trained B, C, and D stimuli were removed from training sessions, and another novel set of stimuli "E, F, and G" (specifically *E_1, E_2, F_1, F_2, G_1,* and *G_2*) was introduced for testing and training. Novel Probe and Novel Combination Probe trials were again scored during the first session following the advancement to evaluate transfer of the identity relation to this set of stimuli. Once performance had reached criterion levels for the EFG set, the EFG stimuli were removed and the next, novel stimulus set (HIJ) was introduced. This removal and introduction procedure was repeated (A_1 through X_1, A_2 through X_2) until performance during Novel

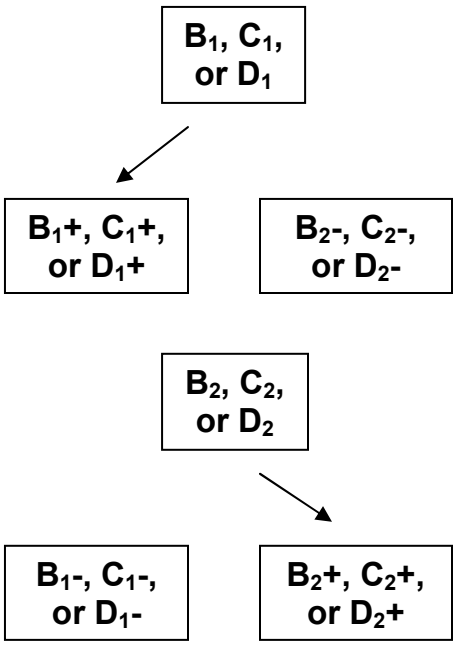


Figure 6. Identity Matching Training Configurations: B_1, B_2, C_1, C_2, D_1, D_2.

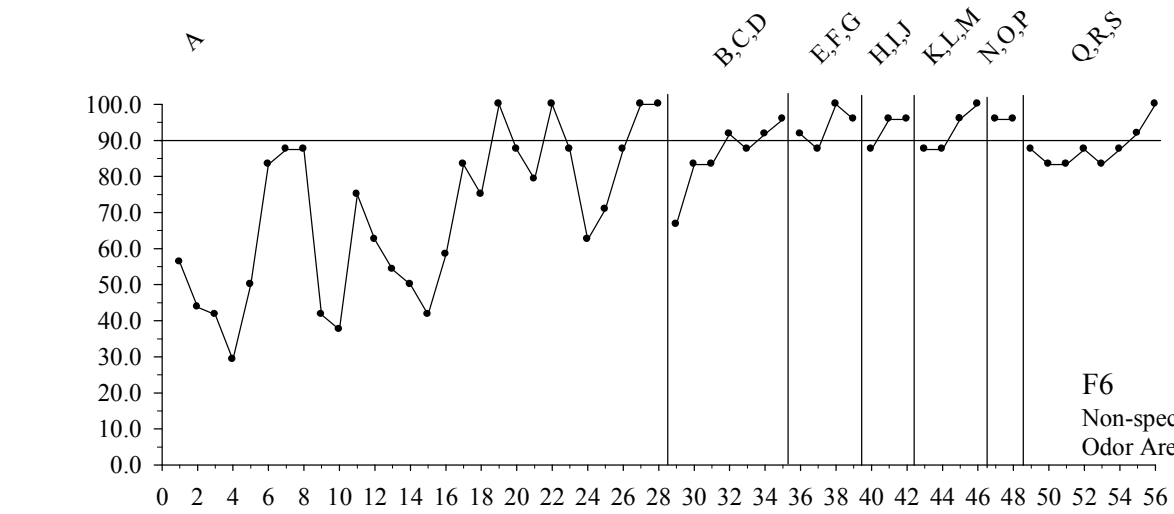
Probe trials for two (or three, subjects G13 and G8) consecutive presentations of novel stimulus sets was statistically significant (binomial test, $p < .05$).

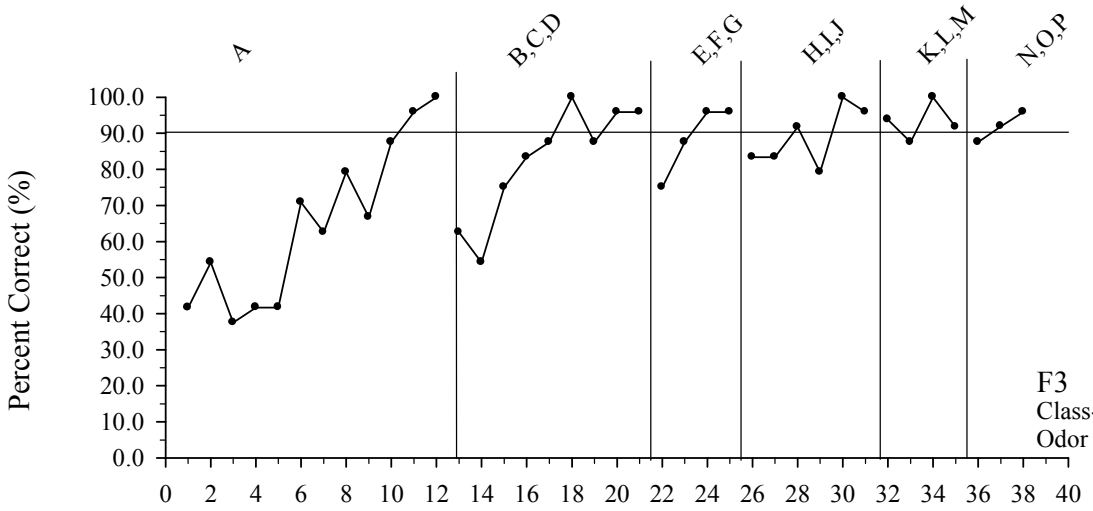
Binomial significance tests were used to evaluate whether or not subjects' performances were significantly different from what would be expected by chance. Each trial presented two possible outcomes, or two comparison stimuli, and each comparison stimulus had a .50 probability of being correct on any given trial. However, correctness of the comparison stimuli was dependent on the sample stimulus for each trial. Thus, if subjects' responding was not under the control of the relation between sample and S+, then responding should have been at chance levels, or 50%. If subjects' responding was under the control of the sample-S+ comparison relation, then responding should have exceeded chance levels (>50%). Binomial significance tests allowed us to evaluate whether subjects' responding was different than what could be expected by chance ($p>$ .05). These tests were calculated by comparing the number of trials with a correct response ("hits") to the total number of possible trials and comparing that probability to .50.

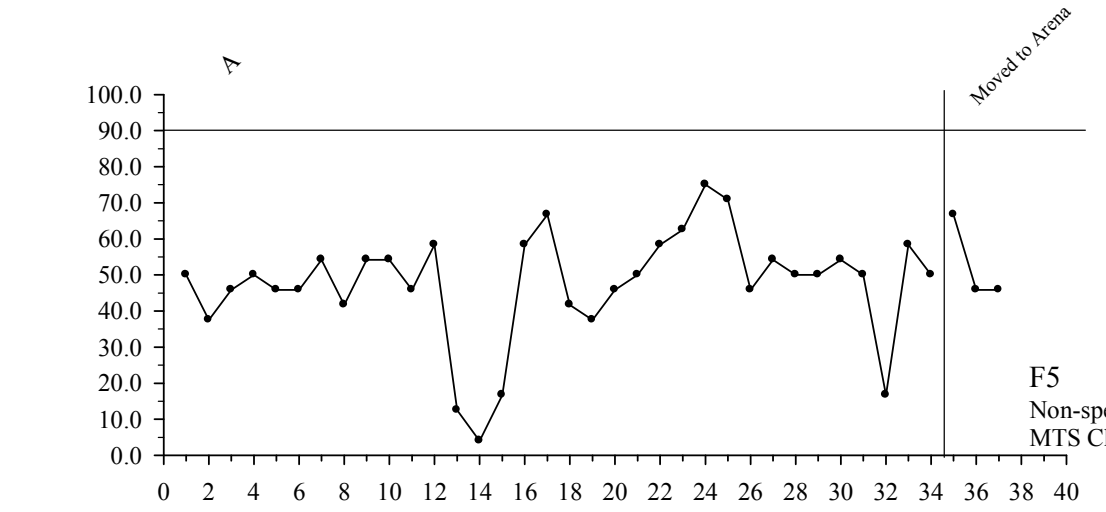
EXPERIMENT 1 RESULTS AND DISCUSSION

The graphs in Figure 7 show percent correct for consecutive sessions for all nine subjects participating in Experiment 1 (generalized identity MTS). Criterion was initially set at 90% for all subjects, as indicated by the horizontal line in each graph. Exceptions were made (criterion of 87.5%, or 21/24 trials) for some subjects to promote timely progression through stimulus sets when responding was highly accurate, but did not quite reach 90% or greater on two consecutive sessions. See Table 3 for details about the

Figure 7. Experiment 1 Percent Correct across Stimulus Set Presentations. Each data point represents percent correct for one session, or the proportion of correct trials divided by all possible trials. Vertical lines (panels) within each graph indicate changes in stimulus sets; each panel is labeled with the respective set. The horizontal line indicates the original criterion (90%). Any criterion level performance that is not followed by a vertical phase change line occurred during a special training procedure.

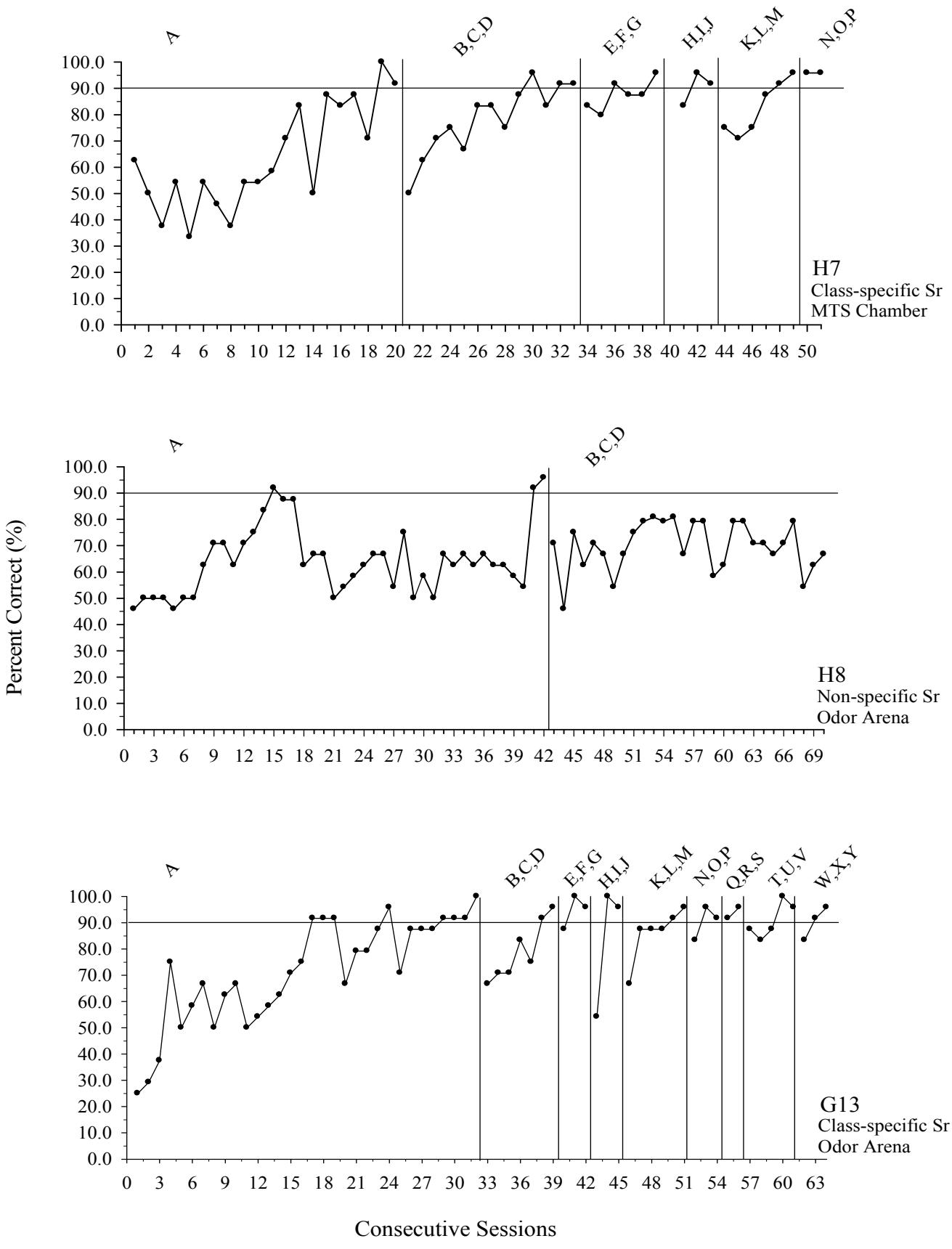


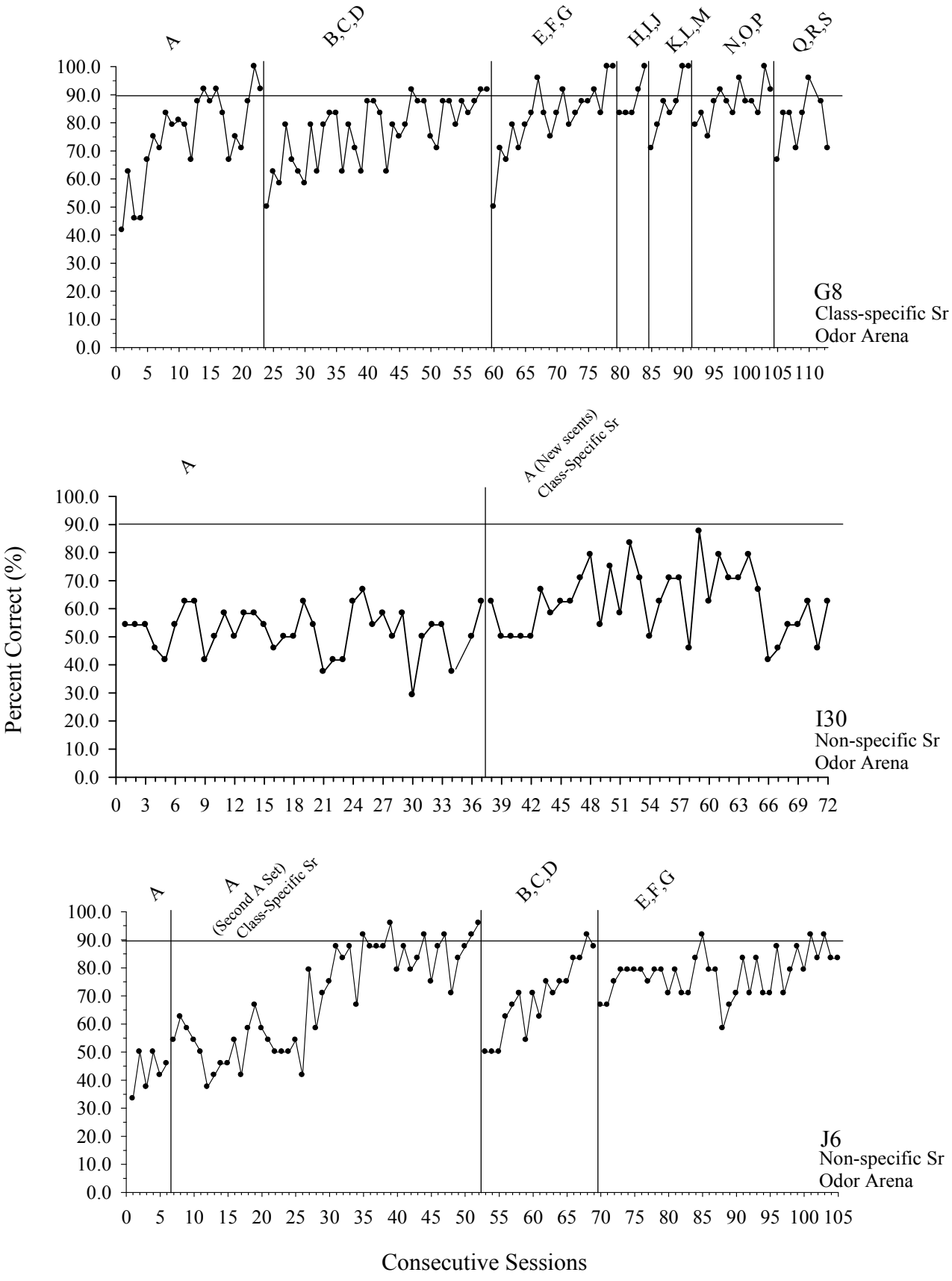




Percent Correct (%)

Consecutive Sessions


A
B,C,D
E,F,G
H,I,J
K,L,M
N,O,P
Q,R,S
T,U,V
W,X,Y
Percent Correct (%)
Consecutive Sessions
H7
Class-specific Sr
MTS Chamber
H8
Non-specific Sr
Odor Arena
G13
Class-specific Sr
Odor Arena


A
B,C,D
E,F,G
H,I,J
K,L,M
N,O,P
Q,R,S
G8
Class-specific Sr
Odor Arena
A
A (New scents)
Class-Specific Sr
I30
Non-specific Sr
Odor Arena
A
A
(Second A Set)
Class-Specific Sr
B,C,D
E,F,G
J6
Non-specific Sr
Odor Arena
Percent Correct (%)
Consecutive Sessions

changed criterion. Also, subjects would often require intervention procedures in order to learn the discriminations for a given set. Occasionally, subjects would reach criterion during sessions when these intervention procedures were being used, however, subjects were not advanced until criterion was met in the absence of such procedures. Table 3 indicates points in training at which special procedures (time out, S- lid fastening) were in use. Table 3 details information about training sets that included procedural changes as well as information about the number of sessions to criterion for each subject across all exemplars.

Sessions to Criterion

Table 3 shows the number of sessions to criterion for each subject across stimulus sets in Experiment 1 as well as the summed total number of sessions and the mean number of sessions for each set and for each subject. Subjects F5, H8, and I30 were dropped from Experiment 1 after lengthy training and after numerous intervention procedures failed to promote learning of the stimulus set. G8 failed to meet criterion for stimulus set QRS, but was advanced to Experiment 2 training due to his age.

The number of sessions to meet criterion for any given stimulus set was variable across subjects, thus mean measures for sessions to criterion are not necessarily representative of all subjects. For example, the mean number of sessions to criterion for set A was 23, whereas subject F3 met criterion for stimulus set A in 12 sessions and subject I30 required 72 sessions, and never met criterion. In spite of these between-subject differences, there appears to be a general downward trend in the number of sessions to criterion across stimulus sets (across exemplars) for those subjects that progressed through several sets, as well as for the mean measures. For example, the

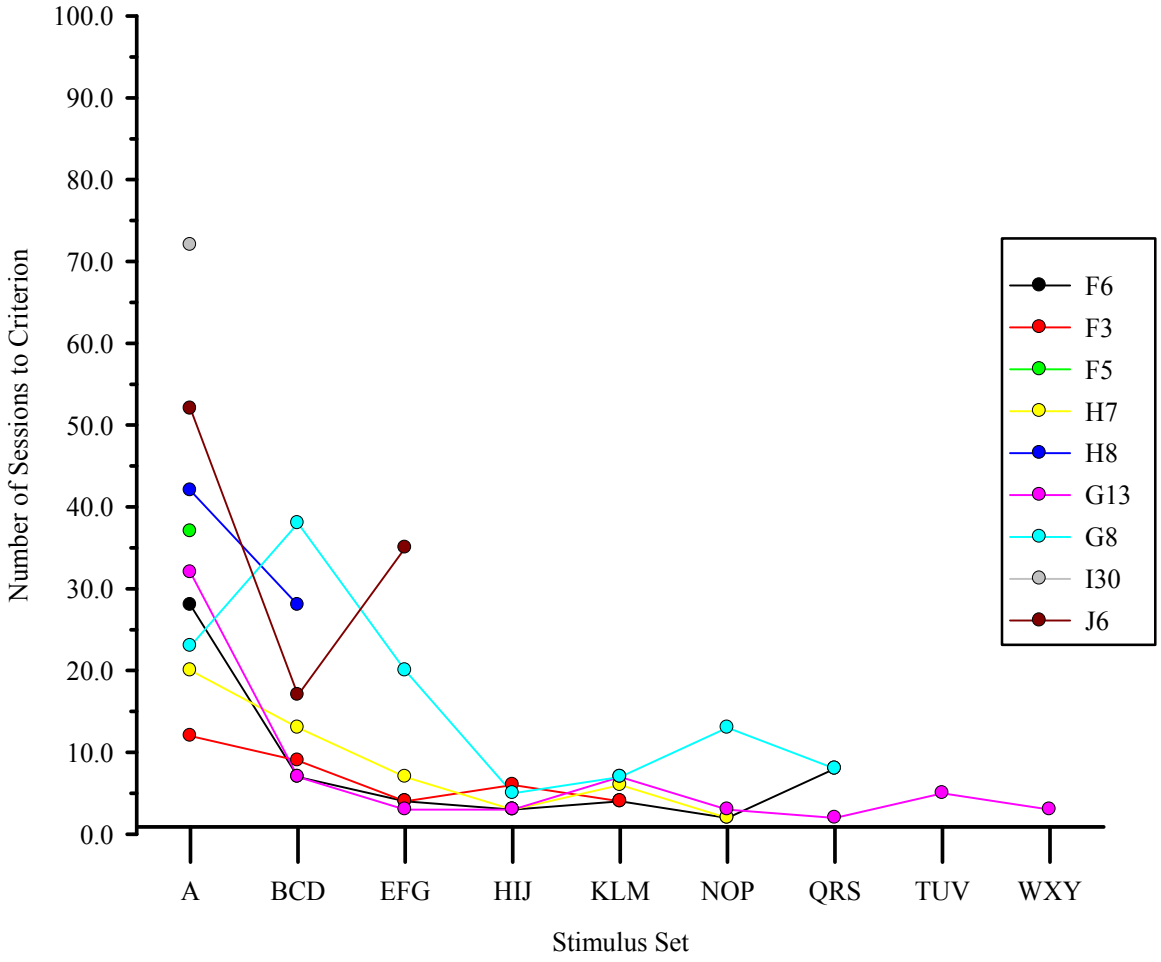
Table 3. Experiment 1 Number of Sessions to Criterion across Stimulus Set Presentations. A superscript numeral one indicates sets in which subjects did not meet criterion for particular stimulus set. A superscript numeral two indicates the modified criterion of 87.5%. Superscript numeral threes indicate that therapy intervention procedures were used during training for that particular set. Any criterion level performances during therapy were not considered for advancement.

Subject / Stimuli	F6	F3	F5	H7	H8	G13	G8	I30	J6	Total	Mean
A	28	12	37[1, 3]	20[3]	42	32[3]	23	72[1, 3]	52	318	35.3
B, C, D	7	9		13	28[1, 3]	7	38		17[2, 3]	119	17
E, F, G	4	4		7		3	20		46	84	14
H, I, J	3	6		3		3	5			20	4
K, L, M	4	4		6		7	7			28	5.6
N, O, P	2			2		3	13			20	4
Q, R, S	8					2	8[1]			18	6
T, U, V						5				5	5
W, X, Y						3				3	3
Total	56	35	37	51	70	65	114	72	115	615	93.9

total and mean number of sessions to criterion for set A ($\sum$= 318, $\bar{x}$ = 35.3) are approximately 63% and 52% greater, respectively, than the total and mean number of sessions to criterion for set BCD ($\sum$= 119, $\bar{x}$ = 17). Therefore, overall or on average, subjects tended to require approximately 37% fewer sessions to meet criterion on the second training set than the first; they showed evidence for savings. In terms of individual subject data, for those subjects who received at least 3 or more exemplars, the number of sessions to criterion drops dramatically following the first training set, sometimes by as much as 75% (F3) and 78% (G8).

Figure 8 also shows the number of sessions to criterion within each stimulus set during Experiment 1 for all subjects and allows for visual inspection of the downward trend hinted in Table 3. It appears as if subjects more readily learned the discriminations (required fewer sessions to criterion) as they progressed through identity MTS exemplars, especially when compared to performances during set A. There also appears to be a large amount of variability across subjects in the number of sessions and number of exemplars required to meet the criterion to advance to Experiment 2 (binomial significance $p<.05$) for novel probes on last two (or three, G13 and G8) probe sessions (See Experiment 1 Method). For example, subject F3 achieved binomial significance and was advanced to Experiment 2 in 35 trials whereas subject G8 required 114 trials, did not achieve statistically significant performance levels, and was only moved because of time and age limitations. Similarly, G13 received several additional exemplars (up to set WXY), exhausted all of the available odors, and was still unable to meet the standard advancement criterion.

Figure 8. Experiment 1 Number of Sessions to Criterion across Stimulus Set Presentations for each subject.



Baseline Matching-To-Sample

Of the nine subjects in Experiment 1, five successfully mastered several baseline stimulus sets (F6, F3, H7, G13, and G8; see Figure 7). Although there were individual differences in the duration of training for each set, these five subjects generally achieved criterion within days or a few weeks of training. Following set A, most of these subjects required several sessions to advance to the next stimulus set. Some subjects (F6, F3, H7, G13) often met criterion in the minimum possible number of sessions - two. For example, subject F6 met criterion for set NOP in two sessions, G13 did the same for set QRS, and H7 during set NOP. F6, F3, H7, and G13 achieved criterion in three and four sessions on multiple other occasions as well. For these animals, baseline discrimination learning was rapid and performance was relatively stable across stimulus sets during Experiment 1. Additionally, these five subjects showed savings across stimulus sets in terms of the number of sessions required to meet criterion, meaning that as exemplars were added to the training history, subjects met criterion in fewer and fewer sessions.

The remaining four subjects (F5, I30, H8, J6) had greater difficulty learning baseline MTS discriminations than the other five subjects (see Figure 7). Despite extensive training, subjects F5 and I30 did not advance beyond the first training stimulus set (A) and therefore never received probes for generalization. Subject F5 was initially trained in the Operant MTS Chamber with non-specific reinforcement, sugar pellets only. His chamber experience lasted 34 sessions and he regularly responded based on stimulus position rather than odor. After

implementing side bias interventions (e.g. S+ located in non-preferred location only) and a 10s time out procedure, F5 was transferred to the Odor Arena for several sessions. The transfer to the Arena, with the variable stimulus positions, was also ineffective and resulted in several sessions of no responding. F5 received a total of 37 sessions with a maximum accuracy (during sessions without therapy interventions) of 67%, equivalent to 16/24 trials. Similarly, I30 was initially trained with sugar reinforcers only and was unable to learn the A discriminations after 36 sessions, S- lid fastening, and a 10s time out procedure (sessions 27-36). Beginning with session 37, I30 received class-specific reinforcers and was assigned two new stimuli, still within the set A format. S- lid fastening remained in effect until session 64. From that point on, his training included a dummy S- intervention that was also ineffective. I30 received a total of 72 trials and performed at a maximum of 67% (16/24) accuracy during a single session that did not include any therapy interventions. Thus, after several types of therapy procedures were unsuccessfully used on both subjects, they were both dropped from the experiment. It is unclear why F5 and I30 were unable to acquire set A discriminations, while the majority of the other subjects did so quite readily. For F5, the combination of the Operant MTS Chamber and non-specific reinforcement may have resulted in a more complex task that would have required many more sessions to master. For I30, responding was rarely above chance levels and only began to climb when S- lid fastening was used. As previously mentioned, the lid fastening intervention often artificially inflated performance measures because responses were more difficult to score. Thus, the scores

observed between session 28 and 64 may have been higher than his true performance accuracy. Perhaps the use of only sugar reinforcers required exposure to longer durations of training or the development of other therapy interventions for mastery to occur. The data for F5 and I30 will not be further discussed.

Like F5 and I30, subjects H8 and J6 also required many more baseline training sessions and therapy intervention procedures to meet criterion than most of the other subjects. H8 achieved criterion for set A in 42 sessions after a change in apparatus and reinforcer condition and the implementation of a time out intervention. Time outs of 10s were used for eight consecutive sessions (sessions 21-28) without any changes in behavior. On session 30, H8 was transferred to the Odor Arena in the hopes of improving his performance. After the apparatus switch, H8 met criterion for set A in 13 sessions. He then spent 28 sessions training in set BCD and was exposed to fastened S- lids for six sessions. Again, no apparent change in performance was observed, and H8 was dropped from Experiment 1 after a total of 70 sessions. J6 required 52 sessions to meet criterion for set A, although he did so without the use of intervention procedures. He then met criterion for BCD in 17 sessions, albeit under the modified 87.5% criterion. At the time of this manuscript, J6 had 46 sessions of training with set EFG and was still being trained on that stimulus set. During EFG training, J6 was exposed to S- lid fastening therapy (session 13 through session 26), but no noticeable changes were observed. Starting with session 27, ginger (F_2) was dropped from his baseline and replaced with marjoram and onion, the other two class 2 stimuli

in set EFG, for the remainder of EFG training. The removal of ginger was based on the observation that most errors were associated with that scent. A general improvement was observed from the initial removal of ginger during session 27 of EFG through session 37. Thereafter, performance began to decline again and was similarly average at the time of this manuscript, hovering around 70-75%. J6's total number of sessions for Experiment 1 was 115. Regardless of their apparent difficulties with baseline discriminations, both H8 and J6 eventually met criterion for at least one stimulus set (A) and also received at least one novel generalization probe session (BCD).

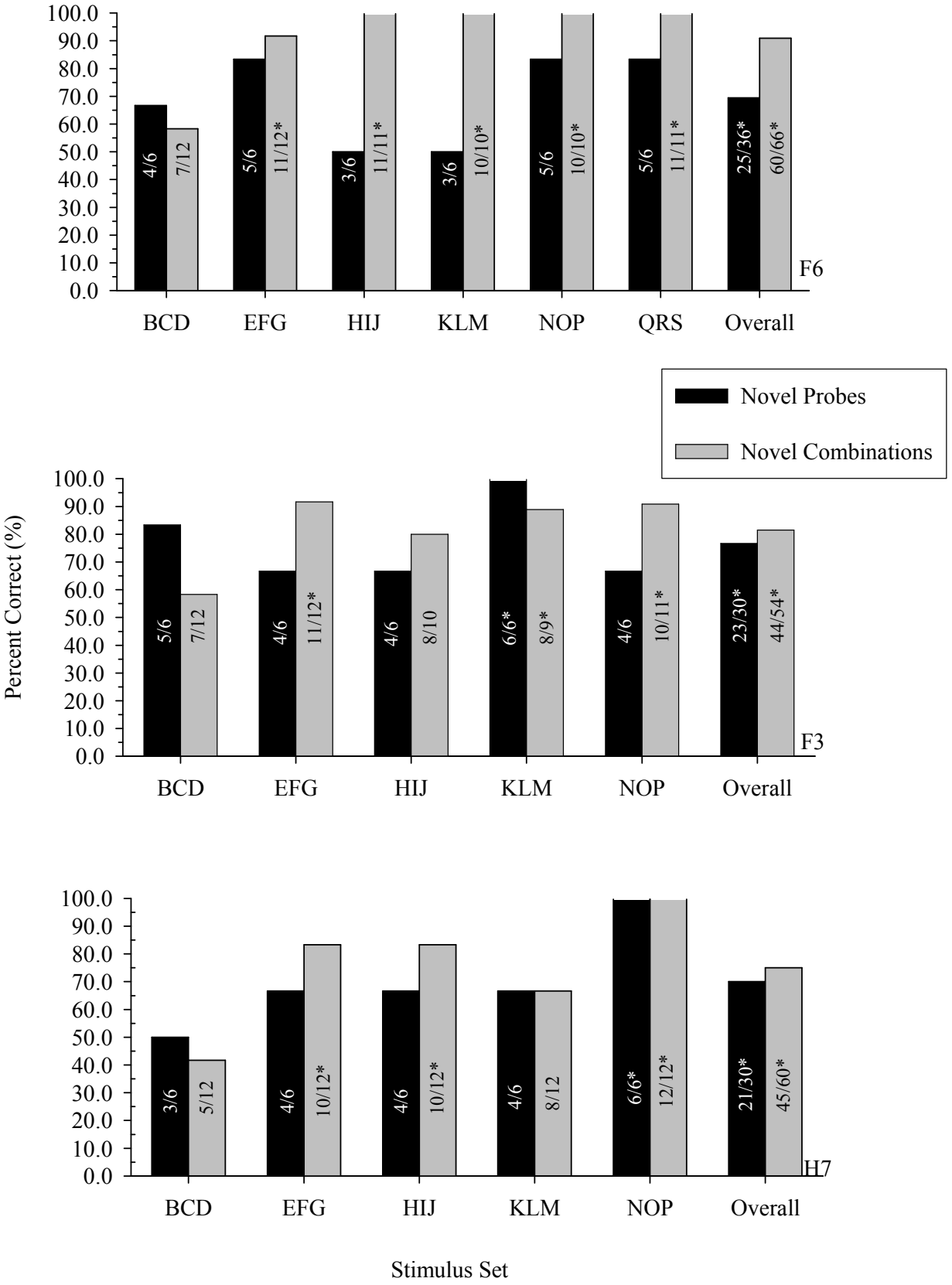
Generalized Matching-To-Sample Probe Results

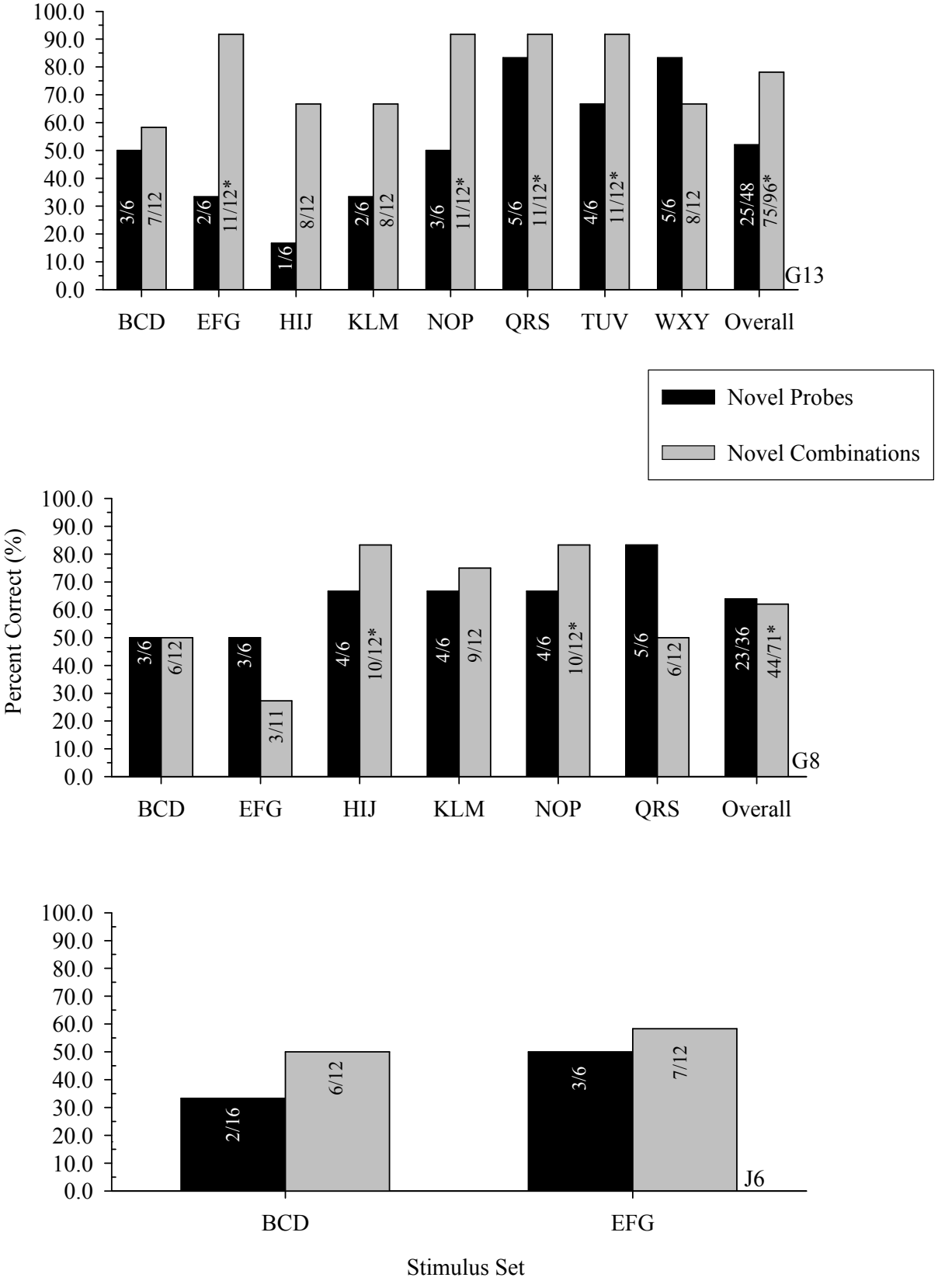
Individual subject performances for those subjects who received generalized identity MTS probes (Novel Probes and Novel Combination Probes) are depicted in Table 4 and Figure 9; this includes all Experiment 1 subjects except F5 and I30. Again, Novel Probe and Novel Combination Probe data are taken during the first session of exposure to a stimulus set. Novel Probe trials were the first six trials during the first session of a new stimulus set (e.g. BCD, EFG). Each trial presented each of six novel stimuli as the sample and the correct comparison. Any stimulus from the opposing class could serve as an incorrect comparison, but was also novel. As soon as each new stimulus was presented as a sample and correct comparison, it was no longer considered novel. For the subsequent trials, Novel Combination Probes, the same six stimuli were presented in various different, novel combinations. Novel Combination Probes were those

Table 4. Experiment 1 Identity Matching-To-Sample Novel Probe and Novel Combination Performance across Stimulus Set Presentations. Asterisks indicate binomial significance ($p<.05$).

Subject / Stimuli	F6	F3	H7	H8	G13	G8	J6
	Novel Probes; Novel Combinations						
B, C, D	4/6; 7/12	5/6; 7/12	3/6; 5/12	4/6; 10/12*	3/6; 7/12	3/6; 6/12	2/6; 6/12
E, F, G	5/6; 11/12*	4/6; 11/12*	4/6; 10/12*		2/6; 11/12*	3/6; 3/11	3/6; 7/12
H, I, J	3/6; 11/11*	4/6; 8/10	4/6; 10/12*		1/6; 8/12	4/6; 10/12*	
K, L, M	3/6; 10/10*	6/6*; 8/9*	4/6; 8/12		2/6; 8/12	4/6; 9/12	
N, O, P	5/6; 10/10*	4/6; 10/11*	6/6*; 12/12*		3/6; 11/12*	4/6; 10/12*	
Q, R, S	5/6; 11/11*				5/6; 11/12*	5/6; 6/12	
T, U, V					4/6; 11/12*		
W, X, Y					5/6; 8/12		
Total	25/36*; 60/66*	23/30*;44/54*	21/30*; 5/60*	4/6; 10/12*	25/48; 75/96*	23/36; 44/71*	5/12; 13/24

Figure 9. Experiment 1 Generalized Identity MTS Novel Probes and Novel Combination Performance. Asterisks denote binomial significance ($p<.05$).


Novel Probes
Novel Combinations
Percent Correct (%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
4/6
7/12
5/6
11/12*
3/6
11/11*
3/6
10/10*
5/6
10/10*
5/6
11/11*
25/36*
60/66*
F6
BCD
EFG
HIJ
KLM
NOP
QRS
Overall
5/6
7/12
4/6
11/12*
4/6
8/10
6/6*
8/9*
4/6
10/11*
23/30*
44/54*
F3
3/6
5/12
4/6
10/12*
4/6
10/12*
4/6
8/12
6/6*
12/12*
21/30*
45/60*
H7
Stimulus Set


100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
3/6
7/12
2/6
11/12*
1/6
8/12
2/6
8/12
3/6
11/12*
5/6
11/12*
4/6
11/12*
5/6
8/12
25/48
75/96*
G13
BCD
EFG
HIJ
KLM
NOP
QRS
TUV
WXY
Overall
Novel Probes
Novel Combinations
Percent Correct (%)
3/6
6/12
3/6
3/11
4/6
10/12*
4/6
9/12
4/6
10/12*
5/6
6/12
23/36
44/71*
G8
2/16
6/12
3/6
7/12
J6
Stimulus Set

trials, usually 12, that included a correct/incorrect comparison combination that had not been previously presented during Novel Probe or earlier Novel Combination Probe trials.

The proportions provided in Table 4 and Figure 9 represent the number of correct Novel Probe or Novel Combination trials compared to the total number of trials given (typically 6 and 12, respectively). Figure 9 shows the percent correct for Novel Probes and Novel Combinations across all stimulus sets and also includes the proportion for each novel probe and novel combination test. Table 4 depicts proportions only. To determine statistical significance for performance accuracy during probes, one-tailed binomial tests were conducted and are represented by asterisks within the table and figure where applicable ($p<.05$). Summed Novel Probes and Novel Combinations across all identity MTS exemplars are also given in both displays. All five subjects who completed several baseline sets (F6, F3, H7, G13, and G8) demonstrated convincing and statistically significant evidence for generalized transfer during several Novel Probes. Of the remaining four subjects, two (H8, J6) received only one or two generalized MTS probe tests – too few to provide a meaningful statistical analysis. The other two (F5, I30) never met criterion and were never given Novel Probe stimuli. Individual subject performances are discussed in detail below.

Subject F6

F6 received six probe tests during Experiment 1 (see Table 4 and Figure 9). His performance during Novel Probes was variable across exemplars, but was generally more accurate in the later sets, NOP and QRS (5/6, 83%), than the

earlier sets, BCD, HIJ, and KLM (3/6, 50%). For example, Novel Probe performance on set EFG (5/6, 83%) was noticeably more accurate than performance on sets HIJ or KLM (3/6, 50%). Performance on Novel Combination Probes also increased following the first set, BCD (7/12, 58%), and remained highly stable across exemplars. All of the Novel Combination Probe performances after the first set were also statistically significant ($p<.05$). The performance range for novel combinations was always above 92% after the first set, and often reached 100%. For the most part, responding in the presence of a novel combination was much more accurate than responding during novel probes, no doubt due to the increased familiarity of the sample stimuli following novel probe trials. Overall, F6 responded correctly on 25/36 (69%) novel probe trials and correctly on 60/66 (91%) novel combination trials, which are both highly significant ($p<.01$, $p<.0001$). Although none of the individual performances on novel probes met binomial criterion for significance, the combined performance during sets NOP and QRS (10/12, $p<.01$) was significant which met the criterion to move to the arbitrary matching phase of the study (Experiment 2).

Subject F3

For F3, performance during Novel Probes and Novel Combination Probes was high during the first probe test for set BCD (4/6, 67%) and remained high throughout testing (see Table 4 and Figure 9). F3 received five probe tests. His performance on Novel Probes ranged from 4/6 to 6/6, but did not increase in a linear fashion across exemplars. F3 achieved statistically significant levels for Novel Probes for set KLM (6/6, 100%, $p<.01$). Only one other subject in

Experiment 1 responded correctly on 6/6 (100%) Novel Probe trials during any one test set (H7). Performance during Novel Combinations for F3 ranged from around chance levels during BCD tests (7/12, 58%) to statistically significant levels during EFG, KLM, and NOP probes (11/12, 92%, $p<.003$; 8/9, 89%, $p<.01$; 10/11, 91%, $p<.005$). In general, his Novel Combination performance was more accurate than Novel Probe performances, but also did not necessarily follow a linear pattern. Summed totals for both Novel Probes (25/36, 69%) and Novel Combinations (44/54, 81%) were statistically significant ($p<.01$, $p<.0001$). Like F6, the summed performance totals for the last two probe sets (10/12), KLM and NOP, for F3 were statistically significant ($p<.01$) and therefore advanced him to Experiment 2.

Subject H7

H7 received five probe tests during Experiment 1 (see Table 4 and Figure 9). H7 showed high levels of accuracy on both Novel Probe and Novel Combinations following the first set, BCD, when performance was around chance levels (3/6, 50%). Following BCD, his performance on Novel Probes increased and remained consistent at 4/6 (67%) for sets EFG, HIJ, and KLM. His peak performance occurred during the final set, NOP (6/6, 100%, $p<.01$). F3 was the only other subject to respond correctly on all six Novel Probe trials for any set during Experiment 1. Moreover, H7 was the only subject to perform at 100% accuracy on both Novel Probes (6/6) and Novel Combinations (12/12) during a single probe session (NOP). Overall, his performance on Novel Combinations increased after the first set (5/12, 42%), plateaued during sets EFG, HIJ, and

KLM (10/12, 83%; 10/12 83%, 8/12, 67%), and was high again during the final probe test for set NOP (12/12, 100%). His performance on Novel Combination trials during sets EFG, HIJ, and NOP were also statistically significant ($p<.01$, $p<.01$, $p<.0002$). Summed totals for both Novel Probes (21/30, 70%) and Novel Combinations (45/60, 75%) were statistically significant ($p<.05$, $p<.0001$). The combined Novel Probe performance from sets KLM and NOP (10/12) was statistically significant ($p<.01$), meeting the criterion to advance to Experiment 2.

Subject G13

Subject G13 received eight generalized MTS probe tests during Experiment 1, the most for any subject tested (see Table 4 and Figure 9). His performance during novel generalization probes was highly variable and did not increase in a linear fashion across exemplars of the identity relation. In fact, if performance across exemplars was plotted as a distribution of scores, an inverted U shape would appear. Performance during the first Novel Probes for set BCD was at chance levels (3/6, 50%) and continued to decline across the next two Novel Probe tests, EFG and HIJ (2/6, 33%; 1/6, 17%). Performance during the next two probe tests increased (2/6, 33%; 3/6, 50%) and then leveled off on the final three probe tests (5/6, 83%; 4/6, 67%; 5/6, 83%). None of his performances during Novel Probes were individually statistically significant. Performance during Novel Combination Probes was much more stable across exemplars, averaging about 9/12 (75%) with a range of 7/12 to 11/12. No apparent upward or downward trends exist across exemplars because G13 consistently scored 11/12 during most of the Novel Combination Probe tests, which is statistically

significant (p=.003). The summed total for Novel Probes was not significant (25/48, 52%, p>.05), but the summed total for Novel Combinations was (75/96, 78%, p<.0001). Unlike most of the other subjects in Experiment 1 (except G8), this subject's performance on Novel Probe tests did not meet our criterion of binomial significance for the last two Novel Probe sessions that allow for advancement to Experiment 2. The combined Novel Probe performance for the last two probe sessions, TUV and WXY, was not statistically significant for G13 (p=.07) and our collection of spices and oils had been exhausted. Thus, for G13, a modified advancement criterion was adopted that required statistically significant performance on the last three Novel Probe tests, rather than two. The combined Novel Probe performance for sets QRS, TUV, and WXY was 14/18 (78%, p=.01), and he was then advanced to Experiment 2. Since then, we have increased the available number of odors in preparation for future subjects with similar circumstances. See Appendix A for the entire stimulus list.

Subject G8

G8 received six generalized MTS probe sessions during Experiment 1 (see Table 4 and Figure 9). During the first two Novel Probe tests for sets BCD and EFG, performance was at chance levels (3/6, 50%). Novel Probe performance during the third, fourth, and fifth probe sets (HIJ, KLM, and NOP) hovered at 4/6 (67%) and finally peaked at 5/6 (83%) during the last probe set, QRS. None of G8's individual Novel Probe performances were statistically significant and neither were the summed totals for Novel Probes (23/36, 64%, p>.05). His Novel Combination Probe performance did not trend upwards as clearly as his Novel

Probe performances; instead, it tended to bounce around irregularly throughout Experiment 1 probe sessions. The average performance during Novel Combinations was 7/12 (58%), although, because of the highly variable performance during individual probe sessions, this measure is of limited value. For example, his Novel Combination performance during the first and last probe sets (BCD and QRS) were equally accurate (6/12, 50%). The sets in between varied from 3/11 (27%) during set EFG to 10/12 (83%) on sets HIJ and NOP, yet the summed totals for Novel Combinations were statistically significant (44/71, 62%, p=.02). It is not clear why performance accuracy on Novel Probes steadily increased while performance accuracy on Novel Combinations wavered. Given that all Novel Probes and Novel Combinations were reinforced, we should expect more accurate response patterns as the trials within probe sessions progressed. We should also expect that as more exemplars were added to his training history, his Novel Combination performance would have been more accurate across sets as well. Neither of these assumptions appears to be entirely applicable for G8 in terms of Novel Combinations. Additionally, like G13, G8 never met the original advancement criterion of binomial significance on the Novel Probes for the last two probe sets (9/12, p=.07). Because of his age towards the end of Experiment 1, the decision was made to advance him to Experiment 2 without achieving the original advancement criterion and without attempting to train another stimulus set (TUV), although additional scents were available. This decision was made because it was deemed that any data collected in Experiment 2 would be of greater value than continuing Experiment 1. For consistency, we used the same

advancement criterion for G8 that was used for G13. Novel Probe performances on the last three sets were significant (13/18, p=.04) and he was moved to Experiment 2.

Subjects H8 and J6

H8 and J6 mastered at least the first training set (A) and received at least one novel generalization probe session (BCD), as shown in Table 4 and Figure 9. Subject H8 received only one probe session for stimulus set BCD. He was also exposed to various intervention procedures in an attempt to maximize stimulus control and facilitate learning. Subject J6 received two probe sessions, sets BCD and EFG, and received similar interventions as the others mentioned, but was also still in Experiment 1 EFG training at the time of this manuscript. Given these restriction, overall H8 and J6 did not show convincing evidence of generalized transfer of matching through Novel Probes or Novel Combinations of stimuli. Subject H8 performed only slightly above chance for Novel Probes (4/6, 66%, p>.05) during BCD tests. However, his performance was much more accurate during Novel Combinations (10/12, 83%, p<.01). In fact, H8 was the only subject in Experiment 1 that achieved statistical significance on either Novel Probes or Combinations during testing for the first stimulus set, BCD. J6 received two probe sessions for sets BCD and EFG and performed at or below chance levels as well on Novel Probes during both probe sessions (2/6, 33%; 3/6, 50%). Performance during Novel Combination probes was similar; at or around chance levels for both probe sessions (6/12, 50%; 7/12, 58%). The summed totals for both Novel Probes and Novel Combinations for J6 were not statistically significant either (5/12,

42%, $p>.05$; 13/24, 54%, $p>.05$). Since H8 and J6 did not receive many probe sessions, interpretation of their performances is limited. Perhaps if further experimentation had occurred with these subjects, their performances during probe sessions would have improved across exemplars, much like their counterparts. After all, their accuracies during the probe tests were not much different than the other five subjects, but it was the length of baseline training that differed and resulted in their elimination from the experiment.

Experiment 1 Summary

In sum, approximately half (five) of the subjects in Experiment 1 successfully, and often rapidly, acquired baseline discriminations and continued on to demonstrate convincing evidence for generalized MTS during probe sessions. The data for these five subjects support and extend the findings of Peña et al. (2006) to include a new apparatus and reinforcer condition in the generalized MTS literature. On the other hand, about half of Experiment 1 subjects struggled with baseline discriminations, often requiring extensive training and interventions. Although individual differences are to be expected, it was unexpected that almost half (four) of the subjects from Experiment 1 would not advance to Experiment 2 and that two of those subjects would never advance beyond the first training set (A). For the two subjects that received probe tests but did not reach Experiment 2 (H8, J6), performance was at or around chance and adds little in terms of evidence supporting generalized MTS. Further testing with J6 will hopefully provide additional probe measures, adding to the generalized matching data set.

Although seven subjects were able to learn baseline MTS discriminations and received at least one probe session, it was not without occasional therapy interventions. These are indicated in Table 3 with a superscript numeral two. As was the case with F5, I30, H8, and J6, some of the more successful subjects in Experiment 1 also required intervention in the form of time outs, S- lid fastening, and dummy S- stimuli in order to master the discriminations. For example, G13 required the S- lid fastening procedure during set A (session 10 through 21) and this intervention was successful. S-lid fastening was also used for H7 during set A from session nine through 12 and was similarly effective. The remaining subjects (F6, F3, and G8) did not require any interventions during Experiment 1.

It is apparent that the procedures (e.g. apparatus, exemplar procedure, reinforcer condition) used in Experiment 1 do not produce consistent performances across individuals; for some, they seem to have facilitated learning and transfer and for some they seem to have hindered learning and the ability to reach criterion on even baseline discriminations. For example, perhaps through the varied stimulus positions of the Arena, some Arena subjects more readily learned discriminations based on the odor of stimuli without the presence of conflicting control by stimulus position. On the other hand, some subjects in Experiment 1 may have struggled because baseline discriminations were not included across stimulus sets as novel stimuli were gradually introduced. This procedure was used in Peña et al. (2006) and appears to be generally more effective across subjects than the procedures used here. In other words, proportionally more of the subjects from the Peña et al. (2006) study were able to

learn and master discriminations and transfer responding to novel stimuli than in the current study. Determining which variables contributed to or restricted learning will be an important feature of related experiments to come.

Regardless, the five successes from Experiment 1 encouraged the inclusion of a second experiment. These five Experiment 1 subjects were the focus of Experiment 2 and the initial goal was simply to challenge them with more complex, arbitrary conditional discriminations and observe their performances. Prior to Experiment 2, there was very little, if any, extant literature investigating arbitrary conditional discrimination learning in rats. Because these five subjects mastered Experiment 1 discriminations readily and convincingly generalized to novel stimuli, we decided that perhaps these animals could also demonstrate even more complex relational behaviors and that it would be valuable to move forward, if only to baseline arbitrary conditional discriminations. To that end, Experiment 2 consisted of arbitrary conditional discrimination training and eventually resulted in emergent equivalence testing for all five animals.

EXPERIMENT 2

Based on the impressive performances observed in some subjects during Experiment 1, Experiment 2 was designed to evaluate whether these same animals could master discriminations of increasing complexity. Experiment 2 was initiated based upon individual subject performance during the last two (or three, G13 and G8) generalized identity MTS probe sessions. Experiment 2 began by training several arbitrary conditional discriminations. At the start, it was unclear if

subjects would learn these discriminations, as there is little literature supporting or even attempting to investigate such behavior in rats. However, given the extensive history of identity matching, the high levels of performance accuracy observed during generalized identity MTS probes, and the experience with class-specific reinforcers (except for F6), it was hypothesized that these subjects would acquire and master the baseline arbitrary conditional discriminations. Based on the assumption that baseline discriminations could and would be learned, a second aspect was added to Experiment 2; testing for emergent equivalence relations. So, if subjects were able to master several baseline discriminations within the first training set, they would then be tested for emergent relational responding through symmetry, transitivity, and equivalence probe tests for that same stimulus set. All five subjects did, in fact, reach equivalence testing stages. The proceedings and results of Experiment 2 are described below.

METHOD

Subjects

Five subjects from Experiment 1 were used for Experiment 2 (F6, F3, H7, G13, and G8). These subjects were moved to Experiment 2 procedures after meeting statistical significance ($p<.05$) during identity MTS probe sessions as previously described.

Apparatus

All subjects in Experiment 2 except one (H7) were trained and tested in the Odor Arena. H7 remained in the Operant MTS Chamber for the entirety of experimentation.

Stimuli

Stimuli used in Experiment 2 were the same as Experiment 1 and are described in the General Methods. All subjects received class-specific reinforcement except one (F6), who received only sugar reinforcers for the duration of Experiment 2.

Procedure

Similar to Experiment 1, in Experiment 2 subjects were trained using the same general MTS procedures, but the sample-correct comparison pairings were determined by the designated stimulus class of each odor (class 1 or class 2), rather than the identity of the odor. In other words, unlike the identity MTS procedure, in which the identity relation between samples and comparisons determined the reinforced comparison choice, arbitrary MTS training trained relations between physically dissimilar samples and correct comparisons (S+) that were designated by the experimenter as class members. Subjects were trained to respond to the correct comparison stimulus (S+) which was different from, but within the same class as, the sample stimulus.

For each set of three stimuli (e.g. ABC, DEF, GHI, etc.) four arbitrary conditional discriminations (e.g. $\mathbf{A_1 \rightarrow B_1, B_1 \rightarrow C_1; A_2 \rightarrow B_2, B_2 \rightarrow C_2}$) were trained using arbitrary match-to-sample procedures. For example, the first A-B conditional discrimination consisted of training conditional discriminations where $\mathbf{A_1}$ or $\mathbf{A_2}$ served as samples and $\mathbf{B_1}$ or $\mathbf{B_2}$ served as comparisons within each session. The second B-C conditional discrimination training consisted of conditional discriminations where $\mathbf{B_1}$ or $\mathbf{B_2}$ were sample stimuli and $\mathbf{C_1}$ or $\mathbf{C_2}$ were

comparisons within each session. Finally, the third A-B/B-C conditional discrimination integrated both conditional discriminations A-B and B-C within the same training sessions. Each set of the conditional discrimination baseline training in Experiment 2 was designed to facilitate the use of probes to evaluate emergent symmetry and transitivity relations. This baseline training used a mixed-node MTS design. The nodal stimulus (in this case, either $\mathbf{B_1}$ or $\mathbf{B_2}$) served as either a comparison, ($\mathbf{A_1 \rightarrow B_1, A_2 \rightarrow B_2}$), a sample ($\mathbf{B_1 \rightarrow C_1, B_2 \rightarrow C_2}$), or both ($\mathbf{A_1 \rightarrow B_1, A_2 \rightarrow B_2; B_1 \rightarrow C_1, B_2 \rightarrow C_2}$). This varied the role of the stimulus position, balanced the functions of sample and comparison within the nodal stimulus, and trained conditional discrimination baselines for later probing of emergent symmetrical and transitive relations. Sample stimuli (A_1& A_2) were presented an equal number of times in each session. Comparison stimuli (B_1& B_2) were correct and incorrect an equal number of times in each session, and locations (hole position for Arena or left and right positions for MTS Chamber) were randomized or counterbalanced, respectively (see Appendix).

Experiment 2 Training, Continued: Arbitrary Match-To-Sample Baseline

Training during the first stimulus set, A-B, included arbitrary conditional discriminations for four different stimuli, two from each class; $\mathbf{A_1}$ and $\mathbf{B_1}$, $\mathbf{A_2}$ and $\mathbf{B_2}$ (see Table 5). Specifically, when the sample, or conditional, stimulus scent was $\mathbf{A_1}$, responses to $\mathbf{B_1}$ were reinforced (S+), while responses to comparison $\mathbf{B_2}$ were not (S-). When the sample stimulus scent was $\mathbf{A_2}$, the correct comparison scent $\mathbf{B_2}$ was reinforced (S+), while $\mathbf{B_1}$ was not (S-) (see Figure 10). Subjects continued training at A-B ($\mathbf{A_1 \rightarrow B_1, A_2 \rightarrow B_2}$) until criterion-level performance,

or two consecutive sessions with overall performance of 90 percent or greater, was achieved. After criterion was met for the first trained pairs, subjects moved to the second, B-C, conditional discrimination training set.

Again, each set of the conditional discrimination baseline training in Experiment 2 was designed to facilitate the use of probes to evaluate emergent symmetry and transitivity relations during later test session. In this second training set, the nodal stimuli ($\mathbf{B_1}$ or $\mathbf{B_2}$) now served as samples during B-C conditional discrimination training ($\mathbf{B_1}$ and $\mathbf{C_1}$, $\mathbf{B_2}$ and $\mathbf{C_2}$). The varied role of the stimulus position balanced the functions of sample and comparison within the nodal stimulus, $\mathbf{B_1}$ or $\mathbf{B_2}$. So, B-C training included arbitrary conditional discriminations for four different stimuli, two from each class; $\mathbf{B_1}$ and $\mathbf{C_1}$, $\mathbf{B_2}$ and $\mathbf{C_2}$ (see Table 5). In this stage, when the conditional stimulus scent was $\mathbf{B_1}$, responses to comparison $\mathbf{C_1}$ were reinforced (S+), while responses to comparison $\mathbf{C_2}$ were not (S-). Similarly, when the sample stimulus scent was $\mathbf{B_2}$, responses to comparison $\mathbf{C_2}$ were reinforced (S+), while responses to comparison $\mathbf{C_1}$ were not (S-) (see Figure 11). Sample stimuli (B_1& B_2) occurred in equal numbers throughout each session. Comparison stimuli (C_1& C_2) were correct and incorrect an equal number of times in each session, and locations (hole position for Arena or left and right positions for MTS Chamber) were randomized or counterbalanced, respectively. Subjects continued training at B-C ($\mathbf{B_1 \rightarrow C_1, B_2 \rightarrow C_2}$) until criterion-level performance, or two consecutive sessions with overall performance of 90 percent or greater, was achieved. After criterion was met for these trained pairs, subjects were further trained using mixed A-B, B-C conditional discrimination sessions.

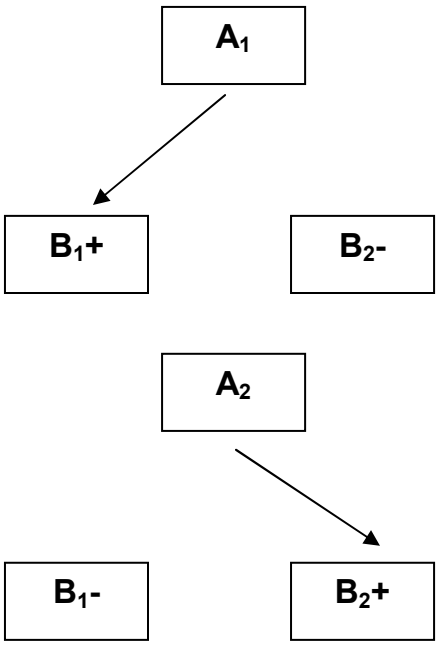


Figure 10. A-B Training Configurations.

Table 5. Experiment 2 Arbitrary Conditional Discrimination Stimulus Stimulus Presentation Order. Asterisks denote stimuli that were created using liquid odorants rather than dry, powder spices.

	F6			F3		H7	
Stimuli	1 (sugar)	2 (sugar)	Stimuli	1 (sugar)	2 (grain)	1 (sugar)	2 (grain)
A	Nutmeg	Dill	A	Clove	Rosemary	Clove	Rosemary
B	Celery	Cinnamon	B	Onion	Sage	Onion	Sage
C	Clove	Ginger	C	Sumac	Thyme	Sumac	Thyme
D	Oregano	Onion	D	Bay	Paprika	Bay	Paprika
E	Thyme	Coriander	E	Marjoram	Cumin	Marjoram	Cumin
F	Mustard	Sumac	F	Turmeric	Oregano	Turmeric	Oregano
G	Cumin	Marjoram	G	Cinnamon	Nutmeg		
H	Garlic	Rosemary	H	Ginger	Mustard		
I	Sage	Turmeric	I	Dill	Celery		
			K	Fennel	Allspice		
			N	Hickory	Savory		
			Q	Spinach	Peppermint*		

	G13			G8	
Stimuli	1 (sugar)	2 (grain)	Stimuli	1 (sugar)	2 (grain)
A	Clove	Rosemary	A	Mustard	Thyme
B	Onion	Sage	B	Sage	Orange
C	Sumac	Thyme	C	Thyme	Oregano
K	Fennel	Allspice	K	Sassafras	Bay
P	Caraway	Sassafras	N	Dill	Allspice
R	Raspberry	Grape*	R	Peppermint*	Grape*
I	Dill	Celery			
L	Beet	Carob			
S	Almond*	Maple*			

Following criterion performances during B-C baseline training sessions, mixed A-B, B-C conditional discrimination training began. Training sessions were mixed in the sense that the role of the stimulus position, comparison or sample, was varied through nodes B_1 and B_2 within each session (**$A_1 \rightarrow B_1$, $A_2 \rightarrow B_2$; $B_1 \rightarrow C_1$, $B_2 \rightarrow C_2$**). For example, within a single session, responses to **B_1** or **B_2** comparisons were reinforced (S+) when **A_1** or **A_2** were samples, as well as responses to **C_1** or **C_2** (S+) given **B_1** or **B_2** as the sample (see Figure 12). Sample stimuli (A_1& A_2, B_1& B_2) occurred in equal numbers throughout each session. Comparison stimuli (B_1& B_2, C_1& C_2) were correct and incorrect an equal number of times in each session, and locations (hole position for Arena or left and right positions for MTS Chamber) were randomized or counterbalanced, respectively. Subjects continued mixed conditional discrimination training (**$A_1 \rightarrow B_1$, $A_2 \rightarrow B_2$; $B_1 \rightarrow C_1$, $B_2 \rightarrow C_2$**) until criterion-level performance, or two consecutive sessions with overall performance of 90 percent or greater, was achieved. After criterion was met for these mixed pairs, Symmetry, Transitivity, and Equivalence (symmetry and transitivity) Probe trials were conducted.

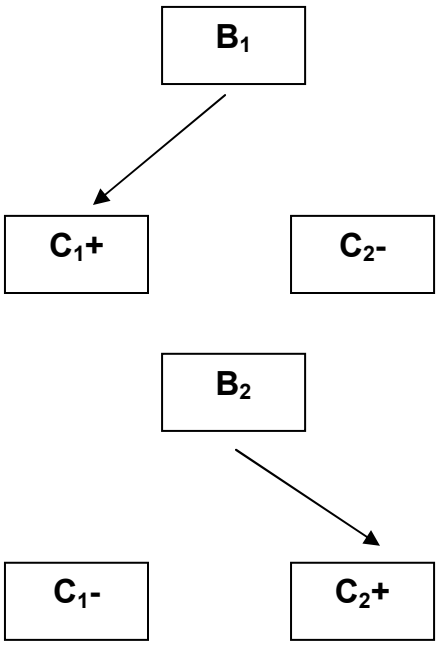


Figure 11. B-C Training Configurations.

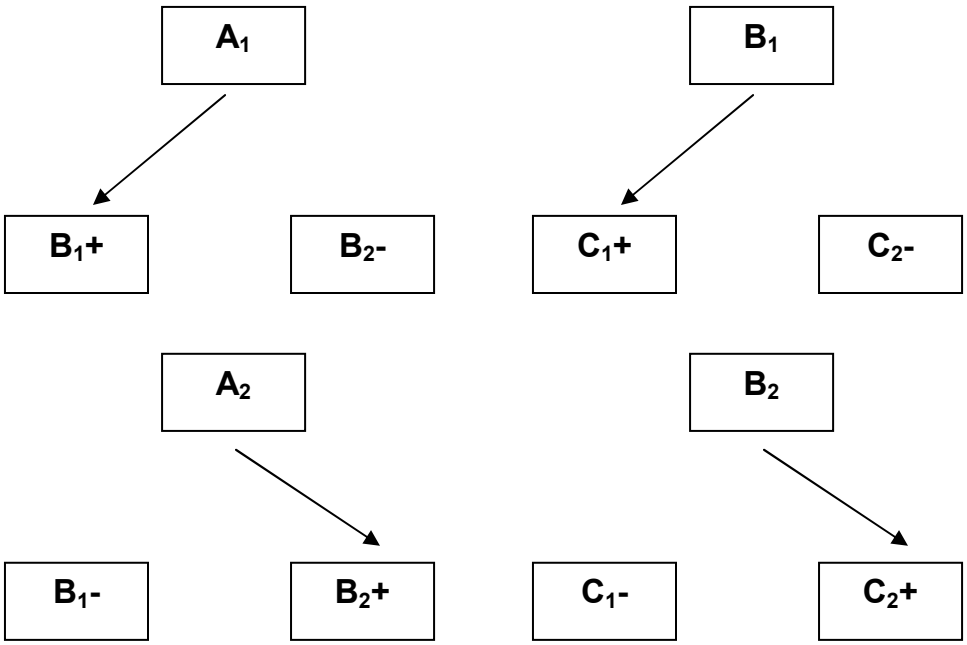


Figure 12. Mixed A-B, B-C Conditional Discrimination Training Configurations.

Experiment 2 Probes: ABC Symmetry, Transitivity, Equivalence Probes

Symmetry, Transitivity, and Equivalence Probe trials were conducted following criterion-level performance during mixed A-B, B-C training. Symmetry, Transitivity, and Equivalence trials occurred twice per session for a total of 16 trials. Novel probes, the first eight trials, occurred only during the first presentation of the symmetrical, transitive, and equivalence configurations. These novel probe trials were unbaited to control for pellet detection, and reinforcers were delivered using tweezers following responses to S+. The second presentation of symmetry, transitivity, and equivalence trials were not considered novel, since reinforcement was delivered during the first presentations. The second set of eight trails occurred randomly along with eight baseline trials from the A-B and B-C training sessions (see Appendix).

Emergent symmetrical relations are said to occur when the functions of the sample and the comparison stimuli are substitutable without any direct training. For the trained A-B conditional discrimination, "A" and "B" stimuli must become members of the same arbitrary class (Class 1 or Class 2) of stimuli in order to show emergent symmetry. Subjects' emergent symmetry performance was evaluated using Symmetry Probe trials in which the previously trained samples ($\mathbf{A_1}$ or $\mathbf{A_2}$) were now the comparison stimuli, and previously trained comparisons ($\mathbf{B_1}$ or $\mathbf{B_2}$) now served as samples. Successful symmetrical responding was demonstrated if, for example, following training of $\mathbf{A_1}$ or $\mathbf{A_2}$ as the sample stimulus and $\mathbf{B_1}$ or $\mathbf{B_2}$ as the correct comparison (S+) (**trained** $\mathbf{A_1 \rightarrow B_1}$, $\mathbf{A_2 \rightarrow B_2}$), subjects then chose A_1 or A_2 as the correct comparison (S+) given

that either B_1 or B_2 was the sample stimulus (*emergent* $B_1 \rightarrow A_1$, $B_2 \rightarrow A_2$) (see Figure 13). Performance during Symmetry Probes trials were evaluated for statistical significance using binomial tests ($p < .05$).

Similarly, through B-C conditional discrimination training, the "B" and "C" stimuli must become members of the same arbitrary class (class 1 or class 2) of stimuli in order to show emergent symmetry. Subjects' emergent symmetry performance for B-C training was evaluated using Symmetry Probe trials in which the previously trained samples (B_1 or B_2) were now comparison stimuli, and previously trained comparisons (C_1 or C_2) were now in the sample position. Successful symmetrical responding was demonstrated if, for example, following training of $\mathbf{B_1}$ or $\mathbf{B_2}$ as the sample stimulus and $\mathbf{C_1}$ or $\mathbf{C_2}$ as the correct comparison (S+) (**trained** $\mathbf{B_1 \rightarrow C_1}$, $\mathbf{B_2 \rightarrow C_2}$), subjects then chose B_1 or B_2 as the correct comparison (S+) given that either C_1 or C_2 was the sample stimulus (*emergent* $C_1 \rightarrow B_1$, $C_2 \rightarrow B_2$) (see Figure 14). Performance during B-C Symmetry Probes was evaluated to determine statistical significance ($p < .05$).

Transitivity Probe trials ($A_1 \rightarrow C_1$, $A_2 \rightarrow C_2$) also occurred during the first presentation of the transitive stimulus configuration. Emergent transitive relations are inferred when responding to one stimulus (A_1) occurs as a function of its symmetrical relation to a second stimulus ($\mathbf{A_1 \rightarrow B_1}$ $\therefore B_1 \rightarrow A_1$), through its symmetrical relation with a third stimulus ($\mathbf{B_1 \rightarrow C_1}$ $\therefore C_1 \rightarrow B_1$). Therefore, all three stimuli, "A, B, and C", must become members of the same class (class 1 or class 2) of stimuli in order to show emergent transitivity. For example, subjects

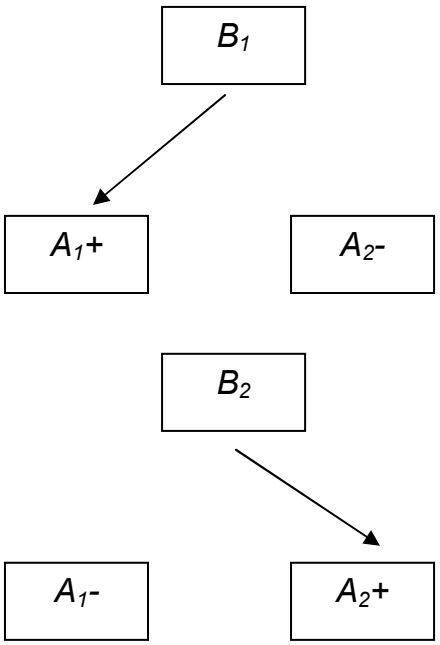


Figure 13. B-A Emergent Symmetry Probe.

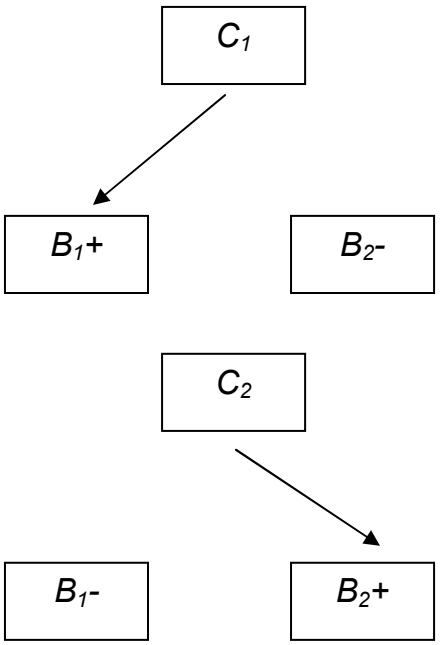


Figure 14. C-B Emergent Symmetry Probe.

were trained to select $\mathbf{B_1}$ or $\mathbf{B_2}$ comparisons given $\mathbf{A_1}$ or $\mathbf{A_2}$ as samples, and $\mathbf{C_1}$ or $\mathbf{C_2}$ given $\mathbf{B_1}$ or $\mathbf{B_2}$ as samples, respectively. Emergent transitivity would be defined as responding to C_1 or C_2 as comparisons given A_1 or A_2 as samples (see Figure 15).

In addition to the Transitivity Probes, the "A" and "C" stimuli also served as probes for emergent equivalency between A_1 or A_2 and C_1 or C_2 and were evaluated using the same procedure from baseline training. During probe sessions, the first two transitivity probes were reinforced, thus the $\mathbf{A_1}$ and $\mathbf{A_2}$ stimuli were trained as samples and the $\mathbf{C_1}$ and $\mathbf{C_2}$ as comparisons ($\mathbf{A_1 \rightarrow C_1, A_2 \rightarrow C_2}$). For example, given $\mathbf{A_1}$ as the sample, responses to $\mathbf{C_1}$ were reinforced (S+), and given the sample $\mathbf{A_2}$, responses to $\mathbf{C_2}$ were reinforced (S+). Immediately following these Transitivity Probes, Equivalency Probes will also be conducted. The Equivalence Probes will include trials where C_1 and C_2 are presented as samples, and A_1 and A_2 as comparisons ($C_1 \rightarrow A_1, C_2 \rightarrow A_2$) (see Figure 16). For example, given C_1 as the sample, responses to A_1 were reinforced and indicate emergent symmetry, and given C_2 as the sample, responses to A_2 were reinforced.

If statistically significant performance ($p<.05$) was observed during these equivalence probes, it would serve as even further evidence for equivalence class formation. Following criterion-level performance for A-B-C probes, the conditional discrimination training and symmetry, transitivity, and equivalency testing were repeated for additional sets of stimuli (e.g. DEF, GHI, etc.) in a multiple exemplar format similar to Experiment 1 training.

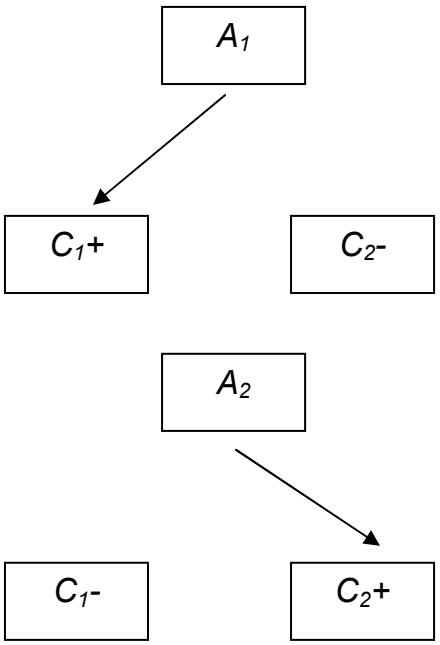


Figure 15. A-C Emergent Transitivity Probe.

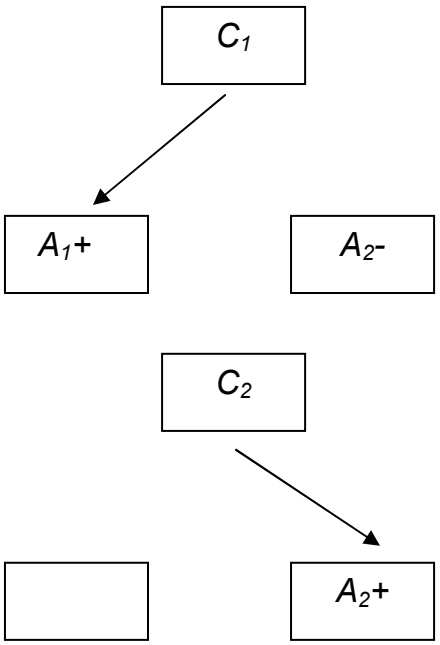


Figure 16. C-A Emergent Equivalence Probe.

Experiment 2 Probes: Across-Class Transitivity and Equivalence

Another component was added to Experiment 2 for two subjects (F6 and F3). Following Probe tests for set GHI, both subjects began a phase that consisted of across-class Transitivity and Equivalence (F3 only) Probes. This extension of the original Experiment 2 method was designed to test whether subjects could maintain high levels of accuracy on discriminations that encompassed both classes of stimuli from A to I, rather than in groups of six stimuli (e.g. ABC, DEF). The first across-class exposure was the A-I Baseline sessions. This training set contained 32 total trials; 28 baseline arbitrary MTS discriminations from stimuli A to stimuli I that were learned in previous phases of Experiment 2, for example A_1-B_1 and E_2-F_2, and four untrained discriminations that spanned across the previously trained stimulus sets, for example C-D and F-G from both classes. The baseline discriminations were given to ensure that subjects were responding with high levels of accuracy on across-class baseline discriminations before giving any across-class Transitivity Probes. The four untrained baseline trials that spanned the trained stimulus sets (AB**C**, **D**E**F**, **G**HI) also provided a measure of class grouping before administering any true probe trials.

The remainder of the across-class sessions consisted of repeated baseline discriminations (16 trials) as well as interspersed untrained Transitivity Probe trials that spanned across the exemplars previously trained (A through I). Probe sessions were conducted in succession beginning with sessions of 28 trials that contained 12 novel Transitivity Probe trials that included the A stimuli as samples and stimuli D, E, F, G, H, and I as potential comparison stimuli (A to D-I

Transitivity). Following achievement of criterion, the next phase of 12 Transitivity Probe trials was given, wherein the B stimuli served as samples and D, E, F, G, H, and I stimuli served as comparison stimuli (B to D-I Transitivity). These sessions also contained 26 total trials. The same process was repeated using sessions of 26 trials with 10 C to E through I Transitivity Probe trials (C to E-I Transitivity).

Even more Probe configurations were designed for F3 after the C to E-I Transitivity set. F3 also received 16 DEF through GHI Transitivity Probe trials (DEF-GHI Transitivity), where stimuli D, E, and F served as samples and stimuli G, H, and I served as comparison stimuli. The only discrimination not included was F-G, since that was not considered transitive in nature. The DEF-GHI sessions contained 32 total trials. Following the DEF-GHI Probe set, F3 was exposed to across-class Equivalence Probes as well, beginning with 12 D-I to A Equivalence Probe trials (stimuli D through I serve as samples, A serves as comparisons), then 12 D-I to B Equivalence Probe trials, 10 E-I to C Equivalence Probe trials, six G-I to D trials, and four H-I to F trials. These sessions consisted of 32, 32, 28, and 24 total trials, respectively.

Experiment 2 Training and Probes: Across-Class Exemplars

Through the across-class Transitivity and Equivalence Probes given to F6 and F3, a procedural conflict was discovered. To counter this, later Experiment 2 exemplars for some subjects were not alphabetically consistent (e.g. KNQ, ILS) with previous exemplars (e.g. ABC, DEF). Specifically, it was discovered that several discriminations given during Experiment 2 probes overlapped with

discriminations learned during Experiment 1 training. These overlaps in training may have allowed subjects to respond based on the exclusionary principle during emergent symmetry testing (e.g. $C_X \rightarrow B_X$, not B_Y), rather than based on the relationships learned through baseline Experiment 2 training. A brief example may clarify the concern. During Experiment 1 BCD training, subjects were trained that when a class 1 stimulus (B_1, C_1, or D_1) was the correct comparison, a class 2 stimulus was always the incorrect comparison (e.g. $B_1 \rightarrow B_1$, not B_2, C_2, or D_2), and vice versa. In the case of set BCD, the incorrect stimulus function, S-, varied across three possible stimuli from the same class (B, C, or D), but S- was always from the opposite class as the sample and correct comparison. The problem arises particularly from the $C_X \rightarrow C_X$, not B_Y discrimination from Experiment 1, where subjects learned that when C_X was the sample and S+, responses to B_Y were never reinforced. Thus, during emergent symmetry testing for the B-C relations of the ABC set (e.g. $C_X \rightarrow B_X$, not B_Y), subjects could choose B_X simply because when comparison C_X was correct during Experiment 1, comparison B_Y was always incorrect and comparison B_X was never an option. Of course, C_Y and D_Y were also trained as S- in the presence of C_X, but these baselines do not interfere with B-C Symmetry Probes. This same issue holds true for the F-E Symmetry Probes during the DEF set, and the I-H Symmetry Probes during the GHI set.

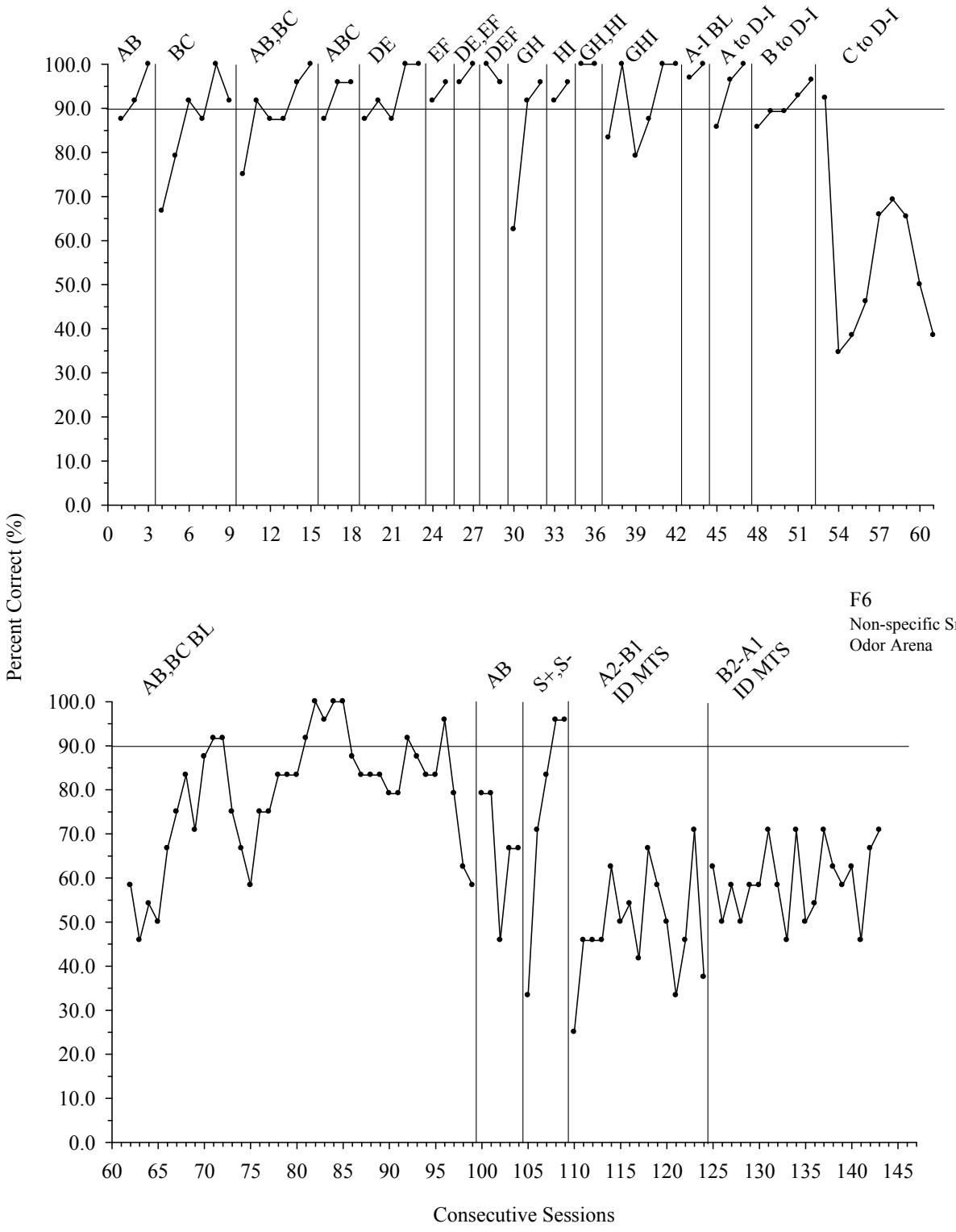
As a result, the behavior observed during emergent Symmetry Probes would not necessarily be indicative of relational responding that results when equivalency has formed between groups of stimuli. Instead, subjects may select

the correct stimulus because, through Experiment 1 training, they have learned that responses to certain stimuli are never reinforced in the presence of certain other stimuli – in this case C_X and B_Y. Therefore, it is possible that subjects could have performed with comparable levels of accuracy on Symmetry Probes without ever learning the baseline arbitrary MTS discriminations from Experiment 2 (B-C, E-F, H-I). To examine the effects of the training from Experiment 1 on responding during Experiment 2 probes, exemplar sets that spanned the Experiment 1 groupings were added. Sets KNQ, KPR and ILS, and KNR were given to subjects F3, G13, and G8 respectively. All procedures remained identical to those used for previous exemplars.

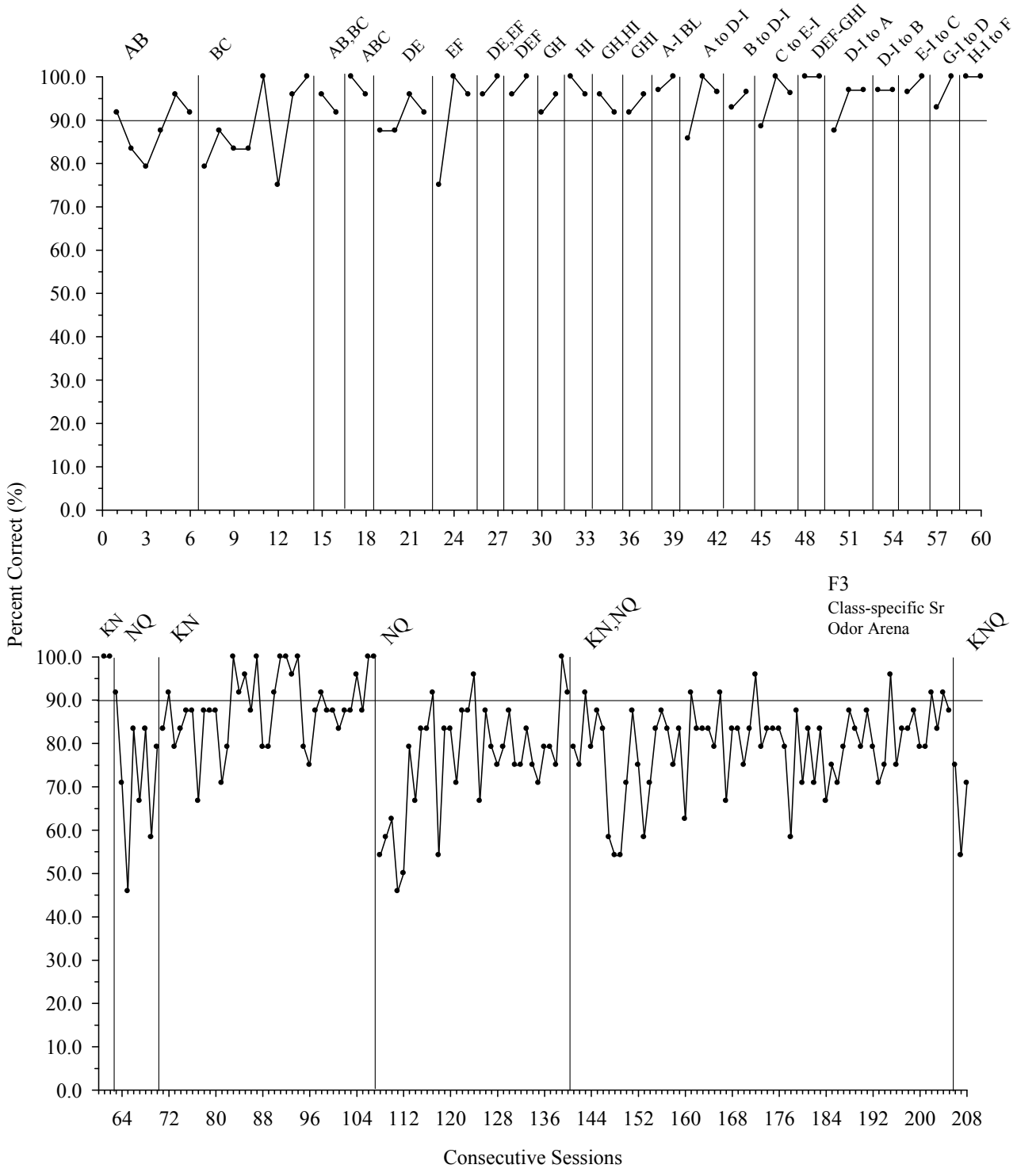
EXPERIMENT 2 RESULTS AND DISCUSSION

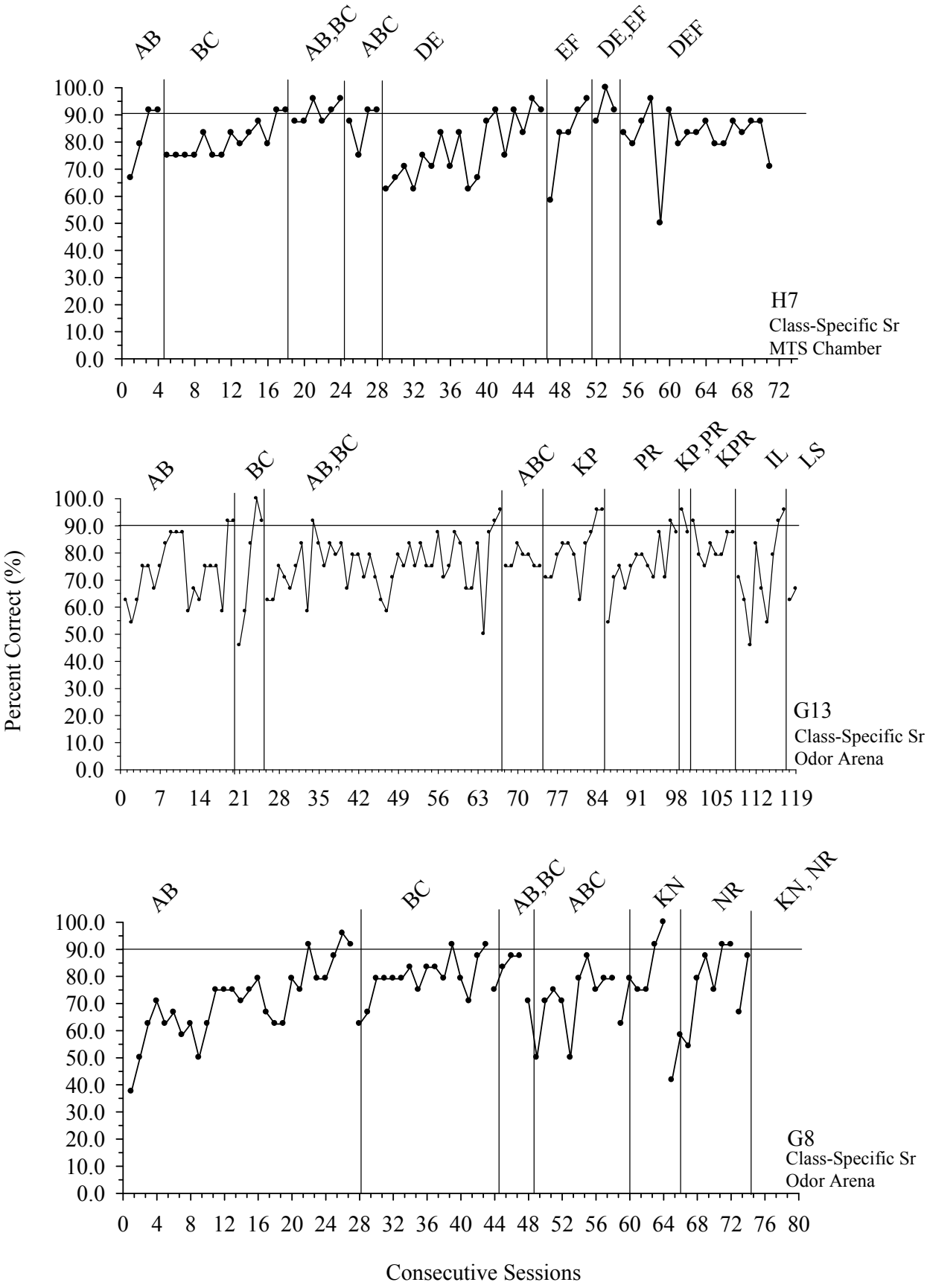
Figure 17 shows session by session, individual subject data across stimulus set presentations during Experiment 2. Each data point represents percent correct for a single session or the ratio of correct trials over the total number of trials in the session. The panels represent changing stimulus sets (e.g. A-B, B-C) and are separated by vertical lines. The horizontal line represents the original mastery criterion of 90%, or at least 22/24 trials, which was in effect for most subjects during Experiment 2. Like Experiment 1, an exception to the original criterion (87.5%) was used for some subjects during some training sets. This change was made for subjects that were having difficulty meeting the 90% criterion, but were regularly responding correctly on at least 21/24 trials (87.5%), or for subjects that were approaching two years of age. This was especially true

Figure 17. Experiment 2 Percent Correct across Stimulus Set Presentations. Each data point represents percent correct for one session, or the proportion of correct trials divided by all possible trials. Vertical lines (panels) within each graph indicate changes in stimulus sets; each panel is labeled with the respective set. The horizontal line indicates the original criterion (90%). Any criterion level performance that is not followed by a vertical phase change line occurred during a special training procedure.



F6
Non-specific Sr
Odor Arena


AB
BC
AB,BC
ABC
DE
EF
DE,EF
DEF
GH
HI
GH,HI
GHI
A-I BL
A to D-I
B to D-I
C to E-I
DEF-GHI
D-I to A
D-I to B
E-I to C
G-I to D
H-I to F
Percent Correct (%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60
F3
Class-specific Sr
Odor Arena
KN
NQ
KN
NQ
KN,NQ
KNQ
64 72 80 88 96 104 112 120 128 136 144 152 160 168 176 184 192 200 208
Consecutive Sessions


AB
BC
AB,BC
ABC
DE
EF
DE,EF
DEF
H7
Class-Specific Sr
MTS Chamber
AB
BC
AB,BC
ABC
KP
PR
KP,PR
KPR
IL
LS
G13
Class-Specific Sr
Odor Arena
AB
BC
AB,BC
ABC
KN
NR
KN, NR
G8
Class-Specific Sr
Odor Arena
Percent Correct (%)
Consecutive Sessions

for subjects G13 and G8, who were over two years old at the time of this manuscript. Additionally, some subjects required intervention procedures such as time out or S- lid fastening to master some discriminations. Any performances that met criterion during these procedures were not considered as such. Any criterion level performance that is not followed by a vertical phase change line occurred when therapy intervention procedures were in place. Subjects were typically required to meet criterion in the absence of any therapy procedure in order to advance to the next stage. Exceptions are noted.

Sessions to Criterion

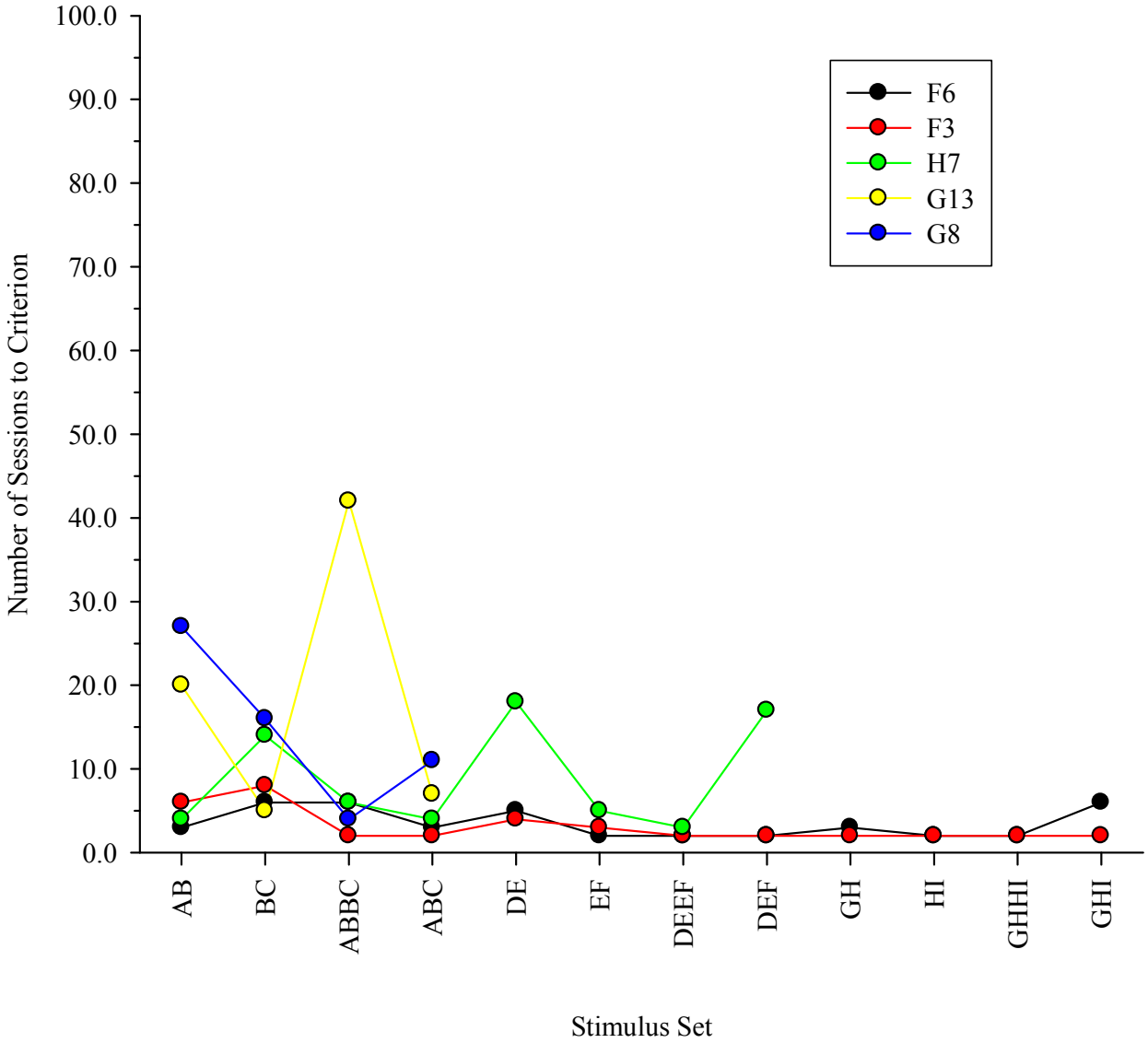
The five Experiment 2 subjects acquired and mastered the discriminations quite rapidly and also showed savings in the number of sessions needed to meet criterion across stimulus sets. So, for the most part, as baseline discriminations were added to their repertoires, criterion was achieved in fewer and fewer sessions. The number of sessions required to meet criterion during training sets in Experiment 2 are depicted in Table 6 and Figure 18. Both displays include individual subject data for the number of sessions that occurred within each stimulus set (e.g. A-B, E-F). In Table 6, the superscript numeral one indicates sets in which criterion was not met, the superscript two denotes the modified 87.5% criterion, and the superscript numeral three indicates that therapy interventions were in effect during the set. Table 6 also includes summed totals

Table 6. Experiment 2 Number of Sessions to Criterion across Stimulus Set Presentations. A superscript numeral one indicates sets in which subjects did not meet criterion for particular stimulus set. A superscript numeral two indicates the modified criterion of 87.5%. Superscript numeral threes indicate that therapy intervention procedures were used during training for that particular set. Any criterion level performances during therapy were not considered for advancement.

Subject / Stimuli	F6	F3	H7	G13	G8	Total	Mean
A-B	3	6	4	20	27	219	43.8
B-C	6	8	14	5	16[2]	132	26.4
A-B, B-C	6	2	6	42	4[2]	78	15.6
ABC	3	2	4	7[1]	11[1]	27	5.4
D-E	5	4	18			27	9
E-F	2	3	5			10	3.3
D-E, E-F	2	2	3			7	2.3
DEF	2	2	17[1]			21	7
G-H	3	2				5	2.5
H-I	2	2				4	2
G-H, H-I	2	2				4	2
GHI	6	2				8	4
K-N		40				40	-
N-Q		40				40	-
K-N, N-Q		65[2]				65	-
KNQ		3[1]				3	-

	F6	F3	H7	G13	G8	Total	Mean
K-P				11		11	-
P-R				13[2]		13	-
K-P,P-R				2[2]		2	-
KPR				8[2]		8	-
K-N					6	6	-
N-R					8	8	-
K-N, N-R					7[1]		-
KNR					TBA		-
I-L				9		9	-
L-S				2[1]		2	-
Total	42	185	71	119	79	729	123.3
Mean	3.5	30.8	8.9	11.9	11.3	-	10.8

Figure 18. Experiment 2 Number of Sessions to Criterion across Stimulus Set Presentations for each subject.



and means across and within all subjects to show individual, overall, and average learning set trends during the course of Experiment 2.

The average number of sessions to meet criterion in Experiment 2 decreased greatly between the first and second stimulus sets (A-B, $\sum$= 219, $\bar{x}$ = 43.8; B-C, $\sum$= 132, $\bar{x}$ = 26.4) and the second and third sets (B-C, $\sum$= 132, $\bar{x}$ = 26.4; A-B, B-C mixed, $\sum$= 78, $\bar{x}$ = 15.6), although not as dramatically or consistently as observed during Experiment 1. The overall percentage of decrease in the mean number of sessions to criterion between sets one and two and two and three was 40%. In other words, on average, subjects required about 40% fewer sessions of training on the second set compared to the first and the third set compared to the second. Although individual subjects differed from the mean in their exact number of sessions required before achieving criterion, the mean downward trend is proportionally representative to the downward trends present in the individual subject data. The number of sessions to criterion for the third and fourth sets (A-B, B-C mixed, $\sum$= 78, $\bar{x}$ = 15.6; ABC, $\sum$= 27, $\bar{x}$ =5.4) decreased as well, on average approximately 65%. After the first stimulus set ABC, the downward trend for number of sessions becomes less obvious and perhaps shows some evidence for a floor effect, as the minimum number of sessions to criterion is two. The large decreases in the number of required sessions to meet criterion between the first few sets (e.g. ABC) and subsequent plateau across later sets (after ABC) is apparent in Figure 18. As for individual subjects, progress through both baseline and probe phases of Experiment 2 was highly variable both within

and across subjects. Again, session by session data for the entirety of Experiment 2 can be seen in Figure 17.

Baseline Arbitrary Matching-To-Sample

All five subjects who transferred to Experiment 2 were able to acquire and ultimately master baseline conditional discriminations for at least two stimulus sets (e.g. ABC, DEF) and for a maximum of four sets. These subjects acquired the baseline conditional discriminations within days or a few weeks of training, although there were some exceptions or instances that required longer training sessions to meet criterion. Some subjects mastered baseline training sets in the fewest possible number of sessions (2), and in general, progress through Experiment 2 baselines was more rapid than progress during Experiment 1. Overall, most subjects learned most baseline discriminations with little difficulty and required little therapy interventions.

Emergent Equivalence Probe Results

All five subjects were also eventually exposed to Equivalence Probe tests, and these data can be seen in both Table 7 and Figure 19. In both displays, individual subject probe performances are divided up across stimulus sets (ABC, DEF, etc.) for the three types of probe trials (symmetry, transitivity, equivalence). Total performances on all probe tests are given for each stimulus set and across all stimulus sets as well. Proportions represent the number of correct trials over the total possible number of trials, in this case four symmetry and two each for transitivity and equivalence trials. Again, all eight Probe trial types were given

Table 7. Experiment 2 Emergent Equivalence Probe Performance across Stimulus Set Presentations. A superscript numeral one represents sets in which subjects did not reach criterion and were either dropped from the experiment or were advanced to the next set. A superscript numeral two indicates sets where subjects met criterion using a lowered version of two consecutive sessions with 87.5% correct rather than 90%. A superscript numeral three indicates instances where subjects received therapy interventions. Asterisks indicate Binomial Significance ($p<.05$).

Subject / Stimuli	F6	F3	H7	G13	G8
	Symmetry Probes (X/4); Transitivity Probes (X/2); Equivalence Probes (X/2) Total Probes (X/8)				
A, B, C,	3/4; 1/2; 2/2 (6/8)	4/4; 2/2; 2/2 (8/8*)	2/4; 2/2; 2/2 (6/8)	3/4; 2/2; 1/2 (6/8)	1/4; 1/2; 2/2 (4/8)
D, E, F,	4/4; 2/2; 2/2 (8/8*)	3/4; 2/2; 2/2 (7/8*)	3/4; 2/2; 2/2 (7/8*)	-	-
G, H, I,	1/4; 2/2; 2/2 (5/8)	4/4; 2/2; 1/2 (7/8*)	-	-	-
K, N, Q	-	2/4; 1/2; 1/2 (4/8)	-	-	-
K, P, R	-	-	-	3/4; 2/2; 2/2 (7/8*)	-
K, N, R	-	-	-	-	incomplete
Total	8/12; 5/6; 6/6* (19/24*)	13/16*; 7/8*; 7/8* (27/32*)	5/8; 4/4; 4/4 (13/16*)	6/8; 4/4; 3/4 (13/16*)	1/4; 1/2; 2/2 (4/8)

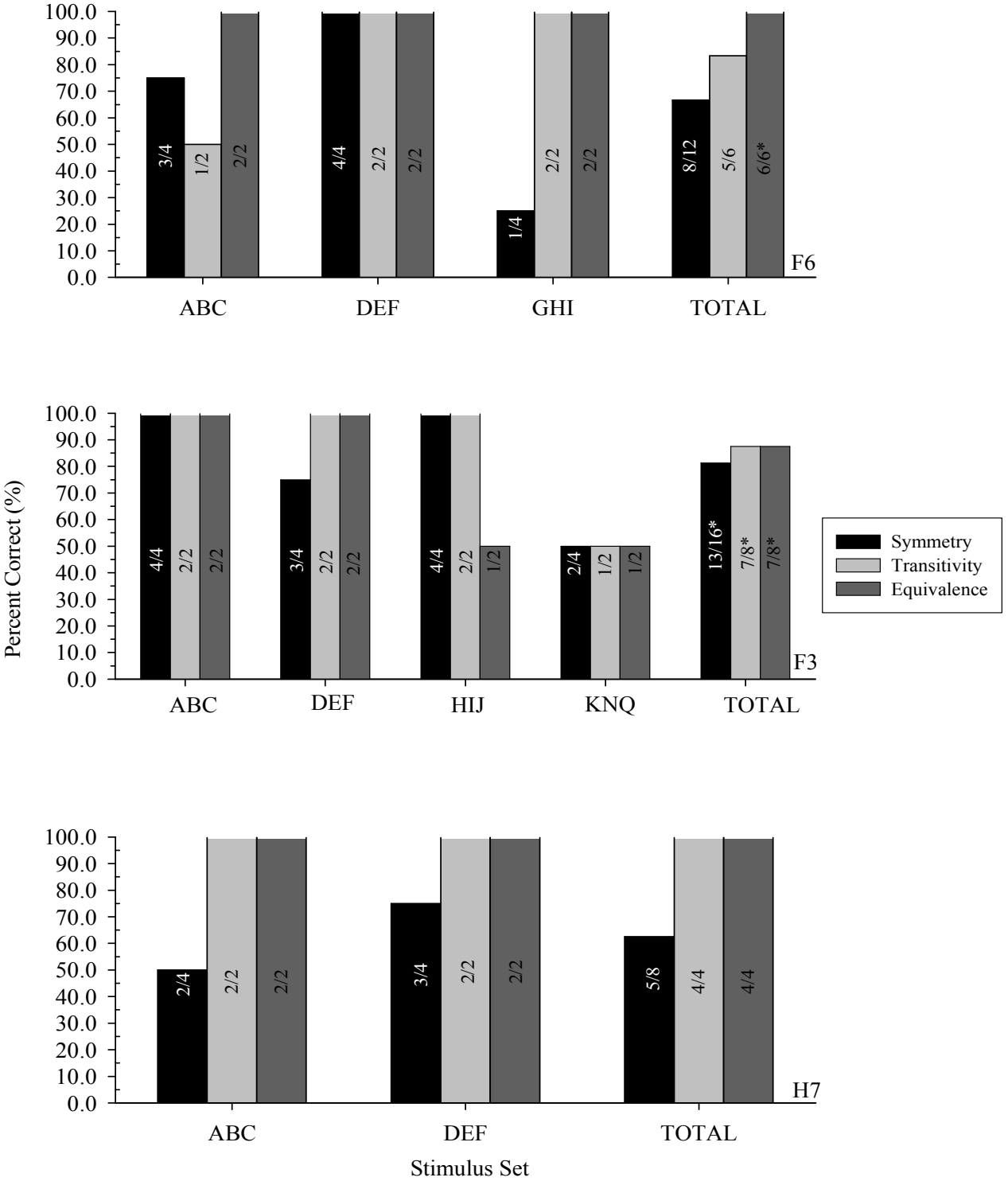
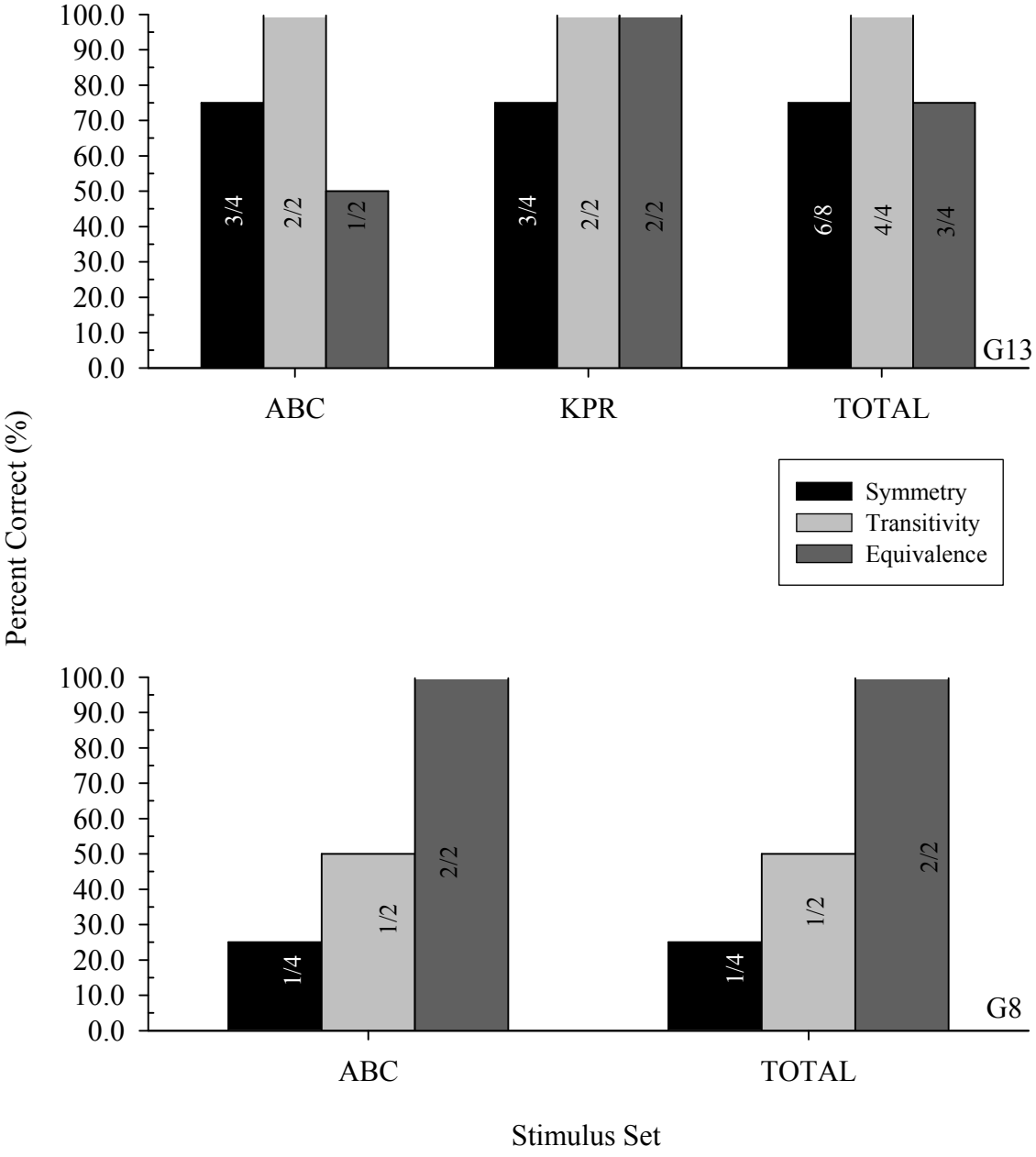


Figure 19. Experiment 2 Percent Correct for Emergent Equivalence Probes across Stimulus Set Presentations.


100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
3/4
2/2
1/2
3/4
2/2
2/2
6/8
4/4
3/4
G13
ABC
KPR
TOTAL
Symmetry
Transitivity
Equivalence
Percent Correct (%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1/4
1/2
2/2
1/4
1/2
2/2
G8
ABC
TOTAL
Stimulus Set

during the first eight trials of each Probe session and comparison stimuli were unbaited to control for pellet detection. Figure 19 depicts percent correct measures as well. In Table 7, the proportions in parenthesis within each cell represent the total correct probe trials over the total possible trials for an entire stimulus set (e.g. ABC, DEF), in this case eight. The last row of Table 7 shows the total proportion of correct probe trials over total probe trials for all symmetry, transitivity, and equivalence probes (across all stimulus sets) as well as the proportion of the total number of correct probe trials over the total possible probe trials for all probes given in Experiment 2. The total number of probes varies across subjects depending on the extent of experimentation. For Table 7, asterisks denote binomial significance ($p<.05$). Individual subject performances during Experiment 2 are discussed below.

Subject F6

Subject F6 proceeded through the initial training sets fairly rapidly without requiring many sessions of training to meet criterion or any forms of therapy interventions (Table 6, Figure 18). Interestingly, the initial switch from the Experiment 1 identity MTS task to the Experiment 2 arbitrary MTS task did not produce any noticeable disruption in performance; F6 scored 21/24 (88%) on the first A-B session and met in three sessions. Comparable learning rates were observed for the other baseline sets as well, with a maximum of six sessions required for sets B-C and A-B, B-C. F6 completed a total of 42 sessions during Experiment 2, with a mean of three and a half sessions to criterion for each set, far less than the across-subject mean.

F6 received three exemplars of Emergent Equivalence Probes; ABC, DEF, and GHI and performed with high levels of accuracy during all three tests, although there was no clear upward trend across exemplars (Table 7, Figure 19). For Probes during set ABC, F6 responded correctly on six out of eight Probe trials (75%), making single errors during the Symmetry (75%) and Transitivity Probes (50%), but performing at 100% during Equivalence Probes. The error during the Symmetry Probes occurred during the B-A discrimination for class 1, while the error during the Transitivity Probe occurred during the A-C discrimination for class 2. The summed ABC Probe performance (6/8) was not statistically significant, but still greater than 50%. During Probes for set DEF, F6 responded correctly on all eight trial types (100%) and the summed Probe performance for DEF was statistically significant ($p<.003$). Performance during the final Probe set GHI declined to 5/8 (63%), with all of the errors occurring during symmetry probes (1/4, 25%). GHI summed Probe performance was not significant. Across the three exemplars, F6's performance for Equivalence Probes was statistically significant (6/6, $p<.01$), as was his summed total Probe performance (19/24, $p<.003$). Although his summed performance for Symmetry (8/12, 67%) and Transitivity Probes (5/6, 83%) were not significant, both were above 50%.

After receiving GHI probes and meeting criterion for the GHI set, F6 was transferred to the across-class phase of Transitivity testing. Figure 17 depicts session by session data for these across-class phases, Table 8 shows the number of sessions required to meet criterion for this additional phase, and Table 9 depicts the Transitivity Probe performance across all sets. F6 responded correctly on all

Table 8. Experiment 2 Number of Sessions for Across-class Transitivity and Equivalence Probes for Subjects F6 and F3.The superscript numeral one indicates that F6 did not achieve criterion during the C to (E-I) stimulus set.

Subject / Stimuli	F6	F3
A-I Baseline	2	2
A to (D-I) Transitivity	3	3
B to (D-I) Transitivity	5	2
C to (E-I) Transitivity	9[1]	3
DEF-GHI Transitivity	-	2
(D-I) to A Equivalence	-	3
(D-I) to B Equivalence	-	2
(E-I) to C Equivalence	-	2
(G-I) to D Equivalence	-	2
(H-I) to F Equivalence	-	2
Total	19	23

Table 9. Experiment 2 Across-class Transitivity and Equivalence Probe Performance for Subjects F6 and F3. All proportions and percentages are taken from the first session within each stimulus set. The proportions and percentages in the first row represent the number of correct Probe trials over the total number given while the percentage in parentheses represents the overall performance during the session in which Probes occurred.

Subject / Stimuli	F6	F3
	Proportion Correct Probe Trials/Total Possible Probe Trials, Percent Correct Probes (Overall Percent Correct for Session)	
A-I Baseline	4/4, 100% (97%)	4/4, 100% (97%)
A to (D-I) Transitivity	9/12, 75% (86%)	9/12, 75% (86%)
B to (D-I) Transitivity	10/12*, 83% (87%)	12/12*, 100% (93%)
C to (E-I) Transitivity	9/10*, 90% (92%)	8/10, 80% (88%)
DEF-GHI Transitivity	-	16/16*, 100% (100%)
(D-I) to A Equivalence	-	11/12*, 92% (88%)
(D-I) to B Equivalence	-	11/12*, 92% (97%)
(E-I) to C Equivalence	-	10/10*, 100% (96%)
(G-I) to D Equivalence	-	6/6*, 100% (93%)
(H-I) to F Equivalence	-	4/4, 100% (100%)
Total Probe Performance	28/34*, 82%	87/94*, 93%

four untrained baseline conditional discriminations within the A-I Baseline set (4/4, 100%). These discriminations included C-D and F-G discriminations for class 1 and 2 that were not explicitly trained during the previous Experiment 2 stimulus sets (AB**C**, **D**E**F**, and **G**HI). Performance during the next three across-class Transitivity Probe sets increased from 9/12 (75%) to 11/12 (83%) to 9/10 (90%). Performance on the last two Probe sets was statistically significant, $p<.003$ and $p<.01$. Summed totals for across-class Transitivity Probes were 28/34 (82%), which was significant, $p<.0001$. Although these performances are impressive, they should not have occurred. Since F6 was trained using non-specific reinforcers, his performance during across-class Probes should have been at or around 50%, as there was no class-specific reinforcer that would have served a unifying function across stimuli within each class. His highly accurate and unexpected performances initiated an investigation into his training history, where it was found that overlaps occurred between Experiment 1 training sets and Experiment 2 Probe configurations. The stimulus sets given to F3, G13, and G8 that span the sets trained in Experiment 1 (e.g. KNQ, KPR, KNR) were designed to test this hypothesis and are discussed within each subject's section.

Following the first session of the C to E-I across-class Transitivity Probe set, F6's performance began to markedly deteriorate from an overall score of 92% on the first session to 34% on the second session. Eight more sessions of the C to E-I set were given without much improvement in accuracy and F6 was transferred back to mixed A-B, B-C training. F6 remained training at mixed A-B, B-C for 40 sessions and received varying degrees of S- lid fastening therapy (100%, 75%,

50%, 25%), but was not able to demonstrate reacquisition of the A-B, B-C set, even with 50% fastened S- lids. F6 was then transferred to the A-B set for five sessions, and finally to a simple (S+, S-) discrimination. By transferring F6 to a simple discrimination, we were evaluating whether his performance decline was due to the difficulty of the conditional MTS task, as a correct/incorrect discrimination is notably easier than a conditional one, or an inability to discriminate odors caused by the subject's age and consequential decrease in olfactory capabilities. F6 mastered the simple discrimination in six sessions; the ability to learn and acquire olfactory discriminations had not been lost. F6 was then given an identity MTS task similar to those learned during Experiment 1; X_1-X_1, not X_2 and X_2-X_2, not X_1, using celery and dill as the stimuli. Once again, performance levels declined to about 50%; the MTS task, even identity MTS, was now apparently too difficult to learn. The final effort to bring responding under stimulus control included another identity MTS task with two different odors, cinnamon and nutmeg, to be sure that the celery and dill were not hindering learning. Again, performance remained around 50%. F6 was dropped from Experiment 2 after 20 additional sessions of identity matching with no evidence of meeting criterion.

Subject F3

Similar to his performance during Experiment 1, F3 progressed rapidly through Experiment 2 training discriminations and did not typically require many sessions to meet criterion (Table 6, Figure 18). The maximum amount of sessions to criterion between sets ABC and GHI was eight (B-C) and F3 often achieved

criterion within two consecutive sessions during the early portions of Experiment 2 (e.g. A-BB-C, D-EE-F, GH, HI, G-HH-I). Like F6, the initial change from Experiment 1 to Experiment 2 had no observable effect on performance; F3 scored 22/24 (92%) during session one of A-B conditional discrimination training and required only six sessions to meet criterion. F3 completed a total of 185 sessions across four separate stimulus sets, with an average of 30.8 sessions to criterion per set, although the final KNQ set inflated both measures considerably (up to GHI, $\sum$= 37, $\bar{x}$ = 3.1; KNQ, $\sum$= 148, $\bar{x}$ = 37).

F3 received four exemplars of Emergent Equivalence Probes; ABC, DEF, GHI, and KLM and responded significantly above chance on three out of the four tests (Table 7, Figure 19), with performance on set KNQ falling to 50%. For set ABC, F3 responded correctly on all eight Probe trials (100%), which was statistically significant, p<.003. During Probes for set DEF, F3 responded correctly on 7/8 trials (88%), with an error occurring during a Symmetry Probe trial. The Symmetry error occurred during the F-E discrimination for class 1 (sugar). The summed Probe performance for DEF was statistically significant (p<.03). Performance during the Probe set GHI remained at 7/8 (88%) as well, with an error occurring during an Equivalence Probe trial, I-G, for class 2 (grain). GHI summed Probe performance was also significant (p<.03).

After receiving GHI probes and meeting criterion for the GHI set, F3 was transferred to the across-class phase of Transitivity and Equivalence Probe testing. Figure 17 depicts session by session data for these across-class phases, Table 8 shows the number of sessions required to meet criterion for this additional

phase, and Table 9 depicts the Transitivity and Equivalence Probe performance across all sets. F3 responded correctly on all four untrained baseline conditional discriminations within the A-I Baseline set (4/4, 100%). These discriminations included C-D and F-G discriminations for class 1 and 2 that were not explicitly trained during the previous Experiment 2 stimulus sets (AB**C**, **D**E**F**, and **G**HI). Performance during the next four across-class Transitivity Probe sets varied from 9/12 (75%) to 12/12 (100%) to 8/10 (80%) to 16/16 (100%), but all remained above 50%. Performance during the B to D-I and DEF-GHI Transitivity Probe sets were both statistically significant, $p<.0002$ and $p<.0001$, respectively.

Following completion of the across-class Transitivity Probe sets, F3 also received across-class Equivalence Probe sets, which simply reversed the sample and comparison functions used during the across-class Transitivity sets (see Experiment 2 Method). Five across-class Equivalence sets were given, and performances during all of them were very accurate (92%, 92%, 100%, 100%, 100%) and all but one set resulted in statistically significant performances ($p<.003$, $p<.003$, $p<.001$, $p<.01$). Performance during the H-I to F Equivalence Probe set was highly accurate (100%), but not significant, since it only contained four possible Probe trials. Overall measures of across-class Probes were also very accurate (87/94) and significant, $p<.0001$.

To control for the potential procedural overlap between Experiment 1 training and Experiment 2 testing, F3 was trained a fourth stimulus set (KNQ) that spanned the classes trained during Experiment 1. Performance during KNQ Probes declined from the highly accurate levels observed for the earlier sets.

Overall Probe performance was at 50% (4/8) with errors occurring during all three Probe types. It's important to note that KNQ Probes were conducted approximately 11 months after set GHI Probes were conducted. Training for KNQ followed the across-class Transitivity and Equivalence phases. Thus, the performance deterioration observed during KNQ probes may be influenced by a variety of factors including the length of training required to reach KNQ Probes (145 sessions), the latency between GHI and KNQ Probes (11 months), the age of the subject (almost two years), and an aversion to hickory, to name a few. Regardless, across the four exemplars (ABC, DEF, GHI, and KNQ), F3's performance for all three Probe types was significant (13/16, $p<.01$; 7/8, $p<.03$; 7/8, $p<.03$). His summed total Probe performance was also significant (27/32, $p<.0001$).

During the second session of training for the N-Q baseline discrimination within the KNQ stimulus set, performance for F3 began to decline in much the same manner observed with F6. Where baseline performances tended to be at or above 90% in the past, F3 was now hovering around 70%. After seven N-Q sessions with little improvement, F3 was moved back to the K-N discrimination. Eight more sessions passed before S- lid fastening was implemented. Prior to this stage of experimentation, F3 had not been exposed to any intervention therapy procedures. At this point, F3 was also beginning to show some signs of aversion to hickory (stimulus N_1), which may have increased the delays to criterion achievement observed for the next 11 months. Criterion for set K-N was achieved in 38 sessions after lid fastening had been completely removed. N-Q training

required another 33 sessions and reinstatement of lid fastening therapy. Criterion was then met for K-N, N-Q after another 65 sessions without the aid of fastened lids, but only under the modified 87.5% criterion. KNQ Probes were finally administered six months after initial training with K-N had started. F3 was trained for three KNQ sessions and then removed from the experiment due to age.

Subject H7

H7 acquired and mastered several baseline training discriminations, albeit at a comparatively slower pace than his counterparts F6 and F3. For some sessions, however, H7 did not require many sessions to meet criterion (see Table 6, Figure 18). Examples of this can be seen during his performance for sets A-B and set ABC (four sessions each), set A-B, B-C (six sessions), set E-F (five sessions), and set D-E, E-F (three sessions). The other three sets required more extended training durations and therapy intervention procedures. The maximum amount of sessions to criterion for H7 during Experiment 2 was 18 during set D-E. The initial change from Experiment 1 to Experiment 2 had some effect on performance during the first A-B session (16/24, 67%), but it was temporary; H7 achieved criterion for set A-B in a total of four sessions. Across Experiment 2, H7 completed a total of 71 sessions, with an average of 8.9 sessions to criterion per set.

H7 received only two exemplars Emergent Equivalence Probes, ABC and DEF, before being dropped from Experiment 2 (Table 7, Figure 19). Performance during both sets of probes was at or above 75%, and an increase in accuracy was observed between the first tests with ABC probes to the second tests with DEF

probes. For set ABC, H7 responded correctly on six out of eight total Probe trials (75%), which was not statistically significant, $p>.05$. Errors occurred during two Symmetry Probes (2/4, 50%), one from each class (grain and sugar) for the C-B discrimination. Correct responses were made during both Transitivity and Equivalence Probes for that set (2/2, 100%; 2/2, 100%). During Probes for set DEF, H7 responded correctly on 7/8 trials (88%), again with the single error occurring during the E-D Symmetry Probe trial for class 1 (sugar). The summed Probe performance for set DEF was statistically significant (7/8, $p<.03$). Although the summed performances for each probe type, Symmetry, Transitivity, and Equivalence were not significant, they were still fairly accurate (5/8, 63%; 4/4, 100%; 4/4, 100%). In addition, the summed performance for both exemplars was significant (13/16, $p<.01$).

H7 often required 'therapy' intervention procedures during Experiment 2 training and testing phases. During baseline training for the B-C discrimination, H7 began to respond to whichever stimulus was encountered first, regardless of correctness, and maintained this type of response strategy for multiple sessions. A 10s (and then 15s) time-out procedure was implemented if H7 responded to the S- stimulus first. Specifically, if H7 made a digging response to the S- stimulus first, the MTS tray was removed from the Chamber for 10s, and later 15s. A clipboard was placed in front of the Chamber for the duration of the time-out. Once 10 (or 15) seconds had elapsed, the clipboard was removed and the tray was reinserted. These steps were repeated for however many times H7 responded to the S-. This intervention (like the fastened S- lids in the Arena) was designed to increase

instances of sampling behavior to the stimuli. During D-E training, we attempted to remove the time-out intervention, but performance declined. Thus, the time-out procedure, which began during the eighth session of the B-C discrimination, remained in place throughout Experiment 2. H7 received DEF Probes and spent 17 sessions in post-Probe DEF training without meeting criterion. H7 was then dropped from Experiment 2.

Subject G13

Progress through most Experiment 2 training discriminations was noticeably slower for G13 than for any of the aforementioned subjects. G13 required many more sessions to meet criterion (Table 6, Figure 18) than F6, F3 (prior to KNQ), or even H7 and required many more total sessions across Experiment 2 as well. The minimum number of required sessions to criterion for G13 was two (K-P, P-R), but typical durations were much longer, with a mean 11.9 sessions to criterion per set. For G13, the initial change from Experiment 1 to Experiment 2 also had a greater effect on performance compared to his counterparts. Following the procedural change, G13 dropped to around 60% accuracy and required 20 sessions to meet criterion for A-B. Progress through B-C was considerably more rapid, but slowed again during A-B, B-C training which lasted for 42 sessions. Overall training rates for sets within KPR were much quicker than for ABC; the entire KPR set lasted 34 sessions. At the time of this manuscript, G13 had completed 119 sessions through set L-S and was still training with that set. Thus far, G13 has not required any therapy procedures, but was often advanced through training sets using the modified 87.5% mastery

criterion due to his increasing age and consistent performances between 87.5% and 100% that did not meet the specified original criterion.

G13 received two exemplars of Emergent Equivalence Probes at the time of this manuscript (ABC and KPR), and is currently training to receive a third exemplar, ILS. G13 scored at or above 75% during both Equivalence Probe sets (6/8, 75%; 7/8, 88%), and achieved statistical significance during KPR Probes (Table 7, Figure 19). During ABC Probes, G13 responded correctly on six of eight Probe trials (75%), with errors occurring during one class 2 Symmetry Probe (grain) and one class 2 Equivalence Probe (grain), 3/4, 75% and 1/2, 50%, respectively. His overall ABC Probe performance was not statistically significant (6/8, $p>$. 05), but was comparable to the performances of F6 and H7 during who also scored 6/8 during ABC Probes.

The second Equivalence exemplar for G13, KPR, was part of the manipulation that was designed to control for the potential procedural overlap between Experiment 1 training and Experiment 2 testing. During KPR Probes, G13 responded correctly on 7/8 trials (88%), again, with an error occurring during a class 2 Symmetry Probe trial (grain); 3/4, 75%. This time, the summed Probe performance for KPR was statistically significant (7/8, $p<.03$). For G13, performance accuracies were not lessened by the KPR set, which spanned the training sets from Experiment 1. This contradicts the observed performance of F3 during KNQ Probes, which was much less accurate than any previous exemplar. However, it is important to reiterate the potential influence of the lengthy training durations required by F3 before KNQ Probes were given. As a result, G13 was

comparatively younger in age during KPR Probes than F3 and had experienced the last Probe exemplar about one month earlier, rather than 11. The overall Probe performance for G13 across both exemplars was highly accurate (13/16, 81%) and statistically significant ($p<.01$). Since only two exemplars were given, the overall Probe performances for each of the three Probe types were not significant, but were accurate nonetheless (6/8, 75%, $p=.14$; 4/4, 100%, $p=.06$; 3/4, 75%, $p=.31$).

Subject G8

Like G13, the progress through Experiment 2 baseline discriminations was less rapid for G8 than for the most of the other subjects (Table 6, Figure 18). G8 also required, on average, more sessions to meet criterion during training sets than F6, F3 (prior to KNQ), and H7 ($\bar{x}=11.3$), but was comparable to G13. The minimum amount of required sessions to criterion for G8 was four (A-B, B-C), but overall, individual training set durations were much longer. For G8, the initial change from Experiment 1 to Experiment 2 had a substantial effect on performance. Following the change, performance accuracies dropped to around 40% and required 27 sessions to meet criterion. Progress through the next three sets (B-C, A-B B-C, and ABC) was much faster, with a total of 31 sessions for all three. At the time of this manuscript, G8 had completed a total of 79 sessions through set K-N, N-R and was still in training for that set. In general, current training rates for sets within KNR appear to be much quicker than for sets within ABC. Similar to G13, G8 did not require therapy intervention procedures during Experiment 2, but was occasionally advanced using the modified 87.5% mastery

criterion due to age and consistently accurate, but not original criterion level, performances.

G8 received one exemplar of Emergent Equivalence Probes at the time of this manuscript (ABC), and is training to receive a second exemplar, KNR. As shown by Table 7 and Figure 19, G8 did not perform as accurately during ABC Probes as his counterparts (4/8, 50%) and did not perform significantly above chance levels, $p>.05$, either. For the ABC Probes, G8 responded correctly on four of eight Probe trials (50%), with errors occurring during all but one class 2 Symmetry Probe (grain) and one class 1 (sugar) Transitivity Probe; 1/4, 25% and 1/2, 50% respectively. Correct responses occurred during both Equivalence Probe trials (2/2, 100%). The performance of G8 during ABC Probes was by far the least accurate of all the Experiment 2 subjects, with the exception of the performance of F3 during KNQ Probes.

The second Equivalence exemplar for G8 will be KNR, which is part of investigation into the potential procedural overlap between Experiment 1 training and Experiment 2 testing. Performance on KNR Probes is predicted to be more accurate than performance during ABC, which will add to the evidence against exclusion-based responding provided by the performance of G13 during KPR Probes.

Experiment 2 Summary

All five subjects were able to acquire and maintain baseline arbitrary conditional discriminations, a significant finding in itself for rats. Through the extensive training of Experiment 1, these subjects were able to progress through

multiple arbitrary MTS discriminations, in many cases, at an unexpectedly rapid rate. Some exceptions include H7, who required intervention techniques to learn and maintain some baselines during sets ABC and DEF, and F6 and F3, who required interventions during the later portions of training. G13 and G8 have, thus far, not required any interventions during Experiment 2, but may as time progresses. Generally speaking, all of the subjects towards the end of Experiment 2 were considerably aged. It is possible that the performance declines observed during Experiment 2 could have been affected by this factor, but it is not completely clear what behavioral effects occur as a function of time and age.

In terms of Equivalence Probe performances, most subjects performed at very high levels of accuracy even during the first exposure to Emergent Equivalence Probes. In fact, all but one (G8) responded correctly on over 75% of the Probes during the first exemplar set (ABC). Additionally, most subjects either maintained high levels of accuracy or increased in accuracy across exemplars and baseline sets. Exceptions occurred for F6 and F3 during their third and fourth sets, respectively, and G8, who was still in training for the second exemplar set (KPR) at the time of this manuscript. Moreover, four out of five subjects had overall Probe performances that were well above chance levels and achieved statistical significance. Additional exemplars with G13 and G8 will add to these figures, but regardless, the performances thus far have been impressive and unprecedented.

The results observed during Experiment 1 were not entirely unexpected, judging by the performances observed in Peña et al. (2006). In contrast, the results observed during Experiment 2 far exceeded our initial expectations about the

extent to which these subjects could learn even baseline conditional discriminations, much less demonstrate relational-type responding during emergent Equivalence Probes. Since no comparable results have yet to be obtained with rats, the findings of Experiment 2 are interesting and require further examination. Furthermore, the results of Experiment 2 were often more accurate than the findings observed with other nonhuman animals (D'Amato et al., 1985; Kastak, Schusterman, and Kastak, 2001; Schusterman and Kastak, 1993; Vaughan, 1988), which is all the more concerning. Thus, the reasons for the highly accurate performances are not at all apparent; to be sure, Experiment 2 has raised more questions than it has answered and many avenues are yet to be investigated. Determining which variables contributed to such accurate performances is a necessary goal for future investigations.

One question that we attempted to answer during this study was the effect of Experiment 1 training on the behavior observed during Experiment 2 testing. The first manipulation addressing this question was the design of stimulus sets that spanned Experiment 1 training sets and comparing performance on those sets to the original sets (ABC, DEF, and GHI). Through this manipulation, it was thought that some conclusions about the effect of overlapping training and exclusion could be made. However, the performances of F3, G13, and G8 have yet to provide convincing evidence in either direction. First, F3's performance during KNQ was disrupted by a multitude of factors including time, intervening experience, age, and a stimulus aversion to hickory. For these reasons and more, it is difficult to interpret the decline in performance during KNQ Probes. Second,

although it was his second exemplar, G13's performance during KPR Probes was more accurate than his performance during ABC Probes, suggesting that responding was not exclusion-based. Data from the current training of G8 in set KPR will add to the existing data from F3 and G13, and will allow us to make more conclusive statements about the hypothesized exclusion-based response strategy. We must also consider the length of time between exposures to training sets during Experiment 1 and 2 and whether or not responding could still be under the control of those early learning experiences.

Another attempt to clarify the questions regarding the effects of pre-training occurred during a third experiment. Experiment 3 was designed and conducted with subjects that had no prior training with identity MTS or otherwise, thereby eliminating the effects of past learning. Since these subjects were completely experimentally naïve, they were also of a younger age at the start of arbitrary conditional discrimination training and Equivalence Probes, thereby eliminating the potential influence of age on learning and maintenance of responding.

EXPERIMENT 3

While the across-class manipulation of stimulus sets in Experiment 2 was a useful way to try to eliminate the possibility of an exclusion principle confound, a third experiment (Experiment 3) was set up to investigate equivalence without any potential interference with training history. In addition, the use of experimentally naïve subjects would also further clarify the effects of subject age. To answer questions regarding the required pre-training necessary for subjects to

perform accurately during Equivalence Probes for Symmetry, Transitivity, and Equivalence, Experiment 3 included the addition of two experimentally naïve subjects who had no previous history with identity MTS.

METHOD

Subjects

Experiment 3 included two additional subjects (I4, I29) who received only arbitrary MTS training, with no previous identity MTS history.

Apparatus

Both subjects were trained and tested in the Odor Arena apparatus.

Stimuli

All stimuli used in Experiment 3 were the same as Experiment 1 and 2, as described in General Methods. Both subjects were trained using class-specific reinforcers.

Procedure

Experiment 3 Training: Arbitrary Match-To-Sample Baseline

Experiment 3 included the same A-B, B-C, and mixed A-B, B-C training as Experiment 2. Thus, the A-B discriminations were the first training set for these subjects. All other relevant procedures and training arrangements were identical to Experiment 2. See Table 10 for stimulus configurations for Experiment 3.

Table 10. Experiment 3 Arbitrary Conditional Discrimination Stimulus Presentation Order. Class 1 stimuli were paired with sugar pellets and Class 2 stimuli were paired with grain pellets.

	I4			I29	
Stimuli	1	2	Stimuli	1	2
A	Allspice	Orange	A	Spinach	Marjoram
B	Bay	Onion	B	Rosemary	Lime
C	Beet	Oregano	C	Thyme	Garlic
			D	Tomato	Sumac
			E	Mustard	Nutmeg
			F	Paprika	Fennel

EXPERIMENT 3 RESULTS AND DISCUSSION

To further test the hypothesis that extensive identity MTS training and experience with multiple exemplars of the identity relation are important (and possibly necessary) for demonstration of emergent equivalence, the third experiment was added to the study; arbitrary MTS training and equivalence testing without prior history with multiple exemplars of the identity relation.

Subjects I4 and I29 are still in training, and thus the results for Experiment 3 are somewhat incomplete and a full analysis is pending. Figure 20 shows session by session data across stimulus set presentations for both subjects. Panels indicate shifts in stimulus sets, sets are labeled above each panel. The horizontal line indicates the original mastery criterion of 90%; however, both subjects were under the modified 87.5% criterion at the time of this manuscript. Each data point represents the number of trials where a correct response occurred divided by the total possible trials for a single session.

Sessions to Criterion

Individual subject data for the number of sessions required to meet criterion across stimulus set presentations in Experiment 3 are depicted in Table 11 and Figure 21. A superscript numeral one indicates when subjects were unable to meet criterion, a superscript numeral two indicates a change in criterion (from 90% to 87.5%), and a superscript numeral three indicates when the use of therapy procedures occurred. Again, if therapy intervention procedures were in effect, no criterion-level performances during those procedures were considered for advancement. Table 11 also includes summed totals and means across and within

Figure 20. Experiment 3 Percent Correct across Stimulus Set Presentations. Each data point represents percent correct for one session, or the proportion of correct trials divided by all possible trials. Vertical lines (panels) within each graph indicate changes in stimulus sets; each panel is labeled with the respective set. The horizontal line indicates the original criterion (90%). Any criterion level performance that is not followed by a vertical phase change line occurred during a special training procedure.

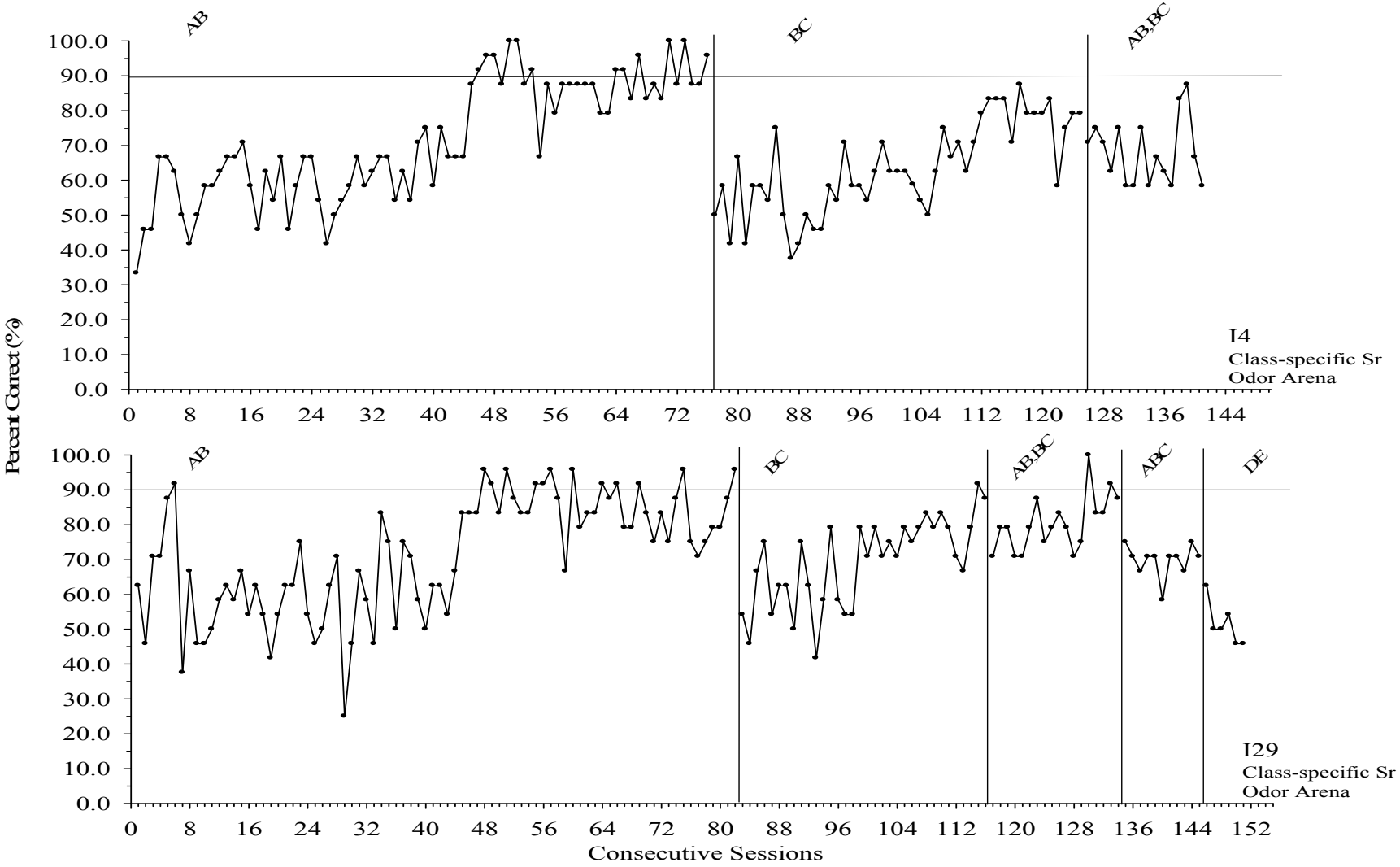
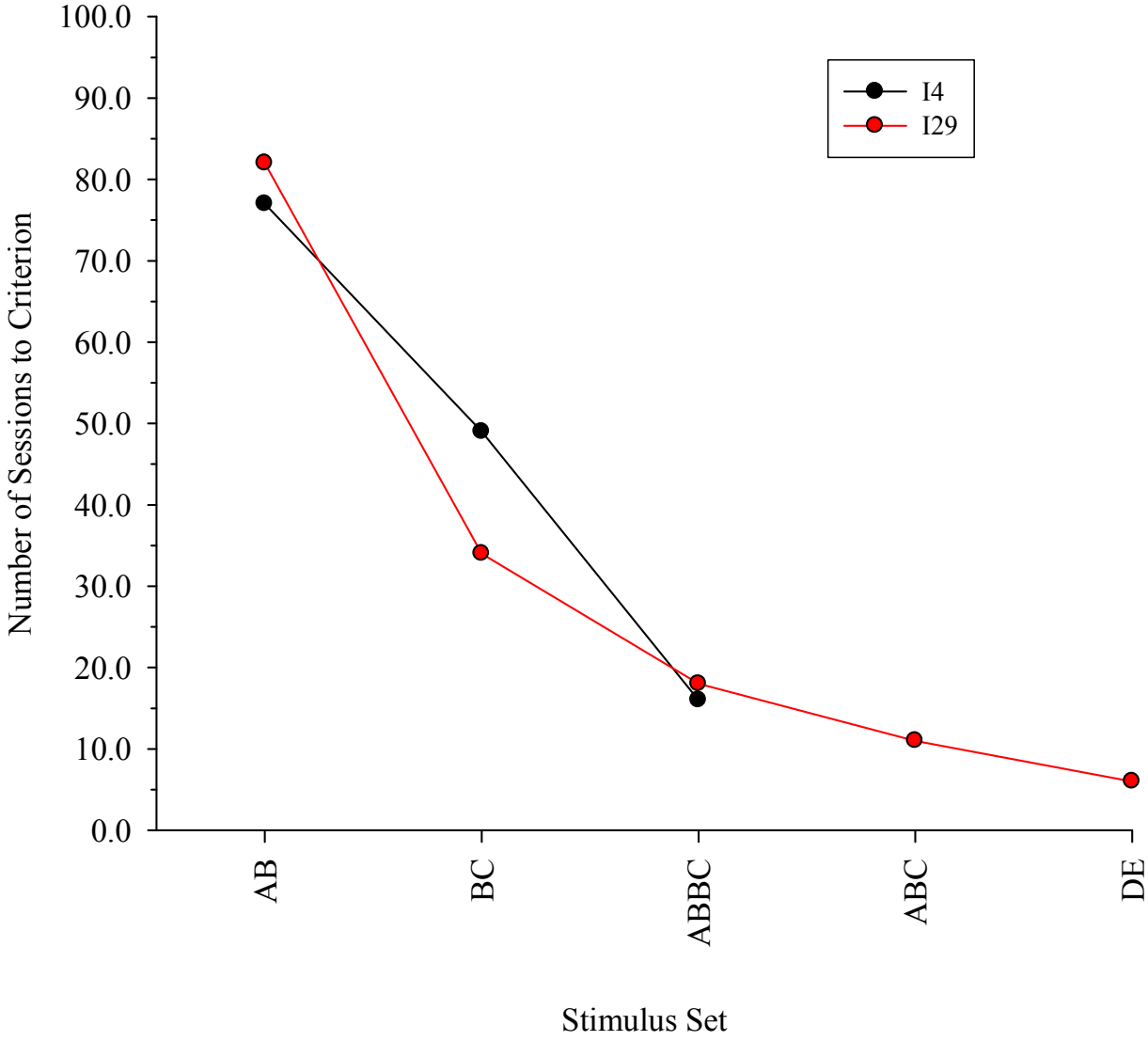

AB
BC
AB,BC
ABC
DE
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
Percent Correct (%)
0
8
16
24
32
40
48
56
64
72
80
88
96
104
112
120
128
136
144
152
Consecutive Sessions
I4
Class-specific Sr
Odor Arena
I29
Class-specific Sr
Odor Arena

Table 11. Experiment 3 Number of Sessions to Criterion across Stimulus Set Presentations. A superscript numeral one denotes that the subject did not meet criterion for the marked stimulus set. The superscript numeral two indicates that the modified 87.5% mastery criterion was in place for that particular set and a superscript numeral three indicates that therapy interventions were used during the marked stimulus set.

Subject / Stimuli	I4	I29	Total	Mean
A-B	77[2, 3]	82[2, 3]	159	64.5
B-C	49[1]	34[2]	83	41.5
A-B, B-C	16[1]	18[2]	34	17
ABC	-	11[1]	11	-
D-E	-	6[1]	6	-
Total	142	151	260	115
Mean	47.3	30.2	-	38.3

Figure 21. Experiment 3 Number of Sessions to Criterion across Stimulus Set Presentations for each subject.



both subjects to show individual, overall, and average learning set trends during the course of Experiment 3.

Subjects I4 and I29 acquired the first A-B baseline conditional discriminations within approximately 80 sessions, considerably slower than some subjects across the entire duration of Experiment 1. In fact, only two of nine subjects from Experiment 1 required over 80 sessions before completing their entire Experiment 1 training by either advancing to Experiment 2 or being dropped from the study. In addition, both subjects were advanced to the B-C training set using the modified 87.5% criterion, rather than the original 90% criterion. The length of time spent in B-C training was about half of the duration spent training with A-B. I29 met the modified criterion in 34 sessions and was advanced to A-B, B-C. I4 never met either criterion during set B-C, yet he was advanced after 49 sessions in an attempt to speed up progress. Currently, I4 is still in training with set A-B, B-C mixed and has completed 16 sessions. I29 met the modified criterion for the mixed A-B, B-C set in 18 sessions, about half of that required for set B-C, and received one set of Equivalence Probes (see below for details). Following 11 sessions of ABC training, I29 was moved, for the sake of time, to D-E without meeting either criterion.

Both subjects showed some evidence for savings in terms of the number of sessions needed to meet criterion across stimulus sets. Although only several stimulus sets were administered to I4 and I29, the overall rate of learning and achievement of criterion during later sets was more rapid than during the first (A-B) set. The average number of sessions to meet criterion in Experiment 3

decreased between the first and second stimulus sets (A-B, $\sum$= 159, $\bar{x}$ = 64.5; B-C, $\sum$= 83, $\bar{x}$ = 41.5) and the second and third sets (B-C, $\sum$= 83, $\bar{x}$ = 41.5; A-B, B-C mixed, $\sum$= 34, $\bar{x}$ = 17). The mean percentage of decrease in the number of sessions to criterion between sets one and two was 36% and two and three was 59%. Although subjects differed from the mean in their individual number of sessions required before achieving criterion, the mean trend is representative of the downward trends in the individual subject data. Decreases between sets three and four (A-B, B-C and ABC) were less profound for I29, with 18 and 11 sessions to criterion respectively.

Baseline Arbitrary Matching-To-Sample

Compared to early training performances for most Experiment 1 subjects, I4 and I29 progressed at a somewhat slower pace. Their performances and durations within each stimulus set resembled the performances and durations of training for the Experiment 1 subjects who did not ultimately advance to Experiment 2 (F5, H8, I30, and J6). As previously stated, some subjects completed the entirety of Experiment 1 (and even several sets of Experiment 2) in fewer than or equal to the number of sessions required for I4 and I29 to master only set A-B. It appears that the arbitrary conditional discriminations were more difficult to learn and required more extensive training durations since they were initiated without any prior training like identity MTS. This notion is not completely unexpected, as identity-type MTS discriminations are more readily acquired by many species than arbitrary MTS discriminations. Even still, the progress through the ABC baseline discriminations for I4 and I29 was fairly slow.

Additional data for both subjects are currently being collected and it is predicted that learning and set acquisition will become more rapid and subsequent baseline training sets will require fewer and fewer sessions to meet criterion. At the time of this manuscript, I29 had received Equivalence Probes for set ABC and was currently training with baseline set D-E. I29's Probe data are discussed below, in Emergent Equivalence Probe Results. I4 was still in mixed A-B, B-C baseline training and had not yet received any Equivalence Probes. His data are discussed in the Baseline Arbitrary Matching-To-Sample section below.

I4

I4 acquired and mastered several baseline training discriminations (see Table 11 and Figure 20). At this time, 142 sessions have been completed through the mixed A-B, B-C set. During the initial 45 sessions of A-B training, I4 averaged approximately 60% correct. Beginning with session 46, I4 received the S- lid fastening intervention in a decreasing fashion; first 100% S- lids were fastened, then 75%, 50%, 25%, and 0%. When session performances were close to criterion, the percentage of fastened lids was decreased. This intervention was very effective; the mean number of correct trials for the final 10 sessions (without S- fastening) was 21.8/24 and the overall percent correct was 91% (218/240 trials). After meeting the modified 87.5% criterion for the last two sessions of A-B, I4 was advanced to B-C training. During the first 30 sessions of B-C, I4 performed around 50%, with a range between 37.5% (9/24 trials) to 75% (18/24 trials). Acquisition began to occur from session 30 through session 49, with a mean of 18.5 correct trials per session and an overall percent correct of 73%

(352/480 trials). After 49 sessions, I29 was advanced to mixed A-B, B-C training without meeting criterion, but was performing consistently around 75% (20/24 trials) for the last 9 sessions. Training with set A-B, B-C is still in progress at the current time. Thus far, I4 has had 16 sessions of A-B, B-C and is averaging approximately 16.3 correct trials per session, with an overall percent correct of 68% (261/384 trials). Since A-B training, I4 has not required therapy interventions again.

Emergent Equivalence Probe Results

I29

Like I4, I29 acquired and mastered several baseline discriminations (see Table 11 and Figure 20) and has thus far completed 151 sessions through set D-E. The initial 45 sessions of A-B training averaged approximately 60% correct for I29 as well. On session 46, I29 also received the S- lid fastening intervention, beginning with 100% fastened, then 75%, and then back to 100% due to decreases in performance accuracy with 75% fastened. After three more 100% fastened S- lids sessions ($\bar{x}$=93%), I29 was transferred back to 75% fastened and then 50%, with some detriments in performance observed. Following performance decline under the 50% fastening, the dummy S- procedure was implemented for 11 sessions, again without much change in performance ($\bar{x}$=19.5). On session 81, the dummy procedure was ended and performance increased to 91% (22/24 trials) during sessions 81 and 82. I4 was then transferred to set B-C.

During the first 16 sessions of B-C, I29 performed around 65%, with a range of 41.7% (10/24 trials) and 79% (19/24 trials). Acquisition of B-C

discriminations began to occur from session 17 through session 34, with a mean of 18.7 correct trials per session and an overall percent correct of 78% (336/432 trials). After achieving criterion, I29 was advanced to mixed A-B, B-C training. Training with set A-B, B-C required only 18 sessions to meet criterion, comparatively less than for A-B or B-C training. No therapy interventions were required for either B-C or A-B, B-C training. I29 was then given ABC Equivalence Probes.

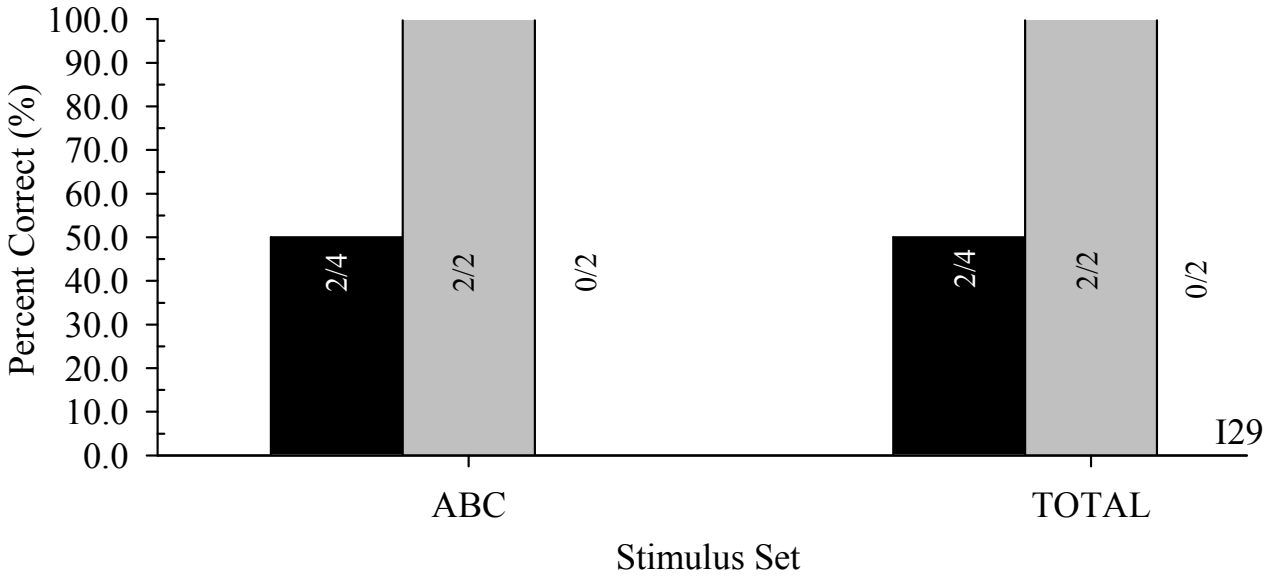
Performances for I29 for Emergent Equivalence Probes are depicted in Table 12 and Figure 22. At the time of this manuscript, I29 had received Probes for set ABC only. In both displays, Probe performances for ABC are shown separately for the three types of probe trials (symmetry, transitivity, equivalence). The proportions represent the number of correct trials over the total possible number of trials, in this case four symmetry and two each for transitivity and equivalence trials. The eight Probe trials were given during the first eight trials of the ABC Probe session. S+ comparison stimuli were unbaited to control for pellet detection. The vertical bars in Figure 22 represent percent correct measures for the three trial types. In Table 12, the proportion in parentheses represents the total correct probe trials over the total possible trials (8) for the entire ABC stimulus set. No asterisks (binomial significance, $p<.05$) are shown in Table 12 because none of the Probe performances were statistically significant.

As Table 12 indicates, I29 did not perform above 50% during ABC Probes. The total Probe performance for set ABC was 4/8 (50%), which was not statistically significant ($p>.05$). Specifically, I29 responded correctly on two of

Table 12. Experiment 3 Emergent Equivalence Probe Performance across Stimulus Set Presentations. The first proportion represents the number of correct Symmetry Probe trials over the total given, the second represents the number of correct Transitivity Probe trials over the total given, and the third proportion represents the number of correct Equivalence Probe trials over the total given. The proportion in parentheses represents the overall number of correct Probe trials over the total given for that set.

Subject	I29
Stimuli	Symmetry Probes (X/4); Transitivity Probes (X/2); Equivalence Probes (X/2) Total Probes (X/8)
A, B, C,	2/4; 2/2; 0/2 (4/8)

Figure 22. Experiment 3 Percent Correct for Emergent Equivalence Probes.



four Symmetry Probes (50%), making errors during the B-A and C-B Class 2 (grain) trials. Correct responses were made during both Transitivity Probe trials (2/2, 100%), while both responses during Equivalence Probes were incorrect (0/2, 0%). I29 did not perform as accurately during ABC Probes as many of the Experiment 2 subjects on their respective ABC Probes, with the exception of G8. Both I29 and G8 responded correctly on four out of eight (50%) Probes, but their individual errors occurred during different trial types.

Experiment 3 Summary

Experiment 3 was conducted in an attempt to answer questions that arose during Experiment 2. One question we attempted to answer was the effect of Experiment 1 training on the behavior observed during Experiment 2 testing and whether or not naïve subjects could demonstrate similar performances without similar identity MTS pre-training. Experiment 3 was the second attempt to clarify the effects of pre-training; the first occurred during Experiment 2 with the stimulus sets that spanned the sets trained in Experiment 1. We were also interested in the detrimental effects of age, and by using experimentally naïve subjects during Experiment 3, it was thought that we could gather more training and testing data than was possible in Experiment 2. Neither of the initial expectations was completely realized through Experiment 3, but our tentative analyses are described below. Since training and testing phases for I4 and I29 are ongoing, we will also be able to further evaluate our questions in the near future.

Both Experiment 3 subjects were able to acquire and maintain baseline arbitrary conditional discriminations for at least three stimulus sets, although the

duration of training for some sets was considerably long. That being said, the amount of training required for I4 and I29 during the sets within ABC was still comparable to the overall length of training for subjects that experienced both Experiment 1 and 2. It is interesting to note that although identity MTS pre-training was not included for I4 and I29, they still required about the same number of training sessions (A-B and B-C) as most Experiment 1 subjects required in order to move to Experiment 2. Perhaps arbitrary conditional discrimination learning and Equivalence Probe performance is dependent on the total number of training sessions and not the type of training in particular (e.g. identity MTS versus arbitrary MTS). In addition to learning baseline discriminations, I29 was also able to advance far enough to receive one exposure to Equivalence Probes, although his performance was not more than would be expected by chance.

As far as the Equivalence Probe performance, I29 did not performed at levels significantly different from chance. Whether the average Probe performance resulted from a lack of identity MTS experience (and exclusion experience) has yet to be determined. Because I29 was the only subject to receive Probes, we cannot rule out the possibility that he was the Experiment 3 equivalent of G8, who was the only Experiment 2 subject who did not score above 50% during ABC Probes. A thorough evaluation of his performance is difficult to make without data from I4 and without any additional performance data for other Probes. Any comparisons to Experiment 2 subjects are also incomplete for the same reasons. Data from ABC Probes for I4 and DEF Probes for I29 will help

clarify these questions and allow us to make more concrete conclusions about the differences between Experiment 2 and Experiment 3 subjects.

Since these subjects were completely experimentally naïve, they were also of a younger age at the start of arbitrary conditional discrimination training and Equivalence Probes, thereby eliminating the potential influence of age on learning and maintenance of responding. At this point, I4 and I29 are still fairly young and will continue training and testing in Experiment 3 until performances decline or other circumstances intervene with experimentation. However, because they are still young, the training difficulties observed during Experiment 3 cannot be attributed to the effect of age.

The results observed during Experiment 3 were not surprising, considering the length of Experiments 1 and 2 and the difficulty of the arbitrary MTS discriminations even for subjects with extensive histories. Beginning training with arbitrary conditional discriminations (four total stimuli in set A-B) may have been too difficult, especially compared to the two stimuli that Experiment 1 subjects began training with during identity MTS set A. If Experiment 1 subjects started with the six stimuli characteristic of all subsequent sets following A (e.g. BCD, EFG), baseline acquisitions may have been more like those observed for I4 and I29. Overall, our goal of clarifying the variables that were responsible for Experiment 2 performances was not achieved via Experiment 3. Instead, the results of the Experiment 2 KNQ, KNR, and KPR Probes appear to be of greater value than the single ABC Probe with I29. Thus, a more definitive conclusion

about which variables contributed to the accurate Equivalence Probe performances remains a necessary goal for future experiments.

Controls

Data for non-baited pellet detection trials are shown in Table 13 and are compared with performance on trials that were baited during sessions that contained non-baited control trials. Table 13 contains the proportion of correct responses during non-baited trials over the total number of non-baited probes given, the percent accuracy for that proportion, and the proportion and percent accuracy for baited trials during the session in which non-baited probes were given. The first set of numbers represents the proportion and the percentage for non-baited trials while the second set of numbers (within parentheses) represents the proportion and percentage for baited trials during the sessions in which non-baited probes were conducted. These two sets of numbers allow comparisons between performances on trials that did not contain a pellet and performances on trials that did. Any significant differences in performance between baited and non-baited trials may indicate tracking the odor of the pellet, if performance was enhanced during baited trials.

For most subjects, non-baited pellet probes occurred over multiple sessions and often across experiments. The only subject that did not receive any non-baited pellet detection probes was I30. Since his performance was never much greater than would be expected by chance (about 50%), it was deemed unnecessary to incorporate pellet probes. If his performance had increased to near-criterion levels, non-baited detection probes would have been conducted. For all other

Table 13. Non-baited Pellet Detection Probe Performance for Individual Subjects across Experiments 1, 2, and 3. The proportion and percentage in the first row represent the number of correct responses during non-baited probe trials over the total number of non-baited probe trials given. The proportions and percentages in parentheses represent performance during baited trials that occurred within the same sessions as non-baited probe trials. Asterisks denote binomial significance ($p<.05$). Subject I30 did not receive any pellet detection probes, which is indicated by the superscript numeral one.

Subject: / Experimental Phase:	F6	F3	F5	H7	H8	G13
Proportion of Correct Responses for Non-baited Trials/Total Non-baited Trials; Percent Correct (Proportion of Correct Responses Baited Trials/Total Baited Trials during Sessions with Non-baited Controls; Percent Correct)						
Experiment 1	35/49*; 71% (293/335; 87%)	38/46*; 83% (276/306; 90%)	4/6; 67% (32/66; 48%)	23/32*; 72% (93/112; 83%)	14/16*; 88% (48/56; 86%)	25/48; 52% (123/144; 85%)
Experiment 2	83/96*; 86% (267/298; 90%)	208/235*; 89% (710/795; 89%)	-	13/16*; 81% (32/36; 89%)	-	73/84*; 87% (383/468; 82%)
Experiment 3	-	-	-	-	-	-
Total	118/145*; 81% (560/633; 84%)	246/281*; 88% (986/1101; 90%)	4/6; 67% (32/66; 48%)	36/48*; 75% (125/148; 85%)	14/16*; 88% (48/56; 86%)	98/132*; 74% (506/612; 83%)

Subject Experimental Phase:	G8	I30	J6	I4	I29
Proportion of Correct Responses for Non-baited Trials/Total Non-baited Trials; Percent Correct (Proportion of Correct Responses Baited Trials/Total Baited Trials during Sessions with Non-baited Controls; Percent Correct)					
Experiment 1	56/75*; 75% (253/309; 82%)	0[1]	22/33*; 67% (122/159; 77%)	-	-
Experiment 2	35/41*; 85% (289/343; 84%)	-	-	-	-
Experiment 3	-	-	-	10/18; 56% (129/150; 80%)	23/38; 61% (227/274; 83%)
Total	91/116*; 78% (542/652; 83%)	-	22/33*; 67% (122/159; 77%)	10/18; 56% (129/150; 80%)	23/38; 61% (227/274; 83%)

subjects, non-baited trials were conducted across different sets to determine if pellet detection was occurring with different odors. For the most part, performances during non-baited trials were not different from performances during baited trails both across and within subjects in all three experiments. During Experiment 1, performance for F6, F3, H7, H8, G8, and J6 during non-baited trials was significantly greater than chance ($p<.05$). All five Experiment 2 subjects responded highly accurately during Experiment 2 non-baited trials as well. All of their Experiment 2 non-baited trial performances were significant. Both Experiment 3 subjects showed a decline in performance levels across baited and non-baited trials, but remained above 50%. Neither subject's performance during non-baited probes was significant ($p<.05$).

It is apparent that some subjects' performance during baited trials exceeded that of non-baited trials. There are several explanations for these differences. First, it should be noted that the sheer number of baited trials, compared to non-baited, tends to somewhat inflate the percentage for accuracy for baited trials. This may account for the observed differences across trial types. It should also be noted that although some subjects' performances during baited trials exceeded those observed during non-baited trials, their performances within each individual session was consistent across both baited and non-baited trials. In other words, when performance was poor during non-baited trials during a given sessions, it tended to be poor during baited as well, and vice versa. It was never the case that performance on baited trials was perfect (e.g. all baited trials correct) while

performance on non-baited controls was not (e.g. sole errors occurred during non-baited trials). Overall, errors and successes typically spanned both types of trials.

It is also important to mention that many non-baited trials occurred during probe trials. This was intended to control for pellet detection during probe tests, allowing for more concrete conclusions about observed performances. Thus, the subjects that received generalized identity MTS and equivalence probe trials during Experiment 1, 2, and 3, also received non-baited probes during the same trials. For example, during Experiment 1, G13 received pellet detection probes only during generalized MTS probes. As a result, his non-baited performance is identical to his Experiment 1 probe performance and thus, the two manipulations are confounded with one another in the sense that performances could have been affected by the lack of pellet odor or the novelty of the odors during generalized probes. Similarly, H7 received most non-baited probes during generalization and equivalence probes in Experiment 1 and 2, although not all. The same caveats are therefore applicable for H7 as well. Although subject F6 received non-baited trials in both baseline and during generalization probes, 36 of the 49 total pellet probes during Experiment 1 coincided with generalized MTS probe trials, with the remaining 12 occurring during baseline training. The discrepancies observed for F6 in Experiment 1 may be due to the larger number of non-baited probes that occurred with generalized MTS probes. Perhaps additional non-baited probes collected during baseline training would have given a better measure of non-baited performances.

For all other subjects (including G13 and F6 during Experiment 2), this was not the case, and several baseline sessions also contained non-baited trials. Performance on non-baited trials during baseline trials novel probes tended to be more accurate than non-baited probes during novel probes. For those subjects who had non-baited probes in multiple different circumstances, the discrepancies are less apparent (See notes about I4, I29 below). Again, compare performance differences for G13 during Experiment 1 and Experiment 2 after inclusion of non-baited trials during baseline in Experiment 2. For future experimentation, such procedures are discouraged for those reasons. Non-baited trials should not necessarily or exclusively coincide with probe trials, as both measures could potentially be affected by the inclusion of the other.

Although most subjects did not show great performance differences across non-baited and baited trial types, I4 and I29 did. These two subjects showed the greatest performance discrepancies between baited and non-baited trials. Since performance during baited trials appears significantly more accurate than during non-baited trials, it may be concluded that these subjects were tracking reinforcers and choosing stimuli based on the pellet odors. For example, I4 responded correctly on 10/18 non-baited trials (56%), but was at approximately 80% correct during baited trials (129/150) that occurred within sessions that contained non-baited controls. Yet, if this apparent advantage was present, we might expect to see more rapid progress or increased levels of accuracy. This was certainly not the case for the subjects mentioned. For example, I4 required many sessions to reach criterion for baseline discriminations and had not reached ABC equivalence

testing at the time of this manuscript. I29 showed similar trends for higher accuracy during baited trials (227/274, 83%) compared to non-baited (23/38, 61%). Although he progressed somewhat more rapidly through baseline discriminations than his Experiment 3 counterpart I4 (and received one equivalence test session), his overall training performance was variable and did not appear to be enhanced by pellet detection. To conclude, if pellet detection was guiding subjects' behavior during experimentation, performances should have been much more accurate and training durations should have been much less lengthy. Also, neither I4 nor I29 received many non-baited probes, and continued testing (more non-baited probes) may reveal less of a difference between non-baited and baited trials.

Occasional inter-observer agreement sessions were conducted and results indicated high levels of agreement across various researchers and for various different subjects. These sessions were used to verify consistent scoring methods across observers. Subjects F6, F3, F5, H8, and G8 each received one inter-observer agreement session. Agreement was 100%, 100%, 95%, 83%, and 100% respectively. Agreement sessions were conducted during Experiment 1 training for F5, H8, and G8 and during Experiment 2 for F6 and F3. Lower levels of agreement were obtained during sessions in the Operant MTS Chamber with F5 and H8 (95%, 83%), partly due to the lack of lids and inherent difficulty identifying and scoring the digging response. The use of lids in the Arena allowed for perfect agreement between multiple observers across three different subjects because the response was much more identifiable. Additionally, subjects G13, I4,

and I29 were trained and tested by individuals other than myself for a large portion of their training, providing further evidence for inter-observer agreement and lack of experimenter cuing. G13 was trained and tested by another researcher throughout the duration of Experiment 1 and the beginning of Experiment 2, at which time he was transferred to me without much disruption in performance. Subjects I4 and I29 were trained and tested by a third experimenter for the initial portions (all of A-B training and beginning of B-C training) of Experiment 3 and were later transferred to me as well. Again, little performance disruption was observed during the transition between experimenters. All three of these animals as well as J6 were transferred to yet another experimenter at the time of this manuscript.

Finally, F3 also received a series of sample reinforcement reduction sessions during Experiment 2 to evaluate whether or not reinforcing responses to the sample stimulus was contributing to learning. Over the course of 11 sessions, F3 received 76 trials without sample reinforcement, 73 of which he responded correctly (73/74, 99%, $p<.0001$). Sample reinforcement was reduced to 75% for the first four sessions without any disruptions in performance. In fact, F3 remained above criterion for all three sessions with 75% sample reinforcement. Sample reinforcement was then further reduced to 50% for one session, again without any performance disruptions in that session. Sample reinforcement was then increased to 75% for an additional session without disruption. Five more sessions with varying percentages of sample reinforcement ended this manipulation, and resulted in 34 trials without sample reinforcement, 32 of which

he responded correctly (32/34, 94%, $p<.0001$). Similar sample reinforcement reduction phases were more formally and thoroughly investigated in Peña et al. (2006) where subjects were exposed to a series of sample reinforcement reduction densities (e.g. 75%, 50%, 0%) also without any observed disruptions in performance or learning. Since the reduction was clearly unimportant to learning in Peña et al. (2006), a more systematic manipulation was not included here.

GENERAL DISCUSSION

In regards to the first major hypothesis proposed, the use of olfactory stimuli, multiple exemplars, and a naturalistic response requirement appears to facilitate learning during MTS procedures that use a multiple exemplar format. The combined results of Peña et al. (2006) and the current study support the assertion that olfactory stimuli and multiple exemplars promote MTS acquisition (both identity and arbitrary MTS), generalization to novel stimuli, and even transfer to untrained discriminations when rats are used as subjects. Most of the subjects in the current study were able to reach advanced stages of training and testing, stages that were not reached by the subjects in the Iversen (1993, 1997) studies that used visual stimuli and did not attempt to train multiple exemplars. By using olfactory stimuli, we have demonstrated that these animals are capable of learning baseline discriminations that were once thought to be too complex. We have also demonstrated that rats can extend what is learned through training to novel or untrained discriminations when using olfactory stimuli and multiple exemplars. The effect of baseline exemplar training on performance can be seen in the required number of sessions to criterion across all three experiments within

this study. The effects of exemplars of the equivalence relations (generalized identity, symmetry, transitivity, and equivalence) can be seen in the performance of subjects across Probe tests, although these trends are less obvious. Finally, the performances from the current study are comparable to the performances of other species using modality-specific stimuli. For example, the duration of Experiment 1 probably approximates the duration of comparable training procedures (e.g. visual or auditory stimuli) for other nonhuman species such as pigeons, primates, and marine mammals.

The results of the current study combined with the results of Peña et al. (2006) also support the use of a naturalistic digging response rather than a more automated response such as a lever press or nose poke. Similar research in the same laboratory (Thomas, 2006) investigated olfactory MTS learning using a more automated response, specifically, breaking a photo beam with the nose for a count of three seconds, but the subjects were slow to learn or unable to learn even baseline discriminations. Although the lid removal response used here is less discrete than breaking a photo beam and therefore requires more experimenter discretion, it apparently promotes MTS learning. Designing a system that incorporates the naturalistic response requirement and an increased degree of automation (e.g. computer scoring) will be important for future investigations.

Our second major question required the incorporation of class-specific reinforcement as a between-subjects variable. This manipulation was initially organized such that half of the subjects were trained using a single type of reinforcer (sugar), while the remaining half received two types (sugar and grain),

which were differentially associated with each class of stimuli (class 1 or class 2). This reinforcer assignment was intended to assess the value of differential outcomes on equivalence class formation, as proposed by Sidman (2000) and as used by Kastak et al. (2001) with sea lions. In the Kastak et al. (2001) functional equivalence study, both subjects performed near 50% prior to the assignment of class-specific reinforcers (phase 2) and near 60% following their removal (phase 4). During the initial class-specific reinforcement introduction (phase 3), performances increased to practically 100% accuracy and again reached 90-100% following reintroduction (phase 5). For our purposes, class-specific reinforcers were not particularly interesting for identity MTS phases; instead, the use of class-specific reinforcers was designed to accelerate advancement to Experiment 2 and to facilitate arbitrary MTS learning during Experiment 2. It was also assumed that the addition of the class-specific reinforcer variable would allow subjects to transfer responding during the Experiment 2 emergent probe trials.

As Experiment 1 progressed, it became apparent that the performance of subjects receiving only sugar reinforcers was much more variable in both accuracy and rate of acquisition than the overall performance of those subjects receiving class-specific reinforcers, even during initial training discriminations (e.g. set A). Thus, with one exception (F6), all Experiment 1 subjects were ultimately transferred to the class-specific reinforcement condition upon indication by the data that performance was consistently at chance levels and all other intervention attempts had failed to affect responding. In this case, it appears as though the inclusion of class-specific reinforcers was a necessary condition to

achieve behavioral control by the stimuli, at least in a timely fashion. Then again, there were no apparent performance differences (rates, accuracy, or otherwise) between F6 and the four class-specific subjects (F3, H7, G13, and G8) that advanced to Experiment 2. Thus, perhaps not as originally proposed, the current investigation seems to still underscore the importance of class-specific reinforcers outlined in Kastak et al. (2001), since most subjects receiving class-specific reinforcers performed highly accurately and since most subjects receiving non-specific reinforcers did not.

Alternatively, class-specific reinforcers were not a necessary condition in Peña et al. (2006), which used only sugar reinforcement, but they were an important influence on class formation in Kastak et al. (2001). Perhaps the format of Experiment 1 was more difficult to master compared to the identity MTS procedures in Peña et al. (2006), where one or two novel stimuli were added to the existing baseline in a staggered fashion rather than all at once and with no baseline trials. It is also possible that the delayed MTS format of the Odor Arena added extra difficulties that were not present in the simultaneous MTS format of the MTS Chamber used in Peña et al. (2006), and the use of differential reinforcement minimized the effects of delay. In comparison though, the highly accurate performance of F6 demonstrates that acquisition and transfer during both Experiment 1 and 2 is possible using only sugar reinforcers. It may be the case that the other subjects initially trained with only sugar reinforcers (F5, H8, I30, and J6) were learning at a slower rate due to individual differences, rather than as a result of the reinforcer condition. For example, consider the relatively slow,

inaccurate performance of those subjects even after being transferred to class-specific reinforcers; only J6 received more than one probe session during Experiment 1. The fact that the duration of training and testing for F6 was not apparently longer than training and testing for any particular subject receiving class-specific reinforcers may also result from individual differences.

Our third main question focused on the differences between two apparatus and also required an additional between-subjects manipulation during the beginning of Experiment 1. Half of the subjects in Experiment 1 were assigned to the same modified operant chamber used in Peña et al. (2006), the Operant MTS Chamber, which varied comparison stimulus positions across left and right sides. The remaining subjects in Experiment 1 were assigned to the Odor Arena which included a highly variable arrangement of possible comparison stimulus positions. It was thought that the varied stimulus locations in the Odor Arena apparatus may eliminate the possibility of control by stimulus position, which frequently affects the performance of nonhumans in MTS procedures (Iversen, 1997). By controlling for location effects, the Odor Arena apparatus was predicted to promote control by stimulus odor. As a result, it was thought that Odor Arena subjects would be less likely to show position preferences and would therefore be more likely to acquire and maintain discriminations than the subjects assigned to the Operant MTS Chamber.

Again, differences between apparatus assignments became apparent as Experiment 1 progressed. Compared to Odor Arena subjects, most of the subjects assigned to the Operant MTS Chamber apparatus demonstrated position biases

and had difficulty mastering even the first MTS discriminations (e.g. set A). An additional subject (H8) was added to the Operant MTS Chamber group to determine whether performance discrepancies were due to individual differences or apparatus differences. The added subject performed similarly to the other Operant MTS Chamber subjects. Thus, all but two subjects, one who performed at high levels of accuracy (H7) and one who was dropped early in Experiment 1 (F5), were ultimately transferred to the Odor Arena apparatus. All subsequent subjects added to the experiment were assigned to the Odor Arena. Like subjects within the non-specific reinforcement condition, most subjects in the Operant MTS Chamber were not able (or less able) to acquire and maintain discriminations, which leads us to conclude that the Odor Arena was more conducive to learning with these procedures. Such a conclusion may not be completely accurate, though, since the subjects in Peña et al. (2006) acquired and mastered identity MTS in the MTS Chamber. Again, perhaps the procedures for introducing novel stimulus sets during Experiment 1 were different enough to complicate learning in the MTS Chamber, but not in the Odor Arena.

It is important to note that none of our subjects were able to learn even one baseline identity MTS discrimination (A) while training in the Operant MTS Chamber with non-specific reinforcement – sugar pellets only. This combination of apparatus and reinforcer conditions was the norm for all of the subjects in Peña et al. (2006), and they were all able to learn identity MTS baselines and even demonstrate generalized identity during novel probes. The remaining procedural difference lies in the format of training and introduction of novel stimuli, as

previously mentioned. It is this difference that probably prevented our subjects from replicating the Peña et al. (2006) results in the Chamber with non-specific reinforcers, and should be taken into consideration when designing MTS procedures. Future investigations should focus on the effect of these two procedural designs and MTS learning in both apparatus.

Amidst the successes of Experiment 1 are four subjects who either failed or were extremely slow to acquire and/or maintain responding during baseline identity discriminations. To observe such vast differences in responding for subjects that received identical reinforcer, apparatus, and procedural conditions is somewhat perplexing, although not entirely unexpected. Differences could be due to the complex nature of the experimental environment, such as the extensive experimenter contact or the competing control of non-stimulus odor cues (e.g. subject odors, slight differences in stimulus odor), individual subject differences (e.g. initial shaping experience, sensitivity to handling, odors, or reinforcers), or the procedural design. For most of these four subjects, additional training sessions or 'therapy' interventions may have brought responding under olfactory stimulus control, but because of time requirements they were not conducted. Of these four, only J6 continues to be tested for identity MTS relations and is currently still progressing slowly. Creating simplified procedures and experimental environments (apparatus, stimuli, etc.) may reduce the number of subjects that do not learn these discriminations.

In sum, we have learned that using olfactory stimuli, multiple exemplars, and a non-automated response often allows for rapid acquisition and mastery of

baseline identity MTS discriminations and transfer of responding to novel stimuli. Second, we have learned that with the procedures used here, some of our subjects were better able to learn and maintain high levels of accuracy when they received class-specific reinforcers. This is not to say that class-specific reinforcers are necessary for identity MTS learning, as demonstrated by F6 and the subjects from Peña et al. (2006), but only that they generally allowed for more rapid rates of learning in this study. Class-specific reinforcers also allowed us to proceed to training with arbitrary MTS procedures with a better chance of transfer during emergent equivalence probes. We certainly did not determine whether or not reinforcers were members of the stimulus classes we designed nor if they were contributing to the impressive performances we observed, a task that future research should carry out. Similarly, we learned that training in the Odor Arena allowed for rapid set acquisition and high levels of accuracy for most of our subjects, whereas training in the Operant MTS Chamber did not. In this case the single exception was H7, who was able to rapidly learn and master identity and arbitrary MTS discriminations in the Operant MTS Chamber. Overall performances were much more accurate and impressive for subjects in the Odor Arena than the Chamber, and it appears as though the variable configurations of stimuli in the Arena (at least for our procedures) effectively controlled for spatial biases typically observed in nonhumans.

Beyond our three main hypotheses about reinforcers, exemplars, and apparatus conditions, we were able to further extend the findings of Peña et al. (2006) because five of our subjects quickly demonstrated generalized identity

MTS. As the kinks in Experiment 1 were ironed out, our main Experiment 2 goal was the training arbitrary MTS discriminations and maybe even testing for emergent equivalence. Given the extensive history of identity matching achieved through Experiment 1 and the use of olfactory stimuli and class-specific reinforcers, we thought that subjects would learn arbitrary conditional MTS baseline discriminations - and they did. Following the rapid mastery of the first few discriminations of Experiment 2 for F6 and F3, it became clear that equivalence tests would also occur - and they did. In fact, Equivalence Probes were given to all of the subjects who reached Experiment 2. No subject failed to acquire at least one stimulus set, and all but G8 advanced beyond the first set (ABC) to receive additional exemplars. Although we predicted that subjects would eventually show some evidence of transfer of responding (perhaps after several exemplars), we did not foresee it occurring at the early stages we observed or with the accuracy we observed. The rare nature of the performances observed during Experiment 2 requires some further consideration and discussion. Why were such performances observed?

Our first conclusion relates to the pre-training received during Experiment 1. Maybe extensive training with exemplars of the identity relation provide the breadth necessary for learning baseline discriminations and for demonstration of emergent relations, as indicated by Peña et al. (2006), Frank & Wasserman (2005), and Schusterman & Kastak (1993). Frank & Wasserman (2005) investigated the necessary conditions for symmetrical responding by varying training procedures such that pigeons received either both identity and arbitrary

MTS training, just arbitrary MTS training, or first arbitrary training with identity discriminations added later. The subjects that received both identity and arbitrary MTS from the outset of training demonstrated symmetrical responding during probe trials, whereas subjects who received only arbitrary MTS or arbitrary and later inclusion of identity MTS did not. The authors concluded that the inclusion of identity matching discriminations from the beginning of training allowed subjects to demonstrate symmetrical responding. Observed differences may result from the amount of identity MTS experience across the groups or perhaps the fact that identity MTS discriminations train variable sample and comparison stimulus functions. Both of these conclusions are potential reasons for the performances of our subjects, as well.

Schusterman & Kastak (1993) expressed similar ideas regarding the required training for demonstration of equivalence responding for Rio, the main subject of their 1993 equivalence study. Rio received extensive generalized identity MTS training and testing (Kastak & Schusterman, 1994) prior to any arbitrary MTS training or equivalence testing during Schusterman & Kastak (1993). The history with generalized identity MTS may have prepared Rio for arbitrary MTS baselines, as she readily acquired the A-B and B-C baseline discriminations with an average of 36.2 and 17.8 errors to criterion, respectively. In addition, Rio was required to meet criterion for 30 arbitrary MTS baselines (e.g. A-B, B-C) before being tested for the corresponding untrained relation (e.g. B-A, C-B, A-C, C-A). The authors concluded that the extensive arbitrary MTS baseline and equivalence exemplars allowed Rio to demonstrate transfer of

responding during emergent equivalence tests. They did not specifically mention the influence of her previous identity MTS training; however, it seems to us that it certainly did not hinder her later performances. Regardless, once she learned the interchangeability of stimulus functions through exemplars of symmetry, Rio was able to perform accurately in the presence of novel equivalence probes, termed derived symmetry. As for our most of our subjects (F6, F3, H7, and G13), responding during the *first* Equivalence Probes was highly accurate, and even significantly above chance for F3. Unlike Rio, the accurate performances for these four subjects were observed before any equivalence relations were trained as exemplars. Perhaps the extensive generalized identity MTS and mixed-node arbitrary MTS training allowed our subjects to transfer during the first set of probes. A replication of the training procedures used in Schusterman & Kastak (1993) would clarify the effects of training differences on emergent equivalence performance.

On the other hand, it seems premature to simply attribute our results to emergent equivalence brought about by the combination of training, olfactory stimuli, apparatus, and class-specific reinforcement (for applicable subjects). Could there possibly be more parsimonious reasons for the performances observed? What other potential sources of control may have been allowing subjects to demonstrate this type of behavior? Obvious potential sources include the role of pellet detection, either by visual or olfactory cues. This potential confound was previously discussed and is not thought to play a significant role in emergent performances given the fact that 1, pellets were buried beneath the sand,

2, translucent lids were placed on the cups (in the Arena) and 3, all probe trials were non-baited. Besides pellet detection, several other external sources of control have been identified, including the fact that Probes were reinforced, the occurrence of the procedural artifact discovered in Experiment 2, and experimenter influences like handling, scoring, and cuing.

A major criticism of many nonhuman equivalence studies is the use of reinforcement during probe tests for generalized identity matching or emergent equivalence relations. The main issue is the fact that once responses are reinforced during these tests, any subsequent performances are technically not novel or emergent because training has occurred. This criticism is applicable for our subjects as well, since all probe responses were reinforced during all three experiments. Reinforcing probes was for the most part, intended to maintain responding, but was also included so that we could observe any performance trends across exemplars. For Experiment 1, there was some indication that some subjects performed more accurately during later exemplars of the identity relation; however, these trends were inconsistent and not universal. In fact, some Experiment 1 subjects responded at or above 67% during Novel BCD Probes. Similarly for Experiment 2, many of our subjects performed highly accurately even during the first exposure to emergent equivalence tests (set ABC), despite the lack of prior exemplar training. The results of future equivalence probes for G13, G8, I4, and I29 will help clarify the effect of multiple exemplars of the equivalence relations. Judging by our current results, we do not encourage the use

of extinction during novel or emergent probe tests especially if it is at the expense of maintaining responding.

Preliminary answers to the overlapping training question were addressed through the KNQ, KPR, and KNR across-class stimulus sets integrated into Experiment 2 for F3, G13, and G8. Again, for these subjects, performances have been variable. Experiment 3 was also designed to answer this procedural question and is similarly inconclusive at this time, as both subjects continue to be trained and tested. What is clear from Experiment 3 is that the baseline performances for I4 and I29 demonstrate that prior identity MTS training is not necessary for baseline arbitrary MTS learning, as both subjects have mastered at least two conditional discriminations. As for emergent probe performance, our preliminary conclusion is that subjects require a considerable amount of baseline training with MTS, whether it's identity or arbitrary, before above chance performances are observed. The type of training required (identity versus arbitrary) is not clear, and we will need to conduct several more Equivalence Probes with G13, G8, I4, and I29 before making a specific conclusion. If Equivalence Probe performances for I4 and I29 do not resemble the performances of the Experiment 2 subjects, then it would appear as though identity MTS training was an influential variable. Alternatively, if their performances are comparably accurate, then perhaps the extent of training, and not the type of training, is important. Future investigations about the required training are necessary, perhaps in a similar format to Experiment 3, but with more subjects.

Although occasional inter-observer agreement sessions were conducted and some subjects were exposed to different experimenters, most sessions occurred without these controls. Thus, another potential variable that has been discussed is the effect of experimenter-subject interactions caused by the use of non-automated apparatus. All subjects had to be manually placed into both apparatus at the start of the sessions, as well as within each trial for Arena subjects. The manual placement of stimuli and movements of the MTS tray (Operant MTS Chamber) may have also influenced responding. An optimal arrangement will reduce the need for heavy experimenter involvement within sessions. Second, and also because the apparatus were not automated, all scoring was conducted by human experimenters. In the Arena, the inclusion of lids and the use of the lid displacement response definition were intended to make responses more discrete, which appeared to be the case. No comparable response manipulations were devised for the Chamber, making scoring standardization more difficult; however, inter-observer agreement sessions (trial by trial) were conducted in both apparatus and all indicated that a consistent scoring criterion was in effect. If needed, the recent addition of video recording for each session will allow for inter-observer agreements to be made for all sessions without the added diversion of having a second experimenter present in the room. Finally, the presence of experimenters in the apparatus rooms brings into question the effect of cuing on subject performance (ala Clever Hans). Our best measures of the effect of experimenter cuing were the subjects that were exposed to multiple experimenters because unfortunately, due to the amount of time required for each

session, it was particularly difficult to conduct regular blind experimenter sessions. Still, for all of these cases, no atypical performances were observed. The addition of the video recording also allows for visual inspection of the behavior of both the experimenter and subject during experimentation, and will be useful for future investigations as well.

Generally speaking, the extent of experimentation achieved during Experiment 2 far exceeded our initial assumptions. Of course there were some problems along the way, all of which have been previously mentioned, but some which deserve a second remark. First, it is important to note that most subjects did not consistently demonstrate highly accurate performances; there was a considerable amount of variability from session to session, even within the same training discrimination. To be specific, the performances of our subjects were somewhat fragile and tended to vary independently of the day of the week, laboratory conditions (e.g. temperature, noise), or animal condition (e.g. weight). This variability tended to increase as a function of subject age, and was observed more frequently in older animals. Second, many subjects required 'therapy' intervention procedures (as previously discussed) in order to learn certain discriminations and some were never faded off of these procedures. Future investigations should consider carefully the external influences that require the use of interventions for discrimination learning (e.g. response cost, apparatus design, reinforcement schedules). Third, several of our subjects developed aversions to certain stimulus odors and we either dropped these odors or replaced them with different ones. Some of the typical aversions included ginger and

hickory. Conversely, some subjects developed preferences for certain odors and showed differential responding to those odors regardless of correctness. Typical preference odors included onion and garlic. Lastly, the extent of experimentation that occurred during Experiment 2 resulted in some subjects being fairly old; in fact, Experiment 2 was ended for most subjects not because of mastery, but because of age and resulting performance deterioration. It was thought that Experiment 3 would result in similar amounts of training and testing as Experiment 2, but with comparably younger subjects and hopefully no performance declines. This has yet to be determined, as both I4 and I29 have received less arbitrary MTS experience than most of the Experiment 2 subjects and continue to be trained.

All things considered, through Experiments 1, 2, and 3 our subjects demonstrated that rats are capable of learning and mastering both identity and arbitrary conditional MTS discriminations, often with relative ease. Some of our Experiment 1 subjects have also shown responding indicative of generalization of the identity relation. Also, some of our Experiment 2 subjects have demonstrated that the trained conditional discriminations were also equivalence relations. These performances suggest that subjects were able to transfer relational responding from a training set to a novel set and, in some cases, form equivalence classes (or something like them) through conditional discrimination training. It is our opinion that these performances were made possible by the use of olfactory stimuli, and that comparable performances would be difficult or impossible to obtain with visual stimuli. This lends more credibility to the notion that it is important to test

subjects in an ethologically relevant manner, with an optimal stimulus dimension. Our results also support the use of conditional discrimination training in nonhuman subjects when investigating stimulus equivalence. Although these procedures may require longer training durations (e.g. fewer trials per session) and extensive experimenter commitments than more automated, standardized procedures, the differences in the resulting data speak for themselves.

In terms of ethological relevance, the results of the current study suggest that rats are able to categorize stimuli into functionally equivalent groups after exposure to differential contingencies across the designated classes (1 and 2). Categorical responding, through generalization and abstraction, is thought to allow for more efficient energy and resource allocation (Schusterman & Kastak, 1993). In the case of generalized identity matching, responding transfers to novel stimuli (Vauclair, 2001), which may allow individuals to adjust to a changing environment and maximize reinforcement opportunities (Schusterman & Kastak, 1993, 1998). Adapting in the face of changing conditions would certainly confer an advantage during the individual's lifetime as well as potentially increase the likelihood of survival and reproduction. Additionally, behavior controlled by equivalence relations may be relevant for social and communicative interactions. These situations include the classification and identification of individuals within the social group (e.g. kin, offspring, mates, competitors), the ability to recognize relationships among individuals (e.g. within dominance hierarchies, coalitions, and between mother and offspring), and interpretation of alarm and food calls (Schusterman and Kastak, 1998; Schusterman et al., 2003). The equivalence

paradigm has been used to interpret and evaluate the social and communicative behaviors of other social species such as sea lions (Schusterman et al., 2003) and given that rats are also highly social, it may be a useful model for investigating the variables associated with their social behaviors as well.

REFERENCES

D'Amato, M.R., Salmon, D.P., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons 7(*Columbia livia*). *Journal of the Experimental Analysis of Behavior, 44*, 35-47.

Dube, W.V. & McIlvane, W.J. (1993). The search for stimulus equivalence in nonverbal organisms. *The Psychological Record, 43*, 761-779.

Dube, W.V., McIlvane, W.J., & Green, G. (1992). An analysis of generalized matching-to-sample test procedures. *The Psychological Record, 42*, 17-28.

Dudchenko, P.A., Wood, E.R., & Eichenbaum, H. (2000). Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation. *The Journal of Neuroscience, 20* (8), 2964-2977.

Eichenbaum, H. (1998). Using olfaction to study memory. *Annals of the New York Academy of Sciences, 855*, 657-669.

Frank, A. J. & Wasserman, E. (2005). Associative symmetry in the pigeon after successive matching-to-sample training. *Journal of the Experimental Analysis of Behavior, 84*, 147-165.

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

Iversen, I.H. (1993). Acquisition of matching-to-sample performance in rats using visual stimuli on nose keys. *Journal of the Experimental Analysis of Behavior, 59*, 471-482.

Iversen, I.H. (1997). Matching-to-sample performance in rats: A case of mistaken identity? *Journal of the Experimental Analysis of Behavior, 68*, 27-45.

Kastak, D., & Schusterman, R.J. (1994). Transfer of visual identity matching-to-sample in two California sea lions (*Zalophus californioanus*). *Animal Learning and Behavior, 22*, 427-435.

Kastak, C.R., Schusterman, R.J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76*, 131-158.

Lipkens, R., Kop, P.F.M., & Matthijs, W. (1988). A test for symmetry and transitivity in the conditional discrimination performances of pigeons. *Journal of the Experimental Analysis of Behavior, 49*, 395-409.

Meehan, E.F. (1999). Class-consistent differential reinforcement and stimulus class formation in pigeons. *Journal of the Experimental Analysis of Behavior, 72*, 97-115.

Peña, T., Pitts, R., & Galizio, M. (2006). Identity matching-to-sample with olfactory stimuli in the rat. *Journal of the Experimental Analysis of Behavior, 85*, 203-281.

Schusterman, R.J. & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record, 43*, 823-839.

Schusterman, R.J. & Kastak, D. (1998). Functional equivalence in a California sea lion: Relevance to animal social and communicative interactions. *Animal Behaviour, 55*, 1087-1095.

Schusterman, R.J., Kastak, C.R., & Kastak, D. (2003). Equivalence classification as an approach to social knowledge: From sea lions to simians. In de Waal, F. B. M. & Tyack, P.L. (Eds.), *Animal Social Complexity* (179-206). Cambridge Massachusetts: Harvard University Press.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127-146.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of Rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37*, 23-44.

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

Sidman, M., Wynne, C.K., Maguire, R.W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52*, 261-274.

Slotnick, B. (2001). Animal cognition and the rat olfactory system. *Trends in Cognitive Sciences, 5* (5), 216-222.

Thomas, R. D. (2006). *Olfactory matching-to-sample in rats using a novel apparatus.* Unpublished master's thesis. University of North Carolina Wilmington, Wilmington, North Carolina.

Vauclair, J. (2001). Categorization and conceptual behavior in nonhuman primates. In Bekoff, M., Allen, C., & Burghardt, G.M. (Eds.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition* (239-245). Cambridge, Massachusetts: MIT Press.

Vaughan, W. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 36-42.

Wasserman, E.A. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin, 113* (2), 211-228.

Zentall, T.R., Galizio, M., & Critchfield, T.S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237-248.

APPENDIX

Appendix A. Olfactory Stimulus Information

Spice	Manufacturer	Oil	Manufacturer
Allspice	Great American Spice Co.	Almond	Durkey
Bay	Great American Spice Co.	Apple	Great American Spice Co.
Beet	Great American Spice Co.	Banana	Great American Spice Co.
Caraway	Great American Spice Co.	Brandy	Durkey
Carob	Great American Spice Co.	Bubble Gum	Great American Spice Co.
Celery	Great American Spice Co.	Butter	Durkey
Cinnamon	Great American Spice Co.	Cherry	Durkey
Clove	Great American Spice Co.	Chocolate	Durkey
Coriander	Great American Spice Co.	Grape	Great American Spice Co.
Cumin	Great American Spice Co.	Lemon	Durkey
Dill	Great American Spice Co.	Maple	Durkey
Fennel	Rocky Mountain Spice Co.	Orange	Durkey
Garlic	Great American Spice Co.	Peach	Great American Spice Co.
Ginger	Great American Spice Co.	Peppermint	McCormick
Hickory	Rocky Mountain Spice Co.	Pineapple	Great American Spice Co.
Lime	Rocky Mountain Spice Co.	Root Beer	Durkey
Marjoram	Great American Spice Co.	Rum	Durkey
Mustard	Great American Spice Co.	Strawberry	McCormick
Nutmeg	Great American Spice Co.	Tangerine	Great American Spice Co.
OJ	Rocky Mountain Spice Co.	Vanilla	Durkey
Onion	Great American Spice Co.	Walnut	Durkey
Oregano	Great American Spice Co.		
Paprika	Great American Spice Co.		
Raspberry	Rocky Mountain Spice Co.		
Rosemary	Great American Spice Co.		
Sage	Great American Spice Co.		
Sassafras	Great American Spice Co.		
Savory	Great American Spice Co.		
Spinach	Rocky Mountain Spice Co.		
Sumac	Great American Spice Co.		
Thyme	Great American Spice Co.		
Tomato	Great American Spice Co.		
Turmeric	Great American Spice Co.		
Worchestershire	Rocky Mountain Spice Co.		

Appendix B. Sample Operant MTS Chamber Data Collection Sheet, Experiment 1, BCD training configuration.

MTS SHEET 1 (BCD) **WEIGHT _____**
RAT H7 ___ **SESSION _____**
DATE _____

TRIAL	LEFT	SAMPLE	RIGHT
1	**ON**	ON	THY
2	**THY**	THY	SUM
3	PAP	BAY	**BAY**
4	SAGE	SUM	**SUM**
5	**PAP**	PAP	ON
6	SUM	SAGE	**SAGE**
7	THY	ON	**ON**
8	**SUM**	SUM	PAP
9	BAY	SAGE	**SAGE**
10	ON	THY	**THY**
11	**BAY**	BAY	SAGE
12	**THY**	THY	BAY
13	SUM	PAP	**PAP**
14	**SUM**	SUM	THY
15	**BAY**	BAY	SAGE
16	ON	PAP	**PAP**
17	SAGE	ON	**ON**
18	**SAGE**	SAGE	SUM
19	PAP	SUM	**SUM**
20	BAY	THY	**THY**
21	**ON**	ON	PAP
22	**PAP**	PAP	BAY
23	THY	BAY	**BAY**
24	**SAGE**	SAGE	ON

%_____ **NOVEL_____** **COMBO_____**

Scoring: Underlined stimuli are S+. Stimulus locations are indicated in Right and Left columns. Sample stimuli are indicated in the center column, labeled Sample. Dots to left of trial number used for tracking trial progress within the session. Correct trials indicated by slash through trail number. Quotation marks " represent sniff response only, with no lid displacement. Incorrect trials have no markings across trial number. Incorrect lid displacement response indicated by circled & numbered (superscript 1) stimulus/hole position, located under S- column. Superscript 2, under S+ column, indicates the second & correct lid displacement response to S+, following an incorrect response to S-. Circled trial numbers indicate probe trials, S+ for double baited, S- for unbaited.

Appendix C. Sample Operant MTS Chamber Data Collection Sheet, Experiment 2, A-B training configuration.

CONDDISC SHEET 1 (AB) **WEIGHT_____**
RAT H7___ **SESSION_____**
DATE_____

TRIAL	LEFT	SAMPLE	RIGHT
1	**ONION**	CLO	SAGE
2	**SAGE**	ROSE	ONION
3	SAGE	CLO	**ONION**
4	**ONION**	CLO	SAGE
5	ONION	ROSE	**SAGE**
6	SAGE	CLO	**ONION**
7	**ONION**	CLO	SAGE
8	ONION	ROSE	**SAGE**
9	**SAGE**	ROSE	ONION
10	SAGE	CLO	**ONION**
11	ONION	ROSE	**SAGE**
12	**SAGE**	ROSE	ONION
13	**ONION**	CLO	SAGE
14	SAGE	CLO	**ONION**
15	ONION	ROSE	**SAGE**
16	**ONION**	CLO	SAGE
17	ONION	ROSE	**SAGE**
18	**SAGE**	ROSE	ONION
19	SAGE	CLO	**ONION**
20	**ONION**	CLO	SAGE
21	**SAGE**	ROSE	ONION
22	SAGE	CLO	**ONION**
23	ONION	ROSE	**SAGE**
24	**SAGE**	ROSE	ONION

%_____ **NOVEL_____** **COMBO_____**

Scoring: Underlined stimuli are S+. Stimulus locations are indicated in Right and Left columns. Sample stimuli are indicated in the center column, labeled Sample. Dots to left of trial number used for tracking trial progress within the session. Correct trials indicated by slash through trail number. Quotation marks " represent sniff response only, with no lid displacement. Incorrect trials have no markings across trial number. Incorrect lid displacement response indicated by circled & numbered (superscript 1) stimulus/hole position, located under S- column. Superscript 2, under S+ column, indicates the second & correct lid displacement response to S+, following an incorrect response to S-. Circled trial numbers indicate probe trials, S+ for double baited, S- for unbaited.

Appendix D. Sample Operant MTS Chamber Data Collection Sheet, Experiment 2, ABC Equivalence Probes.

PROBES SHEET 1 (ABC PROBES) **WEIGHT_____**
RAT H7___ **SESSION_____**
DATE_____

TRIAL	LEFT	SAMPLE	RIGHT	
1	**CLO**	ON	ROSE	B1-A1
2	**ROSE**	SAGE	CLO	B2-A2
3	ON	THY	**SAGE**	C2-B2
4	**ON**	SUM	SAGE	C1-B1
5	**THY**	ROSE	SUM	A2-C2
6	THY	CLO	**SUM**	A1-C1
7	**CLO**	SUM	ROSE	C1-A1
8	CLO	THY	**ROSE**	C2-A2
9	SAGE	CLO	**ON**	A1-B1
10	**THY**	SAGE	SUM	B2-C2
11	**SUM**	CLO	THY	A1-C1 (2ND)
12	ON	ROSE	**SAGE**	A2-B2
13	SAGE	SUM	**ON**	C1-B1 (2ND)
14	**SUM**	ON	THY	B1-C1
15	ROSE	SUM	**CLO**	C1-A1 (2ND)
16	SUM	ROSE	**THY**	A2-C2 (2ND)
17	**SAGE**	ROSE	ON	A2-B2
18	ROSE	ON	**CLO**	B1-A1 (2ND)
19	**ON**	CLO	SAGE	A1-B1
20	**ROSE**	THY	CLO	C2-A2 (2ND)
21	CLO	SAGE	**ROSE**	B2-A2 (2ND)
22	THY	ON	**SUM**	B1-C1
23	**SAGE**	THY	ON	C2-B2 (2ND)
24	SUM	SAGE	**THY**	B2-C2

%_____ **NOVEL_____** **COMBO_____**

Scoring: Underlined stimuli are S+. Stimulus locations are indicated in Right and Left columns. Sample stimuli are indicated in the center column, labeled Sample. Dots to left of trial number used for tracking trial progress within the session. Correct trials indicated by slash through trail number. Quotation marks " represent sniff response only, with no lid displacement. Incorrect trials have no markings across trial number. Incorrect lid displacement response indicated by circled & numbered (superscript 1) stimulus/hole position, located under S- column. Superscript 2, under S+ column, indicates the second & correct lid displacement response to S+, following an incorrect response to S-. Circled trial numbers indicate probe trials, S+ for double baited, S- for unbaited.

Appendix E. Sample Odor Arena Data Collection Sheet, Experiment 1, BCD training configuration.

MTS SHEET 1 (BCD) **WEIGHT_____**

RAT F6___ **SESSION_____**

DATE_____

TRIAL	S+	SAMPLE	S-
1	18	CEL	CINN 12
2	12	GING	CLO 9
3	7	OREG	GING 17
4	8	CLO	ON 6
5	6	ON	OREG 1
6	1	CINN	CEL 11
7	13	CEL	GING 18
8	17	CLO	CINN 5
9	5	CINN	OREG 8
10	2	GING	CEL 7
11	9	OREG	CINN 14
12	11	GING	OREG 15
13	3	ON	CEL 13
14	16	CLO	GING 3
15	10	OREG	ON 4
16	4	ON	CLO 2
17	14	CEL	ON 10
18	15	CINN	CLO 16
19	8	CLO	ON 4
20	16	GING	OREG 17
21	14	CEL	GING 5
22	18	ON	CLO 1
23	13	OREG	CINN 15
24	12	CINN	CEL 16

%_____ **NOVEL_____** **COMBO_____**

Scoring: Dots to left of trial number used for tracking trial progress within the session. Correct trials indicated by slash through trail number. Quotation marks " represent sniff response only, with no lid displacement. Incorrect trials have no markings across trial number. Incorrect lid displacement response indicated by circled & numbered (superscript 1) stimulus/hole position, located under S- column. Superscript 2, under S+ column, indicates the second & correct lid displacement response to S+, following an incorrect response to S-. Circled trial numbers indicate probe trials, S+ for double baited, S- for unbaited.

Appendix F. Sample Odor Arena Data Collection Sheet, Experiment 2, A-B training configuration.

CONDDISC SHEET 1 (AB) **WEIGHT_____**
RAT F3___ **SESSION_____**
DATE_____

TRIAL	S+	SAMPLE	S-
1	5 ON	CLO	SAGE 11
2	14 ON	CLO	SAGE 3
3	16 SAGE	ROSE	ON 13
4	2 ON	CLO	SAGE 9
5	10 SAGE	ROSE	ON 16
6	17 SAGE	ROSE	ON 15
7	15 ON	CLO	SAGE 10
8	11 SAGE	ROSE	ON 12
9	6 SAGE	ROSE	ON 8
10	13 ON	CLO	SAGE 5
11	18 SAGE	ROSE	ON 14
12	7 ON	CLO	SAGE 6
13	8 ON	CLO	SAGE 3
14	4 SAGE	ROSE	ON 7
15	12 ON	CLO	SAGE 15
16	17 SAGE	ROSE	ON 4
17	11 SAGE	ROSE	ON 1
18	9 ON	CLO	SAGE 10
19	3 ON	CLO	SAGE 13
20	1 SAGE	ROSE	ON 5
21	5 ON	CLO	SAGE 8
22	7 SAGE	ROSE	ON 16
23	15 SAGE	ROSE	ON 18
24	18 ON	CLO	SAGE 2

%_____ **NOVEL_____** **COMBO_____**

Scoring: Dots to left of trial number used for tracking trial progress within the session. Correct trials indicated by slash through trail number. Quotation marks " represent sniff response only, with no lid displacement. Incorrect trials have no markings across trial number. Incorrect lid displacement response indicated by circled & numbered (superscript 1) stimulus/hole position, located under S- column. Superscript 2, under S+ column, indicates the second & correct lid displacement response to S+, following an incorrect response to S-. Circled trial numbers indicate probe trials, S+ for double baited, S- for unbaited.

Appendix G. Sample Odor Arena Data Collection Sheet, Experiment 2/3 ABC Equivalence Probes.

SYMMETRY SHEET 1 (ABCPROBES) **WEIGHT_____**
RAT G13___ **SESSION_____**
DATE_____

TRIAL	S+	SAMPLE	S-	
1	16 CLO	ON	ROSE 1	B1-A1
2	6 ROSE	SAGE	CLO 11	B2-A2
3	8 SAGE	THY	ON 14	C2-B2
4	5 ON	SUM	SAGE 15	C1-B1
5	15 THY	ROSE	SUM 13	A2-C2
6	7 SUM	CLO	THY 17	A1-C1
7	4 CLO	SUM	ROSE 12	C1-A1
8	9 ROSE	THY	CLO 2	C2-A2
9	1 ON	CLO	SAGE 4	A1-B1
10	2 THY	SAGE	SUM 8	B2-C2
11	15 SUM	CLO	THY 5	A1-C1 (2ND)
12	14 SAGE	ROSE	ON 10	A2-B2
13	11 ON	SUM	SAGE 18	C1-B1 (2ND)
14	16 SUM	ON	THY 15	B1-C1
15	18 CLO	SUM	ROSE 3	C1-A1 (2ND)
16	17 THY	ROSE	SUM 13	A2-C2 (2ND)
17	6 SAGE	ROSE	ON 1	A2-B2
18	3 CLO	ON	ROSE 9	B1-A1 (2ND)
19	8 ON	CLO	SAGE 11	A1-B1
20	10 ROSE	THY	CLO 7	C2-A2 (2ND)
21	12 ROSE	SAGE	CLO 6	B2-A2 (2ND)
22	4 SUM	ON	THY 10	B1-C1
23	13 SAGE	THY	ON 16	C2-B2 (2ND)
24	5 THY	SAGE	SUM 7	B2-C2

%_____ **NOVEL_____** **COMBO_____**

Scoring: Dots to left of trial number used for tracking trial progress within the session. Correct trials indicated by slash through trail number. Quotation marks " represent sniff response only, with no lid displacement. Incorrect trials have no markings across trial number. Incorrect lid displacement response indicated by circled & numbered (superscript 1) stimulus/hole position, located under S- column. Superscript 2, under S+ column, indicates the second & correct lid displacement response to S+, following an incorrect response to S-. Circled trial numbers indicate probe trials, S+ for double baited, S- for unbaited.