

The question of whether non-humans are capable of conceptual behavior has been of great interest to scientists. While there are several definitions used to identify a concept, the following put forth by Schwartz, Wasserman, and Robbins (2001) seems most appropriate, "The transfer of responding from familiar to novel stimuli is considered by most theorists to be an empirical hallmark of conceptualization" (p. 299). This definition suggests that it is the ability to generalize what was learned to a completely new situation that is the key to demonstrating concepts. For example, a child is said to have learned the concept of shoe if when presented with a shoe she has never seen before (and which is physically dissimilar from those previously experienced), she still identifies it as such. Do other species of animals possess the ability to form concepts, just as the child did in identifying the novel shoe? And, are there differences in learning ability, cognitive processes, and intelligence among different species of animals (Wilson, Mackintosh, and Boakes, 1985)? One of the most basic conceptual behaviors is the ability to respond to the relation of identity or difference between two stimuli. One method developed to study same-different relations in non-humans is the match-to-sample procedure (MTS). As D'Amato, Salmon, and Colombo (1985) ask: "When an animal learns an (identity) matching-to-sample task, does it acquire a matching concept that transcends the stimuli used in original training, or does it learn something much more restricted, a set of specific "rules" that apply only to the training cues and to closely similar stimuli" (p. 35)? The question can be framed in terms of the nature of the specific stimulus control topography involved: is the behavior controlled by the relation of physical identity between sample and comparison, or are less general, multiple topographies involved?

The Match-to-Sample Procedure

In a MTS procedure, a typical trial begins with the presentation of a stimulus to which the subject is required to make some form of orienting response (such as a rat poking a lighted stimulus with its nose). This stimulus is commonly referred to as the sample. Once the subject has made an orienting response to the sample, two or more discriminative stimuli appear that are often referred to as comparisons. An example of an identity MTS trial would be the presentation of a blue square as the sample and a blue square and yellow circle as comparisons, in which choosing the blue square would result in reinforcement and choosing the yellow circle would not.

Successful matching-to-sample in early studies raised several key issues. Were subjects learning the concept of identity, a general matching concept, or were there other solutions? Cumming and Berryman (1961) were among the first to argue that non-humans (specifically pigeons) might not learn the concepts of "identity" or "same" during MTS training, but rather a set of complex rules (or "if...then" rules). To test this, the experimenters used an identity MTS procedure in which pigeons were trained to peck lighted keys. Three different hues were selected for the matching procedure: red, green, and blue. A trial began with a red light (for example) appearing on the center sample key while the other two keys remained dark. An observing response (pecking the center sample key) produced the comparison stimuli on the other two keys (a red light and a blue light, for example). A peck to the comparison identical to the sample resulted in 3 s access to grain. A peck to the incorrect comparison darkened the keys and produced a 3 s timeout. Pigeons were able to match correctly on close to 100% of the trials after only six to nine sessions. Thus, when the sample was blue they pecked the blue comparison, and when the sample was red they pecked the red comparison. After 22 sessions

the experimenters replaced the familiar blue hue with a novel yellow hue as the sample and comparison. What Cumming and Berryman discovered was that when yellow was the incorrect comparison, accuracy remained at criterion levels (90% or higher); however, when the novel yellow stimulus was the sample and the correct comparison, accuracy fell to approximately chance levels (50%) and the subjects exhibited position preferences. Thus, the pigeons did not generalize matching to the new stimulus, but instead their behavior seemed to be under the control of specific stimulus compounds or some other form of relatively specific stimulus control. As an example, Cumming and Berryman (1961) suggested that responding could be accounted for by compounds such as "when the center key is red and the left key is red, peck left" and "when red is the center key and red is on the right, then peck right". Thus, selection was apparently not controlled by general sample-comparison relation, but rather by specific pairings of stimuli. Generalized matching would have occurred if the subjects successfully matched novel stimuli (e.g., selecting the novel yellow comparison when given a yellow light as the sample and yellow and blue lights as comparisons).

Cumming and Berryman also studied pigeons' performances with an Oddity-from-Sample (OFS) procedure because it was noticed that pigeons in the early stages of training almost always pecked the odd comparison. An example of an OFS trial would be the presentation of a blue square as the sample and a blue square and yellow circle as comparisons, where choosing the yellow circle would result in reinforcement. Cumming and Berryman (1965) hypothesized that because the animals begin each trial with several pecks to the sample stimulus, the responses to the sample were extinguished because no direct reinforcement was given. Thus, the pigeons had a lower probability of pecking at that hue when it appeared as one

of the comparisons. If this were the case then, according to Cumming and Berryman, an OFS procedure should produce much higher rates of acquisition.

Acquisition in an oddity-from-sample procedure can be accounted for in three ways. One possibility is that the subject may simply learn to respond to particular configurations of stimuli. For example, they may learn "when red is the center color don't pick red when it is a side color". Alternatively, they may learn "when red is the center key, peck green on the side key, or if green is the sample, peck blue on the side key". The third option is that subjects actually learn to pick stimuli based on the "different" relation. To test these possibilities Cumming and Berryman (1965) used the same hues as before (red, green, and blue); however, after 20 sessions the blue hue was replaced with a yellow hue. If the animals had learned specific configurations when green was presented as a sample and the novel yellow and familiar green were presented as a comparisons, they should perform at chance levels. Their performances should be at chance because presenting green as a sample was a cue for choosing either red or blue and neither was available. However, if the subjects had learned to peck the hue that was different from the sample there should have been no change in performance because they had already learned that when green was the sample, avoid pecking green as the comparison. The experimenters found that on the first session in which the yellow hue was present as a comparison and either red or green was the sample, accuracy dropped to levels that matched performances on the first day of training. Just as in their 1961 experiment, Cumming and Berryman found that the concept of "different" was not controlling behavior. Instead, they argued that subjects were operating under two sets of rules, if "red, pick blue or green, whichever is available" and "if blue, peck green or red", whichever is available. Therefore subjects did not show evidence of generalized control by

the difference relation (i.e., they did not choose the comparison that was different from the sample when the sample was novel).

Cumming and Berryman (1961; 1965) were among the first to argue that successful performances on MTS and OFS training tasks by non-humans might not require abstract conceptual behavior. Before these studies it was generally believed that accurate performances with a few trained stimuli were sufficient to demonstrate that subjects could learn the concepts of identity or difference. Had this been the case, when subjects were exposed to any situation in which stimuli not involved in training were used, their performances should have been well above chance. However, as Cumming and Berryman found, when pigeons were given new stimuli, their poor performances revealed that they had not learned generalized MTS.

Carter and Werner (1978) further elaborated this problem with their discussion of three different models of conditional discrimination learning, the configuration model, the multiple-rule model, and the single-rule model. Their configuration model is similar to that proposed by Cumming and Berryman and holds that the animal's behavior is controlled by the "stimulus situation" or specific configuration of the stimuli. For example, the subject might learn to peck the right key when red is in the center and green is on the left and red is on the right, and to peck the left key when green is in the center, and green is on the left and red is on the right. The subject is not learning to peck the key that is physically identical to the sample, but rather learning particular responses to particular configurations of stimuli. Thus, the subject would be unable to generalize to new stimuli because the configurations would be unfamiliar. For example, if the subject was given a novel trial in which a yellow sample key was presented in the center and yellow was on the right and red was on the left, the subject would be unable to match

the proper stimuli using the configuration model because it had never learned that specific configuration.

Another possible solution to the MTS contingency is the multiple-rule model, which holds that something other than configurations of stimuli are controlling the responding of subjects. Just as Cumming and Berryman (1965) suggested, Carter and Werner argue that subjects might be conforming to "if...then..." rules. There are two components to this rule: (1) there are specific "if... then..." rules that specify which stimulus is correct and (2) there are specific "if...then..." rules that specify which stimulus is incorrect. According to the configuration model, the subject's key peck is controlled by the specific arrangement in which all of the stimuli in a given trial are presented, whereas the multiple-rule model holds that the subjects learn to peck a specific key color only when it is presented in conjunction with a specific sample.

Finally, Carter and Werner argued that generalized matching could only be invoked when it is demonstrated that a single "rule", the identity relation between sample and comparison, controls the animal's behavior. The issue involved in such a single-rule model is that subjects will be able to continually solve novel problems as long as the novel problems adhere to that specific single rule. For example, if subjects have learned the single rule of matching-to-sample they would then be able to generalize this rule when novel stimuli were presented. In their review of the literature, Carter and Werner concluded that there was no convincing evidence for single rule learning in pigeons and argue that perhaps pigeons may learn to discriminate stimuli based on one or more processes depending on the specific stimulus arrangement.

Cumming and Berryman (1961; 1965) and Carter and Werner (1978) showed that extensive MTS training with a small number of stimuli resulted in configuration or multiple-rule

learning in pigeons, rather than generalized matching. Experimenters reasoned that by training with different stimuli on each trial of a session subjects might be more likely to learn generalized matching (e.g., Mishkin and Delacour, 1975; Overman and Doty, 1980; Wright, Cook, Rivera, Sands, & Delius, 1988). This strategy, often referred to as a "trial-unique" procedure or multiple exemplar training, has been successfully employed by a number of researchers.

Mishkin and Delacour (1975) tested 16 rhesus monkeys that were divided into two groups of eight. The experimental group received different pairs of stimuli on each trial (which consisted of 200 small three-dimensional items that varied in color, size, and shape), whereas the control group received the same pair of stimuli on every trial (each object appeared as the sample on half of all trials). The animals were next divided into 2 more groups with one group participating in a match-to-sample procedure (MTS) and the other a non-matching (or oddity) from sample procedure (NMTS). It was found that the group with the fastest rate of acquisition was the experimental NMTS group (with criterion set at two sessions with accuracy at 90% or higher for a maximum of 1,000 trials). In contrast, both control groups showed very slow learning of the baseline. Three animals within the control groups failed to meet criterion even after 1,000 trials (50 sessions), while on average the experimental NMTS group met criterion after only 90 trials (4.5 sessions). This study and subsequent studies (Overman and Doty, 1980) showed that multiple exemplar training could enhance the likelihood of generalized MTS in primates

Because the trial-unique procedure resulted in rapid acquisition in Mishkin and Delacour's rhesus monkeys, Overman and Doty (1980) were interested in whether a trial-unique procedure would facilitate macaques' performances when the delay between sample presentation and comparison presentation was more than a few seconds. Six animals were trained in a sound-

reducing chamber, which was equipped with three rear projection panels. One hundred different picture stimuli were reshuffled each day for a total of 50 pairs and consisted of such items as a shoe, a screwdriver, and a coffee mug. The animals were trained to press the center sample image nine times to receive reinforcement (orange juice). A zero-delay procedure was implemented such that as soon as reinforcement was delivered, the sample disappeared from the center and the matching and non-matching comparisons were displayed on the side panels. Experiment 1 tested the animals' performances after delays that ranged from 5s to 24h with the stimuli used in training but with a trial-unique procedure. More importantly, Experiment 2 tested the animals' ability to match 180 novel stimuli (at varying delays) with each stimulus appearing only once as either a sample or comparison. These novel stimuli were intermixed with familiar stimuli used in Experiment 1. Performances at the shortest delay tested (30 s) averaged above 90 % correct providing a strong demonstration of generalizing matching.

Dube, McIlvane, and Green (1992) noted that there is more than one way that subjects can correctly respond in a trial-unique procedure. It is possible that subjects may choose the correct stimulus by rejecting the comparison that is least familiar. So, for example, if the subject is given a square as a sample and a novel circle and the matching square (which is now no longer novel because it was just seen as a sample) as comparisons, the subject may solve the problem by simply rejecting the comparison that is novel. Dube et al. (1992) also argue that simply giving a probe or test trial with novel stimuli intermixed with those already trained (or baseline stimuli) may not be enough to demonstrate generalized match-to-sample. Subjects may solve the novel problem by rejecting the baseline stimulus for the comparison that is novel. For example, suppose the subject's baseline consisted of squares and circles and the animal is given a probe trial in which the sample is a novel triangle and the comparisons are a square (that was seen in

the baseline) and the matching triangle. The subject may reject the baseline comparison, as opposed to choosing the triangle because of its relation to the sample.

It seems that in order to address these problems, three probe types must be administered to test for generalized identity MTS. The first is a test in which the subject is presented with a novel sample, the matching comparison, and another novel comparison (novel-novel probe type). This type of probe will test whether the animal matches in the absence of any prior experience with the sample or comparison. Because rejecting the novel comparison is a possible solution for this type of problem, additional probe types are needed. A second probe type would be a test in which the sample is a novel stimulus and the comparisons are the matching novel comparison and a familiar baseline comparison (novel-familiar probe type). This probe will help establish whether the animal is solving the problem by rejecting novelty. If the animal were rejecting novelty to solve the MTS problem, he would perform inaccurately on these probe types. In a third probe type the sample would be familiar and the comparisons would consist of the matching familiar comparison and a novel comparison (familiar-novel probe type). This type of probe tests whether the animal is rejecting the familiar comparison for the novel comparison. If the animal is able to perform successfully on each of these probe types, the argument that the subject has demonstrated generalized MTS would be strengthened.

The MTS procedure has been used to test the learning capabilities of many different species. Because there are many ways in which the concept of identity is studied, for the purposes of this discussion, the following review examined a number of experiments and the extent to which they addressed the issues set forth by Dube et. al (1992). More specifically, did each of the studies that demonstrated identity learning in their respective subjects administer several different probe types, for example, Novel-Novel, Novel/Familiar, and Familiar/Novel?

This review is by no means exhaustive; however, it will attempt to examine the relevant studies that claim to demonstrate the concept of generalized MTS among their respective subjects. This review will be organized in a phylogenetic fashion.

Non-human Primates

Non-human primates are thought most likely to demonstrate concept learning and therefore, it seems logical to examine these studies first. Nissen, Blum, and Blum (1948) were among the first to use an identity MTS procedure with chimpanzees. Nissen et. al trained seven chimpanzees initially with two stimuli (a cup and a box) using a Wisconsin General Testing Apparatus (WGTA). The sample object was placed over the center well of the WGTA and covered a piece of food. On either side of the sample appeared the comparison stimuli, with the matching comparison also covering a food item. If the animal chose the matching comparison it was allowed the food item. If the incorrect comparison was selected, the tray was immediately removed and after a delay the trial was presented again. Once the animals had met the training criterion of 10 out of 12 successive trials correct, they were exposed to 11 novel stimuli that consisted of common three-dimensional objects. For the initial days of testing the animals were exposed to 24 novel stimulus combinations in groups of 4 to 12 successive trials each. After each animal received 117 to 129 trials they were exposed to an additional 53 novel stimulus combinations.

Nissen et. al (1948) found that for the initial 24 novel combinations, the animals performed on average at 81.7% correct; and further, when exposed to the remaining 53 novel combinations, the animals performed on average at 90.3% correct. The high level of accuracy suggests that in the initial training something more than configural control had been learned. Nissen et. al demonstrated that introducing a large number of stimuli during training facilitated

the rate at which subjects learned; however, what is lacking in this study is the presentation of the animals' performances on the first trial of each novel combination (correct performances on these trials would help demonstrate that the animals were selecting the correct comparison because it was identical to the sample). Because these data were not presented, it is not possible to infer generalized identity matching.

Oden, Premack, and Thompson (1988) were interested in whether four infant chimpanzees could transfer what they learned in training to novel stimuli when trained with only 2 stimuli and no differential reinforcement from the experimenters. The stimuli used in training consisted of 60 ml measuring cups and brass plated sliding bolt locks. The trial began when the experimenter handed the animal the sample (either a measuring cup or a lock), which the animal was then required to place in a dish outside the testing apparatus. Next, the experimenter placed the matching comparison and the non-matching comparison in the apparatus with the animal. Placing the matching object in the dish with the sample resulted in reinforcement. Once the animal met the training criterion of 10 correct responses out of 12, he/she was given a transfer test. The test session consisted of 12 trials with two novel objects with non-differential reinforcement on each trial. There were three different transfer tests given, an object test, a fabric test, and a food test. Each of these tests presented novel stimuli to the subjects with the fabric and food tests being novel modalities as well. There were a total of six transfer tests given to each subject (two sessions each of objects, fabrics, and food). It was found that on average, the animals took 816 trials to meet the training criterion. On average, correct responding for object transfer was 85.4%, fabric transfer was 84.4% correct, and food transfer was 67.7% correct. The authors argue that the low performance for the food transfer test was due to specific preferences of food items. The experimenters were specifically interested in the chimpanzees'

performance on the first trial of a transfer test. They found that overall performances on Trial 1 of the transfer tests was 77% correct with two of the chimpanzees performing correctly on the first trial of all six transfer tests. Although overall the animals performed only at 77% correct, they were still performing significantly above chance. Thus, in this experiment, chimpanzees met all the criteria required to demonstrate generalized matching-to-sample.

Thompson, Oden, and Boysen (1997), tested whether language-naïve chimpanzees could match identical stimuli as well as judge relations between relations in a conceptual MTS task. The subjects included one juvenile and four adult chimpanzees trained in an apparatus that used a combination of a shelf and computer monitor used to display the stimuli. The samples were presented on a shelf on the experimenter's side of a Lexan window. Digitized pictures of the comparisons were presented on a color touch screen monitor. The animals' choices were recorded by touching one of the comparisons. Reinforcement of preferred candies was delivered via a plastic tube that projected through the window. The physical matching trials were differentially rewarded.

The samples were three-dimensional objects varying in color, size and shape. There were three objects used in the physical matching task for Problem 1, and nine novel objects were used for Problem 2. For Problem Set 1, there were 16 physical matching trials (trials in which the sample and correct comparison were paired based on their physical identity to each other) and 16 conceptual/relational trials per session (stimuli that were matched based on their relation to each other; i.e., tigers and leopards are both cats). Criterion was set at 75% correct for two consecutive sessions. The results of problem set one demonstrated that two animals met the criterion within two sessions. One animal met criterion in three sessions and another in four sessions. The fifth animal was unsuccessful in meeting criterion within four sessions. The

overall mean for the four animals that met criterion was 83.6% correct for the identity matching trials, thus demonstrating that they were matching stimuli based on their physical identity. The tests for problem set two were administered to ensure that the performances demonstrated on problem set one were not due to the particular stimulus sets used. The overall mean performance on physical matching in problem set two was 84.4% correct. For problem set three, individual objects were combined into novel combinations to test whether the animals had indeed learned generalized matching and were not matching based on paired associates. The overall mean performance for physical matching in problem set three was 87.5% correct. Once again, these data seem to suggest that chimpanzees were demonstrating generalized MTS. To ensure this, Thompson et. al also examined the animals' performance on the first trial with new stimuli and found that the overall trial-one performance for all identity matching was 83.3% correct.

Thompson et. al concluded that the chimpanzees were capable of matching items based on their on physical identity. The experimenters also argue that language is not a necessary component to matching since three of the four language naive animals were able to spontaneously match stimuli. Also important to the authors' claim of generalized matching in chimpanzees is the strong performance of the subjects on the first trial in which novel stimuli were introduced (overall performance of 83.3% correct). Thompson et. al (1997) demonstrated that by implementing several different probe types, as well as presenting Trial 1 data, chimpanzees showed evidence of generalized MTS.

Although there is convincing evidence for generalized matching in apes, studies of monkeys have been more controversial. The successful results of Mishkin and Delacour (1974) and Overman and Doty (1980) were noted above, but other studies have reported more limited evidence of generalized matching. For example, Jackson and Pogram (1970) studied the transfer

of color matching to form matching in rhesus monkeys. The subjects in the Jackson and Pogram (1970) study were presented hue stimuli on a panel that consisted of three response keys and a food hopper. The colored stimuli consisted of yellow, green, blue, white, and red lights and were presented in the form of a 1 in diameter circle. The form stimuli (which varied in shape) were white. The sample key was positioned in the center, 1 in above the two keys that presented the comparison stimuli. Monkeys were required to touch the sample to produce the two comparisons. Choosing the matching hue was reinforced with banana pellets and there was an intertrial interval of 2 s. However, if the subjects chose the incorrect non-matching comparison they received no reinforcement and the intertrial interval increased to 10 s. The subjects were required to meet a criterion of 90% accuracy or higher for three consecutive days. Once they had met this criterion they advanced to the form-matching stage.

Of particular interest in the form-matching phase of this experiment was whether the monkeys would be able to transfer what they had learned in the hue-matching phase. The stimuli in this phase were presented in the same manner as the hues. The stimuli consisted of a triangle, a square, an X, and a straight horizontal line. The monkeys were tested for 14 days and each session involved 200 trials. Jackson and Pogram found that when trained with hue stimuli, rhesus monkeys were not able to generalize matching to form stimuli. On average it took the subjects approximately 20 days to reach criterion in the hue-matching component. When advanced to the form-matching component the subjects failed to meet criterion in the 14 days of testing. While there was a learning curve, the subjects only averaged approximately 85% correct by the 14th day. In summary, rhesus monkeys in this study readily learned to match colors; however, they did not generalize matching to a new modality.

Demonstrations of generalized matching in new world monkeys have been problematic. For example, D'Amato, Salmon, and Colombo (1985) tested eight monkeys (*Cebus apella*) on an identity MTS procedure. Three different colored circles (red, yellow, and blue) as well as eight different forms were used as stimuli. Each monkey was trained to criterion with a 2-stimulus set, which varied among the animals. For example, two animals were trained with a circle and a dot, and two animals were trained with a triangle and a red disk. A single orienting response to the sample produced the two comparisons while the sample remained on the screen (simultaneous MTS). A response to the matching stimulus produced a reinforcer while an incorrect response resulted in a 30 s-60-s timeout in which the house light was dimmed. Once the animals met the criterion of two sessions at 90% accuracy or higher with their respective training stimuli, they were given transfer tests. The transfer sessions involved 48 trials of which half were training stimuli and the other half a new 2-stimulus set, all quasi-randomly arranged. The criterion in the transfer phase was set at 70.8% correct or 17 out of the 24 new stimuli correct for sessions two through four. Session 1 of the transfer phase was not included in the criterion because the authors argue that the animals' performances would be disrupted by the novelty of the stimuli. Those animals that failed to meet the criterion in Test 1 were given additional training sessions until they met the criterion limit set in initial training. Once they met criterion again, they were given another test (Test 2).

The experimenters found that only four of the eight monkeys were able to meet the criterion of 70.8% correct or higher in the first transfer test. Of the other four who were given a second test, only two met criterion (but one animal did so when given a third test). Interestingly, all animals except one showed a faster rate of acquisition during Test 2. What is impressive about the four animals that met transfer criterion is that they had only been exposed to two

stimuli before their first transfer test, and even the animal that performed the poorest had only been exposed to six stimuli prior to the third transfer test (in which she met criterion). What would have been of great interest in this study is the animals' performances on the first trial of each transfer test. Without this information it is unclear whether the performances were based on generalized matching, control by specific stimulus configurations, or some type of multiple rule learning.

da Silva Barros, de Faria Galvão, and McIlvane (2002) conducted additional studies of identity MTS in *Cebus apella*. Sample and comparison stimuli were presented in varying locations on a touch screen computer. There were five stimulus sets used. One set was used in training and involved three different shapes. The remaining four sets were used to test for generalized matching. Three of the four testing sets were colors and the fourth consisted of three different shapes. For the training trials, a touch to the sample was followed by the presentation of three comparison stimuli. The sample and the comparisons were presented in a quasi-random order in various locations. When training criteria were met, 12 trials from one of the novel stimulus sets were mixed within 36 baseline or training trials. After the monkeys met a criterion of 18 correct consecutive trials (out of the 48 different trial types), another novel stimulus set was introduced. Results indicated that both monkeys were able to successfully perform near 100% accuracy on the initial test session with some, but not all, of the novel sets. In general, stimulus sets with which the animal had a prior history involving simple discriminations were the ones associated with the strongest performances. Although the monkeys were not able to successfully match all of the novel tests, their high performances provide strong evidence for generalized MTS.

It is evident that procedures of training and testing generalized MTS have varied widely in non-human primates making general conclusions difficult. However, there has been at least one clear demonstration of generalized matching in chimpanzees (Oden et. al, 1988) that met all the criteria set by Dube et. al (1992) and multiple exemplar training procedures have been used with success in several studies with old world monkeys (Mishkin and Delacour, 1975; Overman and Doty, 1980). Difficulties encountered in studies with new world monkeys seem to have been overcome by da Silva Barros et al. (2002). Thus, it would appear that, given the appropriate training and testing conditions, both old and new world primates can acquire the concept of same/different as defined by a generalized matching performance. Other studies suggest that non-primates may also demonstrate generalized matching given the appropriate training techniques.

Marine Mammals

Herman and Gordon (1974) were interested in whether a bottlenose dolphin would be capable of demonstrating generalized matching using auditory stimuli. The subject was tested in its home tank with specialized speakers that were lowered into the tank during testing sessions. There was a start paddle that was placed in the center of a channel, which was designated as the listening area with all of the speakers pointing to this area. At 45 degrees to the left and right of the start paddle were the two stimulus speakers. In front of each of the stimulus speakers were response paddles. The animal was initially trained on a two-sound discrimination with responses to S+ being reinforced and no reinforcement for responses to S-. Each trial began with a tone projected from a control speaker that lasted a maximum of 10 s. The animal was required to enter the channel and press the start paddle to turn off this sound. Next, the sample sound was projected simultaneously from both stimulus speakers and after a short delay the two comparison

sounds were projected sequentially from either speaker. If the animal pressed the paddle that corresponded with the source of the correct comparison, a fish was delivered. If the animal's choice did not match the sample tone, the playback was followed by a 7s inter-trial interval. There were a total of 934 training trials given over a nine day period with each session (approximately 50 trials long) being unique in that no two stimulus sounds were ever paired together more than once within a given session. Once the animal met the training criterion there were 346 novel matching problems (consisting of 5-10 trials each) given over a four-week period. Throughout the four-week testing phase longer delays between sample presentation and comparison presentation were introduced. The delays between sample and comparison presentation ranged from 1 to 120 s.

Herman and Gordon (1974) reported that most Trial 1 errors (as well as regular trial errors) occurred within the first 121 problems. From problem 173 through problem 341 the dolphin produced no errors in matching for the first trial with a particular stimulus combination. The subject performed correctly on 5 out of 6 Trial 1 matches with a 120s delay. This study is important in that it supports the notion that different species may require very different methods to test for the same concepts. Another important issue involved with this study is the presentation of Trial 1 data. By presenting this information, it was evident that the subject generalized matching to new stimuli with almost perfect accuracy.

Kastak and Schusterman (2000) conducted a test of generalized identity matching in two California sea lions in a large outdoor pool. The testing apparatus had hinged boards containing windows in which stimuli were placed. The sample was placed in the center window and the comparisons were presented in the side windows. The sample was presented for an interval of 2-4 s before the comparisons were displayed. Once the comparisons were displayed the subject

was released from a housing station to make its selection. A response was defined as the subject breaking the plane formed by the stimulus box with its nose. If the subject made a correct response, a piece of fish was presented. They trained the animals on an identity match-to-sample procedure with the stimuli consisting of a mixture of abstract and concrete pictures. Once they had mastered matching with the training stimuli, the sea lions were tested for generalized match-to-sample with novel stimulus pairs. The critical stage of the transfer test was their performance on the first trial with novel stimuli.

The transfer procedure began with the presentation of a novel sample, which was always paired with the matching novel comparison and a comparison with which the animal already had a history. It was hypothesized that during this phase the sea lion's responses would tend to avoid the familiar comparisons as opposed to actually matching the sample and the correct comparison (Rio was the only subject to take part in this phase of the experiment). This subject's performance was at 90% accuracy when novel comparisons were the correct choices; however, her performance dropped to chance levels when both comparisons were novel. Indeed, the authors found that Rio was using the stimulus control topography of exclusion to solve the MTS problem (by avoiding the familiar stimulus).

The next phase of training was a trial and error phase in which 40 new stimuli were introduced and trained in pairs along with an existing baseline of 20 stimuli for a total of 60 stimuli. For example, two novel stimuli would be introduced in which one stimulus would be presented as a sample and as the matching comparison and the other stimulus would serve as the alternative incorrect comparison. The next phase of the procedure involved reshuffling the stimuli so that any stimulus could appear with any other stimulus either as the sample or a comparison in a match-to-sample trial. Kastak and Schusterman argued that reshuffling the

stimuli would aid in the acquisition of the MTS procedure because each trial was a new configuration. Thus, "dependence" on the training context would be eliminated so that the animals' response could only be due to the relationship between the sample and correct comparison.

One testing phase of the experiment involved 30 novel stimuli that were arranged into 15 different trials, which were introduced into the already established baseline. The next testing phase also introduced an additional 30 novel stimuli arranged into 15 different trials. Each new stimulus appeared twice as a correct comparison and twice as an incorrect comparison. The data collected from the transfer tests were analyzed in several ways. Of importance was the sea lions' performance on the first trial of the transfer tests in which the animals had a 50% chance of choosing the correct comparison. One subject (Rio) performed at above chance levels with novel stimuli (80% or better). The other's (Rocky) performance was not statistically better than chance with novel stimuli; however, she still averaged 70% correct. The experimenters argued that exposure to multiple exemplar training aided in this transfer by minimizing the likelihood of alternative forms of stimulus control. Also, by exposing the sea lions to many novel stimuli, their neophobic behaviors were extinguished. Thus, Kastak and Schusterman found that the sea lions were capable of transferring their MTS training to novel stimuli.

While the aforementioned studies have provided convincing demonstrations of generalized matching in primates and marine mammals, studies with other animals have been controversial.

Pigeons

Many researchers have continued to study identity matching in pigeons despite the problems noted by Cumming and Berryman (1961) and Carter and Werner (1978). Zentall and

Hogan (1974) used a group design with pigeons to determine whether generalized matching could be inferred in a transfer-of-training procedure. Eight pigeons were trained in a pigeon-testing chamber using two different colored lights (red and green). Half of the pigeons were trained to peck the color that matched the sample stimulus while the other half were trained to peck the stimulus that did not match the sample (odddity from sample task). After 20 sessions of training, both groups were switched from red and green stimuli to yellow and blue stimuli. Also, half of the matching group was shifted to the oddity task while half of the oddity group was shifted to the matching task. Zentall and Hogan found that in initial training the oddity group learned the task faster than the matching group. However, by the end of training there was no difference in performance between the two groups. When the birds were tested for transfer, it was found that over the first five sessions of transfer the non-shifted group performed significantly better than the shifted group. Zentall and Hogan concluded that the better performances when training and transfer tasks are kept the same indicated generalized MTS or OFS. While Zentall and Hogan found a difference in rate of acquisition between the two groups, this effect was not demonstrated in terms of individual animals' performed on the very first transfer trials. Thus, these data and those from similar studies (e.g. Iversen, 1997; Nakagawa, 2000) are not as convincing as those available in primates and marine mammals.

Wright, Cook, Rivera, Sands, and Delius (1988) conducted an experiment that studied whether pigeons trained on a session-unique match-to-sample task (stimuli are presented in novel configurations for each session) were able to transfer their matching ability to novel stimuli. Two pigeons were presented 152 different stimuli per session in a simultaneous match-to-sample procedure with each stimulus appearing only once in each of the 76 trials so that each trial was unique within each session. For comparison, two pigeons were trained using only two

stimuli with each stimulus appearing as the sample and correct comparison for half of the trials and then appearing as the incorrect comparison for the other half of the trials. After each group met criterion (75% correct or higher) they were tested for transfer with novel stimuli. There were 20 transfer trials (with a total of 40 novel stimuli) intermixed within the training baseline that were divided into two transfer sessions (with novel stimuli) with 10 transfer trials per session.

Wright et. al (1988) found that although the rate of acquisition for the trial-unique group was slow, their percent correct rose to about 75% correct over 360 training sessions. The 2-stimulus group met 75% correct on the 16th session of training. What should be noted is that the 2-stimulus group was presented with the same stimuli on all trials, totaling 1,216 presentations over the 16 sessions. Alternatively, the trial-unique group saw each item only once each session, totaling 360 presentations throughout the training sessions. Therefore, the trial-unique group required fewer presentations of each stimulus to meet criterion than did the 2-stimulus group. When tested for transfer with the first novel stimulus set, one of the trial unique subjects showed some evidence for transfer (approximately 70% correct) while the other did not (approximately 55% correct). However, when both animals were tested a third time with these stimuli their performances increased to near 80% accuracy. When tested with the second stimulus set, both subjects performed at 80% correct or higher and interestingly, their baseline scores rose as well. This suggests that by the time the second stimulus set was presented, subjects were able to transfer the matching concept they had learned in training. However, when the 2-stimulus group was tested with the second stimulus set, they performed at near chance levels even though their baselines were nearly identical to the trial unique group. This study again shows that multiple exemplar training aids in the transfer of matching to novel stimuli.

Pigeons have been relatively successful in studies regarding the identity concept using procedures other than MTS. For example, Cook, Katz, and Cavoto (1997) studied pigeons performances on a same/different task using picture arrays as stimuli. The same/different task differs from a MTS procedure in that subjects are not required to choose the comparison stimulus that is physically identical to the sample stimulus but rather subjects are required to make a “same” response (such as pecking a key) if two or more stimuli are physically identical. Alternatively, subjects would be required to make a “different” response (such as pressing another key) if two or more stimuli are physically dissimilar.

Five pigeons were tested in a Plexiglas chamber with stimuli presented via computer on a color monitor. In front of the monitor was a touch screen that detected pecks to the computer screen. A houselight was illuminated at all times unless there was an incorrect response. There were three identical food hoppers in the chamber: one in the center of the front wall and one each on the left and right walls of the chamber. The center hopper was not used in this experiment. Infrared beams were installed in the active food hoppers to detect if the pigeon’s head entered the food hopper.

Experiment 1 examined the rate of acquisition with four different display types. The stimuli consisted of arrays that were displayed against a black background on a computer monitor. The four different display types were: texture, feature, geometric, and object. The “same” texture stimuli consisted of a single item repeated throughout the 24 x 16 array. The “different” texture displays consisted of an 8 x 7 region of the array that was in contrast with the rest of the items (this region either differed in color or shape from the rest of the array). The feature displays were similar to the texture displays, however, for the “different” display instead of an 8 x 7 contrasting region the entire display consisted of items that were different. The

remaining 2 display types were visual search displays. The “same” geometric display used the same elements as the texture displays but with fewer items present. The “different” geometric display was defined by only one element being different as opposed to an 8 x 7 region. The object display type was arranged just like the geometric displays but natural objects were used (birds, flowers, fish, etc.) instead of shapes. Of interest was whether pigeons would learn a generalized same/different response or four separate discriminations. If generalized same/different responding was learned, the acquisition rates should be the same for all four display types. If they had learned four separate relations, there might be different acquisition rates for each display type.

It was found that there was little difference in the rate at which the pigeons learned the four different display types. When the mean performance of “same” and “different” trials was collapsed across the four display types it was found that the pigeons learned the “same” contingencies faster than the “different” ones. An increase in correct responding to the “same” trials occurred approximately 15-20 sessions before there was an increase in correct responding for the “different” trials.

Experiment 2 tested whether pigeons would generalize same/different responding to novel stimuli. Cook et. al argue that if the pigeons had learned a “single rule” in the previous experiment they should demonstrate positive transfer to all four different display types. Novel stimuli for each display type were used to test for transfer and no reinforcement was provided. In the results of transfer tests 1A and 1B, Cook et. al found a mean transfer rate across all display types of 70% correct, which was significantly above chance. Interestingly, it was found that the pigeons were more accurate with the novel “same” displays (76.3% correct) than they were with the novel “different” displays (64.4% correct). Results of transfer test 2A and 2B demonstrated

that the pigeons performed significantly above chance with the novel object displays (67% correct), while there was no significant difference in performance for the “same” (66% correct) and “different” displays (69% correct). The pigeons also performed significantly above chance for the novel object displays from the novel categories (69% correct). What was interesting in this finding is that the pigeons performed just as well with the novel object stimuli as they had with the familiar object stimuli.

Cook et. al found that the most important result from this experiment was that the pigeons demonstrated above chance performance on novel transfer tests across all four display types. They argue that these findings support the conclusion that pigeons used a generalized same/different rule for all display types. Although these data are promising, it should be noted that accuracy was limited (70% correct).

Young, Wasserman, and Garner (1997) showed that increasing the number of items in a visual display facilitated learning a same/different task. They used displays that contained 2 to 16 computer icons that were either the same or different from each other and found that pigeons’ performances improved for the “different” displays when more icons were present. There was no significant difference for the “same” displays as the number of icons varied. Young et. al argued that this variability in the number of icons (termed entropy) was the key to learning a same/different relation.

Rats

Given the success researchers have had in demonstrating generalized identity matching in other species, one might expect to see comparable findings with rats because they are among the most frequently used laboratory subjects. There have been numerous studies in which rats were the subjects with identity matching procedures; however, there is little evidence of generalized

identity matching. For example, consider two studies by Iversen (1993; 1997) of MTS in rats. The apparatus used to test this procedure was a chamber in which one wall had three keys that could be illuminated in array of different colors. In the 1993 study, three rats were trained on a visual MTS procedure with the sample never moving from the center key. Rats were required to poke the illuminated center key with their noses, which produced comparisons on the other two keys. Once the subjects made a choice by poking one of the comparisons, the trial ended and another sample was presented. Iversen found that hundreds of trials were required for the rats to reach performances of 90% correct or higher.

In the testing phase (Iversen, 1997) that began immediately after training, the sample could appear on any of the three keys. As before, the rats were required to poke the sample key with their noses to produce the comparisons. When the sample location was varied it was found that accuracy dropped to just above 60% correct for all subjects. When responses to each key were analyzed separately, two out of the three rats performed at baseline levels (90% correct or higher) when the sample was on the center key. The third rat's performance dropped to approximately 80% correct when the sample was on the center key. To address this issue six 100-trial sessions of center-only sample presentations were given until the subjects had regained their baseline performances. For all three rats, side key sample presentations resulted in chance performances of approximately 50% correct.

In Experiment 2 of this study, the same rats were trained on each sample location separately. All three rats were trained with the sample presented only on the left key (Rat Three was also trained with the sample always appearing on the right key). Once they each met criterion, 12 sessions of center-sample presentation were given before the moving sample procedure was reintroduced. The moving sample procedure was implemented as follows, for

Rats 1, 2, and 3 respectively; 15, 22, and 13 sessions were given in which the sample only appeared on either the left key or the center key. In the next phase of Experiment 2 the sample could appear in any of the three locations. It was found that Rats 1 and 2 developed a position preference such that when the sample was presented on the left key they almost exclusively chose the comparison on the center key. Therefore, forced trials were implemented in which the correct comparison only appeared on the right side key. Rat 3's choices came under the control of the sample position when the sample was presented on the left key and its performance was near 80% correct after 17 sessions. When the untrained right side-key-sample presentation was implemented in conjunction with left and center key sample presentation all three subjects' performance significantly dropped. The subjects were not able to transfer sample control to the right side key, and their performance for left key sample presentation declined as well. When the sample was presented on the center key the performance for Rats 1 and 2 dropped to just under criterion, while Rat 3 remained at the criterion level of 90% correct or higher.

Iversen concluded that rats were unable to transfer sample control when the sample was presented in a different location; thus, each sample location must be trained separately. This study is important in that it provided evidence that rats were learning specific relations between stimuli and stimulus positions in match-to-sample training rather than generalized MTS.

Nakagawa (2000) tested rats on a same/different procedure using visual stimuli. The animals were trained in a "T" maze where a guillotine door was placed at the entrance of the box. Two stimulus cards that displayed letters or shapes were placed at the end of the "T" maze, and the left and right sides of the "T" served as goal boxes where the animals received food reinforcers. Half of the rats were trained that if the two stimuli were the same they should enter the right goal box and if different they should enter the left goal box. The other group of rats

were trained the exact opposite (if same, enter left goal box; if different, enter right goal box). Also, half of the rats were trained with eight stimuli (group F), and half of the rats were trained with 16 stimuli (group M). The rats were presented six "same" trials and six "different" trials in random order. Once they met a criterion of 11 or more correct out of 12 trials, the subjects were trained on novel same/different tasks in which some stimuli were completely new and others were stimuli seen before in training but never in certain configurations. Nakagawa found that there was no significant difference between the groups in the acquisition of learning the same/different tasks. He found that 88% of rats from both groups responded correctly on the first trial of the transfer phase. He thus argues that rats are capable of a generalized identity concept, and indeed, the findings of this study are certainly promising. However, the use of only a limited number of novel stimuli and the lack of individual subject data are limitations to this experiment.

Several experimenters have argued that rats are capable of complex learning if olfactory stimuli are used (Slotnik & Nigrosh, 1974; Lu, Slotnik & Silberberg, 1993; Eichenbaum, 1998; and Slotnik, Hanford, & Hodos, 2000). These authors argue that perhaps the reason for rats poor discrimination performances is that experimenters are not testing them using the modality that is best suited to the species (Iversen, 1993; 1997). For example, Lu, et al. used thirty odors in a conditional go/no-go discrimination task using an eight-channel olfactometer. The go/no-go procedure was implemented such that responses were reinforced ("same") only when two odor stimuli were identical. There was a response area timer of 2 s such that if the rat entered this area within the timed 2 s, it was counted as making a "same" response. This procedure differed from a traditional MTS procedure in that a sample and only one comparison were presented and responses were only reinforced when the two were the same. Each trial began with a sample that

was presented for 1s and was followed by a second odor also presented for 1s. If the comparison stimulus was the same as the sample, then a response was reinforced with .05 ml of water and the trial was recorded as correct. If the rat failed to make a response on an S+ trial the trial was recorded as a miss. If the rats made a response on an S- trial this was recorded as a false alarm. If the animals failed to make a response on an S- trial, the trial was recorded as a correct rejection.

The rats were trained first with 2 odors (butanol and amyl acetate). Next, an additional 200-300 trials were presented in conjunction with a punishment procedure for false alarms. If a response was made to an S- trial, the trial was repeated until the animals made a correct rejection (for a maximum of 3 trials). Additional trials (400) were presented without the punishment contingency. Training was continued the following day but with two novel odors (linalyl acetate and geraniol). The animals were trained with the new odors until they met a criterion of 80% correct for each trial type.

Lu et. al (1993) found that until the punishment contingency was implemented the rats performed at chance levels for approximately 300-400 trials. Performance improved over the preceding 200-300 trials when the punishment phase was introduced and their performance remained stable when the punishment contingency was removed. It was found that the average scores on the 2-odor task in which there was no punishment contingency ranged from 80% to 91% correct. When the two novel odors were introduced into the 2-odor procedure, the rats responded at a high rate of accuracy (80% correct in 20 trials within the first 200 trials) to the S+ stimuli and the responding was inhibited for the S- trial types. All three rats met criterion within the first 200 trials and maintained an accuracy level between 95% and 100% correct for the last 200 trials of that phase.

Next, the animals were exposed to three different odors presented in pairs. Each session contained 180 trials and each S+ combination was presented 30 times while each S- trial was presented 15 times in random order. The criterion limit remained at 80% correct. The stimuli used in the three-odor matching task had previously been used in training; however, all stimuli used thereafter were novel.

When the animals were introduced to the 3-odor matching phase they readily acquired the match-to-sample problems. Two of the three rats were at 90% accuracy or higher within the first session. Rapid acquisition was also shown with the second 2-odor matching-to-sample phase with novel stimuli. Lu et. al (1993) also found that the delays implemented between the presentation of the sample and the comparison had little effect on the rats' performance.

This study clearly provides support for the use of olfactory stimuli when testing the learning and memory capabilities of rats. Each animal was able to readily acquire novel same/different problems while also maintaining their high performances when as much as a 10 s interstimulus delay was introduced. However, this study clearly does not meet the requirements when testing for generalized identity concept. Of utmost importance is the performance of subjects on the initial novel problems, which Lu et al. (1992). do not present. Thus, it is possible that Lu et. al simply observed rapid learning of specific stimulus configurations.

Thus, no studies have demonstrated either generalized MTS or a generalized identity concept in rats. Iversen (1997) showed that rats are susceptible to configuration learning when trying to solve a MTS problem. Others (Lu et. al, 1992; and Nakagawa, 2000) claim that rats can learn same/different relations but have not met all the criteria discussed by Dube et. al (1992).

The present experiment was an attempt to study olfactory MTS in rats using a procedure developed by Eichenbaum (1998). He combined different scents with sand in which the rats

were trained to dig to produce food reinforcement. He found that this method significantly increased the rate at which the animals could learn simple discriminations.

The present study differed from the Nakagawa (2000) and Lu et. al (1993) studies in that a simultaneous match-to-sample procedure was implemented (as opposed to a same/different or go/no-go procedure). A modified version of Eichenbaum's (1988) procedure (using scented sand) was implemented. Subjects were required to first dig in the sample, which initiated the presentation of two comparison stimuli. Multiple exemplar training was used to minimize the likelihood of subjects learning specific configurations. In addition, test types proposed by Dube et. al (1992) were used to assess generalized MTS (novel sample, matching novel comparison, and familiar comparison; familiar sample, matching familiar comparison, and novel comparison; novel sample, matching novel comparison, and another novel comparison).

Method

Subjects

Subjects were four male Holtzman (Sprague-Dawley) albino rats approximately six months old at the beginning of the experiment. Rats were maintained with free access to water, but access to food was restricted to a one-hour period each day (after the testing session).

Apparatus

A modified operant chamber was used in the study. The front and rear walls and the lid to the chamber were replaced with all clear Plexiglass while the sidewalls were stainless steel. The floor of the chamber consisted of stainless steel grids that were 1.3 cm apart. The dimensions inside the chamber were 28 cm long x 26 cm wide x 30 cm high. A 5 cm section was removed from the bottom front of the chamber so that a removable plastic tray could be inserted. The tray used in pretraining consisted of black plastic and measured 28 cm long x 12

cm wide x 30 cm high. There were two 5-cm holes drilled into the top center of the tray that were 8 cm apart. The holes were used to hold clear plastic cups in which the stimuli were presented. The tray used with the two-comparison MTS procedure consisted of clear plastic and had three 5-cm holes drilled into the top. The holes were positioned in a triangular shape so that a simultaneous MTS procedure could be implemented (see Figure 1). The tray used with the three-comparison procedure consisted of clear plastic and had four 5-cm holes drilled into the top (see Figure 2). A handle on the front of the plastic tray was used to insert and remove each of the respective trays between trials. Clear 2 oz. plastic cups were used to present the stimuli that were then placed in the drilled holes.

Procedure

Pretraining: All training and test sessions were held five days a week (M-F) with one session per day and were conducted in the presence of soft (<70 dB) continuous white noise. The first stage of pretraining consisted of allowing subjects to become accustomed to the apparatus. In the next stage, two cups, each containing a 45-mg sucrose pellet, were presented successively to the subjects until they were readily eaten.



Figure 1. Tray used to implement the two-comparison MTS procedure. The hole in the top center was designed to hold the sample stimulus. The two holes in the bottom of the tray were designed to hold the comparison stimuli.

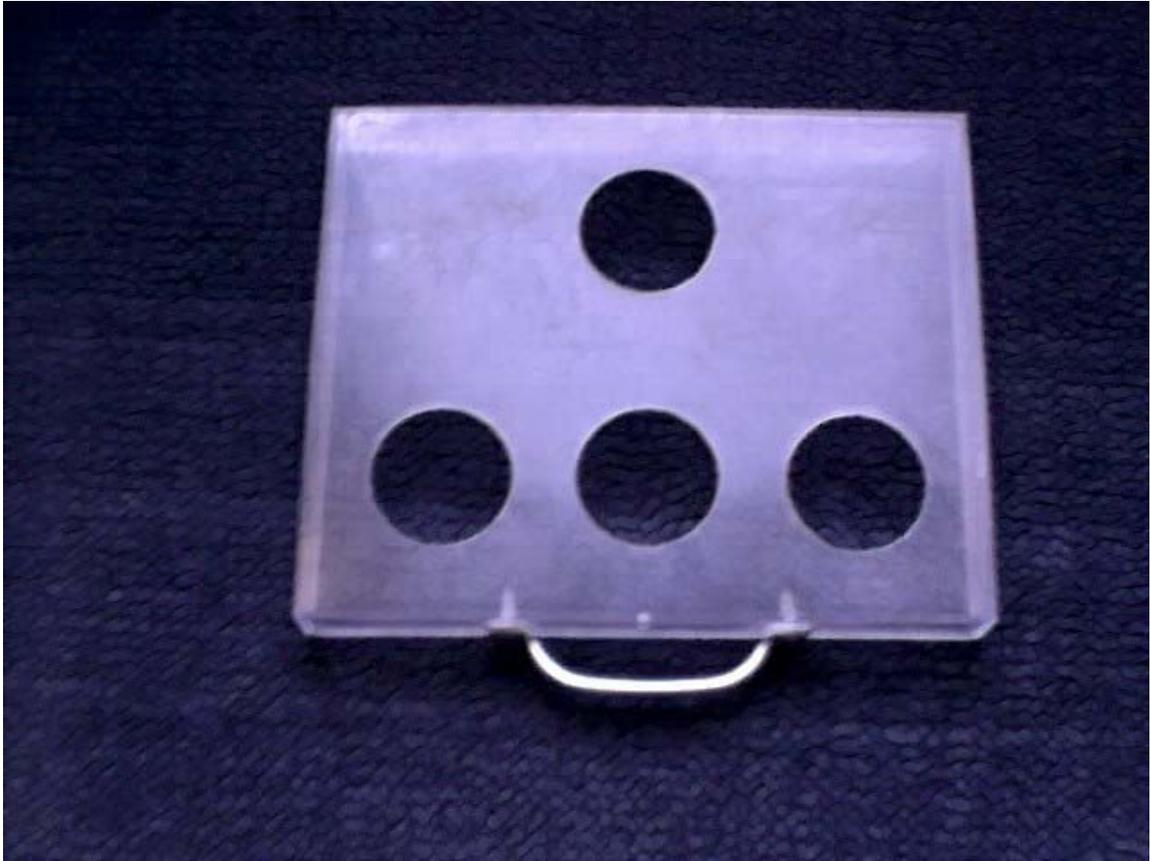


Figure 2. Tray used to implement the three-comparison MTS procedure. The hole in the top center was designed to hold the sample stimulus. The three holes in the bottom of the tray were designed to hold the comparison stimuli.

The next stage of pretraining involved presenting cups filled with sterilized play sand purchased from a local store with a sucrose pellet placed atop the sand. A shaping procedure was used to train digging behavior, such that when the subject was readily eating the pellets from the surface of the sand, the pellets were inserted progressively more deeply. This was continued until the subject was consistently eating pellets buried approximately 1 cm beneath plain sand. The animal was kept in the chamber until he dug in one of the stimulus cups. If an animal did not dig within 15 min, the animal was removed from the chamber and placed in his home cage until the next day's session.

For the purposes of this study, a digging response was defined as one or both paws touching the sand in such a way that sand was displaced. Once the animal was consistently digging in the plain sand, scented sand was introduced. Initially, each animal was presented with two cups that contained the same odor. In one of the cups was buried a sucrose pellet. This was done to test whether the animal was able to detect the pellet (pellet detection procedure). Each animal had approximately 15-20 sessions of pretraining (this includes sessions that were conducted for pellet detection).

Stimuli

Olfactory stimuli were presented by mixing ordinary household spices or liquid synthetic flavorings with sterilized sand (for an exact list and the sequence of odorants used see Table 1). Spices were mixed at a ratio of 10 g per 1000 g of sand (This ratio of spice to sand was chosen because pilot research suggested that it masked the scent of the sucrose pellet). For liquid odorants, the ratio consisted of two drops of liquid (from

Table 1

Sequence of Odors for Each Subject

# Odors	J16	J6	J11	J10
2	Mustard, Cinnamon	Mustard, Cinnamon	Garlic, Celery	Garlic, Celery
3	Sage	Sage	*Cinnamon, Mustard	Cinnamon
5	Celery, Onion	Celery, Onion	Paprika, Coffee	Paprika, Coffee
7	Garlic, Paprika	Garlic, Paprika	*Sage, Onion, Turmeric	Sage, Onion
9	Coffee, Ginger	Coffee, Ginger	*Clove, Coriander	Mustard, *Ginger
11	Marjoram, Thyme	Marjoram, Thyme	Cumin, *Nutmeg, Vanilla	Marjoram, Thyme, Cumin
13	Turmeric, Nutmeg	Turmeric, Nutmeg	Pineapple, Strawberry	Nutmeg, Turmeric
15	Orange, Cumin	Orange, Cumin	Oregano, Walnut	Clove, Coriander
17	Vanilla, Clove	Vanilla, Clove	Coconut, Almond	Vanilla, Oregano
19	Pineapple, Coriander	Pineapple, Coriander	Caraway, *Rosemary	Pineapple, Orange
21	Strawberry, Lemon	Strawberry, Lemon	Lemon, Cherry	Strawberry, Lemon
23		Oregano, Almond	Anise, Brandy	Walnut, Almond
25		Cherry, Root Beer	Allspice, Root Beer	Cherry, Root Beer
27		Walnut, Brandy	Rum, Savory	Coconut, Brandy
29		Rosemary, Caraway	Oregano, Chocolate	
31			Maple, Peppermint	
33			Dill, Bay	
35			Marjoram, Thyme, Sumac, Fennel	
37			Butter, Fenugreek	

*Dropped Odors

standard eyedropper) per stimulus cup (these odorants were added late in the experiment because dry spices were limited). After the addition of odorants, the sand was stirred to evenly distribute the odorant throughout the mixture. The experimenter wore latex gloves during preparation of the stimuli and tweezers were used to place the pellets in stimulus cups. Stimulus cups were filled approximately 3/4 full. In some cases, subjects appeared to develop aversions to particular scents (i.e. they would not dig in either sample or comparison cups with that scent). When this occurred, use of that odorant was discontinued for that animal (see Table 1).

Pellet Detection

A total of 10 rats (three from the present study and seven others) were exposed to a pellet detection test in which they were required to dig in one of two cups that contained the same scent. Within one of the cups, a sucrose pellet was buried approximately 1 cm beneath the sand. Each session contained 20 trials. Performances approximating chance were interpreted as a particular scent successfully masking the odor of the sucrose pellet (see Table 2). Not all odors in the present study were used in the pellet detection procedure; however, other methods were implemented to ensure that the rats were not solving the MTS problem in this way (e.g. double baited trials).

Match-to-Sample procedure:

Two of the rats tested in the Pellet Detection phase of the study continued on to match-to-sample training. Training began with two stimuli for each subject. Subjects J16 and J6 began the MTS procedure with the odors of mustard and cinnamon, and subjects J11 and J10 started with celery and garlic. One of the stimuli was presented in the sample position of the tray (see Figure 1).

Table 2

Pellet Detection for Each Spice Expressed by Percent Correct for the Last Four Sessions

		<u>Sessions</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
<u>Spice</u>	<u>Subject</u>					
Cinnamon	J11		55	40	65	40
Celery	J11		50	55	60	50
Mustard	J6		50	45	50	55
Cinnamon	J6		50	45	50	55
Mustard	J16		45	55	40	55
Cinnamon	J16		50	65	50	60
Coffee	M7		50	55	60	40
Garlic	I7		55	50	50	50
Ginger	J12		60	40	60	45
Onion	J12		50	40	45	45
Paprika	I6		40	35	50	45
Sage	M2		45	50	50	50
Sand (only)	I7		60	55	50	65
Celery	M11		35	65	55	55

The tray was inserted approximately ten cm into the chamber such that only the sample stimulus, which was baited with a sucrose pellet, was accessible to the rat. When the rat completed a digging response in the sample the tray was fully inserted into the chamber (approximately 20 cm), allowing access to the two comparison stimulus cups. One comparison stimulus was the same odor as the sample and this cup always contained a sucrose pellet (S+). The other comparison was a different scent and in this phase of the experiment, was never baited (S-). The tray was removed from the chamber once the subject dug in the correct comparison for approximately 5 s. A correction procedure was used such that if the subject responded to S-, that response was recorded but the subject was allowed access to both comparisons until it dug in S+. If the subject did not dig in either the sample or a comparison cup within 3 minutes, that trial was counted as incorrect and a new trial was presented. The inter-trial interval (ITI) for this experiment was approximately 15 s.

The trial arrangements within sessions were as follows: no stimulus appeared as the sample more than two times in a row, and no comparison appeared in either the left or right position more than two times in a row. Each stimulus appeared an equal number of times as the sample and each stimulus appeared as a comparison in each position an equal number of times (until the number of stimuli disallowed this arrangement—see below). On most trials, only the correct comparison cup (i.e., the cup containing the same odorant as the sample) was baited with a sucrose pellet. However, beginning during the initial Novel Stimuli Phase, two trials (randomly selected across sessions) in each session were programmed with both comparison cups baited. Although the Pellet Detection procedures indicated that rats could not discriminate cups on the basis of olfactory cues provided by the presence of the pellet, these tests were not conducted with all of the odorants that were used in the present study, and the trials on which both comparison

cups were baited served to determine whether rats performances might to some degree be determined by the scent of the pellet rather than the S+ odor. Throughout each session, S+ and S- locations were balanced across the left and right positions.

Reinforcement Reduction:

When the rats met a performance criterion of 75% or higher accuracy for two consecutive sessions, reinforcement in the sample cup was reduced from 100% to 75% reinforcement. Subsequently, when a more stringent criterion of 90% correct for two consecutive sessions was implemented, sample cup reinforcement density was further reduced to 50% (see Table 3) and maintained at this level throughout the remainder of the experiment. Rat J16 was briefly exposed to a reinforcement density of 25% and then 0%, until this manipulation disrupted his performance. Responses to the correct comparison stimulus were always reinforced.

Novel Stimuli, Phase 1:

This phase involved tests for generalized identity matching by arranging sessions in which one or more novel odors were introduced (see Table 3). In these sessions novel odors were presented as samples and were introduced after one or two trials with the familiar stimuli (baseline trials) had been completed. Initially, only a single novel stimulus was introduced, and on the first trial on which it appeared, it was presented as the sample, with one of the baseline stimuli as the incorrect comparison. During this phase, the sample cup was always baited on the first trial involving a novel odor. Subjects were studied with three stimuli until they met a criterion of two consecutive sessions at 90% correct or higher. Each time criterion was met two novel stimuli were introduced.

The first trial involving the novel odors was presented such that a novel odor was the sample and the other novel stimulus served as the incorrect comparison (in certain cases when

Table 3

Outline of Phases

Condition	Criterion
<u>Reinforcement Reduction Phase</u>	
A. 2 Stimuli 100% sample reinforcement	2 consecutive sessions at 75% or higher
B. 2 Stimuli 75% sample reinforcement	2 consecutive sessions at 90% or higher
C. 2 Stimuli 50% sample reinforcement	2 consecutive sessions at 90% or higher
<u>Introduction of Novel Stimuli, Phase 1</u>	
A. Three stimuli MTS test 24 trials (first novel trial includes sample reinforcement)	2 consecutive sessions at 90% or higher
B. Several tests for generalized MTS	2 consecutive sessions at 90% or higher Overall criterion of 5/6 tests
<u>Introduction of Novel Stimuli, Phase 2</u>	
A. Several tests for generalized MTS with no reinforcer in the sample	2 consecutive sessions at 90% or higher Overall criterion of 5/6 tests

* Table continued on next page

Novel/Familiar Familiar/Novel Phase

<u>Condition</u>	<u>Criterion</u>
A. Novel/Familiar Familiar/Novel MTS tests 24 trials (no sample reinforcement both comparison cups baited)	2 consecutive sessions at 90% or higher Overall criterion of 5/6 tests for either probe type

Outline of Three-Comparison Phase

A. Three comparisons 24 trials no (sample reinforcement)	2 consecutive sessions at 90% or higher
12 trials with 2 odors	Overall criterion of 5/6 tests
12 trials with 3 odors	for generalized MTS

*The novel trials in this phase were presented just as those in the Novel familiar Phase of the experiment.

only one stimulus was introduced, the incorrect comparison was selected randomly from among the baseline stimuli). At least one baseline trial separated the trials in which the novel odors were presented. On the next trial involving the novel odors, the novel stimulus that had previously appeared as the incorrect comparison was now presented as the sample and the other served as the unreinforced comparison. Subsequent trials during the session included combinations of the familiar or baseline stimuli and the novel stimuli as samples and comparisons.

For analysis purposes, response accuracy for trials involving a novel stimulus as sample (Novel Probe) were considered separately from trials involving stimuli that had previously been encountered, but had never been presented together in the particular sample/comparison configuration, and from trials involving comparison pairs that had previously been encountered at least once before during the experiment (Baseline trials). Note that a trial was considered a Novel Configuration trial only the first time two odors appeared in a particular sample/comparison configuration--regardless of the particular spatial position of the comparisons (i.e., which stimulus was in the left or right position). Thus, in any given session in which novel stimuli were added, one or two of the programmed trials were analyzed as Novel Probes (depending on whether one or two new stimuli were introduced), trials in which one of the novel stimuli served as a comparison along with one of the baseline stimuli for the first time were analyzed as Novel configurations, and stimulus combinations that had been previously encountered were analyzed as Baseline trials.

Rules for composition of sessions varied somewhat across Phase 1. The number of trials per session varied between 21 and 27 depending on the number of stimuli in the baseline. Initially, sessions were programmed so that each stimulus appeared an equal number of times as

a sample and as a comparison in random order with the same constraints specified in the Reinforcement Reduction Phase. So, for example, when the baseline consisted of three stimuli, each stimulus served as a sample on 8 trials and as a comparison on 16 trials in each 24-trial session. When the baseline was increased to five stimuli, each stimulus served as a sample on 5 trials and as a comparison on 10 trials in each 25-trial session. These same 25 trial types were presented in random order in subsequent sessions until criterion was met and additional stimuli were added. Thus, some possible configurations of the five stimuli did not appear until later in the experiment. With seven odors there were 21 trial types per session. These 21 trials were presented in random order across subsequent sessions until criterion was met. With seven odors, each stimulus served as a sample three times and between five and six times as comparisons. There were only eight of the possible 20 novel configurations presented (which eventually were presented in later phases). When the number of odors increased to nine, there were a total of 27 trials presented. Each stimulus served three times as a sample and six times as a comparison. The 27 trials were presented in random order across subsequent sessions until criterion was met and only 11 of the possible 28 novel configurations were available. When the total number of odors increased to 11, there were 22 trials presented, with each odor appearing twice as a sample and four times as a comparison. At this stage the sessions were organized such that each trial of every session was a Novel Configuration Trial (until no novel configurations remained). A matrix was designed for each animal to depict whether a particular odor had appeared in conjunction with another (see Appendix). After an odor was paired with another, that trial type was not used again until all other combinations had been exhausted. However, odors that had been in an animal's repertoire from the early stages of the experiment had a higher likelihood of being re-coupled. When this became necessary, a particular odor would be paired with an odor

from the beginning of the matrix (in other words the animal had not been exposed to that particular combination since the early stages of the experiment).

When the number of baseline stimuli exceeded 11, all subsequent sessions were composed of 24 trials. At this point each stimulus appeared at least once as a sample in each session. When the number of baseline stimuli exceeded 24, those odors that did not appear within a particular session would appear within the next with no baseline stimulus presented more than twice as a sample. Digging in the sample was reinforced on 50% of the trials throughout the experiment, and novel probes were always baited in Phase 1 (until the subject was transferred to Phase 2—See below), and the correct comparison appeared 12 times in the left and 12 times in the right position on each session. No additional novel stimuli were introduced until a criterion of two consecutive sessions of 90% accuracy or higher was met. The session immediately following criterion level performance was designated a Novel Probe session as described above. After a Novel Probe session, the stimuli introduced in that session were used in all subsequent sessions except where noted.

A second criterion was used to determine whether subjects were providing evidence of generalized matching. Evaluating only Novel Probe trials (the initial trial on which a given stimulus served as a sample), the criterion required correct performance on five out of the most recent six Novel Probes before the animal was advanced to Phase 2. Novel Probe samples were always baited in this phase, so it is possible that accurate performances on novel probes could derive from a “win-stay” response strategy. Also, only the correct comparison cup was baited on novel probes, so accurate performances could be due to subjects detecting the pellet with some of the untested odors. Phase 2 was an effort to examine these two possibilities.

Novel Stimuli, Phase 2:

This phase of the experiment differed from the previous phase in that the initial exposure to novel stimuli did not include sample reinforcement (see Table 3). Sample reinforcement remained at 50% for all baseline trials on all sessions during this phase. As an additional control, all Novel Probe trials during this phase involved placing pellets in both comparison cups, ensuring that subjects could not respond to the correct comparison by detecting the scent of the sucrose pellet. In sessions that did not include novel probes, two randomly selected baseline trials had comparison cups that were both baited. Otherwise, this phase of the experiment progressed as the previous one, with novel stimuli added when a 90% criterion was met. Phase 2 was continued until correct responding on five out of six consecutive Novel Probes was observed.

Novel/Familiar Phase:

This phase of the experiment differed from the previous phases in that Novel Probe sessions always involved two new trial configurations. In one type (Novel-Familiar), the sample was novel but the incorrect comparison was one of the previously encountered baseline stimuli. The other type (Familiar-Novel) involved a familiar sample and a novel incorrect comparison (Phases 1 and 2 generally involved probes that were always “novel-novel”). The initial session with each novel stimulus in this phase included one trial with each probe type. These probe types tested whether the animals were responding on the basis of rejecting a novel stimulus or on the basis of rejecting a familiar comparison. As in the previous phase, Novel Probe samples were never baited and both correct and incorrect comparisons contained pellets (see Table 3). Otherwise, the criterion for introduction of novel odors and the trial and session composition were the same as in the previous phase. This phase was continued until there was correct responding on five out of six consecutive Novel Probes of either type. Only two subjects (J6 and

J10) met the criterion requirements for this phase. After Subject J10 met criterion, however, he had several sessions that were discontinued due to a lack of responding and was subsequently dropped from the study. Subject J11 was advanced to the Three Comparison Phase although he never met criterion in the Novel/Familiar Phase, because of the importance of observing more than one rat in the Three-Comparison condition.

Three Comparison Phase:

This phase of the experiment tested rats' MTS performance when there were three comparisons available. This was of particular importance because in the two comparison tests, rats could be solving the MTS problem by simply choosing the strongest scent. Because the incorrect comparison involved only a single scent, the sample and correct comparison may have produced a detectibly more intense stimulus. In order to assess this possibility, half of the trials in this phase were arranged such that the two incorrect comparisons were the same scent. Thus, each scent within a given trial should have been equally intense. The other trials were arranged such that two different stimuli served as the incorrect comparisons (see Table 3).

For the two animals that advanced to this phase, there were two odors introduced each time criterion was met (two sessions at 90% or higher for subject J6, and two sessions at 88% or higher for subject J11). A probe was given after at least one baseline trial. The sample was not baited while all three comparisons were. Each probe session consisted of two probes that were either novel/familiar or familiar/novel (just as in the previous phase of the experiment). The same criterion was implemented as in previous phases with the exception of subject J11, whose criterion was lowered to two sessions at 88% percent or higher (this was his stable accuracy level). At this stage the experiment was ended due to the advanced age of the subjects (nearing two years). Thus, implementing this phase of the experiment tested whether the animals were

capable of solving the MTS problem when more than one comparison scent was available while also testing whether the animals solved the problem by choosing the most intense scent.

Table 4 depicts the number of sessions each subject was studied for each phase of the study.

Inter-rater Reliability and Blind Sessions:

Throughout all phases of the experiment each animal was exposed to several sessions (J16, 5; J6, 7; J11, 3; J10, 2) in which there was a second investigator present (see Tables 5 and 6). The second investigator independently rated whether the animal dug in the right or left comparison (or center during the three-comparison phase) and was blind with respect to which comparison was correct. The animals were exposed to sessions in which the experimenter was under blind conditions (J16, 4; J6, 7; J11, 5; J10, 0). Several different experimenters conducted the sessions to ensure that no experimenter biases were influencing the animals' selections. Data from these sessions were analyzed separately (see Figure 14).

Results

Figure 3 shows the results from the initial phase of the experiment (Reinforcement Reduction Phase), in which only one conditional discrimination was trained. Each panel shows the percent correct matches across consecutive sessions. The subjects were first required to meet a criterion of two sessions at 75% or higher with 100% sample reinforcement before the sample reinforcement was reduced to 75%. All four animals met this criterion between 15 and 24 sessions. At 75% sample reinforcement, subjects J16 and J11 met the criterion (two sessions at 90% or above) in six and four sessions respectively, while subjects J6 and J10 required somewhat more training (13 and 16 sessions respectively). Subject J16 was the first animal to meet criterion at 50% sample reinforcement (in two sessions) and sample reinforcement was further reduced to 25%, and withdrawn completely for one session. Because the lack of sample

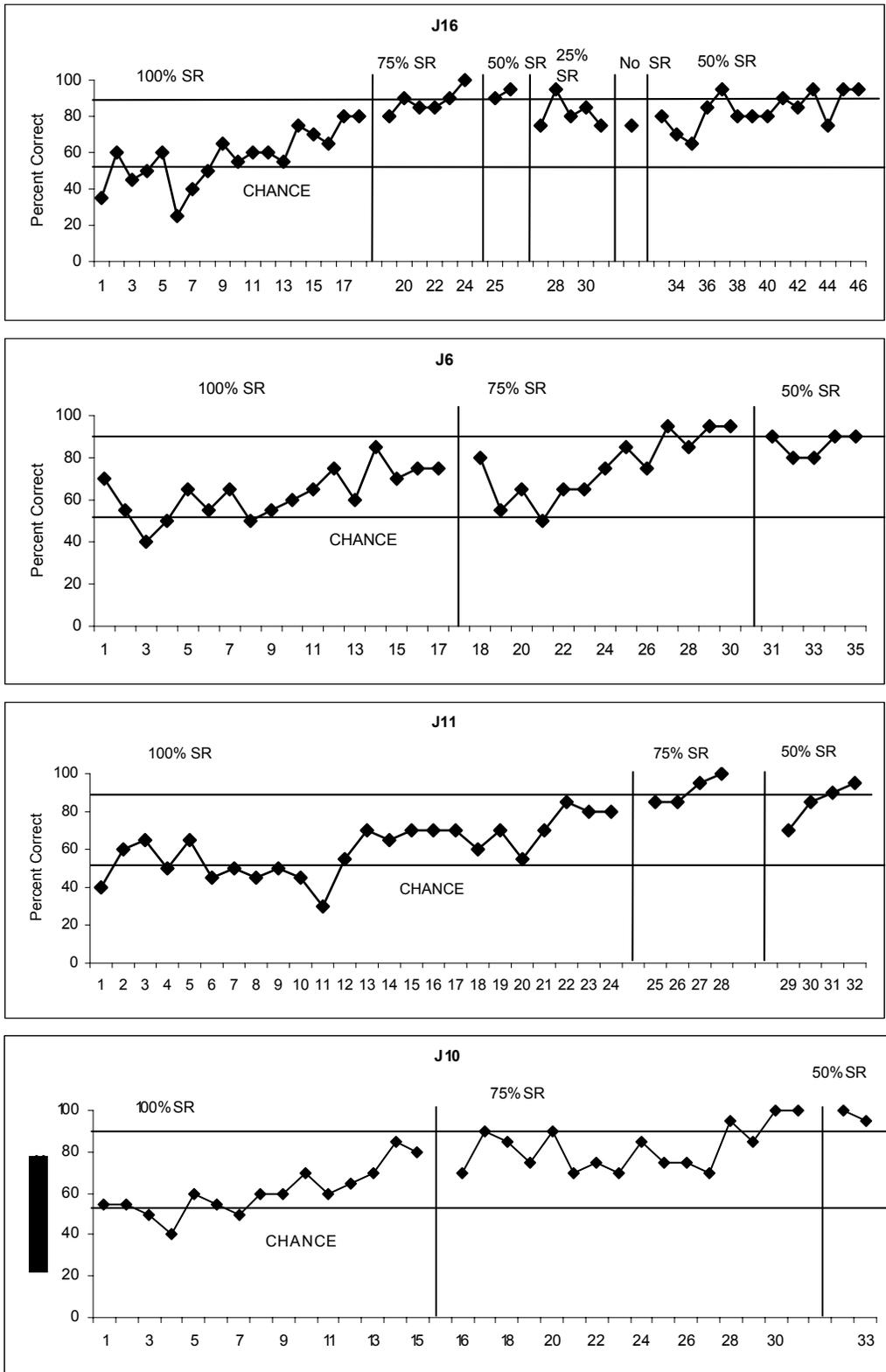
Table 4

Number of Sessions each for Pellet Detection and MTS Procedures

Phase	<u>Subject</u>			
	J16	J6	J11	J10
Pellet Detection	12	17	10	0
Reinforcement Reduction	46	35	32	33
Novel Stimuli, Phase 1	46	27	68	43
Novel Stimuli, Phase 2	68	8	25	6
Novel/Familiar	*	6	22	17
3 Comparisons	*	25	39	*

* Subject did not participate in that particular phase

Figure 3



Number of Consecutive Sessions

Figure 3. Each panel depicts the session-by-session acquisition of MTS with two odors in the Reinforcement Reduction Phase of the experiment expressed as a percent correct for each rat. The percentage above each panel depicts sample reinforcement densities. The top horizontal bar depicts the criterion and the bottom horizontal bar depicts chance levels.

reinforcement was disruptive for this animal, he was taken back to 50% sample reinforcement until criterion was met (14 sessions). Due to J16's difficulties with reinforcement densities of less than 50%, a 50% sample reinforcement density was used for all subjects throughout the remainder of the study. The three remaining subjects met criterion with 50% sample reinforcement within two and five sessions.

Figures 4 and 5 show the individual rat's performances for Novel Stimuli, Phase 1 of the study, as the number of stimuli was increased. Moving left to right, each panel in Figures 4 and 5 shows performances as stimuli were added to the baseline. The top set of numbers (numbers not in parentheses) within each panel indicates the number of correct responses/trials on Novel Probe trials (a novel stimulus used as a sample for the first time), while the numbers in parentheses indicate the percent correct on trials involving Novel Configurations (trials that involved stimuli that were not novel, but had not previously been presented in the particular sample-comparison configuration) during that sub-phase. For example, as shown in Figure 4, as Rat J16 moved from three to seven stimuli (J16 met criterion on his third session with seven stimuli, but due to experimenter error was given six additional sessions before moving on to nine stimuli) he performed correctly on only one of his first five Novel Probe trials, but showed rapid learning with the new stimuli, overall performing correctly on 81% of the novel configurations. Subject J16 met the overall criterion for this phase by performing correctly on the last five out of six Novel Probe trials as he moved from 9 to 13 stimuli. Figure 4 also shows a similar pattern for Rat J6 with high levels of performance on baseline and Novel Configuration trials evident almost immediately, but performances on Novel Probes were not immediately at a high level (three out of the first seven correct). However, as he moved from 11 to 15 stimuli, Rat J6 met criterion by performing correctly on the last 5 out of 6 Novel Probe trials while maintaining 90% or better

Figure 4

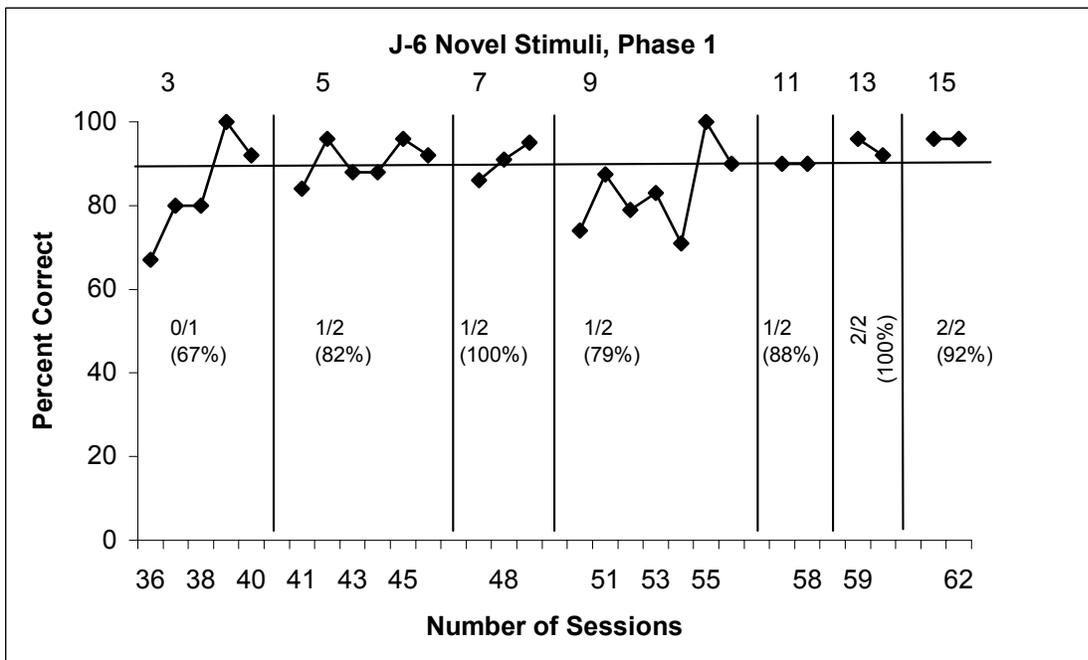
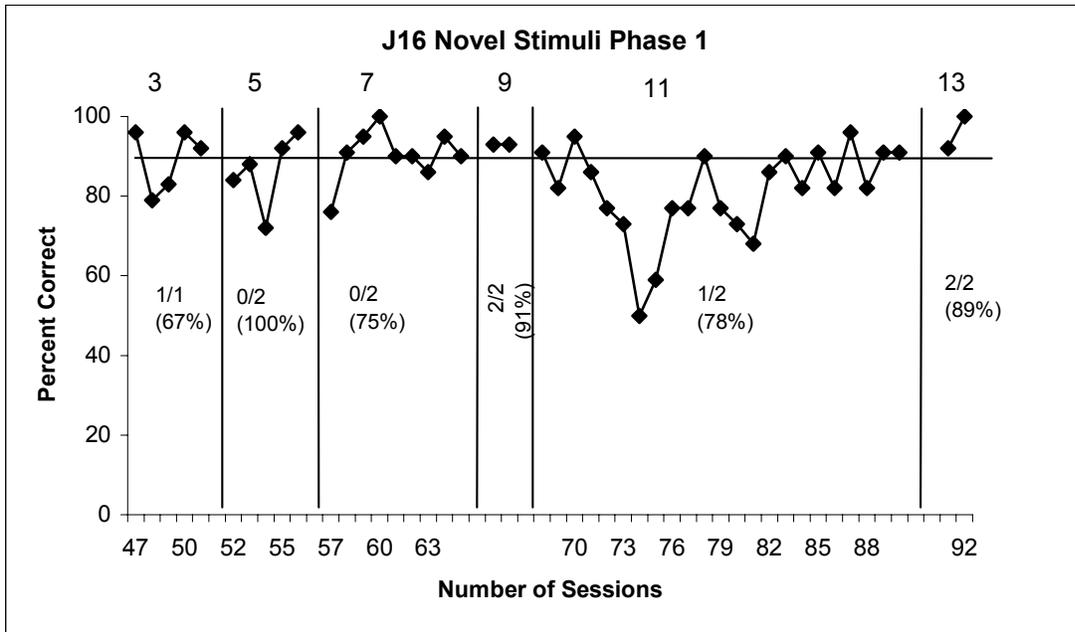


Figure 4. Each graph depicts the session-by session percent correct performance on all trials for subjects J16 and J6 in the Novel Stimuli, Phase 1 of the experiment. The numbers at the top of each panel indicate the number of stimuli. The top numbers within each panel of the graphs depict performances on the novel tests while the bottom numbers in parentheses depict the percent correct performances on the novel configurations. The horizontal line represents criterion.

Figure 5

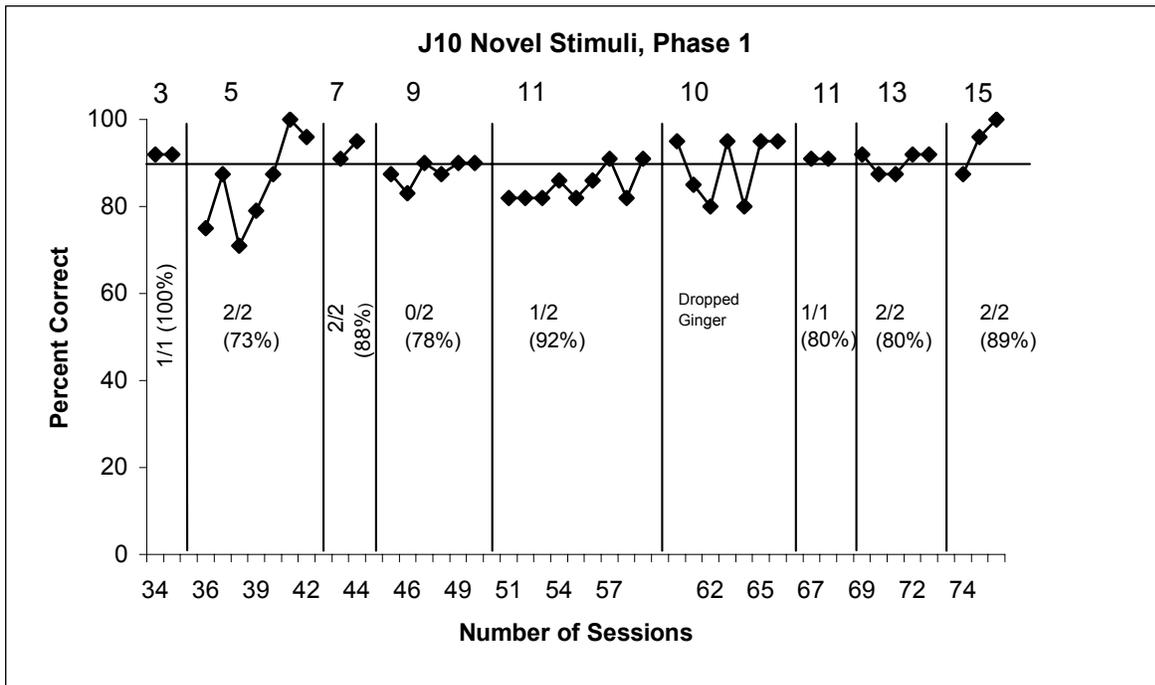
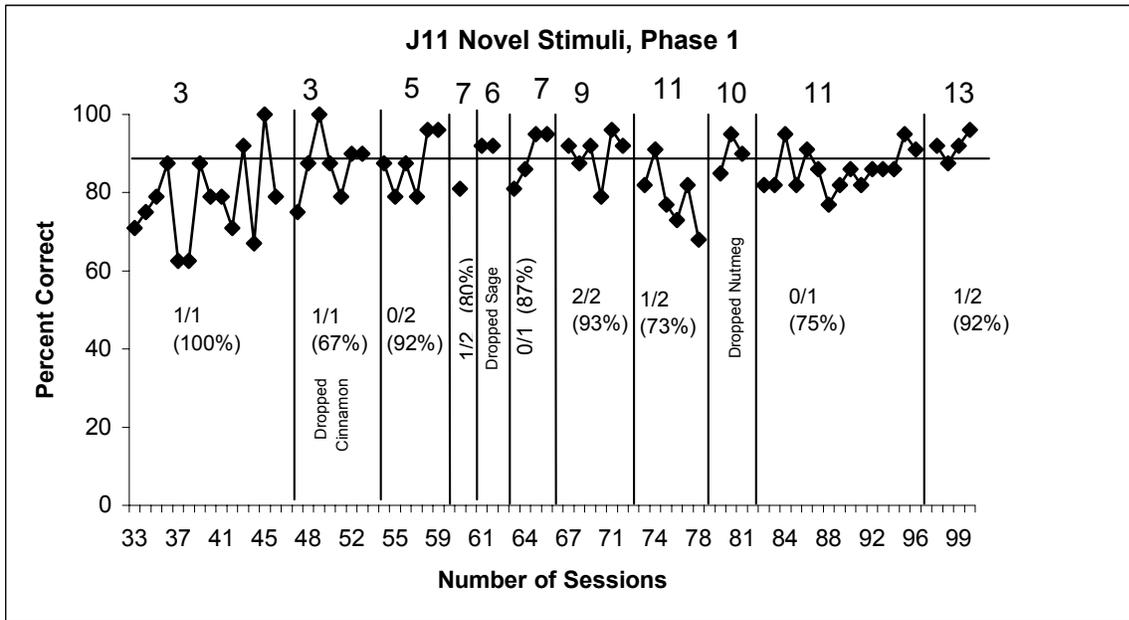


Figure 5. Each graph depicts the session-by session percent correct performance on all trials for subjects J11 and J10 in the Novel Stimuli, Phase 1 of the experiment. The numbers at the top of each panel indicate the number of stimuli. The top numbers within each panel of the graphs depict performances on the novel tests while the bottom numbers in parentheses depict percent correct performances on novel configurations. The horizontal line represents criterion.

correct on baseline and performing at 93% on Novel Configuration trials for the last six sessions. Figure 5 shows performances for Rats J11 and J10. Rat J11 was somewhat slower in developing criterion level baseline performance, requiring 14 sessions of training with three stimuli, and 15 sessions when 11 stimuli were used (during the phase after nutmeg was dropped). This subject was exposed to more test sessions because when certain stimuli were introduced as samples, the rat persistently failed to dig in the sample stimulus. These failures to make the “observing response” were considered incorrect, but after several sessions with each, three odors (cinnamon, sage, and nutmeg) were dropped from this subject’s protocol, and replaced with other stimuli. Due to experimenter error this animal was advanced to Novel Stimuli, Phase 2 without meeting the Novel Probe criterion in Novel Stimuli, Phase 1 (only 4 of his last 7 Novel Probes were correct). Rat J10 showed high levels of accuracy throughout this phase. As he moved from three to seven stimuli, he performed correctly on the first five Novel Probe trials and at 87% on the Novel Configurations, but then responded incorrectly on both Novel Probe trials with 9 stimuli. J10 failed to make the observing response when ginger was the sample, and this odor was removed from his protocol and replaced with cumin. Because of this complication, J10 was advanced from this phase after performing correctly on the last 5 out of 5 Novel Probe trials.

Figure 6 summarizes the results of the Novel Stimuli, Phase 1 conditions, and shows percent correct on Novel Probe trials for the last three criterion test sessions (leftmost bars), and the performance of each animal on all other Novel Configurations (center bars) which included any novel pairing of odors regardless of comparison position. The percent correct on Double-Baited baseline trials for this phase (trials on which both comparison cups contained sucrose pellets) are depicted in the rightmost bars. The results for the Novel Stimuli and Novel Configurations show that subjects J16 and J6 performed at 84% and 87% accuracy respectively,

Figure 6

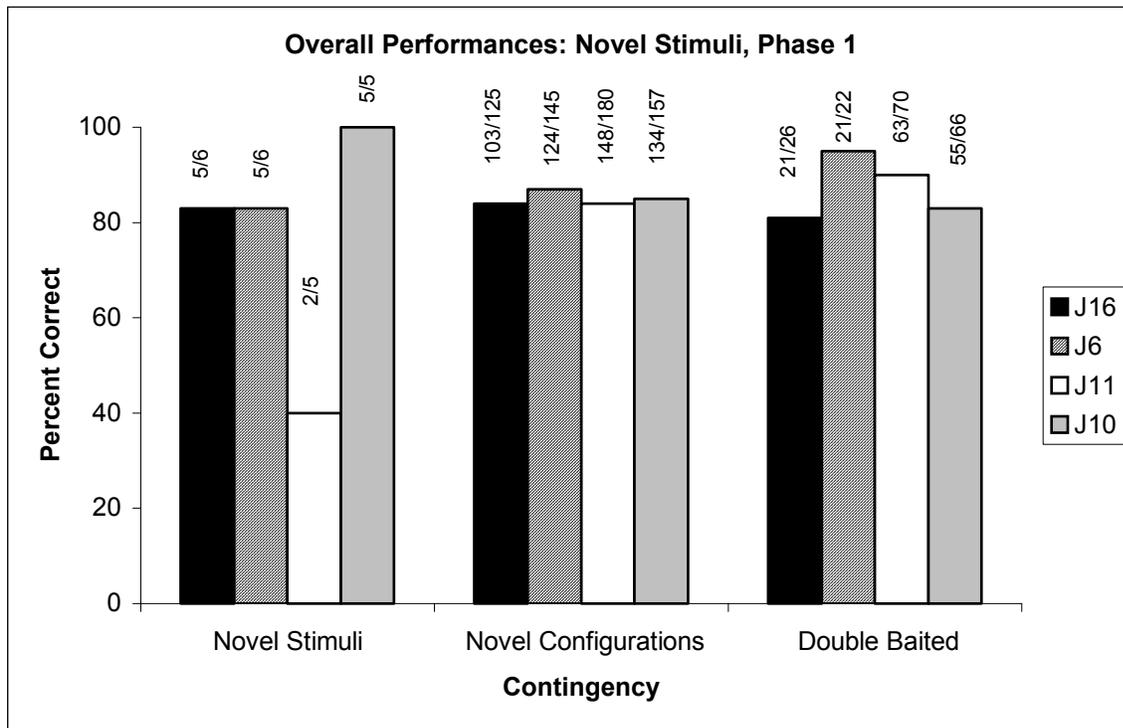


Figure 6. The percent correct performances of each animal for Novel Stimuli, Novel Configurations, and Double Baited trials in the Novel Stimuli, Phase 1 are represented here. The fractions above each bar represent the total number correct for each trial type.

and subjects J11 and J10 performed at 84% and 85% accuracy respectively. Three of the four animals performed well above chance on the Novel Probe trials for an average of 89%.

As evidenced by their performances, all four subjects were able to meet the session-by-session criterion (two sessions at 90% or above), although two of the subjects were advanced from this phase in error before meeting the overall criterion (subjects J11 and J10). In addition, all four animals showed consistently high levels of accuracy on trials involving novel configurations of stimuli. By the end of this phase, subjects were matching at levels well above chance with as many as 15 different odors, and these high levels of accuracy were maintained on double-baited trials. Thus, it likely was the scent of the comparison stimuli rather than the scent of the sucrose pellet that was controlling accurate responding on baseline trials. However, because both the sample cup and the correct comparison cups were baited on Novel Probe trials, it could be argued that the animals were simply digging in the stimulus odor in which digging had most recently been reinforced. In the Novel Stimuli, Phase 2 condition no reinforcer was available in the sample cup on Novel Probe trials to evaluate this possibility. Also in this phase, both comparison cups were baited on Novel Probe trials. Although three of the four subjects performed correctly on at least 80% of these trial types in the previous phase (see Figure 6), these controls were not used on Novel Probes in that phase, so the possibility that subjects' accuracy was based on olfactory control by the pellet in the correct comparison cannot be ruled out on these trials.

The individual results of Novel Stimuli, Phase 2 are shown in Figures 7 and 8. J6 performed correctly on all six of the Novel Probes encountered during this phase and showed high levels of performance on Novel Configurations (115 out of 122 correct) and baseline trials as the number of test stimuli was increased from 17 to 21 (see Figure 7, Panel 2). As Figure 7

shows, Rat J16 met the criterion on Novel Probes when he was advanced to a baseline of 21 stimuli. However, after progressing to 21 different stimuli, baseline performance began to fall off and after 40 sessions without meeting the criterion of two sessions at 90% or higher correct (on average his baseline remained in the low 80's), this rat was dropped from the study. Most likely this animal's slightly lower performance was due to age as he was approximately 16 months old and had completed 160 sessions by this point.

Rat J10 performed correctly on all six Novel Probe trials that involved the novel odors. This subject met the overall criterion by performing correctly on 6 out of 6 tests within the minimum of just three test sessions (see Figure 8, Panel 2). J10 performed correctly on 77 out of 84 Novel Configurations trials and had established a baseline of 21 different odors before being advanced to the next phase of the experiment.

The last rat (J11) also met the overall criterion without sample reinforcement (Figure 8); however, it required several Probe sessions to do so. As he moved from 15 to 25 stimuli J11 was exposed to 12 Novel Probes of which he performed correctly on 9, and ultimately, 5 of his last 6. It will be recalled that Rat J11 was moved to this phase in error without meeting the criteria required of the other three rats, so perhaps it is not surprising that more exposure to Novel Probes was required in his case.

Although he required more sessions to meet the overall criterion than the other animals, J11 was still able to maintain a stable baseline of near or above 90% with as many as 25 different odors and performed correctly on 169 out of 193 Novel Configurations (88%) as well (see Figure 8, Panel 1).

Figure 9 depicts the summary performances for each animal in Phase 2 of the experiment. All four animals met criterion in Phase 2 by obtaining five out of six Novel Probe trials correct.

Figure 7

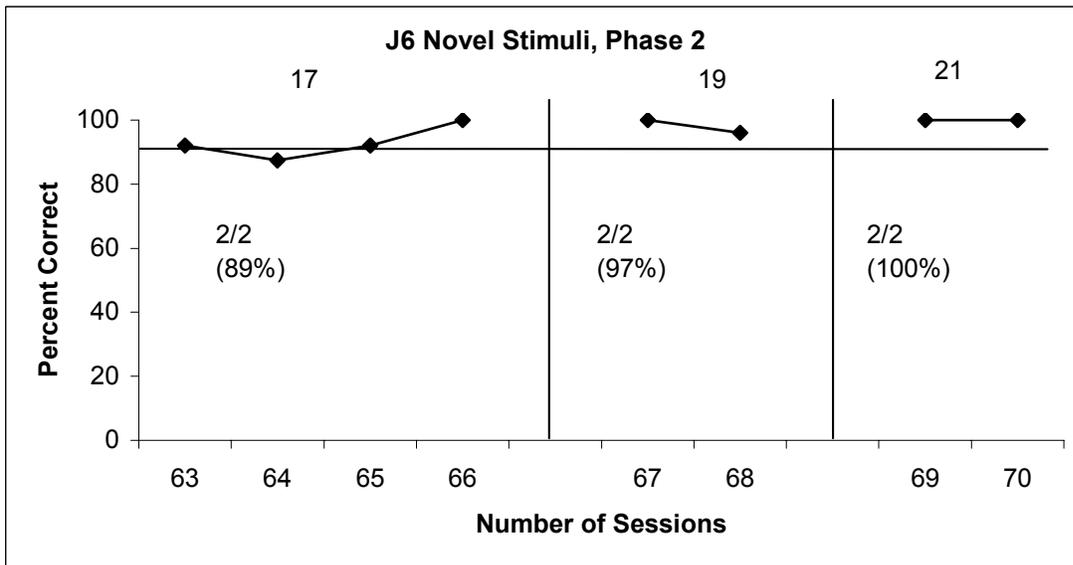
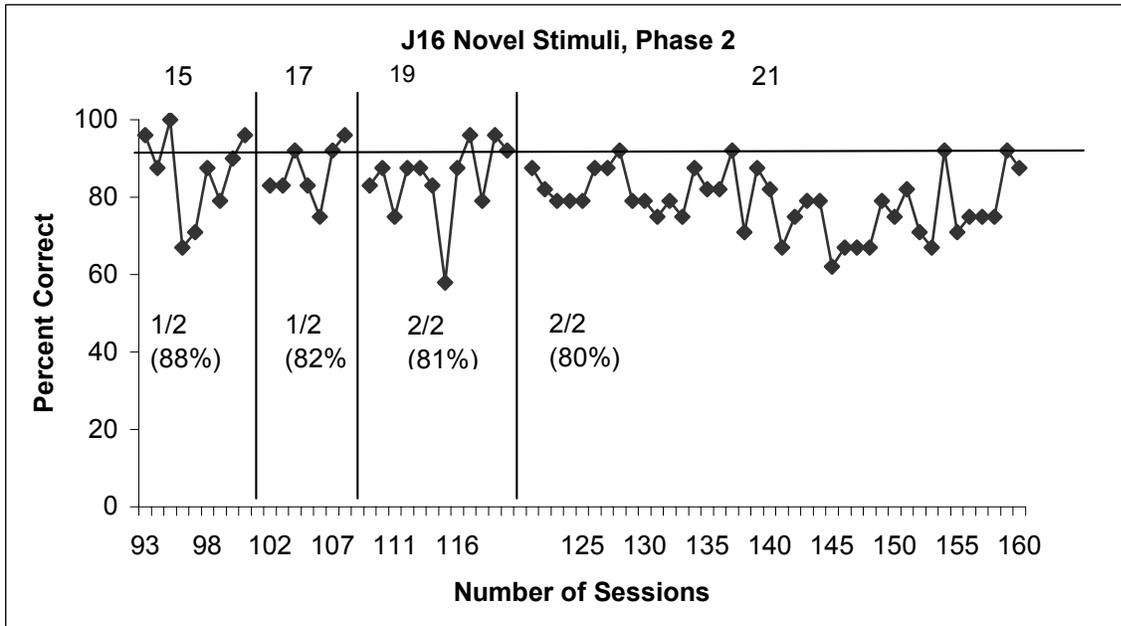


Figure 7. Each graph depicts the session-by session percent correct performance for subjects J16 and J16 in the Novel Stimuli, Phase 2, of the experiment. The top numbers within each panel of the graphs depict performances on the novel tests while the bottom numbers in parentheses depict percent correct performances on Novel Configurations. The horizontal line represents criterion.

Figure 8

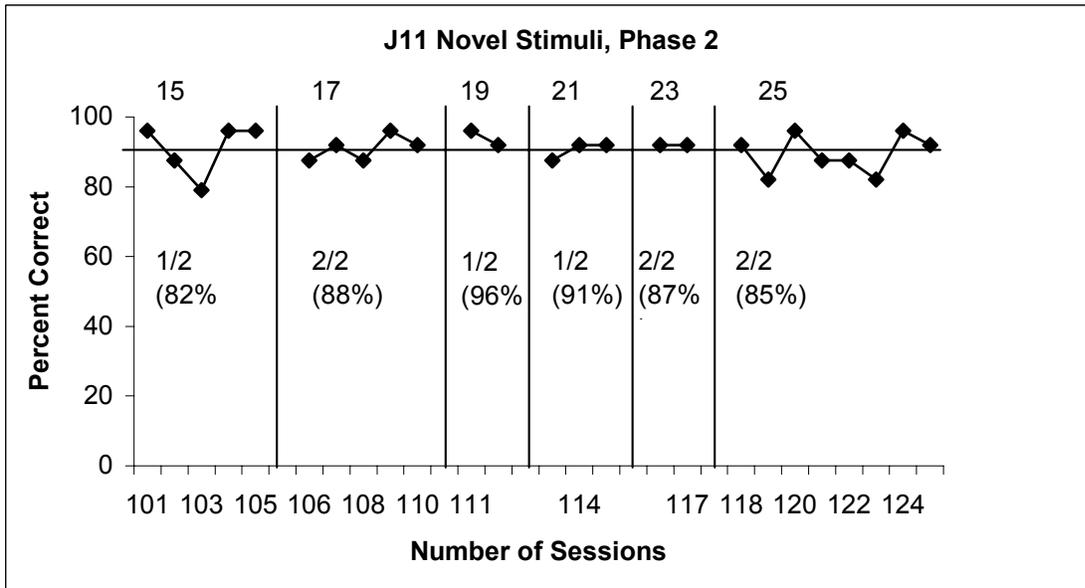
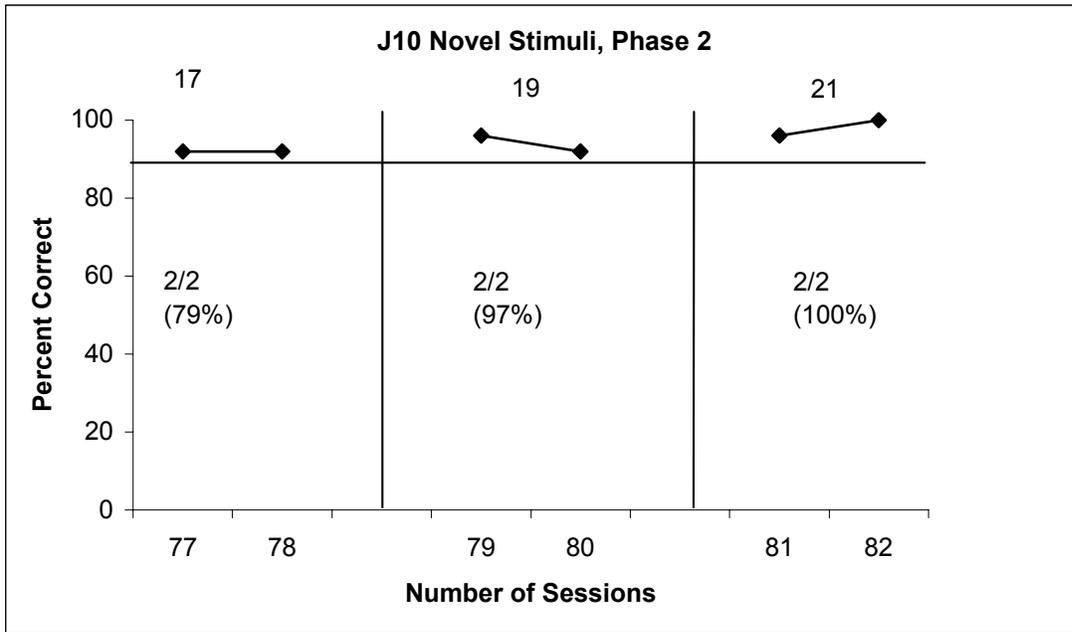


Figure 8. Each graph depicts the session-by session percent correct performance for subjects J11 and J10 in the Novel Stimuli, Phase 2 of the experiment. The top numbers within each panel of the graphs depict performances on the novel tests while the bottom numbers in parentheses depict percent correct performances on Novel Configurations. The horizontal line represents criterion.

Thus, accurate matching with novel stimuli was demonstrated even without reinforcement for responding to the sample, and under conditions in which the odor of the sucrose pellet could not bias responding to the correct comparison. For the Novel Configuration trials all four subjects responded at an accuracy of 80% or higher. In this phase, subject J16 performed at 82% accuracy, J6 performed at 94% accuracy, J11, performed at 88% accuracy, and subject J10 performed at 91% accuracy on Novel Configuration trials. Accuracy on the Double Baited trials for all subjects was comparable to performances with novel stimuli as well as with the overall Novel Configurations.

Although the animals' performances thus far seemed to indicate generalized matching, another possibility could not be ruled out. The Novel/Familiar Probe trials of the experiment were implemented to test whether rats were responding with respect to the novelty of the comparison (e.g., rejecting the novel stimulus, thus, the animal would perform incorrectly on this probe type), and the Familiar/Novel Probes were implemented to further test whether subjects had learned specific configurations of stimuli.

Figures 10 and 11 show the performances for the Novel/Familiar Phase for the three rats tested (note J16 was dropped from the study after failing 40 sessions with 21 stimuli without meeting the set criterion).

Subject J6 met the overall criterion by performing correctly on all 6 tests in this phase in the minimum number of sessions (see Figure 10, Panel 1) and had established a baseline consisting of 27 different odors. Subject J6 was tested for an additional seven sessions after meeting criterion so that filming and inter-rater reliability could be measured.

J11 performed correctly on the first two novel probes in this phase but required several more test sessions. This subject had a total of 35 different odors in his repertoire. As seen in

Figure 9

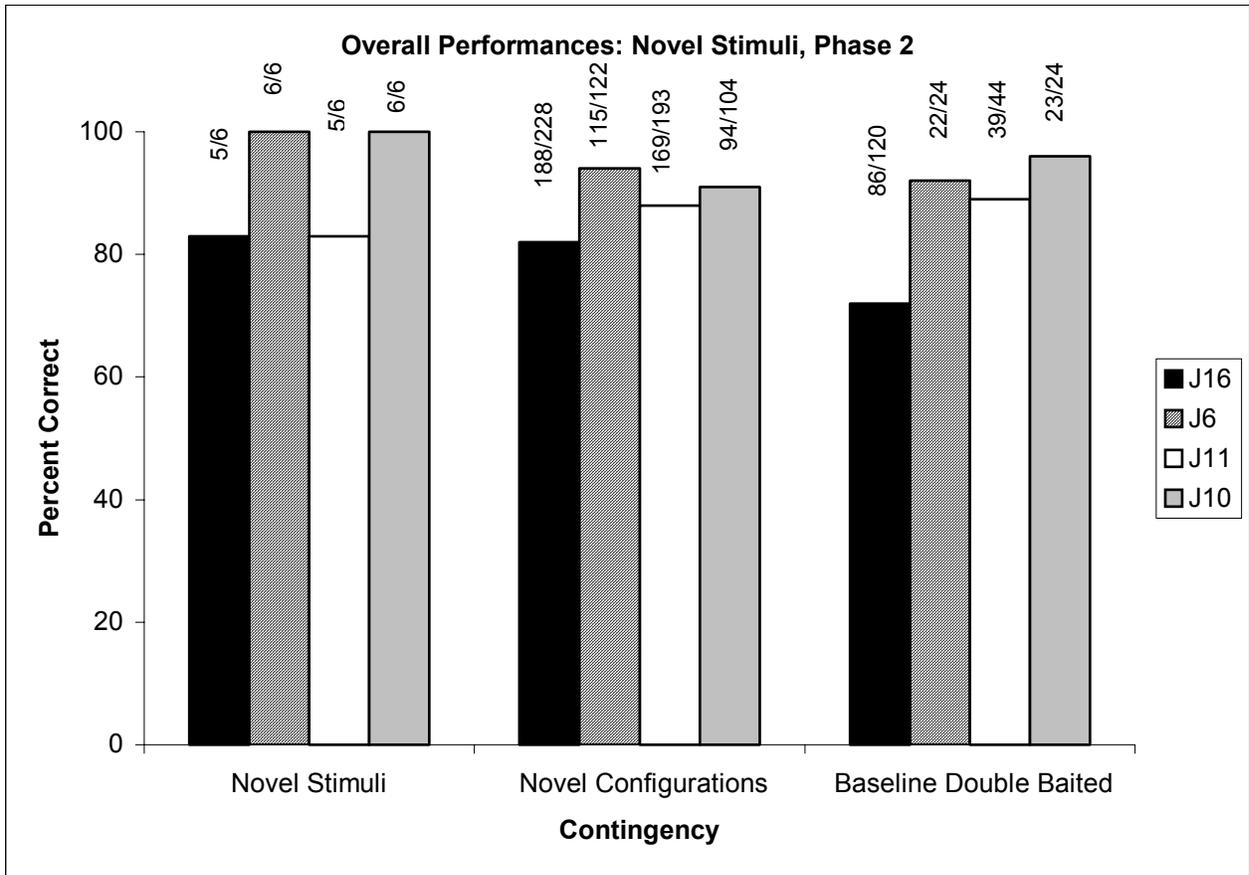


Figure 9. The percent correct performances of each animal for Novel Stimuli, Novel Configurations, and Baseline Double Baited trials in the Novel Stimuli, Phase 2 are represented here. The fractions above each bar represent the total number correct for each trial type.

Figure 10, Panel 2, this subject's baseline faltered somewhat with 29 stimuli but overall he maintained a performance near or above 90% for the remainder of this phase. After performing correctly on the last three out of six tests and after five test sessions this subject was advanced to the Three-Comparison Phase. Although J11 had not met criterion, it was felt critical to assess more than one rat on the Three Comparison condition. Thus, J11 was moved directly to the Three-Comparison Phase.

Subject J10 performed correctly on the first two novel stimulus tests in the Novel/Familiar Phase with 23 stimuli. J10 was taken back to 21 stimuli in error and required three sessions to again meet criterion. This subject performed correctly on all six out of six tests and was responding at near 90% or above with 27 different odors (see Figure 11). Additional sessions were conducted with 27 stimuli for filming and determination of inter-rater reliability. However, after several sessions with 27 stimuli, J10 stopped digging on several trials across several sessions. This behavior continued such that a total of five sessions were discontinued due to lack of responding (not included in Figure 11). It is hypothesized that J10's lack of responding was due to age related issues (he was approximately 18 months old); thus, he was dropped from the study.

Figure 12 shows the overall performances for subjects J6, J11, and J10 in the Novel/Familiar Phase of the experiment. The leftmost bars show that subjects J6 and J10 performed correctly on the last six out of six Novel Probe trials while J11 performed correctly on three out of six Novel Probe trials (this subject, along with J6, was advanced to the Three-Comparison Phase). Each of the animals maintained high performances on the Novel Configurations in this phase of the study, with J6 performing at 100%, J11 at 86%, and J10 at 93%. Subjects J6, J11, and J10 performed correctly on 84%, 68%, and 88% of the Double

Baited trials respectively. Overall, the Novel/Familiar Phase of this experiment was successful in demonstrating that subjects were not solving the MTS problem by rejecting either novelty or familiarity.

In order to provide statistical confirmation of the accuracy of performances with novel stimuli, a one-tailed binomial test was conducted for all novel probe trials beginning with the criterion-level performance of Phase 1 and extending through the end of Phase 2. Three of the four animals showed levels of accuracy that were significantly different from chance (J16, 11/14, $p < .05$; J6, 11/12, $p < .05$; J10, 11/11, $p < .05$). Additional analyses conducted for J6 and J10 adding novel probes introduced during the Novel/Familiar Phase were also statistically significant (17/18, and 19/19, $p < .05$ respectively). Although Rat J11 met the criterion of 5/6 correct on Novel Probe trials in Phase 2, his overall performances on novel probes were not significantly different from chance (Phases 1 & 2: 9/12, $p > .05$; All Phases: 14/22, $p > .05$). Because this subject was advanced from Phase 1 prematurely (after performing correctly on just two out of five novel probes), it could be argued that this limited his chances of performing at statistically significant levels on the novel probes. Another alternative strategy has been suggested as a way in which to solve an identity matching procedure when using olfactory stimuli. In order to test this hypothesis the Three Comparison Phase was implemented. The Three Comparison Phase of the experiment attempted to answer two questions: would subjects be able to maintain their established baselines when more than two comparisons were available, and were they solving the MTS problems by choosing the odor that was most intense? In the Two-Comparison Phase, the sample odor and the matching comparison odor hypothetically produced a more concentrated smell in the testing chamber. Thus, it could be argued that the animal was simply choosing the comparison cup that represented the strongest smell. To control

Figure 10

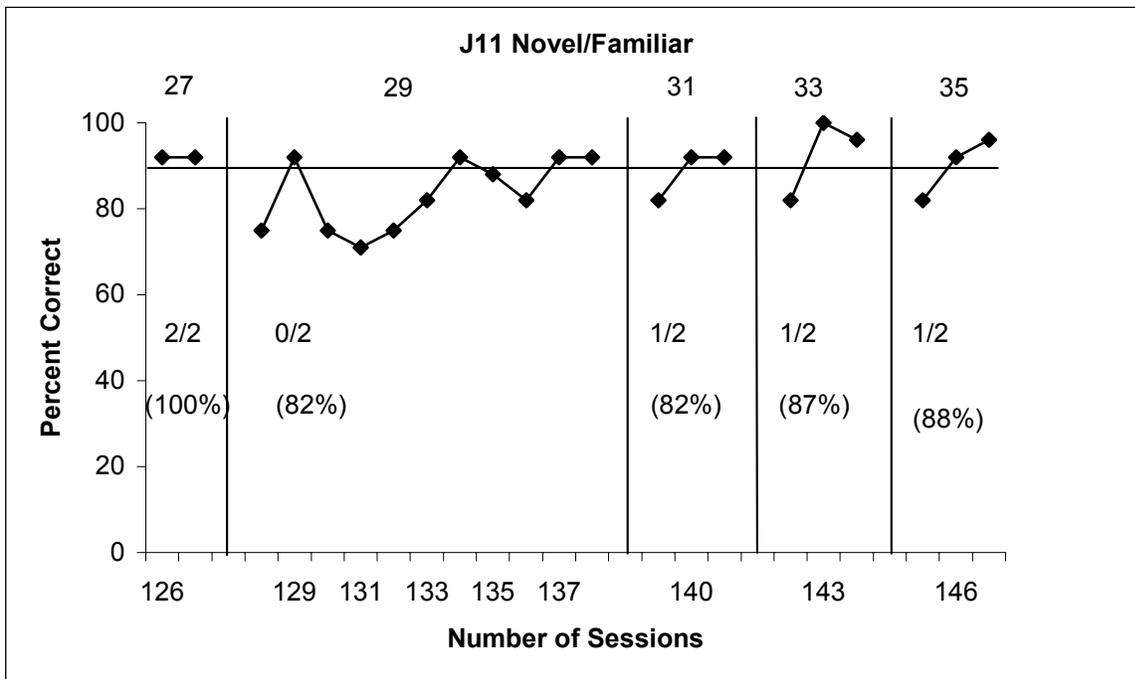
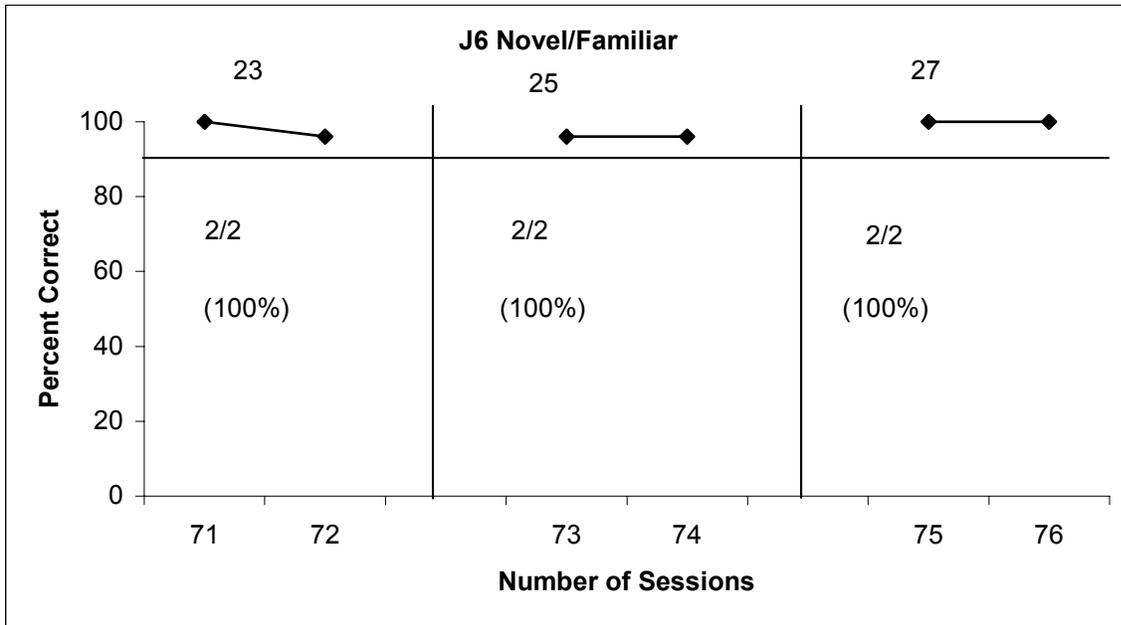


Figure 10. Each graph depicts the session-by session percent correct performance for subjects J6 and J11 in the Novel/Familiar Phase of the experiment. The top numbers within each panel of the graphs depict performances on the novel tests while the bottom numbers in parentheses depict percent correct performances on Novel Configurations. . The horizontal line represents criterion.

Figure 11

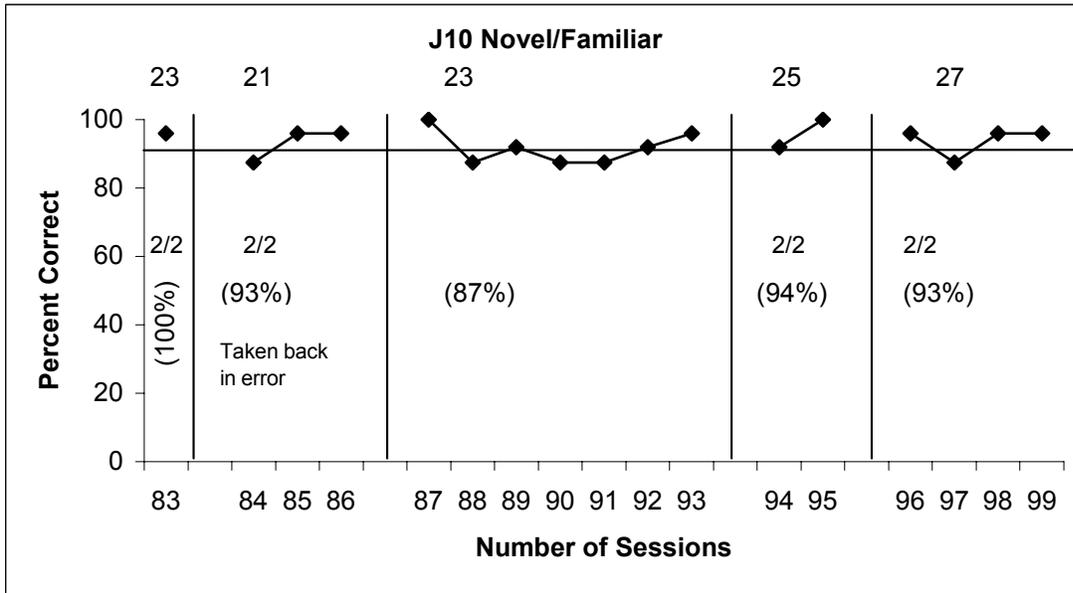


Figure 11. The session-by session percent correct performance for subject J10 in the Novel/Familiar Phase of the experiment is represented. The top numbers within each panel of the graph depicts performances on the novel tests while the bottom numbers in parentheses depict percent correct performances on Novel Configurations. . The horizontal line represents criterion.

Figure 12

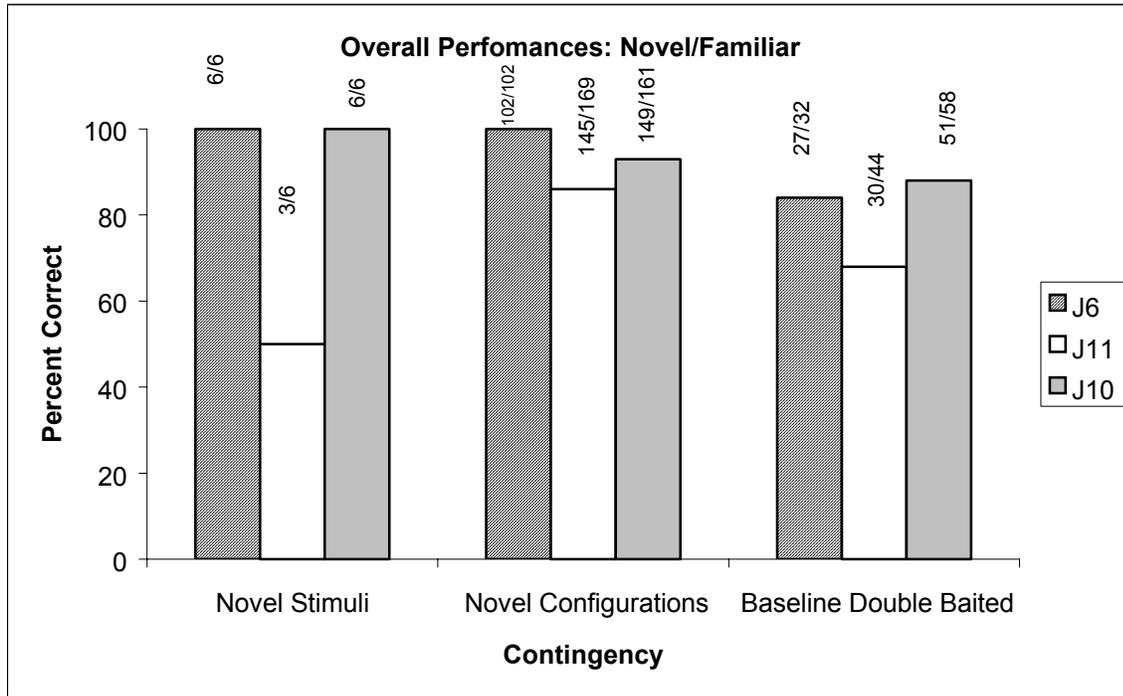


Figure 12. The percent correct performances for subjects J6, J11, and J10 for the Novel Stimuli, Novel Configurations, and Baseline Double Baited trials in the Novel/Familiar Phase are represented here. The fractions above each bar represent the total number correct for each trial type.

for this the Two-Odor Control was implemented. On these trials two of the cups contained one odor, while the remaining two cups contained a second different odor. Only subjects J6 and J11 were advanced to this phase of the study (J11 was moved to the three-choice phase despite not meeting criterion on the previous phase).

The top and middle panels of Figure 13 show the overall percent correct for subjects J6 and J11 on a session-by-session basis, as well as the percent correct for the Two-Odor Control. The bottom panel of this figure depicts their overall performances on the different trial types. Accuracy declined somewhat for J6 when three comparisons were available although his baseline was still well above chance (chance being 33%). This subject required 20 sessions to meet criterion before two novel odors were presented, in which he performed correctly on 1 out of 2 of the tests (see Figure 13, panel 1). J6's overall average baseline performance for this phase was 81% while his overall Two-Odor Control mean for this phase was 78% (see Figure 13, Panel 3). This indicates that this subject was not solving the MTS problem by choosing the odor that was most represented since his Two-odor mean was just slightly lower than his average performance for the entire phase. After five sessions with 29 stimuli the study was ended.

J11 also showed somewhat lower levels of accuracy in the Three Comparison Phase. This subject's performance was above chance levels with an overall phase mean of 79% and an overall Two-odor mean of 75% (see Figure 13, Panel 3). J11's baseline dropped when he developed an aversion to clove and rosemary, which were subsequently removed from his repertoire. This subject still had difficulty meeting the criterion of two consecutive sessions at 90% or above; thus, the criterion was lowered to two consecutive sessions at 88% or above which was still well above chance levels of 33%. J11 required 13 sessions to meet criterion before two novel odors were presented in which he performed correctly on both. After meeting

criterion with 35 stimuli, J11 received one last test in the three-comparison phase performing correctly on both novel trials. His baseline then dropped slightly to 82% and his two-odor control performance was at 75% (see Figure 13, Panel 2).

While subjects J6 and J11 showed a slight decrease from their normal performances it is important to note that they were able to successfully maintain averages well above chance levels with three comparisons, 81% and 79%, respectively. When the trials that were equated for odor intensity were plotted separately, they were near identical to both subject's session-by-session performances. This suggests that neither subject's matching performances were based on odor intensity.

It is clear that because the double baited trials as well as the blind testing trials nearly match the overall phase means for all four subjects, it is unlikely that the animals were detecting the sucrose pellets, or being cued by any of the experimenters. It is important to note that with increasing numbers of odors came increasing numbers of Novel Configurations of stimuli. It is unlikely from these data that the rats had not learned specific configurations of stimuli to solve the MTS problem.

Blind Testing and Inter-rater Reliability:

It was also important to verify that the principle investigator was following the operational definition for a digging response; thus, a second rater was present during several sessions for all subjects. Tables 5 and 6 show the inter-rater agreement for sessions in which there was a separate rater present, as well as the animals' performances on those sessions. There were five sessions for J16 in which there was a separate rater present. One session was conducted with 15 stimuli, while the remaining four sessions were conducted with 21. The inter-rater agreement for all five sessions was 100% (see Table 5).

Figure 13

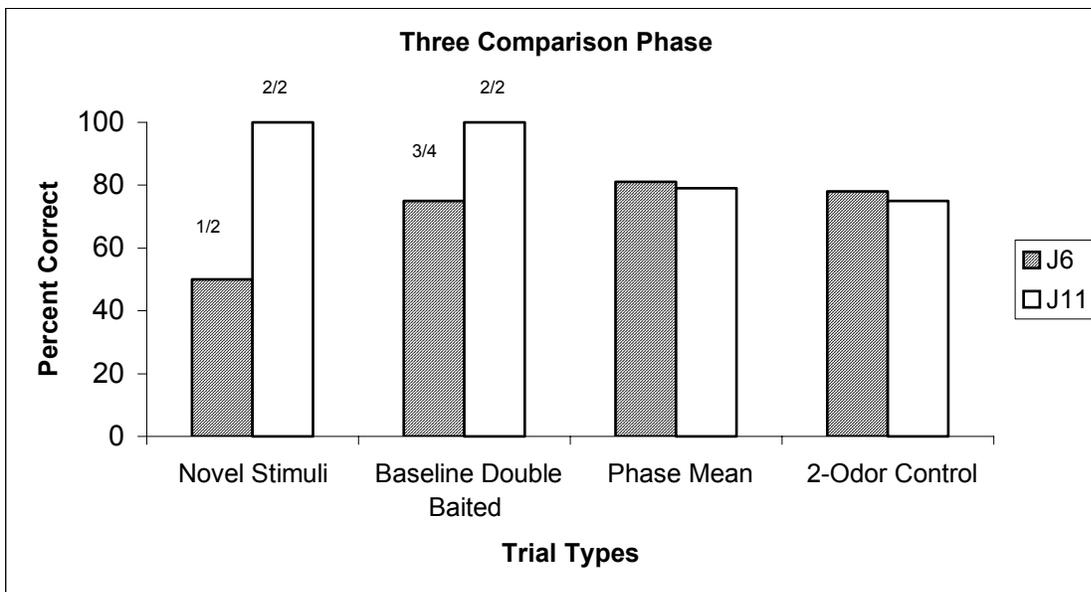
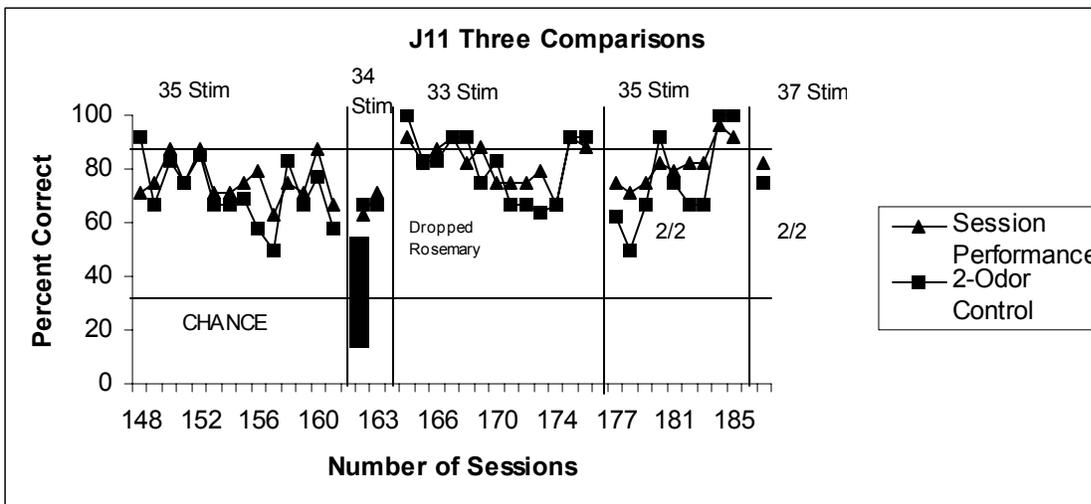
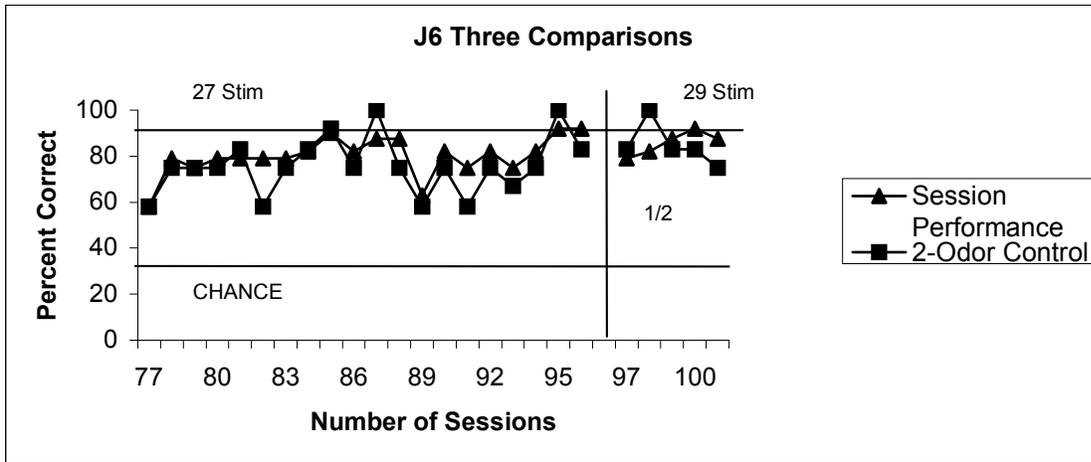


Figure 13. The top two graphs depict the session-by-session percent correct (triangles) and the Two-Odor Control (squares) performances for subjects J6 and J11 in the Three-Comparison phase. The top numbers within each panel of the graphs depict performances on the novel tests. The horizontal line represents criterion. The bottom graph shows the percent correct performances for subjects J6, and J11 for Novel Stimuli, Baseline Double Baited, Blind Tester, and Phase Mean in the Novel/Familiar Familiar/Novel phase. The fractions above each bar represent the total number correct for each trial type.

Table 5 also depicts the number of sessions for subject J6 in which there was a second rater present. There were seven sessions in which this was the case. The average inter-rater agreement was 98% (see Table 5). J11 had a separate rater present for three sessions with the average inter-rater agreement being 97% (see Table 6). Subject J10 had a separate rater for two sessions (one with two stimuli and one with 27). The inter-rater agreement with two stimuli was 90%, while the inter-rater agreement with 27 stimuli was 100% (see Table 6). There were numerous sessions in which testing was conducted with the experimenter blind with respect to the correct comparison position. Subject J16 was tested under such blind conditions for a total of four sessions and maintained an average performance of 81% correct. Subject J6 received a total of seven blind testing sessions averaging 84% correct and subject J11 received a total of five blind testing sessions averaging 74% correct. As seen in Figure 14, performances during the blind sessions were all well above chance accuracy, and performances generally were close to each animal's average performance across the entire experiment (83%, 91%, and 88%, respectively), suggesting that the experimenter was not inadvertently cuing the subjects.

Discussion

Four rats initially started the experiment with a baseline of just two odors. Each animal readily met the criterion of two consecutive sessions at 90% or above before novel odors were added. As the number of odors added to each animal's repertoire increased, their baseline accuracy increased as well. Although high levels of baseline accuracy with multiple stimuli is consistent with generalized matching, this could also be evidence for rapid learning of configurations or of rapid "if...then" relational learning. To rule out this possibility, several experimental manipulations were implemented. The first of such manipulations evaluated the animals' performances on the first trial in which novel odors were introduced to the existing two-

Table 5

J16 Inter-rater reliability

Number of Stimuli	Inter-rater agreement	S Performance
15	100%	100%
21	100%	79%
21	100%	79%
21	100%	88%
21	100%	79%

J6 Inter-rater reliability

Number of Stimuli	Inter-rater agreement	S Performance
17	100%	88%
19	100%	100%
21	100%	88%
27	88%	88%
27	100%	88%
27	96%	92%
29	100%	92%

Table 6

J10 Inter-rater reliability

Number of Stimuli	Inter-rater agreement	S Performance
2	90%	70%
27	100%	90%

J11 Inter-rater reliability

Number of Stimuli	Inter-rater agreement	S Performance
2	100%	90%
25	96%	92%
35	96%	75%

Figure 14

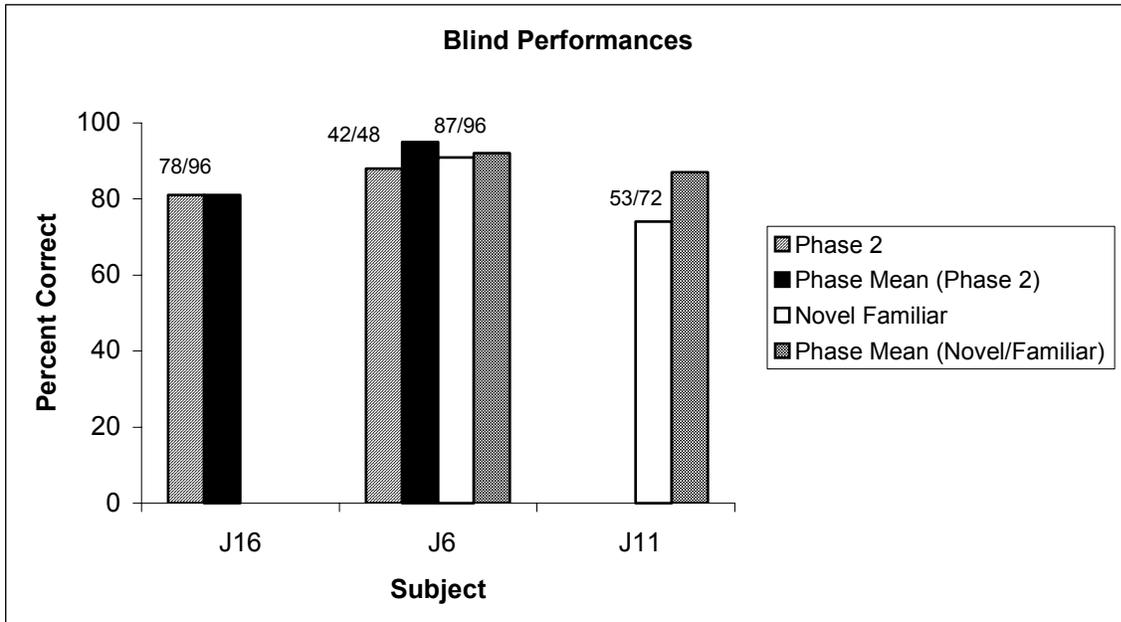


Figure 14. Average percent correct performances for three experimental subjects on blind testing sessions, as well as the average performances for corresponding phases are represented here. The striped bars represent the average for blind sessions in Phase 2, while the solid black vertical bars represent the overall average performances for Phase 2. The solid white vertical bars represent the average for blind sessions in the Novel/Familiar Phase, while the checkered bars represent the overall performances for the Novel/Familiar Phase. The fractions above the bars indicate the number of trials correct.

odor baseline (Novel Stimuli, Phase 1). Three of the four subjects showed highly accurate responding on the novel tests (the fourth subject, J11 was advanced in error after performing correctly on 2 out of 5 tests). All four subjects also exhibited high performances on the Novel Configurations (at 82% or above).

By the end of Phase 1, all four subjects were matching accurately with a total of between 13 and 15 stimuli. It was important to ensure that the rats' behavior was not simply controlled either by establishing a "win-stay strategy" or by detecting the odor of the pellet. Two different methods were implemented to avoid the animals adopting either of these strategies, a baseline of 50% reinforcement in the sample cup and the introduction of Novel Stimuli, Phase 2. In this phase subjects were exposed to novel tests in which there was no reinforcer in the sample and both comparison cups were baited to further rule out the possibility of pellet detection. It was found that all four subjects performed on average at 82% or above on the Novel Configurations. These high performances suggest that subjects had not established a win-stay strategy nor were they simply detecting the reinforcer.

As the experiment progressed and more stimuli were added to the baseline, all four subjects required less time to meet the overall criterion. For example, in Novel Stimuli, Phase 1, the animals required an average of 7.5 test sessions to meet the overall criterion. However, in the Novel Stimuli, Phase 2 part of the experiment the rats required an average of just 4 test sessions to meet the overall criterion. This was interesting because the MTS sessions were progressively becoming more complex. It is important to note that as more stimuli were added to each animal's baseline, there were more instances in which there were completely novel configurations of stimuli. Because the rats were meeting criterion at increasingly faster rates, this suggests that they were indeed learning generalized matching. Finally, statistical analysis of

Novel Probe trials revealed that three of the four rats (J16, J6 & J10) performed at above chance levels, supporting the claim that at least these three provided demonstrations of generalized matching.

However, as Dube et al. (1992) suggested, subjects in an identity MTS procedure could adopt still other strategies to solve an identity matching probe trial. For example, subjects may choose the correct stimulus by rejecting the comparison that is unfamiliar, or they might solve the novel problem by rejecting the baseline stimulus for the comparison that is novel. Thus, the Novel/Familiar Familiar/Novel phase of the experiment was implemented. On test days, the two probe trials consisted of one trial that had a novel odor as the sample and familiar odor as the alternative comparison, while the second probe trial consisted of a familiar odor as the sample while a novel odor was the alternative comparison. Two of the three animals that made it to this phase met the set criterion within the minimum number of test sessions, while the third animal failed to meet the criterion.

For two of the animals the experiment was extended to include a three-comparison task. In this phase half of the trials contained three odors (sample, matching odor, and two different alternative odors) and the remaining half contained just two odors (sample, matching odor, and two of the same alternative odors). This was implemented not only to determine whether they could maintain their baseline performances with three comparisons, but also to test whether they would simply choose the odor that was most represented (two-odor control). Subjects J6 and J11 both had baseline performances that were below their normal performances in previous phases, however, their performances were still well above chance (with averages of 81% and 78% respectively). When their two-odor control performances were analyzed separately, it was evident that their performances were not controlled by digging in the odor most represented

because performances on these trials nearly matched those of their session-by-session performances (J6: 78%, and J11: 75%).

Thus, the present study has found evidence of generalized matching in rats when using olfactory stimuli. Further, high inter-rater reliability and accuracy under blind testing conditions suggests that experimenter bias cannot account for the high success rates in three of the four rats.

The successful MTS performances of the rats in the present study may be due to the multiple exemplar approach (Wright et. al, 1988). The use of olfactory stimuli may have also been important. For example, rats in the present study required fewer sessions on average to meet the training criterion with olfactory stimuli than those in the Iversen (1993) study that used visual stimuli (an average of 37, 24-trial sessions versus 61, 100-trial sessions respectively). Because there were substantial savings that occurred as new odors were introduced, multiple exemplar training was possible as opposed to the training in the Iversen (1993) study in which many sessions were required for the rats to meet criterion with just two stimuli. Rats in the Iversen (1997) study performed poorly when the sample location varied, suggesting they had learned the specific configurations of stimuli. However, by the end of the present study, rats were exposed to numerous Novel Configurations of stimuli within a single session while still maintaining very high baseline performances. This suggests that the rats were not simply learning specific configurations of stimuli like those in the Iversen (1997) study. It is evident that rats require considerably more sessions to meet criterion when visual discriminations are used thus limiting the number of stimuli that may be added to an existing baseline. By using olfactory stimuli, repertoires consisting of a large number of odors were quickly established, which seems essential for generalized matching-to-sample.

The present findings support those of Lu et al. (1992) in which a go/no go procedure was implemented using olfactory stimuli. Lu et al. (1992) found that rats were able to maintain their performances with as many as 30 odors even when a 10 s interstimulus delay was introduced. Lu et al. found that when rats were exposed to the 3-odor phase, two of the three animals performed at 90% or above in the first session. Although the Lu et al. study showed excellent performances on a same/different discrimination task with a large number of stimuli, control by specific “if – then” rules were not ruled out because rats’ performances on the first trials in which novel stimuli were introduced were not presented. The present results extend the Lu et al. findings in that not only were rats able to maintain very high baselines with increasing numbers of stimuli, their performances on the first trial in which a novel odor was presented were just as impressive. Thus, the use of olfactory stimuli may be critical to the successful performances of the present study. However, it should be noted that Nakagawa (2000) has claimed generalized matching in rats with visual stimuli. So the question of whether the olfactory stimulus modality is somehow privileged in rats remains open.

There have been several studies involving various different species of animals in which the claim for generalized matching has been made; however, it seems that only a select few should make this claim. For example, even when held to Dube et. al’s strict criteria, it seems that apes (and perhaps *Cebus apella*), dolphins, sea lions, and even rats (when olfactory stimuli are used) are capable of demonstrating generalized matching-to-sample. Perhaps it is not that only certain species are capable of this task, but that it is the way in which different species are tested. For example, when procedures were used that facilitated learning for a particular species, results were produced that suggested generalized matching. The current discussion has demonstrated

that many of these studies fall short for one reason or another when held to the strict criteria set by Dube et. al (1992), (e.g., Zentall and Hogan, 1974; D'Amato, et. al 1985; Nakagawa, 2000).

There were, of course, several limitations of the present study. For example, each of the rats developed aversions to different odors. Because each of the odors needed to be completely novel when used to test for generalized matching, aversion data could not be collected. Also, the advanced age of the subjects may have affected the performances of at least two of the rats. It is hypothesized that J10 and J16's lack of responding at the end stages of the study was not due to a failure in the experimental manipulations, but rather a result of the subjects' age (20 months and 16 months respectively). All four animals began this experiment when approximately 4-6 months old. Thus, by the end of this study all four rats were approximately two years old. It is possible that any declines in performances were simply due to this factor.

Although the rats in the present study showed evidence of generalized MTS through the use of multiple exemplar training, their performances may most likely be limited to the modality of olfaction. No attempt was made to assess performances with any other modality, and as Iversen (1997) demonstrated, rats required many sessions to acquire just two conditional discriminations when visual stimuli were used. Another limitation was that most sessions were not conducted under blind conditions. The inter-rater reliability and blind testing data collected in the present study, however, support the validity of the present outcomes. Nevertheless, the use of an automated procedure would greatly reduce the likelihood of experimenter error as well as eliminate the possibility of experimenter cuing.

Because it is now known that rats readily acquire olfactory discriminations, perhaps it would be wise to begin with a baseline consisting of more than just two odors. One future direction would be to test rats using the MTS procedure to study stimulus equivalence. Sidman

and Tailby (1982) argue that stimulus equivalence can be defined by three features: reflexivity, symmetry, and transitivity, and an equivalence class can be defined as a group of stimuli that are functionally interchangeable with each other. More importantly, it is argued that reflexivity (or identity matching) may be a prerequisite for equivalence. The use of an olfactory identity matching procedure may be the key in demonstrating stimulus equivalence (as defined by Sidman and Tailby) in rats.