A STUDY OF FUNCTIONAL EQUIVALENCE IN RATS USING CLASS-SPECIFIC REINFORCERS AND OLFACTORY STIMULI

Rebecca M. Rayburn-Reeves

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

Department of Psychology

University of North Carolina Wilmington

2007

Approved by

Advisory Committee

Chair

Accepted by

Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

Functional equivalence has been demonstrated in some non-human animals using a repeated reversal simple discrimination procedure. The current study addressed the issue of stimulus equivalence in rats using a repeated reversals procedure with class-specific reinforcers and olfactory stimuli. Four Holtzman Sprague-Dawley rats were tested to establish functional classes using scented sand as olfactory stimuli in a two-choice modified operant chamber and an odor arena. Stimuli were randomly assigned to one of two sets, and at any given point, one set was designated as positive and one was negative. Responses to the positive set were reinforced until criterion levels of performance at 90% or higher for two consecutive sessions were reached, at which point the contingencies were reversed. Reversals continued throughout multiple testing phases as criterion was reached. The results of the study showed that one subject demonstrated clear evidence of functional classes. That is, after encountering a few reversed stimuli on the first day of a given reversal, correct responses were made to the remainder of the stimuli within that class without explicit reinforcement. However, evidence for functional classes was much less clear for the other subjects in the study. The variables that allowed for one subject to provide evidence and not the others remains unclear and further research using rats and olfactory stimuli needs to be done to assess the differences in outcomes seen in the present study.

ACKNOWLEDGMENTS

I would like to thank Dr. Mark Galizio for his dedication and knowledge as well as his enthusiasm and creativity that helped me realize the importance of this work. I would also like to extend my thanks to the members of my committee who took the time to review the thesis and offered their expertise in order to create a good body of research. My thanks also go to my lab members, especially Brooke Poerstel, whose support and knowledge helped me to succeed.

Special thanks go to my family and friends who gave me great insight into the importance of this type of work and the long discussions had over dinner that brought me closer to the final conclusions of the study. I would also like to thank by boyfriend, Nate, whose never-ending patience helped me get through the tough times during these two years.

Finally, thanks to the Department of Psychology and the Graduate School for allowing me to pursue this project and for offering their support and guidance every step of the way.

DEDICATION

I dedicate this thesis to my mother, Linda, and father, Jimmy, whose love and support helped me get where I am today. I can never thank them enough for their selflessness and dedication to their children and I will never forget how much it has meant to me. I would also like to dedicate this to by brother, Danny. His passion for life and never-ending pursuit of knowledge gave me the inspiration to always reach for the stars. Thank you.

LIST OF TABLES

LIST OF FIGURES

Figure Page

INTRODUCTION

The use of symbols is a phenomenon thought by many to distinguish human and animal cognition (Deacon, 1997). As for humans, the most exquisite example of symbol use can be seen in our language system. Deacon (1997) writes that language is “an outward expression of an unusual mode of thought – symbolic representation” (1997, p 22). Many species have shown to be capable of communicating with one another, and although language is a special type of communication often considered to be uniquely human, inquiries to whether non-human animals are capable of symbolic thought have come to take their place in the scientific research of animal cognition.

One approach to the study of symbolic processes was proposed by Sidman and Tailby (1982). They trained eight normal children, ranging from five to seven years old on a set of conditional discriminations using both auditory and visual stimuli. On a conditional discrimination task, a subject is presented with a stimulus, called the sample stimulus, and is then presented with two or more stimuli, called comparison stimuli. The reinforcement of responses to the comparison stimulus is conditional upon the sample stimulus that is given. Sidman and Tailby first trained children using a conditional discrimination procedure called identity matching-to-sample. In this procedure, a sample stimulus is given, such as a blue circle, and then two or more comparison stimuli are simultaneously presented, one that matches the sample (a blue circle) and one that does not (a yellow circle). The subject is reinforced for picking the comparison stimulus that matches the sample stimulus. If the subject responds to the stimulus that does not match the sample stimulus (the yellow circle), no reinforcement is given. Because the correct comparison stimulus is physically identical to the sample stimulus, this particular procedure is called identity matching-to-sample.

Following successful demonstration of identity matching, Sidman and Tailby then trained children on another type of conditional discrimination called arbitrary match-to-sample. In this procedure, the sample stimulus is not physically identical to either comparison stimulus. This arbitrary relation between sample and comparison is thought to be comparable to the arbitrary relation between words and their referents and thus the procedure is often termed symbolic match-to-sample. The subjects are trained through reinforcement to choose a certain stimulus on the condition that a specific stimulus is presented as the sample. For example, a subject could be trained to choose a red square when presented with a blue circle as the sample stimulus, whereas if the subject was given a yellow circle as the sample stimulus, reinforcement would follow responses to a green square. Upon the condition that the yellow circle is given as the sample, the subject would choose the green square as opposed to the red square.

Sidman and Tailby's (1982) experiment trained children on an arbitrary match-to-sample procedure using four sets of stimuli. Set A consisted of Greek names for letters in the Greek alphabet that spoken to the subjects (auditory stimuli). Visual stimuli were arranged in sets of three labeled B, C, & D, with three stimuli (printed Greek letters) per set. For example, Set B stimuli were pictures of uppercase Greek letters {lambda (B1), xi (B2), and gamma (B3)}. Set C stimuli were made up of lowercase Greek letters {lambda (C1), xi (C2) and gamma (C3)}. Set D stimuli were comprised of different lowercase letters of the Greek alphabet {phi, (D1), sigma (D2) and delta (D3)}. Visual stimuli were displayed through a projector onto nine different circular windows arranged in a circle displaying the comparison stimuli, aside from one placed in the middle in which the sample stimulus was displayed.

During a session, a child sitting in front of the visual apparatus could touch the keys with his or her hands. A sample stimulus (visual or auditory) would appear at the beginning of a trial. The visual samples were presented on the center key and auditory stimuli were presented from a tape. In order for the comparison stimuli to be presented, the children were required to respond to the sample key after the sample was presented. Once the sample stimulus was presented it remained present for the duration of the trial.   If the subject responded to the correct comparison stimulus, a chime sounded, followed by the delivery of a penny. Incorrect responses produced no reinforcement.

After training identity matching to criterion levels of performance, an arbitrary match-to-sample procedure was implemented. The subjects were trained to choose B1 (printed uppercase Greek symbol for lambda) if given A1 (auditory name “lambda”), B2 (printed uppercase Greek symbol for xi) if A2 (auditory name “xi”), and B3 (printed uppercase Greek symbol for gamma) if A3 (auditory name “gamma”). The A stimuli were also trained with the C stimuli in the same way (if A1, choose C1). The next step was to mix AB and AC matching trials within a given session. For example, trial one of a session would present Set A samples with Set B comparisons. The next trial would present Set A samples with Set C comparisons. Both AB and AC pairs were presented an equal number of times within a given session. Next, the subjects were trained on DC relations, where both D and C groups represented visual stimuli. Once the DC relations were trained, AB, AC and DC trials were mixed into a given session. The relations that were trained are referred to as the baseline relations. The final stage of baseline training implemented a gradual decrease in reinforcement probability.

The probe testing phase consisted of trial types that had not been previously presented. The purpose of this testing phase was to see whether new relations between stimuli that had never been trained could emerge without explicit reinforcement. Specifically, Sidman was looking to see if three relations, reflexivity (if A= then A), symmetry (if A=B, then B=A), and transitivity (if A=B and B=C, then A=C) could emerge without direct training. The emergence of all three of these relations is a necessary aspect to what Sidman refers to as stimulus equivalence and can be considered an operational definition of an equivalence class in that the related stimuli have come be functionally substitutable for one another. Reflexivity is an identity relation (A=A) where a stimulus is conditionally related to itself. If the subject can successfully match a novel comparison stimulus with the identical sample stimulus, then this response is taken as evidence for reflexivity. The symmetrical relation deals with the concept of reversibility or a bidirectional relationship. Subjects are trained with A1 as the sample and B1 and B2 as the two comparison stimuli. The idea is that if these two stimuli have become equivalent through the connection of the reinforcement contingency, then a symmetrical relation should hold (Sidman, 2000). This is tested by presenting B1 as the sample stimulus with the comparison stimuli of A1 and A2. If the subjects choose A1 when B1 is the sample, a symmetric relation between A1 and B1 has emerged. The last of the three mathematical relations required for stimulus equivalence is transitivity. With this particular relation, a subject will respond to two stimuli that have become related based on training to a third stimulus. For example, a conditional discrimination is trained with A1 as the sample and B1 and B2 as the comparison stimuli. The other conditional discrimination has B1 as the sample and new stimuli, C1 and C2 as the comparison stimuli. Following such training, transitivity would be demonstrated if the subject is able to, upon testing A1 as the sample and C1 and C2 as the comparison stimuli reliably respond to C1. Such performances would show that A1, B1 and C1 have all become members of a specific class, and these equivalence properties have become widely accepted as providing an operational definition of symbolic behavior.

After training was completed, the testing phase was implemented where symmetry and transitivity probe trials were interspersed with the baseline trials and were not reinforced. Results showed close to perfect performance for most children on every stage of the testing phases. For the symmetric relations, the subjects had been taught only three relations and showed the ability to learn six new relations without having to be explicitly taught. For the transitive relations, subjects were explicitly taught nine baseline relations and 27 new relations emerged during testing. This study was followed by dozens of others that consistently demonstrated the untrained emergence of novel relations following arbitrary match-to-sample training in human subjects (see Sidman 2000 for a review). In effect, the stimuli involved in such training become functionally substitutable and as Sidman notes, these are properties that make symbols so important in human verbal behavior. Sidman describes how these equivalence relations (observed even in children and mentally handicapped adults) give humans the ability to use language as a basis for objects in the world in order to deal more effectively with each other and the environment (Sidman, 1994).

Equivalence in Non-humans

Sidman and his colleagues also searched for emergent equivalence relations in non-human subjects. Rhesus monkeys, baboons and human children were tested for stimulus equivalence by Sidman, Rauzin, Lazar, Cunningham, Tailby and Carrigan, (1982). Both the rhesus monkeys and the baboons used the same apparatus; an automated operant chamber consisting of a house light located at the top of the apparatus, a food dispenser near the bottom, two response keys at either side of the food hopper and five keys located in the middle of the stainless steel inner wall. The design of the apparatus was arranged so that Keys 1-4 occupied the four corners of the panel while the sample key was placed in the middle of the square. After training animals to key press for food reinforcement, the subjects were systematically trained on the conditional discrimination procedure. Responses on the trial initiation key (located underneath the five keys and in between the two food hoppers) illuminated the sample, and a response to the sample was required to produce the comparison stimuli. Responses on the correct comparison key were reinforced with a banana pellet and an inter-trial interval was implemented. The initiation requirement ensured the animal was attending to the stimulus at the onset of the trial. The stimuli consisted of red and green key lights and vertical and horizontal white lines presented on a black background.

Animals were first taught identity matching with either color to color baseline training or line to line. For example, for the color to color baseline trials, if a green hue was presented as the sample, and the two comparison stimuli were red and green hues, the subject was reinforced for responding to the green hue comparison because it matched the sample stimulus. For the line to line trials, if a vertical line was given as the sample, with a horizontal and a vertical line as the two comparison stimuli, the subject was reinforced for choosing the vertical line. Two monkeys had difficulty learning the line to line conditional discriminations and were taught to respond differently to each sample (Sidman et al., 1982). For example, vertical line samples had to be responded to five times before the presentation of comparison stimuli while horizontal line samples had to remain on for a minimum of two seconds before being responded to in order for the comparison stimuli to be presented. Once criterion levels of accuracy were achieved, all animals were given sessions with both hue-hue and line-line conditional discrimination trials mixed.

Arbitrary match-to-sample conditional discrimination procedures were then implemented; such as training vertical line samples with green and red comparisons (green being the reinforced comparison) and training horizontal line samples with red and green comparisons (red being the reinforced comparison). These trials were considered the baseline conditional discriminations. After training A1 (vertical) as the sample stimulus with B1 (green hue) and B2 (red hue) as comparison stimuli, a test for symmetry involved presenting B1 as the sample and A1 and A2 as comparison stimuli. A response to A1 would show symmetry and give evidence for class membership. The same could be said for the red hue and the horizontal line stimuli in terms of equivalence. Experiment 1 showed that for all monkeys, the baseline performances were maintained above 90 percent; however, on the symmetry trials performance never rose above chance levels. Experiment 1 showed no evidence of bi-directional or symmetry relations from any of the non-human primates.

Experiment 2 replicated Experiment 1 using the previous symmetry probe trials of line to hue (vertical and horizontal samples with red and green comparison stimuli) and hue to line (red and green samples with vertical and horizontal comparison stimuli), which were now included as baselines (because the subjects had been exposed to them), while two new conditional discriminations were used as tests for symmetry. Sidman et al. noted that the previous trials that mixed line and hue comparison stimuli did not use the same incorrect comparison stimulus given during the baseline training sessions. If the incorrect comparison stimulus is a necessary aspect to the conditional relationship between the sample and correct comparison stimulus, then changing the incorrect comparison stimulus during probe trials could disrupt performance. The argument for using the same comparison stimuli, both correct and incorrect, was given by Sidman et al. (1982), saying, “a valid symmetry test would require sample and correct comparison to be interchangeable only when the original incorrect comparison was also present”. Instead of giving a vertical sample with red (incorrect) and green (correct) hue comparison stimuli as in Experiment 1, a vertical sample was given with green hue (correct) and horizontal line (incorrect) comparison stimuli, which was the incorrect stimulus that the subjects were initially trained on when presented with the vertical sample. The other conditional discrimination that was changed from Experiment 1 to 2 was when given a horizontal line sample and instead of red and green hue comparison stimuli, the subject was given a red hue and a vertical line. Results again showed no evidence of symmetry in any of the monkeys. Transitivity tests were not given to the non-human subjects.

The third experiment tested children on the relations of reflexivity and symmetry using procedures similar to those used with monkeys because previous success with humans used different procedures in testing, making it difficult to know whether the procedure had been the deciding factor in the differential outcomes of human versus non-human subjects. If children could show symmetry using the same procedure as was given to the monkeys then it was not the procedure that could account for the failure of the monkeys to show symmetry.

Six children, five males and one female approximately five years in age, were tested using an apparatus similar to the one the non-human subjects were tested with. The key arrangement for the presentation of the stimuli was the same as was used with the monkeys, except that there were eight keys arranged in a circle around the sample key. Four of the keys were not used in such a way that the sample and comparison stimuli were presented in the same way as for the monkeys. The same stimuli from Experiment 1 were used for Experiment 3, except that the vertical and horizontal lines were black on a white background instead of white on black. There was no trial initiation key. The children were given training on conditional discriminations by observing an experimenter perform the task for several trials and then being allowed to do it themselves for penny reinforcement.

All subjects were trained on six conditional discriminations (hue-hue (2), line-line (2), line-hue (2)) considered the baseline relations. Hue-line symmetry probe trials were then intermixed among the baseline trials and were not reinforced. Each session contained 72 trials in which the six baseline and two probe trials were presented an equal number of times (eight).

Four of the six children were successful on the reversed sample and comparison trials with performances of greater than 80 percent; however, two of the children failed to show symmetrical relations. Because of the success of the majority of the children to show symmetry, Sidman et al. (1982) concluded that the procedure used in training and testing the monkeys could not be accountable for the failure of the monkeys to show the equivalence relation.

The last experiment was a test to see if equivalence might be observed in a different primate species, baboons. The apparatus used was similar to that used with the monkeys; however, the chamber was larger and instead of five keys, there were only three arranged in a horizontal row. The sample stimuli were presented on the center key with the comparison stimuli presented on the outer keys. The stimuli used were the same as were used for Experiment 1 (lines and hues).

Training of the conditional discriminations included hue-hue and line-line relations which were mixed within a given session. Once line-hue trials were added to the baseline training, the subject's performance declined. Because of this, line-hue training was given solely until the subjects reached high levels of accuracy and then incorporated into the baseline. After the subjects were performing at high levels of accuracy for all trial types in the baseline relations, reinforcement probability was reduced in preparation for the symmetry probe tests.

The results of the testing phase gave no indication of the formation of symmetry after training of conditional discriminations. One baboon (Bab-Sim) was given three different tests and his performance never rose above chance. The first test gave no reinforcement when correct responses were given, so the second test reinstated reinforcement probability at 100 percent. After the second failed attempt, experimenters gave another round of baseline training to criterion levels of performance before testing the baboon a third time. The other baboon showed almost perfect performance on baseline trials whereas for the probe trials, he averaged around 49 percent correct. Based on the five experiments, Sidman et al.'s (1982) research suggested the possibility that non-human animals were not capable of forming equivalence relations such as symmetry and transitivity.

Dugdale and Lowe (2000) attempted to study equivalence in three chimpanzees (Sherman, Austin and Lana), who had more than ten years of experience using a lexigram system. The first phase of Dugdale and Lowe's study was identity matching, where the correct comparison stimulus is identical to the sample stimulus. The stimuli used in phase one was also used throughout the experiment. Four shapes (Y, a zig-zag, a triangle and a cross) and two colors (red and green) were used as stimuli; however in the identity matching phase, shapes were broken down into two groups and tested first, followed by color matching. For example, if a triangle was presented as the sample, and a triangle and a cross presented as the comparison stimuli, responding to the triangle comparison stimulus was reinforced while responding to the cross was not. Much of the reason for breaking the stimuli down into three different training groups was to allow each stimulus to be seen as a sample and as a comparison stimulus. This ensures that failure to show symmetry in subsequent tests isn't due to the novel positions of the stimuli in those testing phases (Dugdale & Lowe, 2000). All three chimpanzees met criterion for set one (triangle and cross identity matching) and set two (Y and zig-zag) and maintained criterion levels of accuracy when reinforcement was decreased from 100 to 20 percent. Color matching probe trials were then interspersed among the shape matching baseline trials. However, performance of all chimps remained at chance levels of accuracy on probe trials. Once reinforcement was implemented into the testing phases, performance improved to criterion levels of accuracy.

Phase 2 introduced an arbitrary matching-to-sample procedure, where set A (Y = A1 and zig-zag = A2) samples were paired with set B (green = B1 and red = B2 hues) comparison stimuli. For example, if the subject was given Y (A1) as the sample stimulus, responding to the green hue (B1) comparison produced reinforcement. When the zig-zag (A2) was the sample, responding to the red hue (B2) comparison produced reinforcement. The subjects failed to learn the relations until a fading procedure was implemented. Colors were assigned to each sample stimulus (Y was presented on a green background; zig-zag on red) and was slowly diminished with each correct response. To prepare for BA symmetry testing, reinforcement was again dropped from 100 to 20 percent and the color background for each sample was faded until only the shapes remained. Subjects were trained on this phase until reaching criterion levels of performance.

Phase 3 introduced the unreinforced symmetry probe trials which were mixed with AB baseline trials. For the probe trials, the stimuli that had previously been samples were now comparisons and visa versa. For example, a green stimulus was presented as the sample while the zig-zag and Y stimuli were now comparisons. Lana, one of the chimpanzees, maintained 90 percent accuracy on baseline trials but performed at chance levels when presented with the unreinforced probe trials. Because of Lana's failure to show symmetry on the unreinforced probe trials, phase four reintroduced reinforcement for both baseline and probe trials. In addition to the increase in reinforcement to all trials within the session, identity matching for set A and B baseline trials were added. Despite the changes made, Lana remained at chance levels of accuracy on the symmetry probe trials.

In Phase 5 the simultaneous matching procedure, under which the sample stimulus remains present when the comparison stimuli are given, was changed to a zero-second delay matching procedure, in which the comparison stimuli are presented as the sample stimulus disappears. Lana remained at chance levels of performance on BA probe trials, even though AB and identity matching for both set A and B trials remained above chance. Sherman's performance results were similar to Lana's in that his performance on reinforced probe trials was at chance levels while the baseline trials remained at criterion levels.

Dugdale and Lowe concluded that because these chimpanzees had extensive training with conditional discriminations in a language-like lexigram system and exemplar training and still failed to show symmetry, that it was unlikely that other non-human animals would be able to show symmetry as well. It was thought that if any non-human animals had the ability to form equivalence relations it would be seen in primates who had extensive training using a lexigram system. Their findings were also consistent with Sidman et al.'s (1982) study which failed to show evidence for equivalence in rhesus monkeys and baboons.

The results of Sidman et al.'s (1982) and Dugdale and Lowe's (2000) experiment supported the possibility that stimulus equivalence might be unique to humans. Some researchers also went further to claim that, because language was so closely tied in symbol use and therefore stimulus equivalence, language is a necessary prerequisite for stimulus equivalence (Horne & Lowe, 1996). If research could be done to show that non-human animals could show equivalence relations then it could be said that language is not necessary for the formation of equivalence relations.

D'Amato, Salmon, Loukas and Tomie (1985) argued that the stimuli used in Sidman et al.'s (1982) study with rhesus monkeys, baboons and children might have put the non-human subjects at a disadvantage. The horizontal and vertical lines used in the study might have been easier for the children to discriminate between than for the non-human subjects. D’Amato, et al. (1985) studied six monkeys (*Cebus apella*) on conditional discrimination tasks using different types of visual stimuli that were thought to be more discriminable for non-human primates than those used in Sidman et al.’s (1982) experiment. The apparatus used for the study was similar to that of Sidman et al’s (1982) study. The animals sat facing a wall that had four windows, one at each corner of a wall and one in the center that presented the stimuli. The subjects had to press a trial initiation switch a total of ten times for the presentation of the sample stimulus to appear in the center window. D’Amato et al. used a zero-second delay between the sample and comparison stimuli, which meant that the sample stimulus disappeared as the comparison stimuli were presented. Responses to the correct comparison stimulus produced reinforcement, while incorrect responses initiated a 60s timeout period.

The stimuli used in this experiment (a triangle A1, a red hue B1, a dot C1, a vertical line A2, a plus sign B2, and a circleC2) were randomly paired for each subject so that two monkeys were trained with the triangle and the dot as sample stimuli, while two others were trained with the triangle and the dot as comparison stimuli. For all subjects, the triangle and the red hue were conditionally paired. For example, if the triangle was presented as the sample stimulus, and the red hue and the vertical line appeared as samples, responses to the red hue went reinforced while responses to the vertical line did not. Subjects were trained to criterion levels of performance on the original order of their stimulus pairings. For example, two subjects were trained to choose the vertical line comparison stimulus when given the dot as the sample stimulus and to choose the red hue comparison stimulus when given the triangle as the sample. After reaching criterion performance levels, the correct comparison stimuli during training were presented as sample stimuli, while the sample stimuli became comparison stimuli. Two different types of tests for symmetry were given. The first test was called a BA+ test, where responses to the comparison stimuli that had previously been samples were reinforced. For example, the vertical line stimulus was presented as the sample stimulus and the subjects were reinforced for responding to the dot comparison stimulus. The second symmetry test was called a BA- test in which the comparison and sample stimuli were switched as in the BA+ test; however, responding to the comparison stimuli that had previously been sample stimuli produced no reinforcement. Subjects were reinforced for responding to the comparison stimulus that had not been paired with the sample stimulus during training. Results of the BA+ and BA- tests were compared and if the difference in the number of correct responses between BA+ and BA- was significant (if BA+ had significantly more correct responses than did BA-);, it was taken as evidence for symmetry.

No significant evidence of symmetry was found for any of the subjects on Task 1. Some subjects began to show preferences for certain stimuli and almost exclusively chose those stimuli, so task two rearranged the stimuli for four of the subjects and repeated the procedure from Task 1, but again trends toward symmetry responding were not statistically significant. The lack of evidence supporting symmetry in non-humans might have been due to the “competing sources of control” such as learning specific stimulus configurations rather than relations between sample and comparison stimuli (D’Amato et al., 1985).

Experiment 2 tested for transitivity across two conditional discrimination sets. Tasks 1 and 2 from Experiment 1 were used in baseline training sessions until the subjects reached criterion (a high level of performance indicated by judgment of the researchers) on both Tasks. Two transitivity tests were given in Experiment 2 as in Experiment 1 with symmetry. For example, one subject was trained to pick the red comparison stimulus (B1) when given a triangle as the sample (A1) and to pick plus (B2) when given red (B1) as a sample. The first transitivity test gave A1 (triangle) as the sample stimulus. Responses to the plus (B2) comparison stimulus produced reinforcement while responding to the other comparison stimulus did not (T+). The subject was also tested on another transitivity test (T-) in which responses to the plus produced no reinforcement. The two tests were compared as in the previous experiment. Results indicated that performance was high for the baseline conditions as well as performance on the positive transitivity tests (T+) while performances on the negative transitivity tests (T-) were well below chance levels. This was taken as strong evidence for transitivity for all subjects participating in Experiment 2 in which the differences in performance between the two tasks were significant.

Experiment 3 replicated Experiment 2's training and testing phases using pigeons instead of non-human primates as subjects due to the fact that transitivity had not been shown in animals other than non-human primates. All four pigeons had extensive training on zero-second conditional discriminations. The apparatus used was similar to the non-human primate's apparatus and the same stimuli were used as well. The pigeons were trained on Tasks 1 and 2 of the previous experiments until high levels of performance were achieved. Both Tasks were then alternated between sessions until meeting the same criterion as before to move on to the testing phase. The transitivity tests were very similar to those of Experiment 2, using both T+ and T- sessions. Baseline performances remained high during the testing phases; however, results on the transitivity tests differed from those of the non-human primates in that no evidence of transitivity was found.

D'Amato et al.'s (1985) results on the symmetry tests with cebus monkeys were consistent with previous research that has failed to find evidence for symmetry in non-humans. The successful outcome in Experiment 2 using monkeys tested on the transitivity relation lends support to the idea that the animals were capable of using the nodal stimulus as a connection between the other two stimuli tested during transitivity probe sessions. Failure to find evidence of transitivity in Experiment 3 using pigeons might have been due to inevitable procedural differences between the monkeys and the pigeons. Subsequent studies have further explored the extent to which pigeons might show the emergence of transitive relations between stimuli.

Meehan (1999) suggested that the lack of stimulus class formation in most non-humans might be due to the lack of class-specific reinforcers or differential reinforcement for different stimuli. Using 16 pigeons for his study, Meehan used two differential outcomes, varying in type, amount, location, delay and delivery. After the pigeons were magazine trained on both outcomes and were reliably pecking at all the stimuli shown, they were trained on a symbolic match-to-sample procedure with a zero-second delay. Five sample responses were required to produce the comparison stimuli (FR5). The stimuli consisted of 11 pictures of animals, either alone or in groups and were randomly assigned to one of four classes (A, B, C or D). The pigeons were first given AB training, where responses to B1 comparison stimuli following A1 as the sample produced one outcome (O1) and responses to B2 stimuli following A2 as the sample produced a different outcome (O2). Half of the pigeons were taught the conditional discriminations without the implementation of a correction procedure until reaching criterion performance levels (Procedure 1), while the other half were given only five sessions, but were given the correction procedure (Procedure 2). The pigeons in Procedure 1 were given BC and CB probe tests once meeting criterion levels of performance, while those in Procedure 2 were moved to CB and BC probe testing sessions once they completed five sessions with AB and BA training. For both Procedure 1 and 2, half the pigeons received differential outcomes while the other half received both outcomes randomly. Because the pigeons with class-inconsistent reinforcement from both groups did not develop stimulus control for the AB and BA relations, four of the pigeons were dropped from the study and the other four were retrained with differential outcomes.

Testing of transitive relations (AC and CA) was divided into four sessions. The first session tested 24 trials of the AC and CA relations and was followed by 12 96-trial baseline sessions in which all four (AB, BA, BC and CB) baselines were mixed. Next, between-class and within-class identity relations were tested. For a between-class example, if given A1 as the sample stimulus, the comparison stimuli would be A1 and A2; A1 being the reinforced stimulus. For within-class identity matching, if given A1 as the sample, the comparison stimuli would be A1 and B1; with reinforcement given for responses to A1. Four between-class identity relation sessions were tested first, followed by 12 96-trial baseline sessions. Then four within-class identity relations were tested and followed again by 12 96-trial baseline sessions. Next CD relations were trained with the same differential outcomes as with the previous sets until criterion levels of performance were attained. Following the training of CD, a testing phase of four DC relations was implemented. Six baseline relations followed these sessions, with the added feature of incorporating CD into the baseline. Next, both AC transitive and reflexive relations were retested, followed by another six baseline sessions.

The next testing phase researched DB and BD (transitive) relations which were mixed into a 24 trial session, followed again by the six baseline sessions. After the baseline sessions, came another testing phase, where they tested AD and DA (transitive) relations, followed by another round of baseline sessions. The last thing to be trained was off-baseline training with DO relations. In this phase, D1 was given as the sample with a white comparison key, that when pecked, produced the first outcome (O1). If D2 was given as the sample, pecking the white comparison key produced the second outcome (O2).

Results for the AC relations gave “clear evidence of the emergence of transitive-like relations” (Meehan, 1999, p102), maintaining around 75 mean percent correct. For both within and between-class reflexive relations, the pigeons showed high levels of performance. The results for testing phase consisting of the “symmetric-like, composite (symmetric and transitive-like) and transitive-like relations” showed no evidence of symmetric-like or composite relations and only weak evidence of transitive-like relations shown by a couple of the pigeons. Following the off-baseline training of the D stimulus, stronger evidence for composite relations were found, but no clear evidence of symmetric-like relations were shown. When retesting the AC and the within and between reflexive-like relations, the pigeons overall performance began to weaken, but maintained close to 80 percent correct. Meehan concluded that equivalence-like classes had been produced through the use of class-consistent reinforcement with a conditional discrimination procedure. However, 75 mean percent correct as a criterion level for concluding that equivalence relations emerged may be problematic. For example, if a subject were to correctly respond to C1 when given A1 as the sample 100 percent of the time, but only respond to C2 in the presence of A2 50 percent of the time, the mean percent correct would be 75. Responding at chance levels of performance for one of the two scenarios would fail to show evidence of the emergence of equivalence relations. Given the mixed results found in Meehan's (1999) study, more research needs to be conducted to strengthen the claim that pigeons are capable of showing the emergence of equivalence relations.

Schusterman and Kastak (1993) suggested that, earlier failures to find equivalence in non-humans might have been due to the fact that the subjects were not exposed to multiple exemplar training, where the stimuli from the trained relations (baseline) are incorporated into the testing phases as incorrect comparison stimuli. Their study trained two sea lions (Rio and Rocky) on 30 different sets of visual stimuli with three members in each set (A1, B1 and C1; A2, B2 and C2…A30, B30 and C30) using conditional discriminations in a match-to-sample procedure (Schusterman & Kastak, 1993). The apparatus used included three wooden boards with three window boxes (one on each board) used to hold the visual stimuli. The first eight sets of stimuli (A1-A8, B1-B8, and C1-C8) were three-dimensional objects painted black and presented on a white background, while the rest of the 23 sets were two-dimensional figures (such as an ant, a tulip, a moon, a star, etc) painted black on a white background. The sample stimulus would be presented in the center window by a research assistant. After the sample stimulus was presented, two comparison stimuli would be presented simultaneously by two research assistants. The sample stimulus remained present for the duration of the trial once it was shown. After 2-4 seconds, the subject was allowed to respond to one of the comparison stimuli by poking his/her nose at the stimulus.

The procedure began by training the first two sets of AB relations. For example A1 (a ring) and A2 (a plus sign) were used as samples, with B1 (a baseball bat) and B2 (a square) as comparison stimuli. Responding to B1 in the presence of A1 produced reinforcement while responding to B2 in the presence of A1 produced no reinforcement. Once these two relations were trained to criterion levels of performance, Rio was given the next six sets (A3-A8, and B3-B8) one at a time (A3 & B3, then A4 & B4…etc), using stimuli from the previous sets as incorrect comparisons. For example, Rio was given A3 as the sample with B3 as one comparison stimulus and either B1 or B2 as the incorrect comparison. Because Rio had already been trained to relate B1 with A1 and B2 with A2, she could use the principle of exclusion to rule out those stimuli as the correct comparisons. After reaching criterion on the first eight sets, Rio was trained on the rest of the 22 sets (A9-A30 and B9-B30), two sets at a time, with the incorrect comparisons being the stimuli that she had previously been trained on and were now incorporated into the baseline. After reaching criterion levels of performance on all 30 sets of stimuli, the next step was to randomly remove 12 sets from A and B to be used for symmetry testing (BA relations). The testing phase was divided into two parts, with six stimulus pairs for each part. After completing the BA symmetry testing phase, Rio was trained on BC relations, trained with two pairs at a time. Once each of the two pairs was learned, they were incorporated into the baseline to be used as comparison stimuli while the next two were trained. Once again, 12 sets of the BC relations were randomly removed to be used in testing for symmetry (CB), transitivity (AC) and equivalence (CA) with stimuli from the baseline used as incorrect comparisons. For example, when testing for a transitive (AC) relation, if given A1 as the sample, the subject would be given C1 as a comparison stimulus along with any other stimulus from the C bracket (C2-C30). The transitive relation would be considered to have occurred if the subject could reliably respond to C1 when given A1 as the sample. Before testing for the equivalence relations, BC training was given again using each of the six sets used for both testing phases. For example, six of the 12 relations were trained to criterion again, and then used in testing CB relations, followed by another training phase for the other six BC relations, which were then used for a second test of CB relations. After completing the testing phases, the relations from the testing phases were again incorporated into the baseline. The same procedure was used in testing AC transitivity relations (two testing phases made up of six sets each) to criterion levels of performance. After that, CA equivalence relations were tested using the same sets as the AC training sets. Once the 12 sets were tested for CA equivalence, the remaining 18 sets were tested on CA equivalence (but not to criterion levels of performance), broken down into three groups of six sets.

The results of the BA symmetry testing phases were that Rio responded correctly to three out of six of the relations on the first exposure to each relation. On the second test, Rio passed five out of the six relations (not significantly better than chance). For the CB symmetry tests, Rio scored five out of six on both testing phases, reaching statistically significant levels of performance. On the AC transitivity tests, Rio made only one error between the two testing phases, showing evidence for transitivity. Finally, for the CA equivalence relation tests, Rio made four out of six correct responses on the first test and no incorrect responses on the second, indicating sufficient levels of performance indicative of equivalence. When Rio was first exposed to the symmetry testing probe trials (CB), his performance was not significantly above chance; however, he was still trained to criterion levels of performance on the relation. Because so many stimuli were used in the experiment, Rio was exposed to many different symmetrical relations. Although Rio initially did not show evidence of symmetry (based on the first exposure to stimuli in novel configurations), his performance improved with each new conditional discrimination indicating the ability to learn the abstract relation of symmetry. This type of training is called multiple exemplar training because multiple examples of symmetry are exposed to the subject, first as a probe and then trained through reinforcement to criterion levels of performance, to help demonstrate the idea of the symmetrical relation. Due to the high levels of performance from Rio on the symmetry, transitivity and equivalence relations after multiple exposures of the equivalence relations, Schusterman and Kastak concluded that the use of multiple exemplar training is one way to help facilitate the acquisition of equivalence relations in non-human animals.

Functional Equivalence

Although Schusterman and Kastak (1993) showed evidence for stimulus equivalence using the conditional discrimination procedure, other studies using this procedure have either been able to provide evidence for transitivity or reflexivity and not symmetry or failed to show evidence for any of the three relations required to claim equivalence. Because of the limited success of demonstrating stimulus equivalence in non-human animals, some researchers have argued that Sidman's definition of stimulus equivalence (in terms of conditional discrimination training and the emergence of reflexivity, symmetry and transitivity) is too stringent a test. Vaughan (1988) attempted to study a form of equivalence, termed functional equivalence, through simple discrimination training. Dube & McIlvane (1993) defined functional stimulus classes requiring that, "all class members share a common stimulus function and… variables applied to one class member may affect the others without explicit conditioning" (pg. 764). With functional equivalence, arbitrary stimuli come to be reacted to similarly on the basis of a common behavioral function (Jitsumori et al., 2002; Kastak & Schusterman, 2001; Sidman, 1989; Vaughan, 1988). Functional equivalence has also been called functional reassignment on the basis that when a new function is conditioned to some members of the class, the new function will transfer to the rest without explicit training (Vaughan, 1988).

Functional equivalence can be shown through a simple discrimination procedure involving a series of repeated reversals which offers another way to test for equivalence. In this procedure, stimuli are arranged into two separate Sets (Set 1 and Set 2). At the onset of the study, one class is designated as the positive or reinforced class and the other is designated as negative or non-reinforced. Each trial consists of presenting two stimuli, one from each set, simultaneously. Subjects are trained to respond to the members from Set 1 and not to respond to members from Set 2 until reaching criterion levels of performance. Once subjects reach criterion, the contingencies are reversed making Set 2 positive and Set 1 negative. Each time criterion is reached the contingencies are reversed. The purpose of the experiment is to see whether, after a series of reversals, when some members of a set are reversed, subjects are able to show a transfer of function to the rest of the members of that set without explicit reinforcement.

The most famous success story for the ability to establish functional equivalence was with pigeons using simple discriminations and the repeated reversals procedure (Vaughan, 1988). Vaughan believed that "behaviorally and mathematically valid equivalence relations could arise in an alternative procedure without demonstrating the presence of identity, symmetry and transitivity" (Vaughan, 1988, p. 42). Vaughan trained pigeons with 40 slides of pictures containing trees of various types. Each slide was randomly assigned to one of two groups. Three pigeons were used in this study. For each pigeon, the slides were reassigned to one of the two groups as a control measure to make sure that there wasn't some hidden association among the particular members of a set. Once the two groups were determined, a simple discrimination procedure was implemented. Each session consisted of 40 trials, where two stimuli, one from each class, were presented simultaneously. One of the sets was designated as positive (S+; responses to these stimuli were always reinforced) while responses to the others were not reinforced (S-). After discrimination training sessions were completed, pigeons began to reliably respond to stimuli in the reinforced set and not to the negative or non-reinforced set. Once pigeons reached a specific criterion of responding to the positive set, the contingencies were reversed. Now responses to stimuli in the previously negative set were reinforced and vise versa. The pigeons, at first, responded to stimuli in the previously positive class because they had been reinforced for doing so in the past; however, eventually the responding shifted to the stimuli in the now positive set. Once the pigeons reached criterion levels of performance, the contingencies were again reversed. Repeated reversals continued each time criterion to the designated positive set was reached. The measure used to calculate performance was rho (the probability of ranking a positive over a negative) which remained below .5 or 50 percent for the first 40 slides (or 20 trials). After a series of repeated reversals the performance rose to above chance levels, suggesting that transfer of function had occurred between members of a particular set. On the first day of a new reversal, the pigeons would respond to the class that had been most recently reinforced. When no reinforcement was given for responses to the first few stimuli, the pigeons began responding to stimuli in the other set of that class, even though the rest of the 20 positive slides had not been explicitly reinforced. The procedure established stimulus classes through control of a common response between members of the class. This suggested that the pigeons had formed functional classes based on a reinforcement contingency. The stimuli within a particular class had become equivalent and substitutable for one another in the sense that after some members of a class were trained on a new function, the other class members would follow (Vaughan, 1988).

Delius, Jitsumori & Siemann (1995) conducted a systematic replication of Vaughn using only four stimuli. The stimuli were grouped into two pairs in which one stimulus from each pair was positive (reinforced) while the other stimulus was negative (non-reinforced). For example, stimulus A (a red hue designated as A1) and stimulus K (a green hue designated as A2) were paired with A being the reinforced stimulus and K being non-reinforced. The other pair consisted of B (a blue hue designated as B1) and L (a yellow hue designated as

B2). The stimuli were presented as pairs (A/K & B/L) simultaneously from a projector onto two windows in an operant chamber. During the training phase, the pigeons were reinforced for choosing A (A1) and B (B1) stimuli until they met criterion performance levels, at which point, the contingencies were reversed. Responding to the stimuli that were previously positive (A and B) now produced no reinforcement while responding to the stimuli that were previously negative (K and L; A2 and B2 respectively) were reinforced. After learning the new task (reliably choosing K and L), the contingencies were again reversed. Thirty five reversals were completed during the training phase.

During the testing phase, only one pair (A and K) was presented for the first few trials of a session until the subjects were responding to the reinforced stimulus (A) with at least 85 percent accuracy. Once this was achieved, the second pair was introduced into the session. If the subjects responded to B (most recently the incorrect stimulus) instead of K, it was taken as a demonstration of functional reassignment or a transfer of function in which A and B had been placed into one class and K and L into another. This procedure was repeated a number of times.

After the first testing phase, subjects were given reversal sessions in which stimuli A and B were shown through one window (half red and half blue) while K and L (green and yellow) were shown in the other window. The reasoning for this phase of the experiment was to pit the members of each class in close proximity to each other in the hopes of strengthening the association between the members of the classes. After completing a number of reversals, another testing phase was implemented using the same procedure as before (showing one pair for the first few trials and then introducing the second pair).

Results of the testing phases gave evidence for functional equivalence in that, once the pigeons had learned the new reversal with the first pair (A+ and K-), they were reliably able to respond correctly to the second pair (B+ and K-) without having to be explicitly taught the reversal for that pair.

In contrast to Dugdale and Lowe (2000) and Sidman et al.'s (1982) failure to show stimulus equivalence with non-human primates testing conditional discriminations in a match-to-sample procedure, Tomonaga (1999) was able to successfully demonstrate functional equivalence in chimpanzees using the Vaughn procedure. One subject (Chloe) was used for the experiment and tested on two two-item groups of visual stimuli. For example, the first group consisted of two stimuli, a green and red hue (green = A1; red = A2), while the two stimuli in the other group consisted of a picture of a star and a picture of a snake (star = B1; snake = B2). The training phase consisted of Chloe being trained to respond to the stimuli in each group sequentially (first star (B1), then snake (B2) & first green (A1), then red (A2)). Responding to the incorrect stimulus (snake or red) first produced no reinforcement and a time out procedure was implemented. Responding to the correct stimulus (star or green) first produced an auditory 'click' sound while both stimuli remained on the screen. If she then chose the other stimulus from that pair after choosing the correct stimulus, a chime would sound and would be followed by reinforcement.

The first phase of Experiment 1 trained Chloe on the sequential responding procedure. For example, trial 1 presented the star and the snake and Chloe was reinforced for responding to the star first, followed by a response to the snake. Each session consisted of 96 trials with each set mixed randomly within the session. Once Chloe reached a criterion of 90 percent or higher on a given session, a preliminary test for functional class formation was given. The test consisted of a single session with 104 trials in which eight probe trials were intermixed within the 96 baseline trials from the training session. For the probe trials, the two sets were rearranged so that the incorrect stimulus from each set was switched to the other set. For example, the previous pair, star-snake, became star-red, while green-red became green-snake. The purpose of this cross-over test was to see if Chloe had formed a class between the members of each set. If both star and green had formed a class, then she should continue to choose both of these stimuli first. Chloe was not reinforced for the probe trials but responded correctly if she chose star-red and green-snake. Chloe got 100 percent correct on the baseline pair (star-red), which she had been trained on; however, her performance on the probe trials (green-snake) remained at chance levels. After the cross-over test, Chloe was given another 96 trial baseline session.

Chloe was then tested on between session reversal blocks in which each block consisted of three phases; pretest, test, and posttest. The pretest phase trained a reversal of one of the two sets to criterion levels of performance (red-green). The test session consisted of 114 trials in which 12 of the trials were probe trials where the second set (snake-star) was reversed and presented. The twelve probe trials were intermixed within the baseline of the first reversed pair (red-green), with the exception that the first 16 trials consisted only of the trained reversed pair (warm-up trails). The posttest implemented differential reinforcement for each set, depending on whether the order of the pairs was reversed (red-green) or original (green-red). Once all three phases were completed, the contingencies were reversed again (back to the original order) and another block began.

During the warm-up trials for each test session, Chloe’s performance remained accurate (an average of 96 percent on first 16 trials of each block), however, she was inconsistent with performance on the remaining trials. The only pair in which Chloe showed above chance performance during the test phase was when the order was switched from reversed to original and only for the shape pair (star-snake). The only improvement over the blocks was for the color pair when the contingencies changed from reversed to original (green-red). She showed below chance performance when the shape pair was switched from original to reversed (snake-star) and only slightly above chance for the color pair when going from original to reversed (red-green), however, performance was not significantly above chance. Chloe was overall more accurate when the contingencies switched from reversed to original than when they were switched from the original to the reversed order. Although she was fairly accurate at switching the shape pair from reversed to original, her chance performance on the color trials suggested a lack of class formation between the stimuli from each set (star and green & snake and red).

Experiment 2 implemented a within session reversal procedure. Each session consisted of 96 trials broken down into three blocks of 32 trials. The first 32 trials presented the pairs in the original order in which Chloe had been trained (green-red (A1-A2 & star-snake (B1-B2)). The second 32 trials presented the reversed order for the pairs and the last block returned the pairs to the original order. Trials one and two of each block presented one baseline pair (green-red) called a pretest reversal. Trial three was a probe trial presenting the other pair (star-snake), and the rest of the trials for each block presented both pairs randomly and an equal amount of times. If the first session began with the original ordered pair, then the next session began with the reversed order. A total of 64 sessions were implemented. Chloe's performance was below chance on the first two session blocks but gradually improved to an average above 95 percent correct on the probe trials for the last two sessions. This experiment suggested that differential responding and within session reversals help to facilitate functional class formation. Once the contingency for one pair was changed, responding to the other pair also changed without being explicitly reinforced.

Schusterman and Kastak (1998) tested a California sea lion (Rio) for evidence of functional equivalence using a simple discrimination procedure. Rio was previously successful in establishing 30 three-member sets of stimuli using a match-to-sample procedure. In the current study, the 30 sets were divided into fifteen pairs of two three-member sets (1 & 2, 3 & 15, etc) and two of the 30 sets were chosen for the first simple discrimination procedure (A28, B28, C28, A29, B29 and C29). The subject was given two stimuli simultaneously; one from each set (B28 & B29), and allowed to choose between the two. Which ever stimulus Rio chose was reinforced and that set was labeled the positive set. For example, if she chose B28, all members (A28, B28 and C28) were reinforced for being responded to. Rio was then trained to criterion levels of performance (90 percent in a block of 10 trials) for the first simple discrimination (B28 & B29). After reaching criterion, Rio was tested on one of the remaining two pairs (A28 & A29) to see if the stimuli had become members of the same class. After meeting criterion on the second pair of stimuli, Rio was tested on the third pair (C28 & C29). Following completion of the second transfer test, the subject was given the same procedure for the remaining 14 pairings. For each transfer test, trial one was a novel pairing and was considered the "critical measure of the equivalence response" (Schusterman & Kastak, 1998, p. 1091).

A control condition was implemented following testing of equivalence relations using novel stimuli. Stimuli were randomly assigned to one of two three member classes and two stimuli, one from each set (X1 and X2), were chosen for the first simple discrimination. The stimulus Rio responded to was reinforced and the members from that set were designated as positive (X1, Y1 and Z1). The control tests were given under the same conditions as the previous tests (experimental condition); upon reaching criterion levels of performance. If Rio scored above chance on the control tests with novel stimuli that had not been trained as equivalent, it would indicate that something other than the formation of equivalence classes had occurred with the previous stimuli (inadvertent cueing, etc).

After a total of 10 transfer tests, Rio's performance on the control condition had not improved from chance. Rio responded correctly to only three out of 10 transfer tests for each of the transfer tests for the control condition, whereas for the experimental condition, Rio made 28 out of 30 correct responses on the transfer tests. For example, on the first transfer test for the experimental condition in which the previous pair (B28 and B29) was given and trained to criterion levels of performance (Rio was reliably choosing B28), the second pair (A28 and A29) was presented and Rio responded to A28. The difference between the testing and control conditions differed significantly. Schusterman and Kastak concluded that Rio had successfully demonstrated equivalence class formation using the simple discrimination procedure.

Initially, Sidman did not agree that these repeated reversals experiments provided an adequate demonstration of equivalence relations because of the failure to test for reflexive, symmetric and transitive relations in this procedure. Hayes (1989) argued that the critical aspect for demonstrating equivalence relations is the emergence of performances that cannot be accounted for by a history of direct reinforcement. The argument against functional equivalence as a valid demonstration of the equivalence relation is that performance based on contingency class arrangements is directly trained through the repeated reversals procedure. One way to assess whether classes formed via repeated reversals should be considered evidence for the equivalence relation is to also test for stimulus equivalence using conditional discriminations to observe the emergence of reflexivity, symmetry and transitivity. If subjects can demonstrate functional equivalence but fail the tests for stimulus equivalence, it would suggest that the two forms of class membership are fundamentally different (Dube et al. 1993). However, if acquisition of functional equivalence can transfer to a successful demonstration of stimulus equivalence, it would provide evidence that the two processes may share a similar behavioral prerequisite (Dube et al. 1993). Subsequent research has suggested that the two phenomena are, at least, very closely related.

Kastak, Schusterman and Kastak (2001) replicated the Vaughan procedure with sea lions with the added step of testing for reflexivity, symmetry and transitivity. The same two female California sea lions (Rio and Rocky) discussed above were studied in a follow-up to the stimulus equivalence studies reviewed earlier. A two-choice apparatus was used with visual stimuli that consisted of plywood boards with black painted symbols on the front of the boards painted on white backgrounds. Two sets were arbitrarily assigned using 20 different symbol stimuli. Set 1 used coded letters of the alphabet (A-J). The ten stimuli used to designate Set 2 were coded as modified numbers ranging from one to ten (the ten was a zero with a line through the middle as to create the least amount of similarity between the stimuli in order to reduce confusion).

At the beginning of the study, the letter set was designated as positive while the number set was designated as negative. Forty trials were presented per session, with ten individual trials presented four times each. The specific stimuli were therefore presented a total of four times per session and were counterbalanced for left and right positions. A typical trial would consist of the simultaneous presentation of a stimulus from each set. For example, A1 (the letter A) and A2 (the number 1) would be presented. Responses to the positive set (the letter set) produced a 0.5-s acoustic tone serving as a conditioned reinforcer, followed by a reinforcement of either a capelin or a herring fish. A vocal response of “no” followed a response to the S- stimulus. Once the subjects’ performance rose to above 90 percent, the contingencies were reversed and the stimuli from the previously positive set became negative while the negative set became positive. This phase continued until the same criterion was met for responding to the positive stimuli and then the contingencies were again reversed. Several such reversals were continued throughout the experiment.

The experiment was divided into seven phases, the first phase being a training phase to get both subjects familiar with the procedure and the rest of the phases designated as testing phases. Each testing phase included a certain number of reversal sessions, but differed with respect to certain aspects of the procedure. The first reversal phase presented stimuli in matched pairs. For example, a trial might consist of the letter A and the number 1or the letter B and the number 2 and so on for all ten stimuli in each set. Because the first testing phase began with the first reversal, the letter set was positive when testing began. Responses to the correct stimulus were followed by a 440-Hz tone and fish reinforcement described above. Once criterion of 90 percent or better for two consecutive sessions was met, the contingencies were again reversed. After a series of repeated reversals, performance improved only slightly and Phase 2 was implemented.

During Phase 2 the stimulus pairs were scrambled so that each member of one set had an equal probability of being paired with each member of the other set on a trial. For example, the letter ‘A’ could be paired with the number ‘5’ or any other number. This raised the amount of possible pairs from ten to 100 and was implemented as an attempt to facilitate functional class formation through multiple exemplar training. Scrambled pairs remained in effect throughout the experiment once implemented. This change improved performance slightly, but at a non-significant level.

The third phase of the repeated reversals procedure used different reinforcement for correct responses to each of the classes as well as different tones for each conditioned reinforcer. A number of studies have shown the benefits of using class-specific reinforcers in discrimination learning (Dube, et al., 1993; Meehan, 1999; Sidman, 2001; Tomonaga, 1999). Rio received a 587-Hz tone followed by capelin reinforcement for responses to the letter set, while Rocky received a 293-Hz tone followed by a herring when she correctly responded to S+. Incorrect responses were treated the same as during previous phases. Rio needed a total of nine reversals to reach criterion of 90 percent correct or better for two consecutive sessions and Rocky had 15 reversals and only reached criterion of 90 percent for one day; however, phase four was implemented because of the significant improvement in performance from both subjects.

The next phase removed the class-specific reinforcers. Introduction of the class-specific reinforcers was proposed as another way for the arbitrary stimuli to become associated together. When the class-specific reinforcers are removed, it is assumed that the specific reinforcement has to drop out of each set as a member of that particular set. Reinforcement reverted back to that of phase two, where correct responses were reinforced with a 440-Hz tone and either of the two fish reinforcements. Performance dropped to chance levels when the class-specific reinforcers were removed. After a few reversals and the decline in performance, the class-specific reinforcers were reinstated. Once again, if set A (the letter set) was designated as positive, responses to set A produced a specific acoustic tone followed by a specific fish reinforcement.

The last phase of the first experiment introduced reversals within a session instead of between sessions as was the case for the previous phases. In this phase one or more reversals could occur within a given session. Criterion of 90 percent or higher for the previous ten to 14 trials, however, had to be met before a reversal could be implemented within the session. With the help of class-specific reinforcers, both subjects demonstrated functional equivalence during the experiment. This involved responding to stimuli that were negative the last time they were presented and did not evoke responses. Kastak et al. (2001) found evidence of class formation because after the first few trials of a reversal, where stimuli from the negative class were now positive, the rest of the stimuli from that class were responded to without having to be directly reinforced for each member of that class explicitly. The shift in responding can be considered a type of functional reassignment in that the reassignment of the stimuli from being positive to negative (and vise versa) produces a different response from the subject.

Once the functional classes were established, Experiment 2 tested whether the functional classes could transfer to a conditional discrimination procedure and show the three relations defining stimulus equivalence. Subjects were first taught baseline conditional discriminations. Rio and Rocky were trained using 12 novel stimuli which were divided into four sets with three members per set. Each set was labeled one through four and each member in each set was lettered ‘A’ through ‘C’ making Set 1consist of A1, B1 and C1, Set 2 have A2, B2 and C2, Set 3, A3, B3 and C3 and Set 4 consisted of A4, B4 and C4. With the MTS training, one member from a particular set (A1) would be used as the sample (placed in between the two squares used in experiment one) while two comparison stimuli, one from that particular set (B1) and the other from another (C2), were placed on either side of the sample. Responses to the comparison stimulus that belonged to the set in which the sample stimulus belonged (B1) was reinforced with a specific fish. Each set had either a herring or a capelin given to all members of that particular set. For example, set 1 was always reinforced with a herring while set two was reinforced with a capelin. The conditional discriminations were trained until performance reached criterion levels of 90 percent or higher. Along with the MTS training of novel stimuli, reversal sessions were continued with the two classes that had formed from experiment one in phase six of testing.

For the testing phase, two members from each of the two classes from experiment one were randomly selected as testing stimuli (letters ‘E’ and ‘I’ and numbers 4 and 8). There were 40 trials per session in which 28 of the 40 trials were the baseline trials trained at the beginning of the experiment while 12 trials were intermixed using the stimuli from Experiment 1. During the MTS probe trials (the 12 trials mixed within the session) both letters chosen were used, one as the sample and the other as one of the comparisons, while a number (not 4 or 8) from the other class was chosen as the incorrect comparison. After two sessions, baseline trials were decreased from 28 to 16 and the testing trials increased from 12 to 24. Subjects were tested until reaching criterion of 90 percent on test trials for one session, which were then used as training stimuli because they were no longer novel. The second transfer test used two more letters and numbers chosen from experiment one to test conditional discriminations. Transfer test three replicated one and two with two more stimuli chosen from both classes. Test four used all stimuli from tests one, two and three until meeting criterion and test five used the remaining stimuli from experiment one that hadn’t yet been used in the second experiment. The final testing phase used new combinations from the stimuli from test five with the stimuli from tests one through four generating 96 novel stimulus pairings. Results indicated that performances for all six transfer tests showed significant transfer from functional class formation to conditional discriminations. The results suggest that it is possible for non-human animals to show emergent equivalence relations after conditional discriminations and that functional class formation could serve as a sufficient foundation for the demonstration of equivalence relations defined by reflexivity, symmetry and transitivity.

Although evidence for functional equivalence has been shown in several species (primates, sea lions & pigeons), little research using rats as subjects has been done in this area. Rats are an integral part of research geared towards biomedical advancements in the fields of neuroscience and psychopharmacology. A demonstration of equivalence in rats or mice could give neuroscientists an animal model to look at certain parts of the brain that allow for equivalence to emerge. To date, only a few studies have examined equivalence using rats as subjects.

For example, Bunsey and Eichenbaum (1996) developed a way to test rats on equivalence relations using olfactory stimuli and a series of conditional discriminations using an arbitrary match-to-sample procedure. Using ground rat chow and ordinary play sand, 20 rats were initially trained to dig in the mixture to obtain pieces of cereal used as reinforcement. Once the rats were reliably digging in the sand, they were trained on conditional discriminations. The stimuli included six different ordinary household spices (cocoa, turmeric, coffee, salt, onion & nutmeg) which were mixed into the mixture of sand and rat chow and were arranged into two groups of three (A1, B1, C1 and A2, B2 and C2).

Training of the conditional discriminations was broken up into two different phases. The first phase trained AB relations. A1 was presented as the sample in the rat's home cage and the rat was allowed to dig in the sand to obtain reinforcement. Once the rat obtained the reinforcement, the two comparison stimuli were presented, neither of which were physically identical to the sample. Subjects were either presented with A1 or A2 as the sample stimulus, while B1 and B2 were presented as comparison stimuli. If the subject was presented with A1 as the sample stimulus, they were reinforced for responding to B1 and not B2 and responding to B2 in the presence of A2 produced reinforcement. Sessions were comprised of 14 trials. Once the rats attained a criterion of responding to 11 out of 14 trials correctly for three out of four consecutive days, they were moved to the second training phase.

Phase 2 implemented BC training, with B1 and B2 used as samples and C1 and C2 used as comparison stimuli. Subjects were reinforced for responding to C1 in the presence of B1 and C2 in the presence of B2. The same criterion from Phase 1 of training was used for Phase 2. However, in preparation for the testing phases in which probe trials were not reinforced, reinforcement probability was reduced from 100 percent to 30 percent. Upon reaching criterion levels of performance, the testing phases began.

The first testing phase tested for transitivity. A1 and A2 were again used as sample stimuli, while C1 and C2 were used as comparison stimuli. Sessions were composed of 14 trials in which the baseline relations (AB and BC) were presented on 10 of the trials, while the other 4 trials were probe trials and were intermixed within the baseline trials and were not reinforced. If subjects reliably responded to C1 in the presence of A1 and C2 in the presence of A2, it was taken as sufficient evidence for transitivity. Results revealed a preference for this outcome by a ratio of 2:1, which was considered significant; however, a percentage of 67%, although above chance, does not give clear evidence of the transitive relation due to the chance that one of the two conditional discriminations (A1:C1) was learned, while the other conditional discrimination (A2:C2) remained at chance performance.

Before beginning the symmetry testing phase, subjects were retrained on the AB relations with the stimuli which were previously sample stimuli (A1 and A2) used as comparison stimuli, while B1 and B2 were presented as sample stimuli. All subjects required some training on this phase before reaching criterion. Once criterion levels of performance were achieved, subjects were tested on CB relations where C1 and C2 were presented as sample stimuli, while B1 and B2 were presented as comparison stimuli. This phase consisted of a single session composed of 14 trials; seven of which presented C1 as the sample and the other seven presented C2 as the sample. The results of the symmetry testing phase indicated a strong preference for B1 when given C1 as the sample and B2 when given C2 as the sample by a ratio of 3:1 (75%). Given the results, Bunsey and Eichenbaum (1996) concluded that rats were capable of showing equivalence relations when using olfactory stimuli. However, one of the main differences between this study and previous studies on equivalence was the dependent variable used. Traditionally, tests for equivalence concentrate on whether or not the subject responds to the S+ or S- stimulus and data is based on accuracy of choice (either the subject responds correctly or incorrectly). Bunsey and Eichenbaum used time as the dependent variable such that the response to either stimulus was irrelevant to the amount of time spent at each stimulus before a response was made. In effect, subjects could potentially respond incorrectly to all probe trials but spend more time around the correct stimulus and the data would show a preference for the correct stimulus. Because of this difference in the dependent variable, it remains unclear whether the subjects actually responded to the correct stimulus or just spent more time around it before responding to the incorrect stimulus. Although using time as the dependent variable makes interpretation of the Bunsey and Eichenbaum paper problematic, their study provided some support for the notion that rats might be capable of equivalence with olfactory stimuli.

Dube, et al. (1993) studied functional equivalence in five albino rats using auditory stimuli and successive discrimination procedures. Concurrent or successive go/no-go discriminations were selected for the study instead of simultaneous discriminations which were typically used to test for functional equivalence because simultaneous discrimination procedures can produce responses controlled by specific stimulus control topographies and not by the common function between stimuli within a set. For example, subjects could be responding based on selecting the S+ stimulus or rejecting the S- stimulus. Successive discriminations remove the possibility of this type of control by stimulus response topographies.

Six stimulus tones were presented for Experiment 1 using different frequencies to designate specific stimuli and these stimuli were separated into two, three-member stimulus sets at the beginning of the study. Frequencies of 550, 1730, and 5980 were designated as Set

1 (A1, B1 and C1, respectively) and were designated as positive at the onset of the study, while frequencies of 970, 3180, and 11100 designated Set 2 (A2, B2 and C2, respectively) and designated as negative. Only three stimulus sets were chosen for this study in order to assess whether smaller stimulus sets (compared to 20 stimulus sets in Vaughan's 1989 study) could improve acquisition of successive reversals where the rate of learning was rapid enough to generate reliable tests for contingency classes (Dube et al. 1993). Class specific reinforcement was also given depending on which set was designated as positive for each reversal. Stimuli in Set 1 were reinforced with sucrose/water solution while stimuli in Set 2 were reinforced with a 45mg sucrose pellet. Class specific reinforcement has been argued to help distinguish stimuli in one set with those in another and aid in the acquisition of equivalence classes.

Two standard operant conditioning chambers (28 x 22 x 20 cm) were used in which a reinforcer magazine was presented on one end wall and two levers (Lever 1 and Lever 2) were positioned such that Lever 1 was located adjacent to the reinforcer magazine while Lever 2 was located on the opposite wall. The auditory stimuli were projected through a speaker located above Lever 1. Three subjects were designated to Apparatus 1 while the remaining two were designated to Apparatus 2.

The subjects were first given initial discrimination training in which they were shaped to press Lever 1 without auditory stimuli and were reinforced for lever pressing with either the 45 mg pellets or sucrose solution. Once the subjects were habituated to the chamber and responding to the lever, the subjects were divided into two groups. Group 1 (subjects R1, R2 & R3) were given initial training with only one lever (Lever 2 was absent) while Group 2 (R5 & R7) were given initial training with two levers in order to study whether acquisition of discriminating the auditory stimuli was quicker using one or two levers.

For Group 1, subjects were trained to respond to Lever 1 only when an auditory stimulus was present (regardless of which tone was presented) and not to respond in the absence of a tone (tone vs. no tone discrimination). All subjects acquired the discrimination within 40 sessions (75 trials per session). Next, subjects were trained to discriminate between different auditory stimuli. Set 1 was designated as positive and, although any response to stimuli within Set 1 was reinforced, only valid responses (responses within five seconds of stimulus onset) were considered for the acquisition criterion. Subjects were required to respond at 80% or greater to Set 1 stimuli, not exceed 33% of valid S+ responses on lever presses to S-, and finally, the highest number of responses to S- stimuli had to be less than 50% of the lowest number of responses to S+ in order to meet criterion. Subjects met criterion in 49, 41 and 118 sessions, respectively, at which point, contingencies were reversed (Set 1 became negative and Set 2, positive). Criterion to move to another reversal was identical to the previous criterion needed to move to reversal discriminations. Rats R1 and R2 completed six and four reversals and requiring 257 and 359 sessions, respectively. Rat R3 stopped responding during the first reversal and was removed from the study. Only one of the three subjects (Rat R1) who had initial training with one lever showed any savings during consecutive reversals (fewer sessions were required to reach criterion as the number of reversals increased). Although Rat R2 completed four reversals, no savings were observed as the subject was exposed to more reversals.

Two rats were given the two lever initial training procedure (Rats R5 & R7). The two lever procedure was a chaining procedure in which a response to Lever 2 produced stimulus B1 (1730 Hz) which played continuously until the subject crossed the chamber and responded to Lever 1 (opposite wall) to receive sucrose solution reinforcement. Acquisition of the initial training procedure took four and five sessions for Rats F5 and F7, respectively. Next, subjects were given tone discrimination training in which tones were added to the chaining procedure one at a time, beginning with stimulus B2 (Set 1 was designated as positive at the onset of the discrimination training). For example, a session consisted of 100 trials; 50 trials with B1 presentations and 50 with B2 presented in random order. The remaining stimuli were added one at a time in the order, A2, C1, A1 and C2. Criterion to move to reversal discriminations was that valid responses had to be made for 80% or more of the S+ trials and no more than 33% of the S- trials for three consecutive sessions. Rats R5 and R7 met criterion in 116 and 97 sessions, respectively. When switched to reversal discriminations, Rat R5 completed ten reversals and showed savings as more reversals were given; however, Rat R7 only completed three reversals and showed no evidence of savings as more reversals were given.

Throughout the entire study, only two of the seven rats demonstrated evidence for functional equivalence in subjects needed fewer sessions to reach criterion as a function of the number of reversals. As Dube et al. note, changes in procedure are needed before continuing the current study on functional equivalence, such as greater salience between auditory tones and more research on the importance of differential reinforcement on the establishment of functional classes. Another factor that requires further inquiry was the difference in acquisition of auditory discriminations using one or two levers. The study did, however, show that rats are capable of acquiring and reversing auditory discriminations using six different stimuli.

Rats and Olfaction

An important aspect of doing any type of research is to understand as much about the subjects in the study as possible. For example, much criticism arose out of Sidman’s research comparing children with rhesus monkeys and baboons because of the procedural advantages thought to account for the success of the children and the failure of the primates. Using horizontal and vertical lines as stimuli were considered more natural for the children because of the environment in which they grew up whereas the primates would not have had as much experience with the stimuli. Most of the research on non-human subjects uses visual stimuli; however, some species are stronger with senses other than vision which should be taken into consideration. For animals geared towards vision as their primary sense, research on functional equivalence has been successful in humans, pigeons and primates, but research on animals with a different primary sense have been less successful. For example, rats have traditionally been very poor at tasks in which visual stimuli have been used compared to similar tasks using auditory or olfactory stimuli.

For example, Iversen (1993) attempted to establish generalized identity matching (reflexivity) in three rats using visual stimuli. The stimuli were projected onto a row of three different keys positioned horizontally on an opaque wall inside an operant chamber. Rats were initially magazine trained to receive reinforcement from two food hoppers located on either side of the comparison keys (the two side keys). Two stimuli were used for the experiment; a blinking light and a steady light. The sample stimulus always appeared on the center key. Once the rat responded to the sample stimulus (either a blinking or a steady light) by poking the key with its nose, two comparison stimuli were presented while the sample stimulus remained lit. The subject was reinforced for responding to the comparison stimulus that matched the sample stimulus. On every trial, the sample stimulus had an equal chance of being either the steady or the blinking light, with the restriction that either stimulus could not be presented more than three consecutive times. Each session was composed of 100 trials; however, a correction procedure was in effect such that incorrect responses produced an inter-trial interval and the trial was repeated and added to the total number of trials per session. For example, if the rat responded incorrectly to 20 trials out of 100, the total number of trials per session increased from 100 to 120. Both position preferences (probability of responding to one comparison key over another) and stimulus preferences (responding to the steady light over the blinking light or visa versa) were recorded and adjusted when necessary. For example, if a subject developed a position preference (responded to the left comparison key more than the right), the correct comparison was presented on the right key more often until the side preference was diminished. The same correction procedure was in effect for the stimulus preference (if subjects responded to the steady light more frequently, the blinking light was given as the sample stimulus on more trials than the steady light was).

Results showed all three rats acquired high levels of accuracy on the match-to-sample procedure; however, it took a mean total of 25 sessions with 100 trials per session to achieve 80 percent correct for all rats and a mean total of 50 sessions to achieve accuracy levels of 90 percent or higher. All rats showed a stimulus preference of the steady light in the first few sessions, which diminished over time with the implementation of the correction procedure indicated above. Other sources of control besides matching-to-sample were reported for all rats in the earlier sessions, such as the tendency to respond to the comparison key that was correct on the previous trial, regardless of the stimulus presented (repeat errors). For this reason, repeat errors were not reinforced, while responding to the alternate key on the next trial was reinforced and the repeat errors eventually dropped out. Due to the amount of alternate sources of control besides matching-to-sample, all three rats needed many sessions to attain high levels of accuracy.

Following the 1993 study in which he was able to demonstrate high levels of accuracy with rats using conditional discriminations, Iverson (1997) conducted a follow-up study to see if generalized identity matching could emerge when the sample location was varied. The visual stimuli used in the follow-up study were identical to the ones used in the first study (steady and blinking lights). The apparatus was also the same one used in the previous study, with the three response keys arranged in a horizontal fashion and a food hopper located underneath the center key. Three rats were trained to criterion levels of performance as in the previous study where the sample was always presented in the center key and the two comparison stimuli were presented on either side. For all rats, an average of 60, 100 trial sessions were required to achieve 90 percent correct or higher criterion, considered baseline training.

As soon as the rats reached criterion, they were given probe sessions in which the sample stimulus could appear on any of the three keys with equal probability. Each session contained 96 trials in which the sample appeared on each of the three keys a total of 32 times per session. No correction procedure was used. Before beginning a probe session each rat was given 16 baseline trials (where sample always appeared in the center key).

During the warm-up trials, all rats maintained high levels of accuracy. During the probe sessions (where the sample location was varied), however, performance declined to an average of 60 percent, indicating a lack of acquisition of the identity relation. The accuracy was consistently maintained slightly above chance (60 percent) because for the trials in which the sample appeared in the center during the probe sessions (baseline), rats maintained 90 percent or higher; however, when the sample appeared on the left or right key, the performance was at or below 50 percent. The majority of the time that the sample appeared on one of the side keys, the rats would consistently respond to the center key when the comparison stimuli were presented. Due to the fact that when the sample stimulus was moved to a novel location, performance was disrupted; the response was not under the control of the sample. Therefore, Iverson (1997) concluded that rats had not acquired the identity relation using the match-to-sample procedure with visual stimuli.

The inability of rats to demonstrate generalized identity matching using visual stimuli could be interpreted in a couple of ways. Either rats are incapable of generalized identity matching or some procedural variable, such as the type of stimulus used, limited the subject's ability to acquire the discrimination. Rats rely extremely heavily on their ability to smell, although they also rely on their ability to hear, especially at high frequencies. Olfaction in general had previously been described as "a primitive sense used to guide basic activities such as finding a mate, finding food, identifying young and avoiding danger" (Slotnick, 2001). New research is finding that, although it is one of the first senses to evolve in many animals, olfaction has the most direct link of any sense to the regions in the brain that control higher-order functions such as the prefrontal cortex, the hypothalamus and limbic structures (Zald & Rauch, 2006, p. 43). Olfaction was previously avoided as a modality to test with animals mainly because of the inability to standardize scent as one can standardize visual (wavelengths of light) or auditory (decibels of sound) stimuli. With recent advancements in procedures employing both standardized and non-standardized olfactory stimuli, rats have shown significant improvement in performance on complex tasks such as conditional discriminations. For this reason, the current study will use olfactory stimuli to test for functional equivalence.

Because of the amount of time it took Iverson (1993, 1997) to demonstrate identity matching in rats using visual stimuli, as well as the fact that the rat's primary sense is olfaction, Pena, Pitts & Galizio (2006) developed a study using olfactory stimuli to test identity matching in rats. Pena noted that in order for identity matching to be successfully demonstrated, novel stimuli need to be used during testing phases. For example, if subjects are trained to respond to a green hue (versus a red hue) when given a green hue as the sample, testing of the identity relation should transfer to novel stimuli so that when presented with a blue hue as the sample, with a yellow and a blue hue as comparison stimuli, subjects should reliably respond to the blue comparison hue. Because Iverson (1997) had not used novel stimuli to test for the identity relation, he had not successfully demonstrated generalized identity matching.

The olfactory stimuli used in Pena et al's (2006) study were originally used by Bunsey & Eichenbaum (1996) in which ordinary household spices (cinnamon, onion, paprika, etc.) were mixed with play sand. Subjects were required to dig through the sand to obtain the reinforcer. Four rats were used for the Pena et al. (2006) study. The apparatus used was a modified operant chamber in which the front side was a clear Plexiglas wall with a section at the bottom cut out for a tray containing the stimuli to be inserted. The stimulus tray contained three holes arranged in a triangle in such a way that when the tray was inserted into the chamber the sample stimulus could be presented first and then followed by the two comparison stimuli.

Rats were initially trained to dig in the sand to obtain reinforcement. Once the rats were reliably digging in the sand the training phase was implemented. Two stimuli were used for the first phase of training. Two rats received cinnamon and mustard while the other two received garlic and celery. Each session contained 24 trials. A trial would begin with the presentation of the sample stimulus (cinnamon or celery, for example) which was reinforced with a sucrose pellet. Once the rats attended to the sample, the comparison stimuli were presented; one that matched the sample (cinnamon or celery) and one that did not (mustard or garlic). Subjects were reinforced for responding to the comparison stimulus that matched the sample stimulus. A correction procedure was in effect such that if the subject responded to the incorrect comparison stimulus first, the trial continued until a response to the correct comparison stimulus occurred. The initial training phase continued until subjects achieved a criterion of 75 percent correct or higher for two consecutive sessions. Rats reached criterion levels of 75 percent or higher within 15 to 24 sessions with 24 trials per session.

Once subjects reached criterion levels of performance, reinforcement of the sample cup was reduced from 100 to 75 percent and the match-to-sample training continued. Reinforcement probability was again decreased from 75 to 50 percent once the subjects reached criterion of 90 percent correct or higher for two consecutive sessions and remained at 50 percent for the duration of the study. Each subject varied in the number of sessions to reach criterion in the sample reduction phase of training, ranging from 4 to 16 sessions. The reasoning behind reducing sample reinforcement was the idea that the rats were just digging in the comparison scent that had just been reinforced (as the sample) and not because the scent matched the sample.

The first testing phase was a test for generalized identity matching in which a novel stimulus was introduced into the match-to-sample procedure as a sample and correct comparison stimulus, along with one of the stimuli that had been used in training (baseline stimulus). For example, sage was presented as the sample with sage and a familiar scent (cinnamon or mustard) as the comparison stimuli. Once sage had been exposed to the rat, it was incorporated into the baseline of familiar scents. Each time criterion was met (90 percent or higher for two consecutive sessions) two more novel stimuli were introduced into the mix. Novel probe trials were counted as the first exposure to a novel stimulus in the sample and correct comparison positions. Novel combination trials were when a novel stimulus was used as the incorrect comparison stimulus compared with a familiar sample and comparison stimulus. Rats maintained high levels of accuracy in performance (85-90 percent) as new stimuli were incorporated into the testing phases. Because this could be taken as evidence for rapid acquisition of stimulus-stimulus combinations and not as support for the identity relation, analysis of the novel combination probe trials was needed. For all rats, performance remained high (80 percent or higher) on trials where a novel stimulus was presented as the sample and when novel stimuli were used as incorrect comparisons, suggesting that the rats responses came under control of the identity relation. Pena et al's successful and rapid demonstration of the identity relation in rats using olfactory stimuli versus Iverson's (1993, 1997) attempt using visual stimuli lends credit to the claim that it is important to use the best modality of perception (olfactory stimuli is the rat's primary sense), whenever possible, in order to study the extent to what a subject is capable of showing.

After Pena et al. (2006) successively demonstrated generalized identity matching using olfactory stimuli with rats, Galizio, Miller, Ferguson, McKinney & Pitts (2006) conducted a study using simultaneous discriminations with rats using olfactory stimuli in order to obtain acquisition of the repeated acquisition procedure within a single session. They used a non-spatial procedure, which is commonly used to assess learning and memory effects of certain drugs. The same apparatus (a modified operant chamber) used in this experiment was used in Pena et al's (2006) study previously mentioned. Eight olfactory stimuli (celery, cinnamon, garlic, ginger, mustard, onion, paprika & sage) were mixed with play sand and the subjects could dig to obtain the reinforcer. The stimuli were placed in a tray which was inserted into the chamber and the subject could respond to either stimulus presented.

During pre-training, subjects were given magazine training and trained to dig in the sand to obtain reinforcement. Once the subjects were reliably digging in the sand, repeated reversal training began. For each session, subjects were exposed to two different spices; one designated as S+ and the other S-. A response was defined as displacement of the sand in a digging motion with the paws and/or penetration of sand with the snout (Galizio et al., 2006). Subjects were reinforced for responding to the stimulus that had most recently been associated with extinction. The stimulus that had most recently been associated with reinforcement was designated as S- for that trial. Discrimination reversals could therefore be studied because the subject's response came under the control of the odor that was most recently associated with S-. All subjects were required to reach a criterion of 90% or higher for five consecutive sessions.

Once the subjects reached criterion, the repeated reversal/performance phase was implemented in which a different discrimination was added to the repeated acquisition sequence during each session. This discrimination remained invariant for all sessions (performance component). A total of 24 trials were given per session with 16 of the trials designated as the acquisition component (stimuli are variant across sessions) and the remaining eight designated as performance (invariant across sessions). Sessions were divided into blocks of four to assess improvement of the acquisition component across the session.

Results showed that pre-training required 7-14 sessions at 20 trials per session. More sessions were required to reach criterion during the repeated reversal phase (15-39) for all rats, and the last phase (introduction of the performance component into the repeated reversal performance phase) required an additional 12-51 sessions to reach criterion. Previous experiments using non-spatial procedures to study RAP procedures failed to provide within session acquisition and thus have not been used to assess drug effects on performance due to this reason. This study showed the advantage of using olfactory stimuli for non-spatial RAP procedures with rats because of the rapid rates of learning for both the performance and acquisition components in order to assess the effects of particular drugs on learning and memory. Each subject quickly acquired the new acquisition discrimination for the session, many times learning the discrimination within 4 trials and maintaining high levels of accuracy for the performance component. This experiment is another demonstration that it is important to account for the primary sensory modality used by the subjects under study.

Because of the success of the repeated reversals procedure to train equivalence relations, the current study was a systematic replication of Vaughan's (1988), Dube et al.'s (1993) and Kastak et al.'s (2001) study using rats as subjects. The current study also used class-specific reinforcers to help facilitate acquisition because of the amount of success the previous studies have had in using them. It was hypothesized that, if rats were capable of equivalence, it would be shown using the repeated reversals procedure as a prerequisite for conditional discrimination testing, using class-specific reinforcers and using olfactory stimuli. Because of the success of Pena et al's (2006) study with rats using a modified operant chamber to conduct the conditional discrimination procedure to test generalized identity matching, the same apparatus was used in the current experiment to conduct a simple discrimination procedure to test for stimulus equivalence. However, another apparatus, the Odor Arena, was recently added to the laboratory and because it minimizes the problem of rats' learning position preferences, it was used as well. The unique arrangement of the apparatus provided opportunities of acquisition of functional classes that the chamber was unable to show based on the limited arrangement of the stimuli presented to the subjects.

METHODS

Subjects

Six Holtzman (Sprague-Dawley) male albino rats with ages varying from 3 to 6 months when testing began were used for this study. Each rat's food intake was restricted to approximately 15g per day, and with occasional adjustments if excessive weight was gained or lost. They were allowed *ad lib* access to water. Rats were individually housed in metal mesh cages inside the animal lab in the Social and Behavioral Sciences Building at UNCW and were given a twelve hour light/dark cycle. Access to the daily ration of food was given 15 to 120 minutes after testing.

Apparatus

For the purposes of this experiment, two different apparatus were used in order to test the rates of acquisition of the functional classes. One apparatus, the Odor Arena, is a circular platform made of wood, acrylic surface and Plexiglas. It is 37 in. in diameter, surrounded by aluminum siding that borders 11 ½ in high and 120 in. in perimeter, enclosing the circular platform (see Figure 1). Underneath the platform sits a swivel mechanism that allows the apparatus to be rotated if needed. The Odor Arena contains 18 different 5cm holes spaced 5 ¼ in. apart between each hole arranged in a circular pattern; a larger, outer circle and a smaller, inner circle. A number system was used to label each hole position with 1-12 beginning in the north position and going clockwise, labeling the outside circle and 13-18 labeling the inside circle (see Table 1).

The other apparatus was a two-choice, simple discrimination modified operant chamber (28 cm long x 26cm wide x 30 cm high). The front and back walls of the chamber, as well as the hinged door at the top of the box, were made with clear Plexiglas

in order to observe the rat during the session. The walls are metal, and the bottom is comprised of 12 metal rods. The top door of the box is hinged to allow the rat to be placed inside the chamber and taken out when needed. In order to insert the stimulus tray, a 5cm section was cut out of the bottom of the front wall. The black, plastic stimulus tray is a 25cm

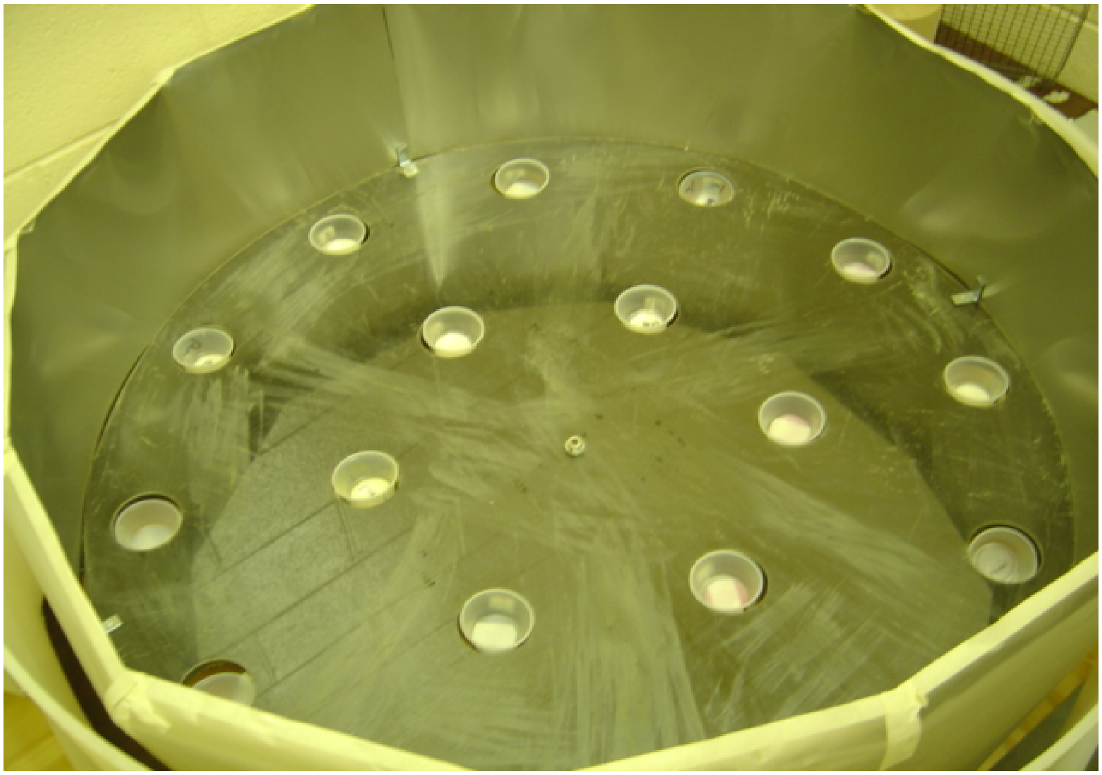


Figure 1

A picture of the Odor Arena is depicted here. All 18 holes are filled with 2oz plastic cups. Measurements are 37 in. in diameter, surrounded by aluminum siding that borders 11 ½ in high and 120 in. in perimeter, enclosing the circular platform.

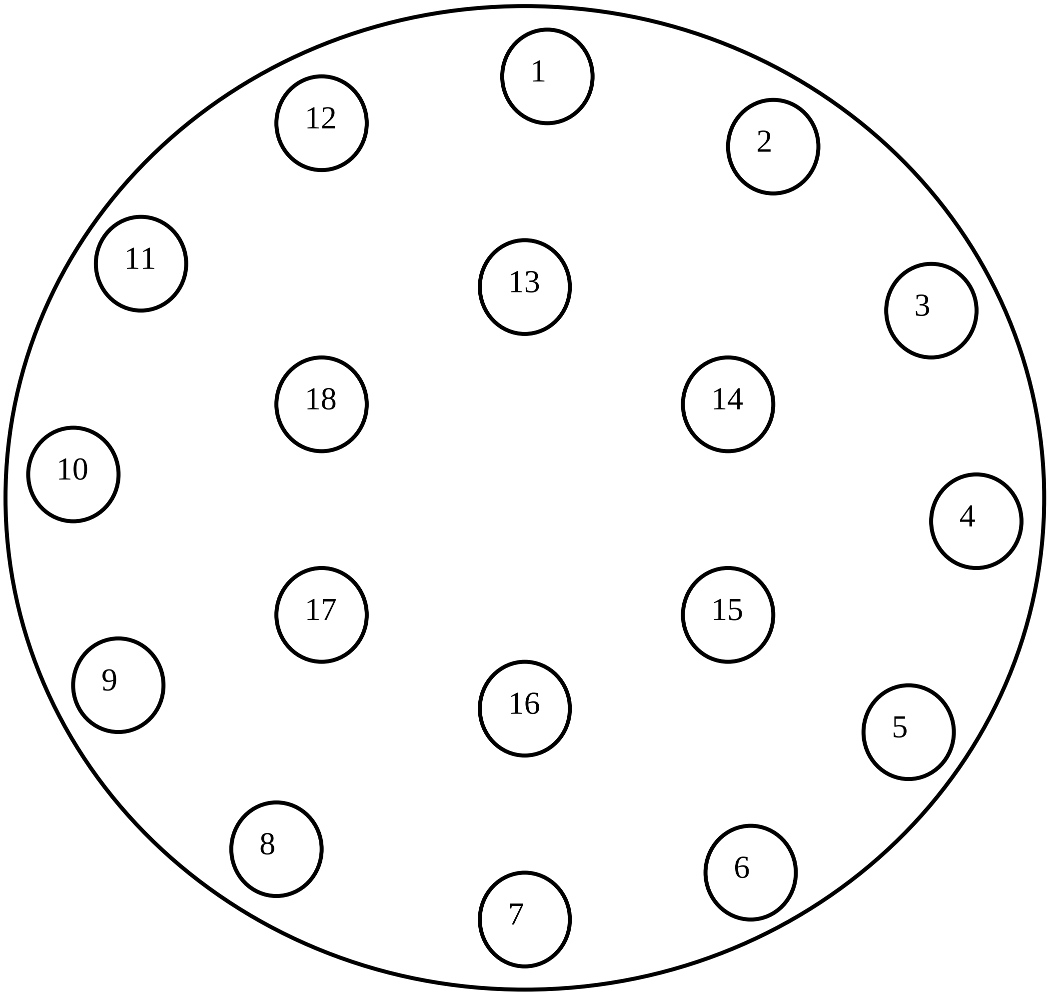


Table 1

Diagram of Odor Arena: (figure not drawn to scale)

Figure 2

A picture of the Two-Choice apparatus is depicted here. The stimulus tray is located at the bottom of the picture in black. The clipboard (located beside the apparatus) was placed in front of the front clear Plexiglas to create less distraction before the stimuli were inserted.

x 6cm x 1.5cm, with two 2cm holes drilled out, horizontal to one another, in order to fit two 2oz. plastic stimulus cups (see Figure 2). The tray enters 6cm into the apparatus when inserted completely. For both procedures, a webcam was used to video record each session. The webcam was placed on the wall above the Odor Arena with the computer monitor adjacent to the

Arena in which the webcam showed the video. A "start session" button was used to start the recording once the animal was placed into the Odor Arena, facing north, and a "stop recording" button was pushed as soon as the trial was completed. For the two-choice simple discrimination operant chamber, the webcam was turned on as soon as the two stimuli were placed into the tray with the clipboard in front of the stimuli, blocking the rat from observing the setup. Once the rat attained the reinforcement from the correct stimulus cup, the trial was ended and the webcam was turned off. This was repeated for each trial until the session ended. At the end of the session, the data was combined and saved into a folder on the computer for storage. The sessions were labeled according to date and put into order from least to most recent.

Stimuli

A total of 20 different odor stimuli, ten per set, were used in this study. For each of the six rats, a unique combination of the scents was randomly assigned into one of two sets (A1—J1 & A2—J2). For a list of stimuli for each rat, see Tables 2 through 5. The spices were bought from the Great American Spice Company and mixed with white, fine-grained, sterilized play sand (Home Depot). The ratio of spice to sand was 1/100g respectively. Each mix of sand with a particular spice was contained in Tupperware tubs that were sealed in order to keep the scent as fresh as possible. This ratio of spice to sand had been previously determined as the least amount of spice necessary to mask the smell of the pellet, based on pilot research from another study (Pena, et al.'s, 2006). The stimulus cups were filled to

Table 2

Stimulus Class Arrangement: Rat H5

Class	1	2
A	Cinnamon	Turmeric
B	Onion	Paprika
C	Cumin	Oregano
D	Rosemary	Mustard
E	Bay	Celery
F	Garlic	Thyme
G	Ginger	Clove
H	Caraway	Sage
I	Sumac	Marjoram
J	Nutmeg	Fennel

Table 3

Stimulus Class Arrangement, Rat F8

Class	1	2
A	Cinnamon	Turmeric
B	Garlic	Ginger
C	Sage	Paprika
D	Celery	Onion
E	Rosemary	Nutmeg
F	Bay	Clove
G	Caraway	Cumin
H	Sumac	Oregano
I	Thyme	Fennel
J	Mustard	Marjoram

Table 4

Stimulus Class Arrangement, Rat F7

Class	1	2
A	Cinnamon	Turmeric
B	Onion	Rosemary
C	Sage	Paprika
D	Bay	Fennel
E	Garlic	Nutmeg
F	Caraway	Cumin
G	Ginger	Must
H	Celery	Oregano
I	Sumac	Marjoram
J	Thyme	Clove

Table 5

Stimulus Class Arrangement, Rat F4

Class	1	2
A	Cinnamon	Turmeric
B	Ginger	Mustard
C	Cumin	Oregano
D	Caraway	Nutmeg
E	Celery	Onion
F	Garlic	Clove
G	Bay	Rosemary
H	Thyme	Fennel
I	Sumac	Marjoram
J	Sage	Paprika

approximately 2cm below the rim for both procedures. The translucent, plastic condiment lids (Fabri-kal) were also labeled on the perforated edges as to what specific scent each lid corresponded to. The lids were stored in separate containers and the cups were stacked side by side in one large container, with the same scents stacked on top of each other. Each 2oz plastic condiment cup (Fabri-kal) was labeled with a specific spice on the bottom of the cup, which was used for that specific spice only. Holes were poked through each lid using a small sewing needle in order to allow the rat to smell the scent more efficiently. The holes for each lid were standardized in size and pattern, as well as the number of holes per lid. For a session with 20 trials, 40 different cups and lids were used; each trial containing a unique set; a cup, the specific scented sand and a lid for that specific spice. Once a session ended, the scented sand used for the day was thrown away and the cups were recycled to the bottom of the pile, for that particular scent, to be used again in later sessions. The lids were also put back in their specific containers on the bottom of the stack.

Two different types of reinforcement were used in this study; one for each stimulus set. For all rats, some stimuli (in set 1 (A1—J1)) were baited with 45mg white sucrose pellets (Noyes) and some stimuli (in set 2 (A2—J2)) were baited with 45mg grain pellets (Noyes). Each specific reinforcer was completely buried approximately 1cm below the surface of the sand using metal tweezers for both apparatus. Each researcher typically wore white lab coats and standard latex gloves during the setup of the experiment as well as during the experiment. Throughout each session, 70dB of white noise was played in order to reduce the level of noise that might have come from outside the lab room, which could have distracted the animal from the experiment.

Procedure

Testing was conducted approximately five days a week, Monday through Friday, between 9:30am and 12:30pm. Each rat was tested only one time each day.

Preliminary Training

Before beginning formal testing, each rat was handled for 15min per day for five days. All rats were introduced to their specific apparatus without stimuli to habituate them to the new environment for three sessions. The rats were initially trained to eat both sugar and wheat pellets inside empty 2oz plastic cups in their respective apparatus for three sessions or until the rat consistently attained the sugar pellet when the tray was inserted into the apparatus. In the next stage, pellets were exposed on top of unscented play sand. Next, scented sand was implemented into the training, with the appropriate type of pellet for that specific scent. For example, cinnamon was paired with a sugar pellet for Rat F7 because it was in the sugar pellet set and turmeric was paired with a wheat pellet. Once the rat responded to the scented sand by eating the specific reinforcer, a systematic method of burying the pellet in the sand deeper on successive trials began until the pellet was buried 1cm below the surface of the sand and the rat was able to consistently dig in the sand to obtain it. For the two-choice apparatus, before the stimuli were placed into the tray, a wooden clipboard was placed in front of the stimuli on the tray to block the rat from observing the setup. Once the stimuli were placed into the holes, the clipboard was taken away and the tray was inserted. This extra step was included as an extra precaution in order to have the rat attend to the scent of the sand before making a response.

Training in Odor Arena

On each session, the cup positions were randomly selected using one of ten data sheets with 20 random sets of numbers in each session and for each rat generated to ensure that the rat wouldn’t pick up on a pattern of the scented cups in certain positions. Restrictions for the random numbers in the Odor Arena of stimuli were that all 18 holes had to be used in a session and that two reinforced scents could not be in the same place twice in a row.

Once the rats were consistently and rapidly removing lids and consuming pellets in scented sand, the next level of preliminary testing began. Once the scents were made for the session, the reinforcers were put in each of the stimulus cups in class 1 with metal tweezers. The tweezers were also dipped into the non-reinforced scented sand cups to make sure that the scent of metal from the tweezers was not a cue to reinforcement. Lids were then placed on the cups they matched and the first set of stimuli, (A1 and B1) were placed in the arena. Every hole was filled with an empty 2oz plastic cup, except for the two holes that the first two stimuli filled for trial one. Once the stimuli were arranged in the Odor Arena, 70dB white noise was turned on and the rat was brought into the room. The rat was placed into the Arena at the beginning of the session and remained in the box for the duration of the session. Once the stimuli were in place, the rat was taken out of his home cage (next to the Arena) and placed in the middle of the Arena facing north. The rat was permitted to navigate through the Arena and respond to the cups. A response was defined as any displacement of the lid with the paw or nose that broke the barrier of the cup and exposed the scented sand. In order to assess whether animals were sampling the comparison stimuli, stimulus approach responses were scored; defined as when the animal approached within 1 cm of a specific spice cup but did not displace the lid. If the rat made a lid response to an S- (the non-reinforced stimulus), a correction procedure was implemented such that the trial continued until an S+ (reinforced stimulus) response was made. When the rat had responded to S+ and turned away from the cup, the trial was terminated. Lid responses to S- were scored as incorrect. Once the rat responded to S+, whether or not he responded to it first, the trial ended and the rat was taken out of the Arena and placed back into his home cage for an inter-trial interval of 30sec, while the next trial was set up. The experimenter then replaced the two scents with the next pair of scents that were placed in different positions in the Arena, indicated by the next trial number.

This pattern continued until all 18 trials were completed and the rat was put back into his home cage. At this time, the last pair of stimuli, as well as the empty 2oz cups, was moved and the Arena was wiped down with Envy disinfectant spray and paper towels in order to clean the apparatus before the next rat was run.

Hole positions for the Odor Arena for each trial were randomly determined (Randomizer.org). Depending on how many trials per session there were, the 18 numbers (for the hole positions) for each set of stimuli were randomly shuffled (1, 8, 16, 5, 14, 3, 1, etc.) and each number was used for a particular trial. For example, if a session contained 20 trials, the 18 numbers would be shuffled to produce 20 numbers ranging from 1 to 18. This was done for both sets of stimuli. The only restriction was that S+ could not be in the same hole position two times in a row.

During the first training session, rats were first given only six stimuli each or 3 pairs of stimuli. For any rat with three pairs of stimuli, 18 trials were programmed so that the rat was exposed to each pair an equal number of times (6). For example, in the Odor Arena, the stimuli were cinnamon, onion and sage for class 1 (A1, B1 and C1 respectively) and turmeric, rosemary and paprika for class 2 (A2, B2 and C2). Trial 1 used A1, cinnamon, and A2, turmeric, as a simple discrimination pair that was randomly placed in the odor arena in one of the 18 locations (specified above).

There were two main differences between the procedures used for the Odor Arena and the Two-Choice chamber. First, a pseudorandom trial sequence was used for the Two-Choice operant chamber. Because stimuli could only be in one of two places, left or right, S+ was presented on the right side as many times as it was presented on the left side in each session. Also, each pair was presented an equal number of times, with the added constraint that each pair would also be counterbalanced left and right. Other than these constraints, the stimuli were presented in random order.

The other difference between the Odor Arena and the Two-Choice chamber was that lids weren't used in the Two-Choice chamber. A clipboard was placed in front of the front Plexiglas of the Two-Choice chamber to block the rat from observing the stimuli being placed into the tray for presentation. Once the stimuli were arranged in the tray, the clipboard was removed and the tray was inserted. Because lids were not used in this apparatus, the response definition was different as well. A dig response was defined by any displacement of the sand by the paw or nose. Once the rat responded to the correct stimulus, the trial ended and the tray was removed from the box while the clipboard was placed back in front of the Plexiglas for an inter-trial interval of 30sec while the next trial was prepared. The first pair of stimuli was removed from the tray and the next pair placed in their correct positions for that trial. When all 18 trials were completed, the session ended and the rat was removed from the apparatus and returned to the home cage. The apparatus was wiped down with disinfectant (Envy) and a paper towel for the next rat. The other rats in this apparatus were run exactly the same.

Analysis of performances in both apparatus was based on the percentage of correct responses. A criterion of two or fewer incorrect responses for two sessions in a row was required to move on to the next phase in training.

Training: Phase 2

When criterion levels of accuracy were reached with six stimuli (three pairs), two more pairs of stimuli were added. With five pairs, the total number of trials increased from 18 to 20 in order to allow for the ten stimuli to be exposed an equal number (4) of times. After reaching criterion, two more stimulus pairs were added, making a total of seven pairs. With seven stimuli, a total of 21 trials were given per session with each stimulus presented a total of three times. After meeting criterion, the final three pairs of stimuli were added to the training phase which decreased the number of trials per session from 21 to 20 in order for all stimuli to be exposed an equal number of times (2) per session.

Testing Phase 1: Repeated Reversals of S+ and S- with Matched Pairs

After the initial training sessions were implemented, the same procedures were used for both apparatus for the remainder of the study. After a criterion of 90 percent or better for two consecutive sessions was reached, contingencies for the two classes were switched so that responses to set 1, the previously reinforced stimulus class, were no longer reinforced, and responses to set 2 were reinforced with a grain pellet. This phase continued until responding to set 2 met criterion of 90% or better for two consecutive days, and at that point the contingencies were again reversed: set 1 (A1—J1) again became the reinforced set and set 2 (A2—J2) became the non-reinforced set. Class-specific reinforced repeated reversals continued for three reversals, until the next phase of testing occurred.

Testing Phase 2: Repeated Reversals with Scrambled Pairs

The second phase introduced shuffled pairs into the repeated reversal testing phase. For example, A1 (cinnamon) was paired with any stimulus from class 2 (A2—J2), instead of just being paired with A2. Throughout the rest of the study, the pairs remained scrambled in order to allow for the maximum amount of facilitation possible for the rat. Because there were ten stimuli in each set, scrambling the pairs was based on shuffling 10 numbers (1-10) twice (one for each set). For example, if there were 10 stimuli for Set 1 (cinnamon, onion, sage, celery, rosemary, bay, caraway, sumac, thyme & mustard) and ten numbers (1-10) shuffled (8, 4, 9, 3, 7, 10, 1, 2, 6, & 5) the odor corresponding to the first shuffled number (8) would be paired with the spice corresponding to that number (sumac); making sumac the first scent used for the session. For Set 2, using ten stimuli (turmeric, clove, fennel, ginger, garlic, oregano, paprika, marjoram, cumin & nutmeg) and again the numbers one through ten shuffled (6, 5, 9, 1, 10, 4, 8, 3, 7, & 2), the first number in the sequence (6) would correspond with the sixth spice in the list (oregano) and that spice would be paired with sumac to create the S+ and S- for trial one of a given session. This was done for all stimuli in each set.

Testing Phase 3: The Delius-like procedure

After completing several reversals (five, three four and five for Rats H5, F8, F7 & F4 respectively), acquisition was shown to be slow and the reversed contingency was not being learned within a single session. Further, it was clear that responding was not controlled by the functional relationship shared by each member of a set. Due to this slow rate of acquisition, the procedure was modified in an attempt to speed up reversal acquisition and improve the chances of gaining a reliable test for the equivalence relation. The modified procedure was borrowed from Delius et al.'s (2000) study in which only two stimulus pairs were trained and tested. The first reversal under this new procedure (the Delius-like procedure) presented only two stimuli from each set for an entire 20 trial session. For example, only Cinnamon (A1) and Onion (B1) from Set 1 and Turmeric (A2) and Rosemary (A2) from Set 2 were trained under the reversed contingency. The Scrambled Pairs procedure was still in effect, so each stimulus from one set could be paired with either stimulus from the other set on a given trial. A criterion of ten out of the first twelve trials as well as five out of the last five trials was required in order to introduce the remaining stimuli. Because each session contained 20 trials, if subjects met criterion of ten correct responses within the first twelve trials of the session, the remaining trials (13-20) were reserved for the remaining eight pairs. The first post-reversal exposures to each of these eight pairs were considered probe trials to test for functional equivalence. If the subject did not meet criterion during first 12 trials, the session continued with only the two pairs and probes were not tested for that session. Once probes were tested, the session ended and subsequent sessions for that reversal contained all ten stimuli until criterion was met. For each subsequent reversal, two different stimuli were trained before moving to probes in order for all stimuli to have been given the same function.

Testing Phase 4: The Extended Delius Criterion Phase

After completing a number of reversals under the Delius-like procedure, a more stringent criterion for the two pairs before moving to probes was implemented (the Extended Delius Criterion). This was done in order to increase confidence that the subjects had acquired the reversed contingency for both stimulus pairs before being moved to probes. Under the Delius-like criterion phase, it was unclear if the behavior was under the control of the reversed contingency for the two pairs. During a couple of probe sessions, subjects were reaching criterion on the first day of a reversal change by chance due to the fact that they responded to the correct stimulus first without being exposed to the S- stimulus on the majority of the 12 trials. It is thought that more rapid acquisition can occur when subjects are exposed to both the S+ and S- stimuli versus being exposed to one or the other. When subjects are exposed to only one of the two stimuli during a given trial, it remains unclear whether the control of the response is under the control of the stimulus the subject was exposed to or both of the stimuli. For example, if the subject responded to the S+ stimulus first, the response was reinforced and the trial ended without the subject being exposed to the S-. If the subject was exposed to the S- first, it is possible that even if the subject did not respond to the incorrect stimulus, they responded to the S+ stimulus without sniffing the scent. During this procedure, two pairs were still trained before moving to probes; however, performance had to meet a criterion of 90% or higher for the 20 trials within the session using two pairs for the first day followed directly by the previous criterion under the Delius-like procedure for the following day.

Pellet Detection Trials

It is always important to implement probe trials into the experiment in order to make sure that behavior is under the control of the scent of the sand and not the scent of the pellet. To control for this possibility, two trials per session will be included to bait both cups with a pellet or to bait neither cup with a pellet. These trials are crucial to understanding whether or not the rat can use the smell of the pellet to guide responding. If the rat performed at or below chance levels when both cups are baited, experimenters will take this as evidence that the pellet is being smelled by the rat. If the rat only responds to the S+ when both cups are baited, or if neither cup is baited, we can assume that the scent of the pellet is not sufficient to guide responding.

Blind Experimenter Sessions

Blind Experimenters were implemented in an effort to control for experimenter cuing. For Rat H5, once significant above chance probe performance was observed, subsequent probe trials were run under a different experimenter to control for the possibility of the initial experimenter cuing the rat as to which stimulus was correct for that trial. During the blind experimenter sessions, the initial experimenter set up the trial and placed both stimuli into the tray; allowing for the blind experimenter to insert the tray into the chamber without knowing which stimulus was correct. Once the subject made a response, the initial experimenter recorded the response and set up for the next trial.

Probe Trials

Probe trials, the first exposure to each stimulus pair a reversal, were the main tests for functional equivalence because they allowed us to assess function transfer without direct conditioning. Evidence of functional equivalence came only when subjects responded to stimuli following a reversal based on how they already responded to some of the stimuli within that set. For example, if cinnamon and onion were part of the same set and cinnamon had been reinforced on a given session, subjects should respond to onion based on their experience with cinnamon and not based on their most recent experience with onion.

RESULTS

Figure 3 shows the session by session performances for Rat H5. The horizontal axis depicts the session number and the vertical axis shows the percent correct for each session day. The horizontal line intersecting some of the data points indicates the 90% criterion level. Each dotted line indicates the separation between contingency reversals and the number of each reversal is within the panel. The solid black line indicates the separation between pre-training and the Vaughan-like reversal training. Subjects were given three pairs of stimuli until meeting the consecutive, two day, 90% criterion and then moved to five, seven and finally ten training pairs with each switch being done after meeting criterion for the previous phase. During pre-training, all subjects began with set one designated as S+ and set two designated S-. The blue dotted line indicates the separation between Vaughan-like and Delius-like reversal training. For the Vaughan-like training (left of the blue line), which ever set was designated as S+ was reinforced until meeting the 90% criterion and then the contingency was reversed and the remaining set was designated positive while the previously positive set was S-. On The first day of each new reversal under the Vaughan-like procedure, subjects were given all ten pairs of stimuli and tested for class formation on the first ten pairs of stimuli. Under the Delius-like procedure, the subjects, on the first day of a reversal, were given only two pairs of stimuli until reaching the same criterion before being exposed to the remaining eight pairs and testing for class formation. The sessions in which only two stimulus pairs were given are indicated by the open circles. The red dotted line separates the Delius-like training with the Extended Delius Criterion phase in which subjects were required to meet a 90% criterion on day one and the previous criterion of ten out of twelve and the last five correct on the following day.

As Figure 3 shows, Rat H5 (Two-Choice apparatus) completed a total of 205 sessions and required 25 sessions to complete pre-training. After pre-training was completed and the first reversal was implemented, Rat H5's performance dropped to 20% on the first day of the reversal. The below average performance for the first day of the first reversal indicates a lack of class formation of the stimuli in each set. Performance rose steadily during each consecutive session and reached the 90% criterion in eleven sessions. Subsequent reversals show that performance on the first day of a reversal never dropped below chance levels after the first reversal and maintained steady acquisition until criterion was met. For example, performance on the first day of the second and third reversals was at 65%. Rat H5 completed five reversals under the Vaughan-like procedure, averaging 9 sessions per reversal to reach criterion.

All subjects were switched to the Delius-like procedure because no above chance performance was observed on the first day of a reversal and it was clear that the subjects were not learning the discrimination on the first day. When switched to the Delius-like procedure, Rat H5 completed a total of 21 reversals and averaged 6.38 sessions per reversal to reach criterion. After completing 7 reversals, Rat H5 never dropped below 65% on any session and rapidly achieved criterion levels of performance for each reversal. Because of Rat H5's more rapid acquisition during reversals as well as the amount of reversals completed compared to the other subjects, Rat H5's data is shown by itself.

Figures 4 through 6 show the results of each of the remaining subject's (Rats F8, F7 and F4) session by session data. In Figure 2, Rat F8 in the Two-Choice apparatus completed 15 reversals; Rats F7 and F4 (Figures 3 and 4, respectively) in the Arena completed 14 and 16 reversals respectively.

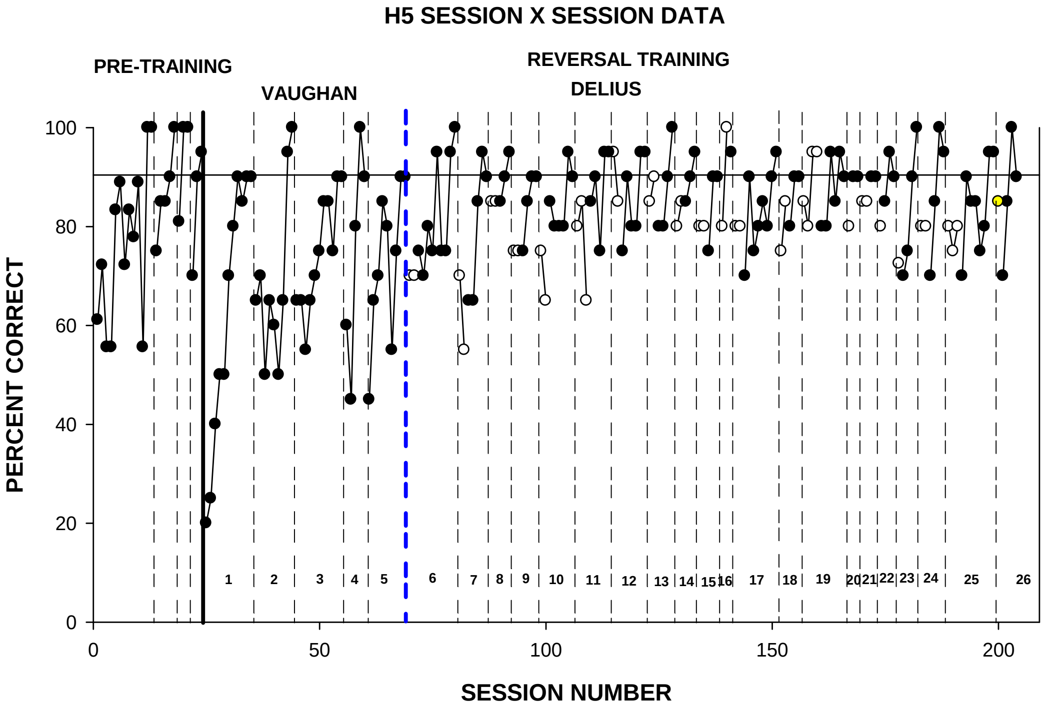


Figure 3

The session by session data for Rat H5 is depicted here. The vertical solid black line separates pre-training from reversal training. The blue dotted line separates the Vaughan-like and Delius-like phase. The black dotted lines separate successive reversals and the numbers at the bottom of each panel indicate the reversal number. The open circles indicate sessions in which only two stimulus pairs were trained while the closed circles indicate when all ten pairs were presented for a session.

The total number of sessions completed by each subject was 241 (Rat F8), 219 (Rat F7) and 238 (Rat F4). During pre-training sessions, Rat F8 took 31 sessions before being moved to reversal training, whereas Rats F7 and F4 took 23 and 15 sessions respectively. Rat F8 was given only three reversals under the Vaughan-like procedure and averaged 20.67 sessions per reversal to reach criterion. In the Arena, Rat F7 had four reversals under the Vaughan-like procedure averaging 17 sessions per reversal to reach criterion levels of performance while Rat F4 completed five reversals under the Vaughan procedure and maintained an average of 16.6 sessions per reversal. For the Delius-like reversal training Rat F8 had 12 reversals, completing each reversal with an average of 10.58 sessions per reversal. Rat F7 completed ten reversals under the Delius-like training, averaging 12.7 sessions to criterion while Rat F4 completed 11 reversals and averaged 10.72 sessions per reversal.

In order to show a more concise depiction of the effects of successive reversals on the acquisition of each reversal, Figures 7 and 8 summarize reversal learning rates by presenting the number of sessions required to reach the 90% criterion. The vertical axis depicts the number of sessions to reach criterion and the horizontal axis shows the number of reversals. The solid black line indicates the separation between the Vaughan-like and Delius-like reversal training. Overall, every rat showed a reduction in the number of sessions taken to reach criterion as the number of reversals increased. Figure 7 shows Rat H5's (Two-Choice) data showing an average of 8.2 sessions to criterion under the Vaughan-like procedure and an average of 4.47 sessions to criterion for the Delius-like procedure. During this particular procedure, Figures 7 and 8 show only those sessions in which all ten stimuli were presented. The initial sessions of reversals under this procedure in which only two stimuli were trained were omitted from this particular analysis. Rat H5 required approximately ten sessions to reach criterion on the first three reversals and showed only slight improvement during the remaining reversals under the Vaughan-like procedure.

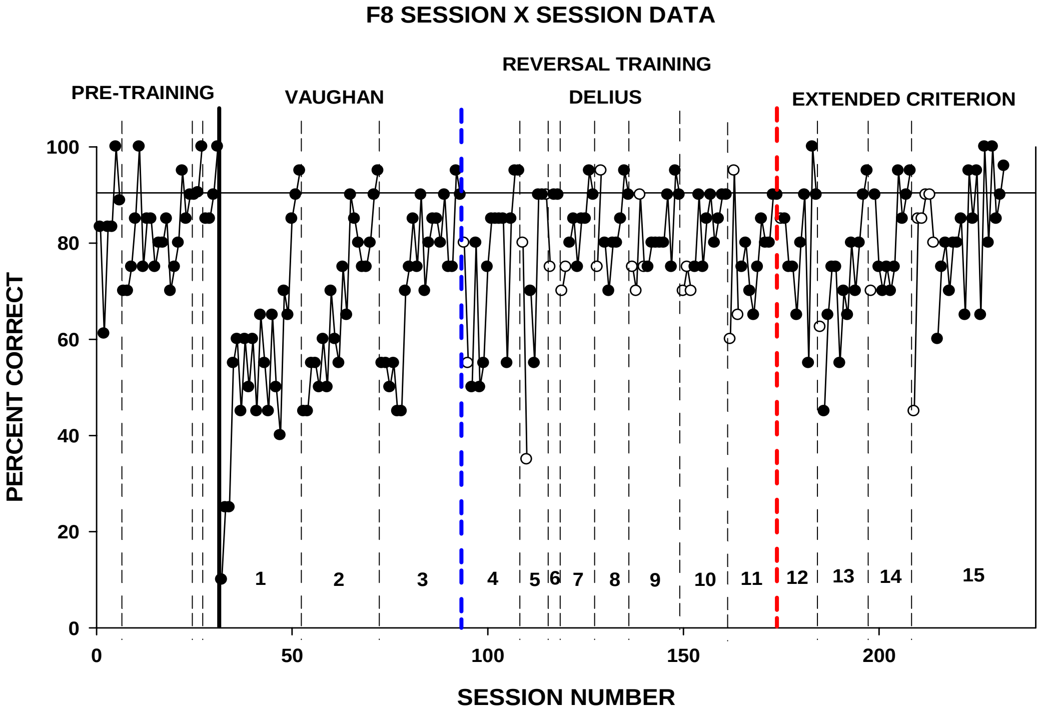


Figure 4

The session by session data for Rat F8 is depicted here. The vertical solid black line separates pre-training from reversal training. The blue dotted line separates the Vaughan-like and Delius-like phase. The red dotted line separates the Delius-like and Extended Delius Criterion phases. The black dotted lines separate successive reversals and the numbers at the bottom of each panel indicate the reversal number. The open circles indicate sessions in which only two stimulus pairs were trained while the closed circles indicate when all ten pairs were presented for a session.

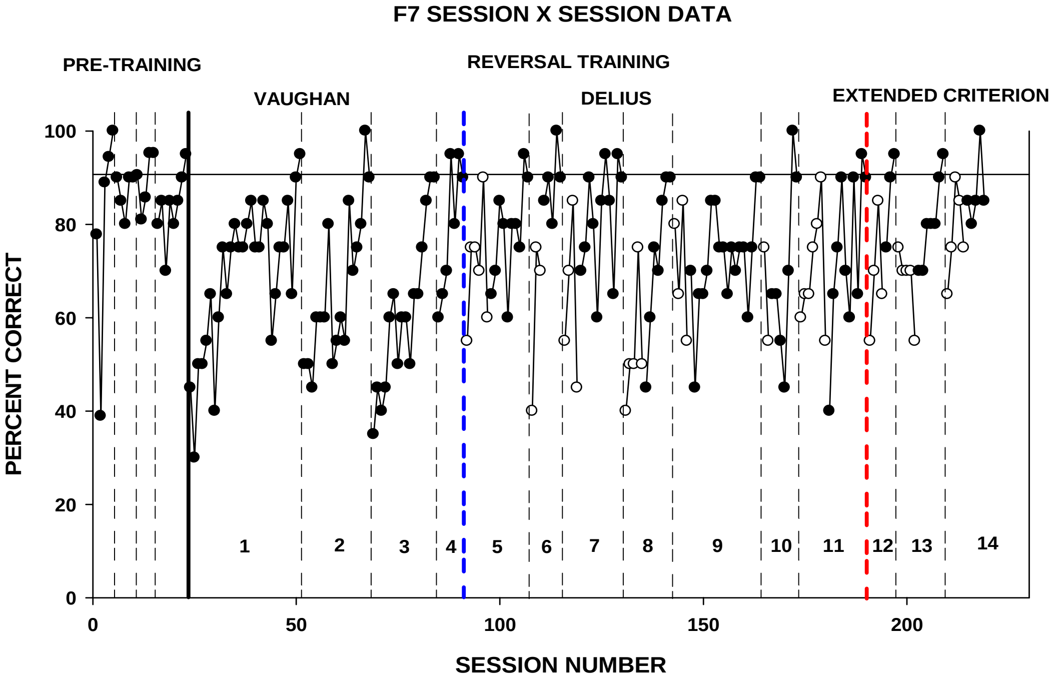


Figure 5

The session by session data for Rat F7 is depicted here. The vertical solid black line separates pre-training from reversal training. The blue dotted line separates the Vaughan-like and Delius-like phase. The red dotted line separates the Delius-like and Extended Delius Criterion phases. The black dotted lines separate successive reversals and the numbers at the bottom of each panel indicate the reversal number. The open circles indicate sessions in which only two stimulus pairs were trained while the closed circles indicate when all ten pairs were presented for a session.

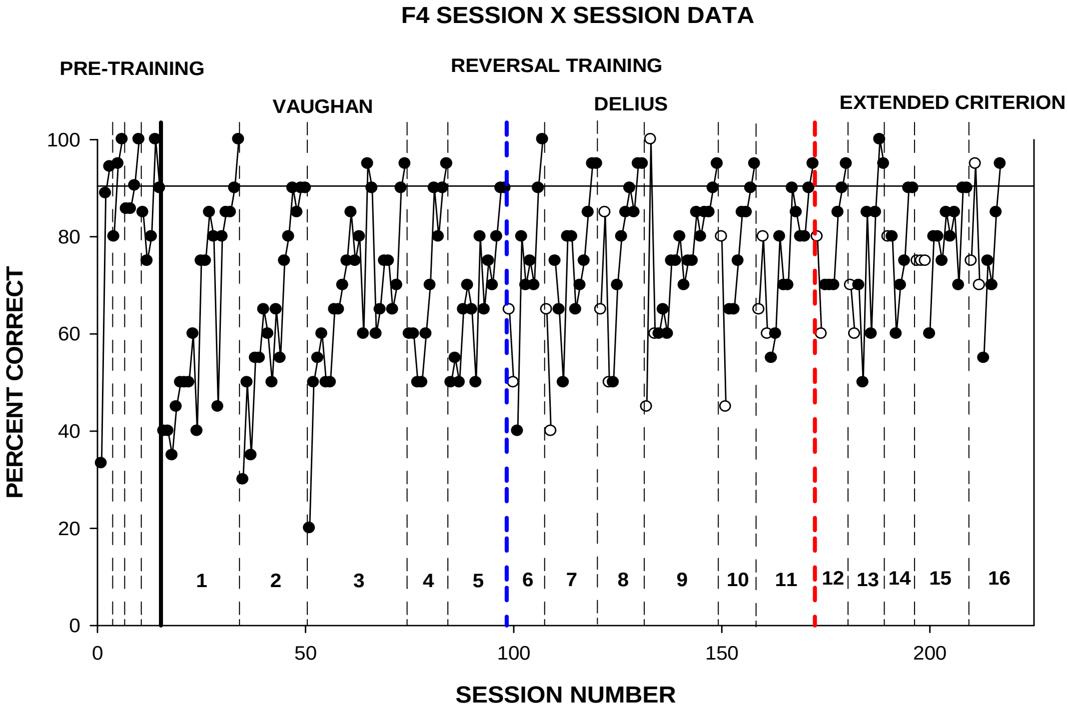


Figure 6

The session by session data for Rat F4 is depicted here. The vertical solid black line separates pre-training from reversal training. The blue dotted line separates the Vaughan-like and Delius-like phase. The red dotted line separates the Delius-like and Extended Delius Criterion phases. The black dotted lines separate successive reversals and the numbers at the bottom of each panel indicate the reversal number. The open circles indicate sessions in which only two stimulus pairs were trained while the closed circles indicate when all ten pairs were presented for a session.

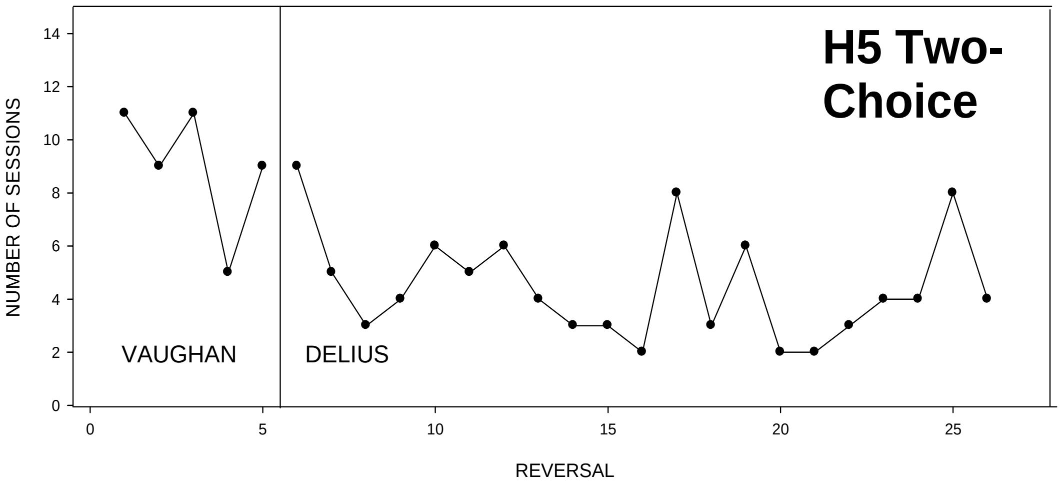
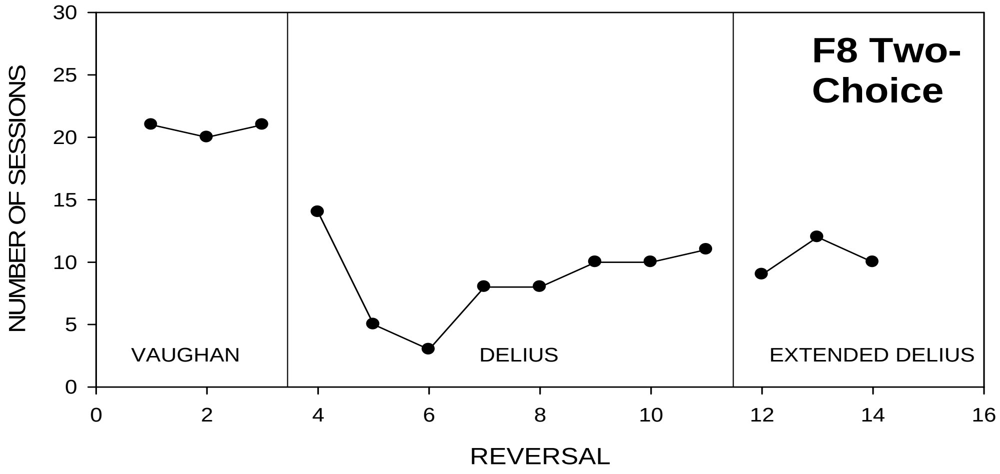
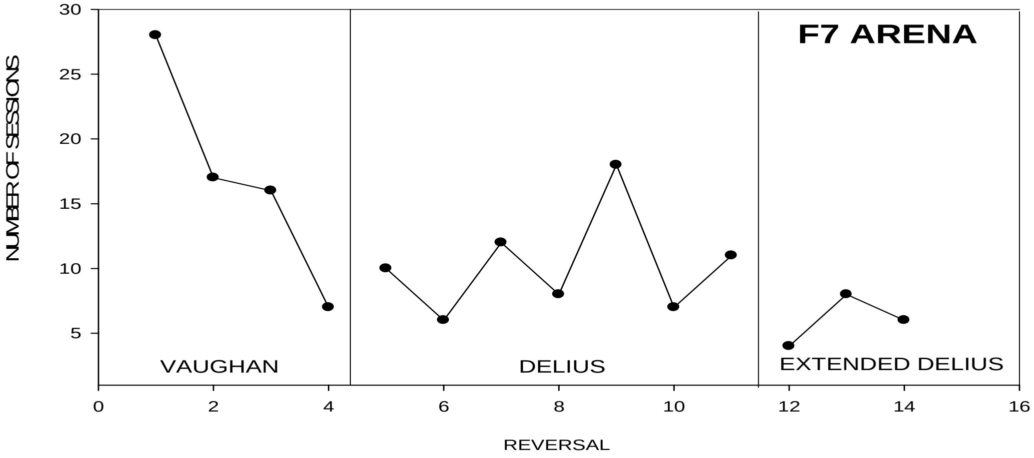


Figure 7

The number of sessions taken to reach criterion for Rat H5 (Two-Choice apparatus) is depicted here. The solid vertical line indicates a separation between the Vaughan-like and Delius-like phases.


F8 Two-Choice
NUMBER OF SESSIONS
REVERSAL
VAUGHAN
DELIUS
EXTENDED DELIUS
0
5
10
15
20
25
30
0
2
4
6
8
10
12
14
16


F7 ARENA
NUMBER OF SESSIONS
REVERSAL
VAUGHAN
DELIUS
EXTENDED DELIUS
5
10
15
20
25
30
0
2
4
6
8
10
12
14
16

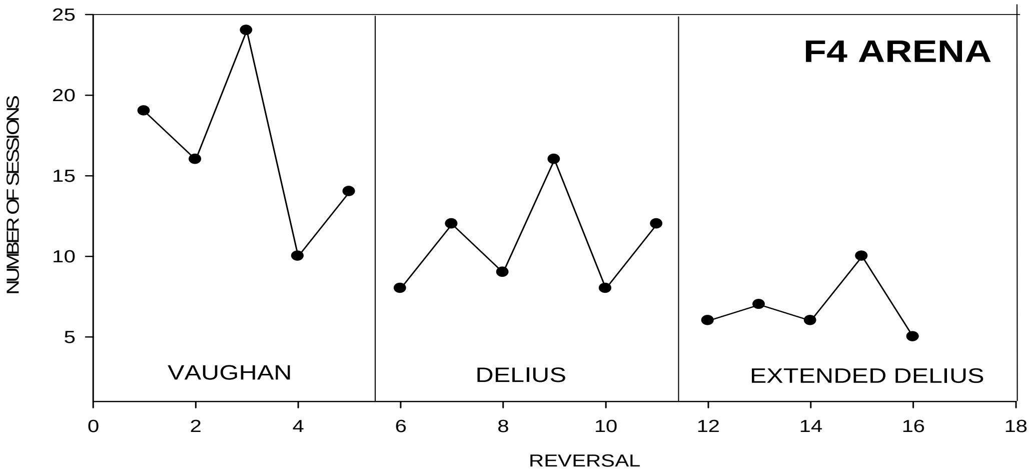

F4 ARENA
NUMBER OF SESSIONS
REVERSAL
VAUGHAN
DELIUS
EXTENDED DELIUS
5
10
15
20
25
0
2
4
6
8
10
12
14
16
18

Figure 8

Each graph depicts the number of sessions taken to reach criterion per reversal for subjects F8, F7, and F4. The vertical slid lines indicate a separation between each phase (Vaughan-like, Delius-like and Extended Delius Criterion phases).

The first reversal under the Delius-like procedure also showed a slow rate of acquisition; however, subsequent reversals showed considerable improvement which was maintained throughout the remainder of the study with only a few exceptions. Reversals 17, 19 and 25 showed an increase in the number of sessions to reach criterion. One possible explanation could be the specific reinforcer (grain) given during these reversals. Throughout the entire experiment, H5 never required more than 11 sessions to reach criterion on a reversal.

Figure 8 depicts the remaining three rat's sessions to criterion data. Rat F8 in the Two-choice apparatus averaged 20 sessions to reach criterion under the Vaughan-like procedure and showed no decline in the amount of sessions needed under this procedure. A drop to 7.6 sessions to criterion under the Delius-like procedure may have been due to more exposure to reversals or because of the acquisition of two of the ten pairs prior to exposure to all ten pairs. A mean of 10.33 sessions was required under the Extended Delius Criterion phase which remained stable across reversals under this procedure. Rat F7 showed a strong general decrease in the number of sessions to criterion across four reversals under the Vaughan-like procedure. The first reversal required nearly 30 sessions, but the fourth and final reversal required only seven. The overall average for the Vaughan-like procedure in which four reversals were completed was 17 sessions per reversal. For the Delius-like procedure, Rat F7 averaged 10.43 sessions, completing seven reversals and showing a decline in the number of sessions needed to reach criterion during this procedure. The Extended Delius Criterion phase in which three reversals were completed averaged 5.67 sessions. During the Delius-like phase there was no real up or downward trend in the data and it showed a fair amount of variability, depicting a lot of bounce from reversal to reversal. Some of this bounce might be accounted for by the differential reinforcement which was alternated for each reversal; sucrose pellets were given during the completion of every even reversal while grain pellets were selected as reinforcement under all odd reversals. As Figure 8 shows, more rapid acquisition was associated with reversals in which correct responses were reinforced with sucrose pellets. Arena trained rats showed more variability in sessions to criterion than subjects in the Two-Choice apparatus. Rat F4 in the Arena was exposed to five reversals under the Vaughan-like procedure and averaged 16.6 sessions to reach criterion with considerable variability throughout this phase. During the Delius-like procedure in which six reversals were completed, Rat F4 averaged 10.83 sessions to reach criterion and the data again showed a large amount of variability. During the last phase in which the Extended Delius Criterion was implemented into the procedure, the subject completed five reversals and averaged 6.8 sessions to reach criterion. Overall, there was a general decrease in the number of sessions required to reach criterion levels of performance for Rats F7 and F4 in the Arena. For all four subjects, improvement was observed across reversals, but performance never reached an optimal state of consistently reaching criterion in a single session.

Figures 9 and 10 show the results from probe trials as a function of the number of reversals. The vertical axis depicts percent correct and the horizontal axis shows the reversal number. The top panel in Figure 9 shows the probe trials for Rat H5. The leftmost panel shows the Vaughan-like procedure in which all ten stimulus pairs were presented when the contingency was reversed. This panel indicates subject’s probe data for the ten trials (probe trials) on the first day of each reversal under the Vaughan-like procedure. For each trial, one stimulus from each set was presented for the first time under the reversed contingency. As Figure 9 shows, Rat H5 was exposed to five probe sessions during the Vaughan-like procedure and performance remained at chance levels ($p > .05$), indicating a lack of class formation for the stimuli within each set. During the first reversal, Rat H5 responded to all eight stimuli incorrectly, indicating that ……………………

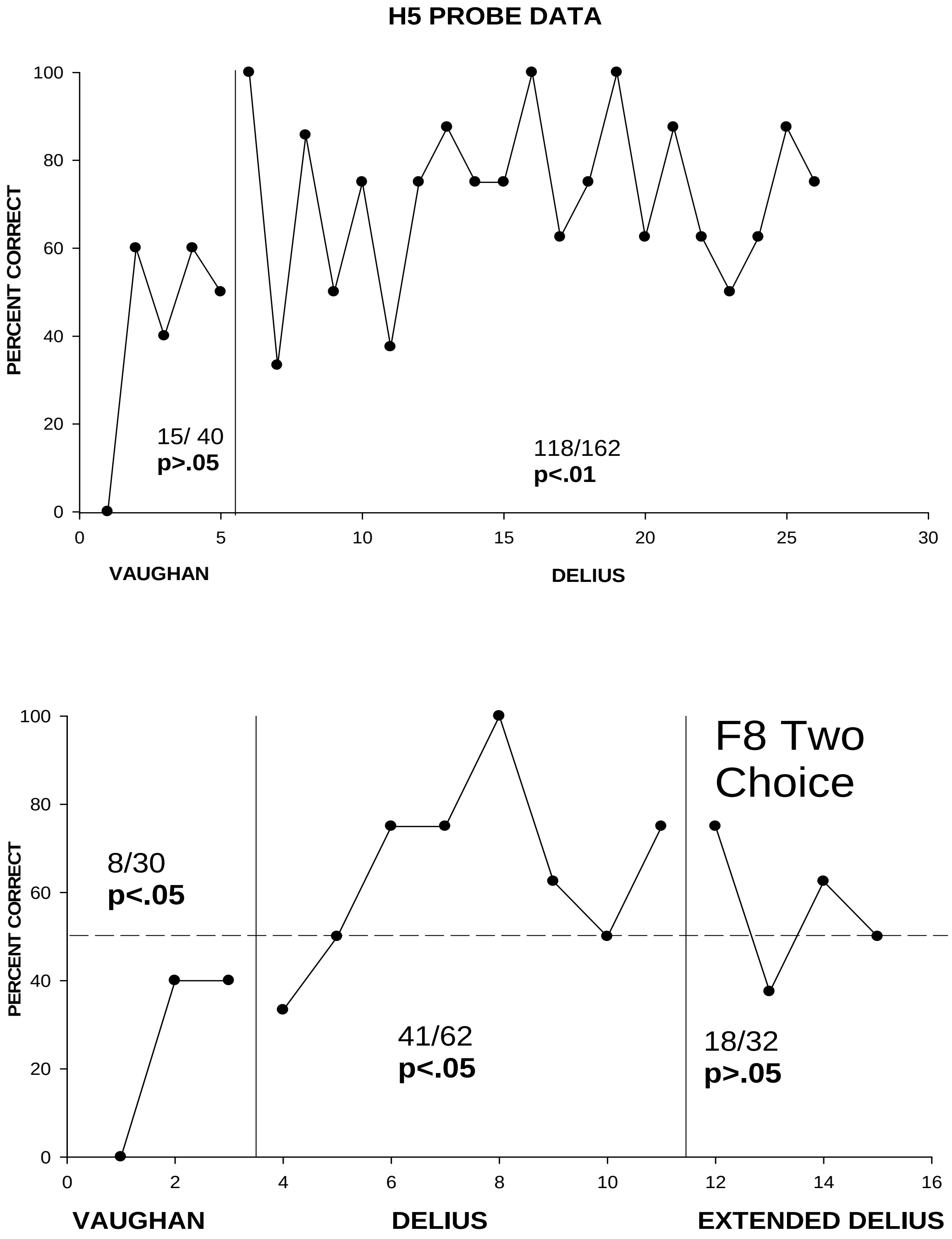


Figure 9

Each graph depicts the probe data for subjects H5 and F8 run in the Odor Arena for the Vaughan-like, Delius-like, and, for Rat F8, Extended Delius phases. The vertical solid lines indicate the separation between different phases. The top numbers below the data indicate the number of correct probe trials compared to the overall number of probe trials given during that phase. The bottom numbers indicate whether the top numbers are statistically significant or not.

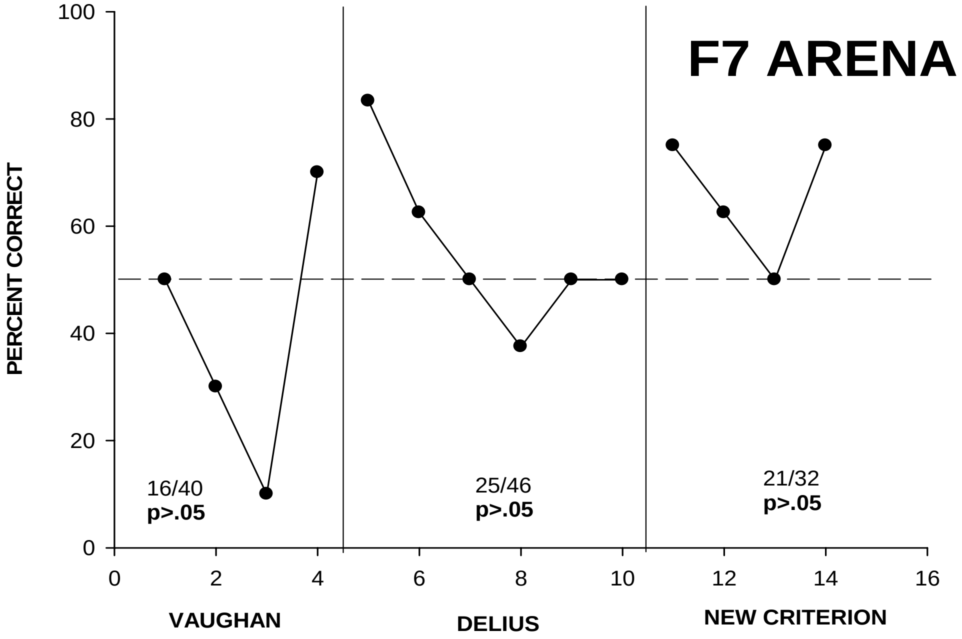
performance was under the control of the last exposure to each of the stimuli the subject was given. Subsequent reversals showed a change in performance during the probe sessions, indicating a shift in control of the responses to each of the stimuli from the last exposure to random selection. Performance for Rat H5 remained at chance levels for the duration of the Vaughan-like procedure, responding correctly to 15/40 probe trials.

The second panel for Rat H5 shows the next series of reversal probe trials under the Delius-like procedure in which the subjects were first trained on two of the ten stimulus pairs to criterion before being tested on the remaining eight stimulus pairs. In this phase, the criterion to move to probes was that the subject had to respond correctly to ten out of 12 trials on the two pairs as well as correctly responding to the last five of those 12 correctly before the remaining eight stimulus pairs were introduced. When the procedure was implemented, Rat H5 showed immediate improvement on the probe sessions and remained at above chance probe performance after most subsequent reversals. As can be seen, Rat H5 showed a lot of bounce in the data during the first few reversals of this phase, which might have been caused by the differential reinforcement given for consecutive reversals; however, during the remainder of the probe sessions, performance became more consistent, remaining above 60% and reaching 100% on two occasions. During the Delius-like procedure, performance dramatically improved to an overall 118/162 correct responses on probe trials, indicating significant above chance probe performance as depicted by a binomial test ($p < .01$) and a shift in control from random selection to control by a few stimuli within a set whose new function transferred to other members of the set without direct reinforcement. This shift in control is indicative of, and crucial to, functional equivalence because stimuli become related based on a specific function that each member of a set shares in common with the other members of that set; when the function to one or more members of the set is changed, that change applies to all members of the set. The difficulty of shifting control from responses based on previous history with a specific stimulus to previous history with another stimulus or other stimuli whose function has been shown to have changed is evident in the lack of research showing evidence for functional equivalence in non-human animals and in the remaining subjects whose data is not as promising. Based on the significant above chance performance under the Delius-like procedure, Rat H5 provided strong evidence of functional equivalence.

The bottom panel in Figure 9 shows that Rat F8 was exposed to only three probe sessions under the Vaughan-like procedure in which performance was poor. When the Delius-like procedure was implemented, however, performance on probe sessions showed a progressive improvement and, on the binomial test, averaged significantly above chance ($p < .05$). The third panel for Rat F8 (seen on the bottom graph of Figure 9 and for both graphs on Figure 10) indicates the Extended Delius Criterion in which probe performance did not improve and actually declined to chance levels ($p > .05$).

Figure 10 shows the results of the percent correct on probe sessions as a function of the number of reversals from Rat F7 and Rat F4 in the Odor Arena. Each panel shows the same procedural change as in Figure 7. Rat F7 had four exposures to the Vaughan-like procedure and showed chance level probe performance (16/40; $p > .05$). When switched to the Delius-like procedure, the subject’s performance remained at chance levels, producing results indicative of a lack of class formation (25/46; $p > .05$). The final procedure implemented (Extended Delius Criterion) provided some improvement on percent correct on probes; however, this was not significant to generate above chance levels of performance for Rat F7 (21/32; $p > .05$).

Rat F4 was exposed to five sessions under the Vaughan-like procedure, but performances were generally at chance levels (22/50; $p > .05$). When switched to the Delius-like and the Extended Delius Criterion procedures, probe performance showed no consistent improvement (20/47; $p > .05$) and (16/40; $p > .05$) respectively. Overall performance for Rat



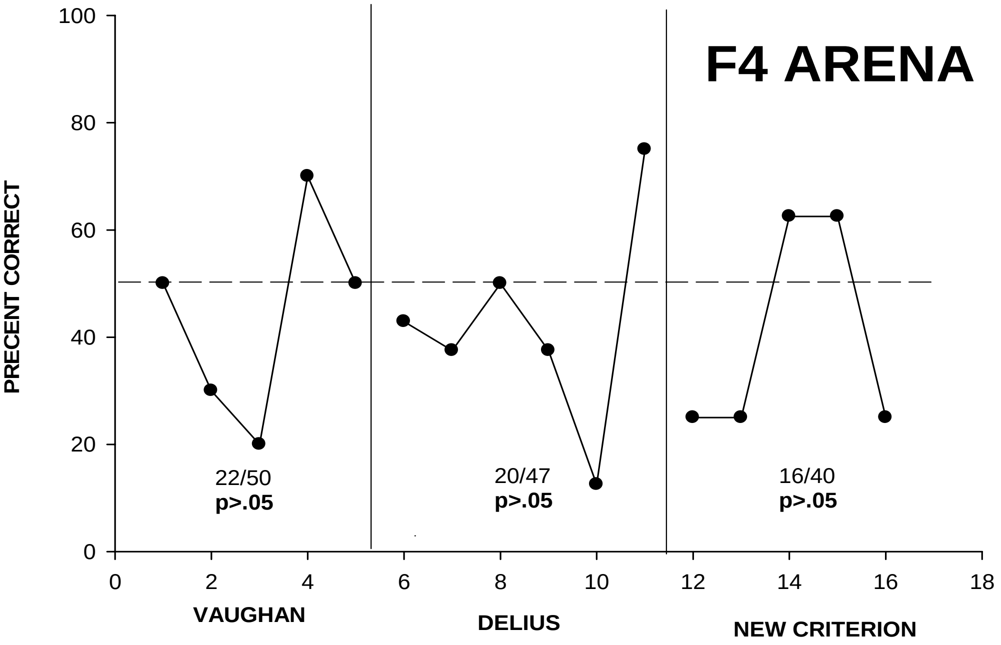


Figure 10

Each graph depicts the probe data for subjects F7 and F4 run in the Odor Arena for the Vaughan-like, Delius-like, and Extended Delius phases. The vertical solid lines indicate the separation between different phases. The top numbers below the data indicate the number of correct probe trials compared to the overall number of probe trials given during that phase. The bottom numbers indicate whether the top numbers are statistically significant or not.

F4 was poor and showed no consistent improvement on probes as a function of the number of reversals.

Figures 11 and 12 depict the average number of correct responses to the first eight trials on the first day of each reversal (probe trials) for all rats. Because more exposure to stimuli within a set whose function had changed should produce better performance on the remaining stimuli within a set, performance should improve as the subjects are exposed to more trials under the Delius and Vaughan-like procedures. For Trial 1, subjects had only been exposed to two of the ten pairs of stimuli, providing only two examples of the reversed contingency. As each new trial was given, more of the stimuli from each set provided information about the remaining stimuli within the set in terms of what responses produced reinforcement. All eight trials were split into blocks of four, where each block included two trials (Block 1= Trials 1 and 2; Block 2 = Trials 3 and 4; Block 3 = Trials 5 and 6; Block 4 = Trials 7 and 8). The solid line indicates the average correct for reversals one through five for each block when the Vaughan-like procedure was in effect. The dotted line indicates when we switched to the Delius-like procedure and the subjects were given sessions with two pairs of stimuli under the new contingency before testing the remaining eight.

Figure 11 shows the data for Rats H5 and F8 in the Two-Choice apparatus. For Rat H5, performance under the Vaughan-like procedure remained stable across blocks at around 40%. Under the Delius-like procedure the data remained consistently high; between 60 to 80% correct, but the highest performance was in the fourth block, consistent with a functional equivalence account. Rat F8 showed variability from block to block under the Vaughan-like condition with the last block showing an increase in average performance on the last two trials. For the Delius-like condition, stable performance across blocks was shown (mean = 62%). Performance was higher for both Rat H5 and Rat F8 under the Delius-like procedure than under the Vaughan-like procedure for all blocks. Importantly, performance was highest on Block 4 for Rat H5 under the Delius-like condition, which is what would be expected if

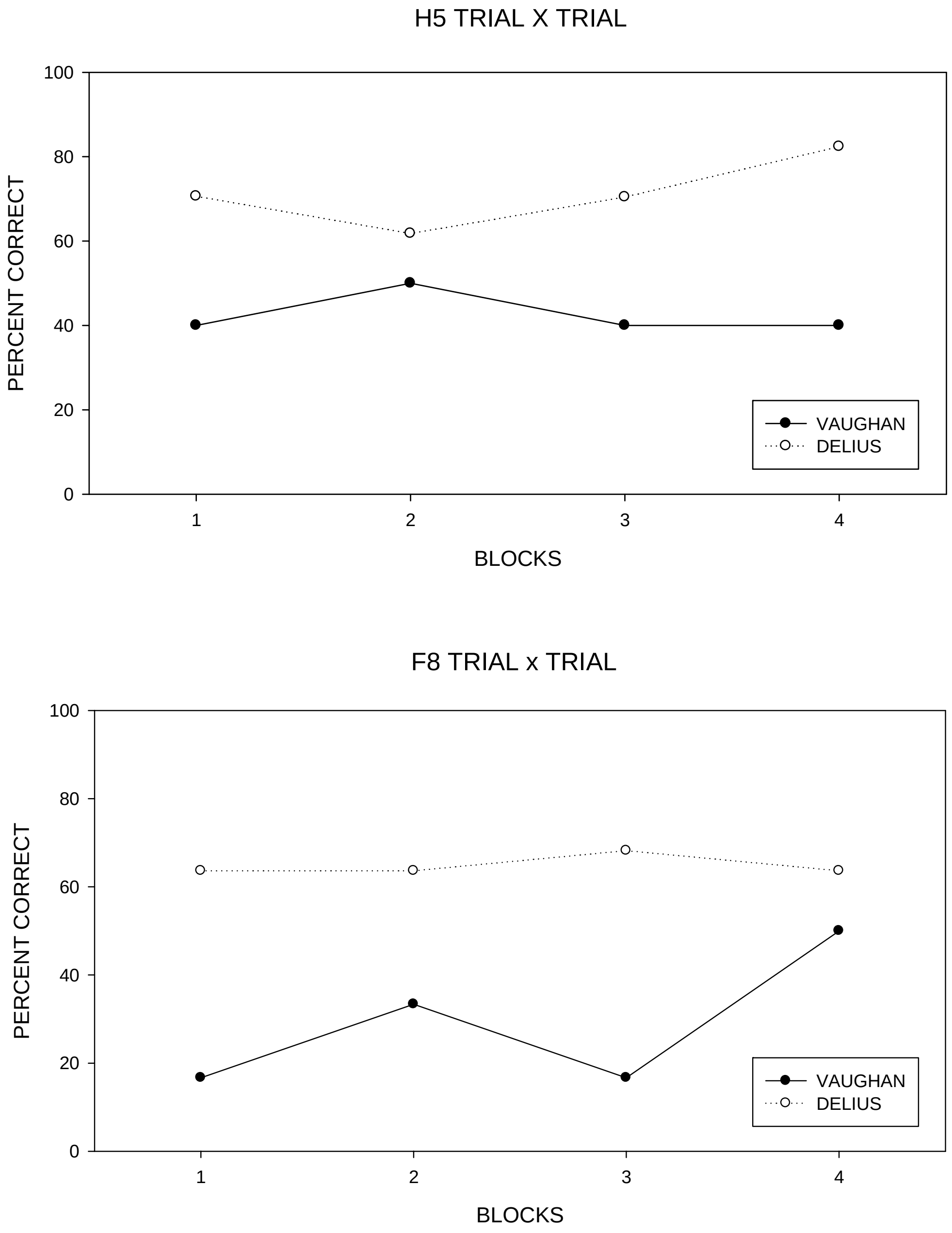


Figure 11

Each graph depicts the trial by trial data (the average number of correct responses to the first eight trials on the first day of each reversal (probe trials)) for Rats H5 and F8. Each block represents two trials. The solid black circles indicate probe trials under the Vaughan-like procedure. The open white circles indicate probe trials given under the Delius-like procedure.

functions trained to stimuli in the beginning of the session are transferring to class members by the end of the session.

Figure 12 shows the data for Rats F7 and F4 in the Arena. Rat F7 showed a decreasing trend from block to block under the Vaughan-like procedure, showing an average of 50% for the first two blocks and decreasing to 25 % on the last block. Performance on the Delius-like procedure shows high performance on the first block (69%) followed by a decrease in performance on the second. From Blocks 2 to 4, however, there is an increasing trend in the data in which the subject's performance rose from 40% on block two to 75% on block four. Rat F4 showed an increasing trend under the Vaughan-like procedure. Block one averaged 10% and performance steadily improved to 40, 50 and finally 59% from Blocks 2 to 4, respectively. Under the Delius-like procedure, Rat F4 showed an increase in average performance from Blocks 1 and 2 to Blocks 3 and 4, where Blocks 1 and 2 averaged around the same percentage (37%) while Blocks 3 and 4 averaged 50%. For both conditions, as Rat F4 was exposed to more trials, performance improved. That is, Rat F4 responded correctly on more trials at the end of the probe session than in the beginning for both the Delius-like and Vaughan-like procedures.

Figure 13 shows control data for Rat H5. The left bar shows percent correct for probe trials, the middle bar shows percent correct on probe trials using blind experimenters and the last bar indicates percent correct on probe trials when neither the S+ or S- stimulus was baited with pellet reinforcement. The dotted line indicates chance levels of performance. The vertical axis depicts the average percent correct for each of the control conditions. The first bar indicates when the subject was run under the Delius-like condition in which only one experimenter ran each probe session. As noted previously, overall probe performance was significantly above chance (118/162; $p < .01$). Once Rat H5's performance began to improve to above chance levels, control procedures were implemented. The two control procedures were done to assess whether the behavior or high performance was maintained through the

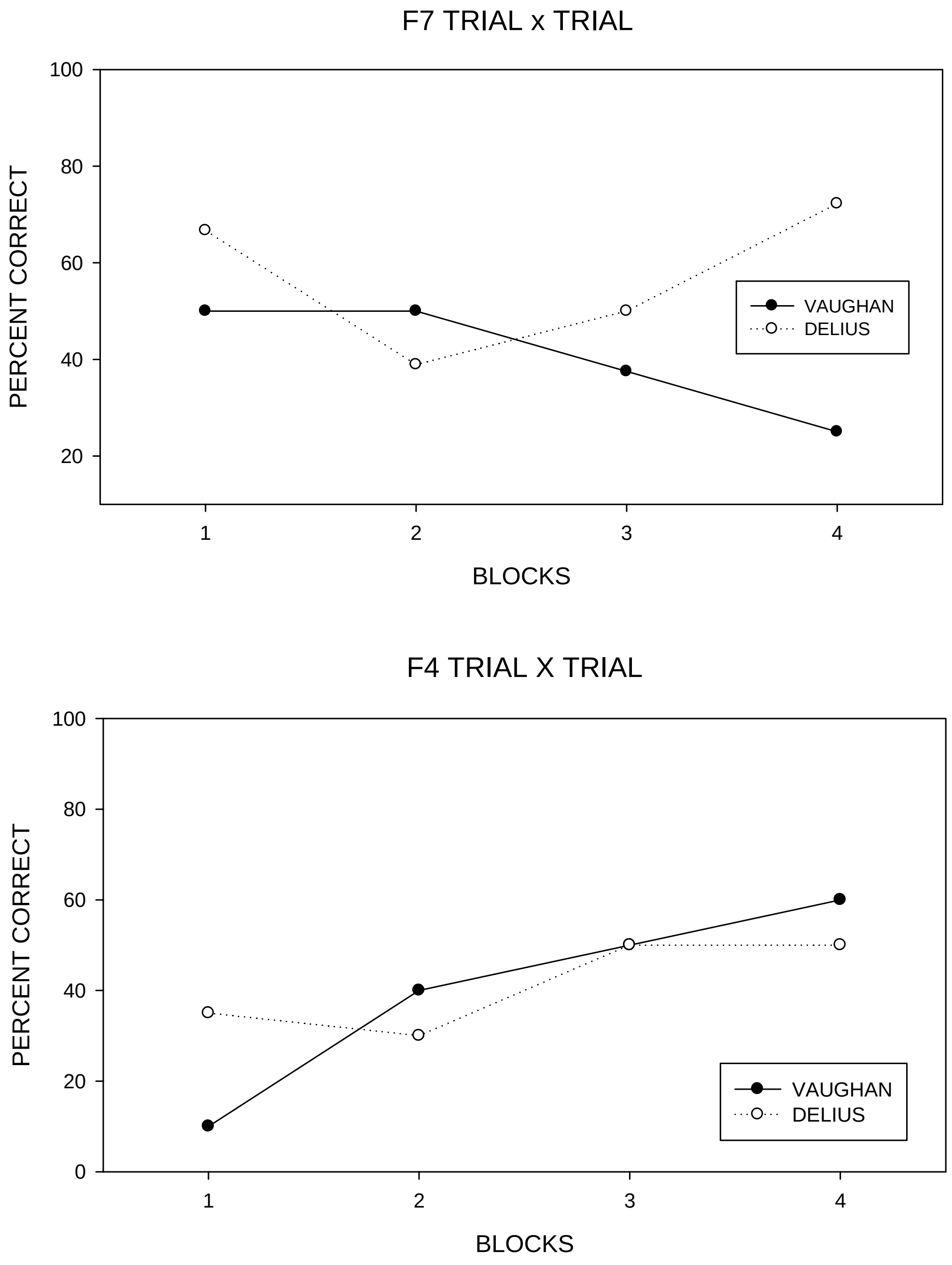


Figure 12

Each graph depicts the trial by trial data (the average number of correct responses to the first eight trials on the first day of each reversal (probe trials)) for Rats F7 and F4. Each block represents two trials. The solid black circles indicate probe trials under the Vaughan-like procedure. The open white circles indicate probe trials given under the Delius-like procedure.

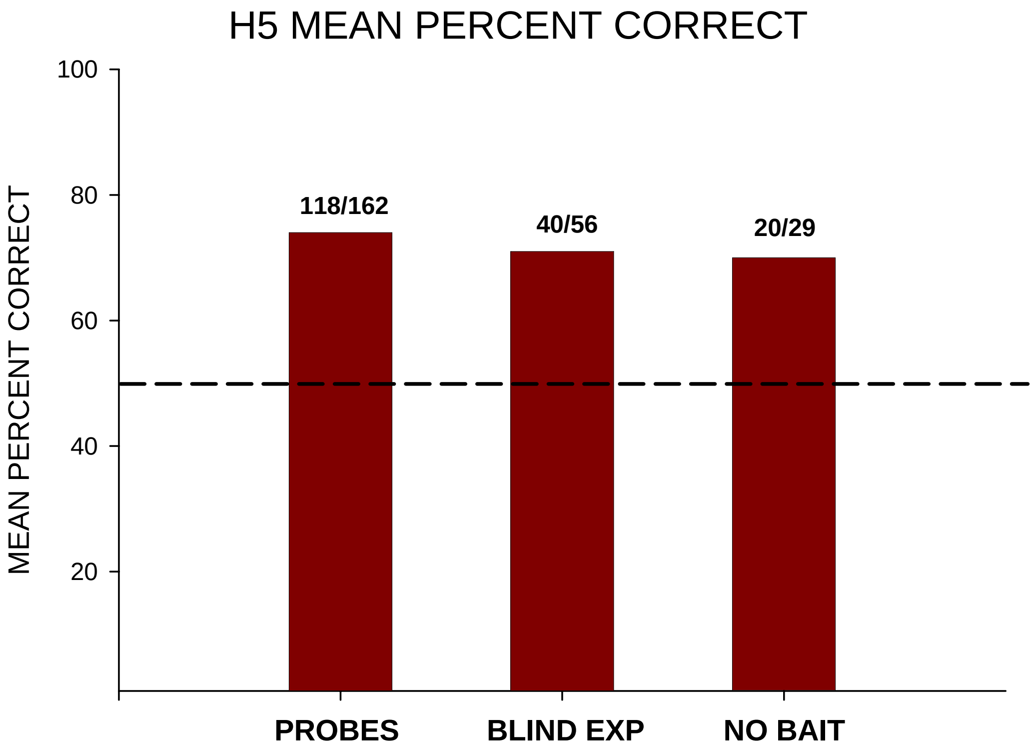


Figure 13

The mean percent correct on control trials for probe data, blind experimenters and pellet detection on probe trials is depicted here. Numbers above the bars indicate the number of correct trials compared to the overall number of trials given. The dotted horizontal line indicates chance levels of performance.

equivalence relation or through other means of control such as cuing from the experimenter or the scent of the pellet. Probe sessions conducted by an experimenter who was blind to which stimulus was S+ remained significantly above chance (40/56; p < .01). The last bar indicates the percent of correct responses when neither the S+ or S- stimuli were baited in order to evaluate the possibility of the smell of reinforcement as a cue. If trials in which neither stimulus cup was baited with reinforcement were at chance levels, it would suggest that control of responding was due to the scent of the pellet and not due to the discrimination of stimuli and emergence of functional equivalence classes. Performance on these unbaited trials was significantly above chance (20/29; p < .05). Overall, Rat H5 remained above chance on all three conditions indicating control by the equivalence relation.

Figure 14 shows the baseline probe data for Rats H5, F8 and F7. During the sessions in which subjects were exposed to all ten stimuli and were maintaining high rates of accuracy for a given reversal, some trials were designated such that neither stimulus was baited with reinforcement. Rat H5 was exposed to 40 trials in which no stimulus was baited and responded correctly to 38 of them. Rat F8 was exposed to 62 trials in which he responded correctly to 57 of them and Rat F7 was exposed to 17 trials in which he responded correctly to 14 of them. All subjects maintained high levels of accuracy during these trials, indicating that control over performance was not maintained by food pellet detection. Rat F4 was not given any trials in which neither stimulus was baited on account of his poor performance throughout the study.

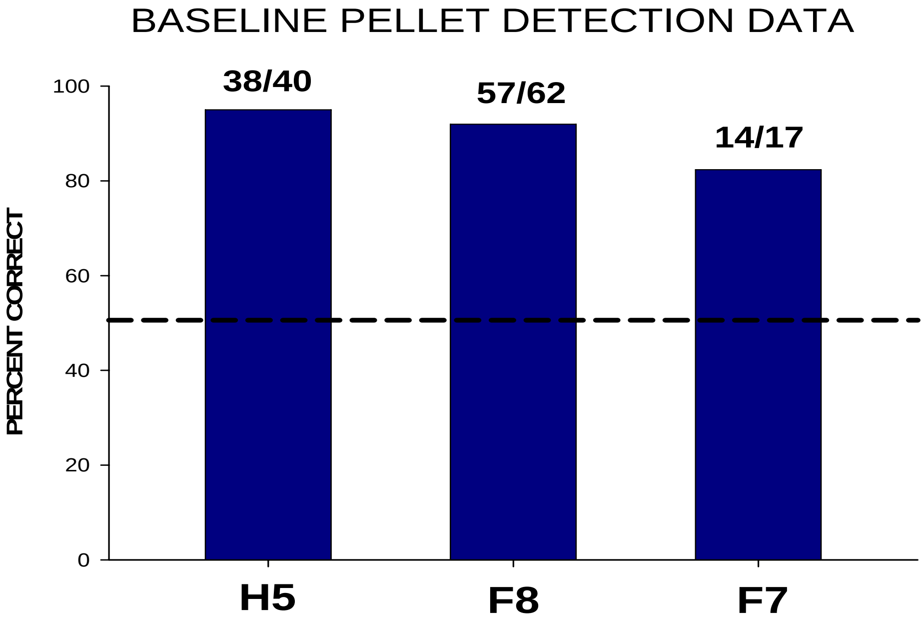


Figure 14

The Baseline Pellet Detection data for Rats H5, F8 and F7 are represented here. Numbers above the bars indicate the amount of correct responses compared to overall responses. The dotted horizontal line indicates chance levels of performance.

DISCUSSION

The first question was whether rats would show any improvement (savings) as they were exposed to more reversals. If the subjects did not show any improvement from reversal to reversal, they would not have been expected to perform well on the probe data or show any evidence of stimulus class formation. Requiring fewer sessions to reach criterion seems to be a necessary step in the acquisition of functional class formation. The results of the session by session data show that, although rates of improvement varied across subjects, every subject (Rats H5, F8, F7, & F4) showed improvement and required fewer sessions to reach criterion as the number of reversals increased. For all rats, acquisition was slow for the first reversal, but subsequent reversals showed a decrease in the number of sessions required to reach criterion.

When analyzing the probe data for Rat H5, performance is shown to improve as a function of the number of reversals given. During the first probe session, H5 responded incorrectly to all eight probe trials, which is indicative of control by the previous contingency; by the function most recently assigned to each stimulus and with which the subject had the most recent history. The next four probe sessions showed performance at chance levels, indicating a shift in control from the most recent history with each stimulus to a response to the first stimulus the subjects were exposed to (random selection). That is, immediately upon encountering a reversed contingency stimulus, control to all stimuli broke down. When switched to the Delius-like procedure, (reversals 6 - 26) performance on probes was consistently above chance, indicating another shift, this time from random selection to control by the equivalence relation. The subject was responding to stimuli that had most recently been incorrect based on other stimuli within that set whose function had been shown to transfer.

The control data for Rat H5 showed that when the subject was tested by a blind experimenter and when neither the S+ nor the S- stimulus was baited with reinforcement, above chance performance was maintained. If performance had declined during the sessions in which a different experimenter was running the session, control of performance might be attributed to cuing by the original experimenter. If trials in which neither stimulus was baited resulted in a decline in performance, it would remain unclear whether performance was due to the equivalence relation or to food pellet detection. Because there was no change in performance during the control trials, the high performance can be attributed to control by class membership and not by either food pellet detection or control by cuing. In conclusion, Rat H5 provided unequivocal evidence of functional equivalence.

The initial probe data for Rat F8 showed some evidence for functional equivalence indicated by above chance performance under the Delius-like procedure, but when switched to the Extended Delius Criterion, no significant above chance probe performance was observed. When analyzing probe data by combining both the Delius-like and the Extended Delius Criterion procedure, overall results showed significantly above chance performance, indicating control by the equivalence relation. However, clearly this was due to the high early performances and unlike H5, accurate probe performances were not maintained by F8. The decline in performance on the final procedure indicates a loss of control by the equivalence relation and could have been due to old age or a decline in olfactory sensory perception. Control procedures were also implemented for Rat F8 in the form of trials in which neither the S+ nor the S- stimulus was baited. During these trials, performance remained consistent with trials in which only the S+ stimulus was baited, indicating a lack of control by pellet detection.

During probe sessions, neither Arena subject (F4 and F7) reached significant above chance probe performance under any of the three procedures, indicating a lack of evidence for functional equivalence. Specifically, Rat F4 maintained chance levels of performance throughout the experiment and showed no signs of class formation even though session by session data show a reduction in the number of sessions to reach criterion. Trial by trial data showed improvement in performance on probe sessions as a function of the number of trials presented during those sessions and this might be taken as support for class formation. Rat F7 showed improved probe performance during the last procedure (Extended Delius Criterion), although accuracy did not reach statistical significance. Rat F7 is still being tested in order to assess the improvement during this phase.

When analyzing the differences seen in each subject's session by session, probe and trial by trial data, three main outcomes were observed. The first outcome from Rat H5 indicates strong evidence for functional equivalence seen in the significant above chance probe performance. The second outcome includes two subjects (Rats F8 and F7) in which the probe data show some evidence for class formation, but the evidence was weak due to the variable nature of the probe performance. The last outcome shown for Rat F4 depicts a failure to find evidence for equivalence due to consistent probe performance at chance levels.

It is possible that the different apparatus used to assess the acquisition of equivalence might have played a part in differences in performance throughout the experiment. For example, due to the fact that the Two-Choice apparatus had only two places for the stimuli to be inserted into the chamber, subjects might have had more exposure to both the S+ and S- stimuli during each trial compared to the Odor Arena where the stimuli could be placed too far apart to be exposed to both. The relatively greater success in the Two-Choice may be due to this exposure. Also, because of the spatial difference in the placement of stimuli within both apparatus, subjects might have responded to the incorrect stimulus in the Odor Arena simply because no other option was available near to the subject. Another possible reason why differences in performance were observed for each subject might be the difference in how the response to the stimulus was made. Subjects were required to flip a lid in the Odor Arena before digging in the sand to obtain reinforcement whereas in the Two-Choice apparatus, subjects were only required to dig. For a number of trials in the Odor Arena, subjects were observed to flip the S- stimulus and not dig in the sand and these trials were counted as incorrect. Other factors with the Odor Arena were that the subjects were transferred to a new Arena which might have disrupted performance due to the fact that the subjects had to become acquainted with the new apparatus before resuming their normal behavior of flipping and digging in the sand.

An interesting pattern of acquisition was seen for Rat F7 during consecutive reversals in which even reversals required fewer sessions to reach criterion than odd reversals and can possibly be accounted for by the differential reinforcement that was given during each reversal. During the odd reversals, the stimuli from set two were reinforced with a grain pellet and the stimuli from set one were under extinction. During the even reversals, set one was designated as positive and the stimuli from that set were reinforced with sucrose pellets. It is possible that faster acquisition rates during even reversals were a product of the sucrose reinforcement given to the stimuli within set one.

Although these differences can account for some of the discrepancy between data outcomes, results for individual rats were still variable for the same apparatus. Rat F8 in the Two-Choice resembled Rat F7 in the Odor Arena more than Rat H5 in the Two-Choice. Therefore, the differences in the apparatus cannot solely account for the differences in performance for each subject.

The results of the present study can be compared to previous experiments assessing functional equivalence. For example, the results of Dube et al.'s (1993) study were very similar to the findings of the present experiment. As in Dube et al.'s study, not all of the subjects provided strong evidence for functional equivalence. One of the subjects in Dube et al.'s study was maintaining above chance performance, but began to decline as more reversals were given and accuracy dropped to chance levels. One subject in the present study (Rat F7) showed exactly the same pattern of behavior as the subject in Dube et al.'s study. Another subject in Dube et al.'s study stopped responding during the first reversal and had to be dropped from the study. Although no rats were dropped from the present study, one subject (Rat F4) showed no significant above chance probe performance and was therefore retired earlier than the other rats. The variability in outcomes observed for each subject in the present study is strikingly similar to Dube et al.'s findings (one subject showed convincing evidence, two subjects showed some evidence but it was weak, and one subject failed to show any evidence). Dube et al. also used two different apparatus to assess the differences in acquisition of the functional classes based on initially training stimulus discriminations using either one or two levers. A similar procedural approach was taken in the present study such that two apparatus were used to observe whether one might generate faster rates of acquisition than another. Similarly, both studies showed variable results, requiring further investigation on the matter.

There were also some differences between the present study and Dube et al.'s study. In Dube et al.'s (1993) study, where the subjects were rats and different sets were differentially reinforced, only three pairs of stimuli were used compared to the present study in which ten pairs were used. Another difference in the two studies was that Dube et al. used successive discriminations with auditory stimuli in which tones were presented and the subjects had to respond by pressing a lever only to the tones designated as positive for that reversal. The reasoning behind using fewer stimuli within a set was in response to Vaughan's (1988) study using 20 stimuli per set. Dube et al. were interested to observe whether smaller stimulus sets produced a higher rate of acquisition on successive reversals to the point where probe tests for functional equivalence was possible. Interestingly, both Delius et al. and the present study also argued in favor of smaller stimulus sets compared to Vaughan's study. The results of their study showed that smaller sets might have contributed to the success of the study because the subjects were able to transfer functions seen from one stimulus to others within the set without having to be explicitly reinforced for doing so. Because auditory stimuli were used, acquisition was slower than in the present study; however, more trials per session were given and the subjects eventually showed significant above chance probe performance.

The pattern of acquisition and improvement across sessions seen in Rat H5's data is consistent with Vaughan's findings with his pigeons; however, the subjects in the present experiment never reached criterion as quickly as Vaughan's subjects. One possible explanation for this difference is the number of reversals given to each of the subjects. The typical rat has an average of approximately two years whereas the lifespan of a pigeon can reach up to 16 years or more. Therefore, the amount of reversals we were able to run was limited compared to the pigeons in Vaughan's study. Vaughan was able to complete hundreds of reversals and obtained optimal performance during the final stages of the experiment. Vaughan was also able to use 20 stimuli per set and assess the differences in performance based on the serial position of the trial. Trials near the end of the first presentation of the 20 stimuli within each set show better performance than trials in the beginning, which is similar to the present experiment's results for the trial by trial probe data using ten stimuli per set for some rats.

One of the main differences between the present experiment and Delius et al.'s study was that Delius et al. used pigeons and only two stimuli per set. Because only two stimulus pairs were used, only one pair was tested during probe trials for each reversal and therefore the subjects had a 50% chance of responding correctly and probes per reversal were either at 0 or 100%. Although Delius et al. were able to generate a larger number of reversals due to having only to learn a single discrimination, more probe trials were needed to attain significantly above chance performance due to the 50% chance that the subject would respond correctly to the probe trial. Because the present study used ten pairs of stimuli instead of only two, fewer reversals were generated; however, probe sessions allotted eight trials in which functional class formation was assessed. Using eight trials in which probes were assessed moved chance to requiring four correct responses out of eight and creating a possible probe performance to be anywhere from 0/8 to 8/8.

Similar to Vaughan's study with pigeons, Kastak et al. (2001) were able to demonstrate evidence for functional equivalence in their two subjects (sea lions) in part because of the fact that sea lions can live over 20 years and during that time the subjects were run under multiple experiments that gave them practice on conditional discriminations. Subjects showed rapid rates of acquisition during reversals and significant above chance performance on probe trials.

Although one of the subjects in the present experiment provided strong evidence of class membership, we were unable to transfer the subject to a conditional discrimination procedure in order to test for reflexivity, symmetry and transitivity which Sidman (1998) argued were necessary tests for the equivalence relation. Unfortunately, the amount of stimuli used in the present experiment only allowed for a small number of reversals.

The present outcome showed that functional equivalence can be demonstrated in rats. However, based on the variable data observed during the present experiment, future research may need to generate more reversals in order to achieve optimal performance (two sessions at 90% or higher) before moving to probe sessions. Because the most promising evidence for functional equivalence was observed for one subject who was run in the Two-Choice apparatus, more subjects should be run in this apparatus in order to assess the importance of the apparatus to performance acquisition. The lack of a sufficient amount of reversals to achieve optimal performance in the present study indicates a need to reduce the amount of stimuli within each set from ten to five or seven. Given that rats usually live up to two years in the laboratory, fewer stimuli within each set allows for more rapid acquisition of each reversed contingency and more chances to observe the acquisition of class formation during probe trials. The use of olfactory stimuli in the present experiment seemed to improve the rate of acquisition of the discrimination of stimuli between sets and on successive reversals and more research needs to be done using the best modality of perception for which ever subject is chosen for a given study.

REFERENCES

Bunsey, M. & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, 379, 255-257.

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

Deacon, T. W. (1997). The Symbolic Species: The Co-evolution of Language and the Brain. New York, NY

Dube, W. V., Callahan, T. D., & McIlvane, W. J. (1993). Serial reversals of concurrent auditory discriminations in rats. *Psychological Record*, 43 (3), 429-440.

Dube, W. V. & McIlvane, W. J. (1993). A search for stimulus equivalence in nonverbal organisms. *Psychological Record*, 43 (4), 761-774.

Dugdale, N. & Lowe, F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior,* 73, 5-22.

Galizio, M., Miller, L., Ferguson, A., McKinney, P. & Pitts, R. (2006). Olfactory repeated discrimination reversal in rats: effects of chlordiazepoxide, dizocilpine and morphine. *Behavioral Neuroscience*, 120 (5), 1175-1179

Horne, P. J. & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 185-241.

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.

Jitsumori, M., Siemann, M., Lehr, M & Delius, J. D. (2002). A new approach to the formation of equivalence classes in pigeons. *Journal of the Experimental Analysis of Behavior*, 78, 397-408.

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.

Keller, F. S. & Schoenfeld, W. N. (1950). *Principles of Psychology*. New York: Appleton-Century Crofts.

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

Pena, T., Pitts, R. C. & Galizio, M. (2006). Identity matching-to-sample with olfactory stimuli in rats. *Journal of the Experimental Analysis of Behavior*, 85, 203-221.

Retrieved February 12, 2007, from the internet database. http://www.graphpad.com/quickcalcs/binomial1.cfm

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 Behavior*, 55, 1087-1095.

Sidman, M. (1994). Equivalence Relations and Behavior: A Research Story. Boston: Authors Cooperative.

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

Sidman, M., Rauzin, Lazar, Cunningham, Tailby & Carrigan (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.

Tomonaga, M. (1999). Establishing functional classes in a chimpanzee (Pan Troglodytes) with a two-item sequential responding procedure. *Journal of the Experimental Analysis of Behavior*, 72, 57-79.

Zald, D. H. & Rauch, S. L. (2006). The Orbital Cortex. In J. L. Price (Ed.), Connections of Orbital Cortex (p. 43). Oxford University Press.

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.