Odor Based Behavioral Tasks Confounded by Distance Dependent Detection: Modification of a Murine Digging Paradigm

EMILY KATZ, OLIVER ROTHSCHILD, ANDRIANA HERRERA, SOFIA HUANG, ANNA WONG, YVETTE WOJCIECHOWSKI, AIDA GIL, QI JIANG YAN, and ROBERT P. BAUCHWITZ

St. Luke's-Roosevelt Institute for Health Sciences, Columbia University, New York, NY

This work was supported by a grant from the FRAXA Research Foundation.

Correspondence regarding this article should be addressed to R. Bauchwitz, Department of Neurology, St. Luke's-Roosevelt Institute for Health Sciences, 432 W. 58th St., Room 411, New York, NY 10019 (212-523-8869; rpb3@columbia.edu)

[Suggested Running Head: Murine Olfactory Sequence Tasks]

ABSTRACT

Tests of murine higher order cognition are important in order to fully explore the effects of genetic alterations and potential therapies in models of human mental retardation and other deficits of intellect. In this study, inbred and F1 hybrid strains of mice were assessed in a five odor sequence task previously described for rats which employs buried food rewards. Data are provided indicating that the use of buried rewards in mouse olfactory digging tasks should be approached with significant caution. An alternate test of sequence dependent working memory (2-sequence task) using a post choice, drop-in food reward is described. After achievement of stable average plateau scores, a streak and slump pattern, rather than near errorless performance, was observed. Therefore, mice do not appear to effortlessly acquire sequences of odors in a working memory paradigm in which rewarded odors are constantly changing.

INTRODUCTION

Mouse models of human learning disorders are becoming increasingly heavily studied as a means to understand the basis of intelligence and its deficits (McIlvane & Cataldo, 1996). It has been argued that more complex cognitive tasks than commonly used will be required to properly assess whether mouse models show a deficit comparable to that seen in humans with mental retardation (McIlvane & Cataldo, 1996; Strupp & Levitsky, 1990). For example, humans with Fragile X Syndrome (FXS), the most common cause of inherited mental retardation, are capable of near normal performance on simple matching to sample tasks, but show a more pronounced deficit when the task becomes more difficult, such as recalling sequences of information (Dykens, Hodapp, & Leckman, 1987; Hodapp et al., 1992; Kemper, Hagerman, & Altshul-Stark, 1988). In our study of mouse models of FXS, we have shown that performance deficits are negligible or absent in basic tests of murine cognition such as the Morris water maze, novel object task, and 8-arm radial maze (Yan, Asafo-Adjei, Arnold, Brown, & Bauchwitz, 2003). Therefore, a more rigorous, complex cognitive task was sought to test FXS mouse models.

Rodents have generally performed poorly in tests of higher order cognition, such as learning set formation, when compared with primates, birds and other animals (Warren, 1973). One possible reason for this difference may be the mode of sensory stimulus employed (Jennings & Keefer, 1969; Nigrosh,

Slotnick, & Nevin, 1975). Rodents have relatively poor visual acuity (Artal, Herreros de Tejada, Munoz Tedo, & Green, 1998), and when tested on visual stimuli, they perform poorly compared to birds or primates. Olfactory based tests may allow a more accurate measure of rodent cognitive abilities (Nigrosh et al., 1975). In light of the results of odor-based tests of cognition, it has been proposed that olfactory learning in rodents can serve as a model for studies of cognition in humans, just as visual learning does in other primates (Slotnick, 1994).

Early rodent olfactory learning tasks required discrimination between two odors (Jennings & Keefer, 1969). Subsequent studies suggested that rats could learn how to choose odors even when those were changing across tests, i,e. the rodents had acquired learning sets (rule based performance), often within a relatively short number of trials (Nigrosh et al., 1975; Slotnick, Hanford, & Hodos, 2000). Acquisition of learning sets is important to the assessment of intelligence, as it is thought to require development of a cognitive skill which allows interproblem learning (Reid & Morris, 1993; Slotnick, 2001; Warren, 1973). A variation on a matching-to-sample test was developed to test working memory for continually changing sequences of olfactory information (Fortin, Agster, & Eichenbaum, 2002). In the Fortin et. al. task, rats were exposed to a series of five odors in a sequence denoted A-B-C-D-E. They were then tested in discrimination tasks in which the rat would be rewarded with buried food for choosing the single odor which came earlier in the sequence of five, (i.e. A vs C,

D or E; B vs C, D or E; or C vs D or E). When we employed a version of this task scaled for mice, we found that modification was needed to prevent the mice from using olfactory cues rather than memory to achieve high performance. Furthermore, we observed that presenting a subset of all possible pairs of odors could allow the rodent to achieve high scores without requiring recall of the full 5 odor sequence. An alternative sequential memory procedure for mice is described here, as well as general data relevant to olfactory cueing in other tasks such as radial mazes. The data presented here indicate that mice may not have a special preparedness to recall sequences of olfactory information. Implications of our tests for the debate over a special role for olfaction in learning set acquisition in rodents are discussed.

EXPERIMENT 1

Method

Subjects . FVB/NJ, C57BL/6J and C57BL/6J Fragile X (*fmr1-tm1Cgr*) mice were obtained from Jackson Laboratories (Bar Harbor, Maine). Male C57BL/6J *fmr1-tm1Cgr* mice ("ko" or "*fmr1*") were bred to wildtype ("wt") female C57BL/6J mice to produce females heterozygous for the *fmr1-tm1Cgr* mutant allele. Male FVB/NJ mice were bred to heterozygous C57BL/6J *fmr1-tm1Cgr/+* female mice to produce litters with approximately half wt and half *fmr1* mutant males. At weaning (3 weeks of age), tail and toe clippings were taken to genotype and identify the animals. Females, and noticeably runted male mice, if any, were discarded and the males housed in litter specific groups of up to five. Food and water were supplied ad libitum until one week prior to testing. At approximately five months of age, sixteen adult males weighing approximately 32 grams each were taken from a barrier facility to a room for cognitive testing, at which time they were housed individually. Light cycle was on beginning 7AM for 12 hours. All testing occurred during the light cycle. Treatment of the subjects was approved by the St. Luke's-Roosevelt Institute for Health Sciences IACUC and in accordance with APA ethical standards.

Apparatus. Odor cups (Dixie 3 oz, 4.5 cm diameter, Fort James Corp., Norwalk, CT) were made by removing all material above 1.4 cm. An exposure

tray to hold a single cup was made by cutting two pieces of corrugated polypropylene (Laminacorr, Cornwall, Canada) to 8 x 6.5 cm. The outline of a cup was traced on one piece and the circular plastic under the cup excised. The two pieces were taped together in order to hold one cup (Figure 1). A two cup testing tray was made by cutting a corrugated polypropylene sheet to 17.5 x 8.5 cm. The 17.5 cm length was sliced on one side at 9 cm from one end (8.5 cm from the other) with a razor blade. The plastic was then bent to 90 degrees at the slice and taped in place with Manco HP260 Crystal Clear tape. With the 8.5 cm piece as the base and the 9 cm piece projecting vertically as the back, a 8.5 x 1.5 cm strip of plastic was taped on the base, 5 cm from the back, to provide a restraint for the three test cups (Figure 1).

Odorants. The gravel in each cup had a distinct odor, generally from a spice mixed into it. The odorants employed were, in order: ginger, chili, ground cloves, garlic powder, mace, turmeric, coriander, ground sage, curry, paprika, oregano, cardamom, tarragon, parsley, garam masala, mustard, cumin seed, cayenne pepper, cilantro, cinnamon, basil, course ground black pepper, dill weed, rosemary (crushed with mortar and pestle), thyme, poultry seasoning, onion powder. Nutmeg was excluded after preliminary odorant detection trials suggested that several mice were averse to it. Three to four taps of spice from a 1.5 ml tube (approximately 25 - 35 mg) were mixed into the gravel with a

toothpick; the resultant spice odor was detectable by human operator. The term "odor" will henceforth be used to refer to the odorant spices mixed into gravel.

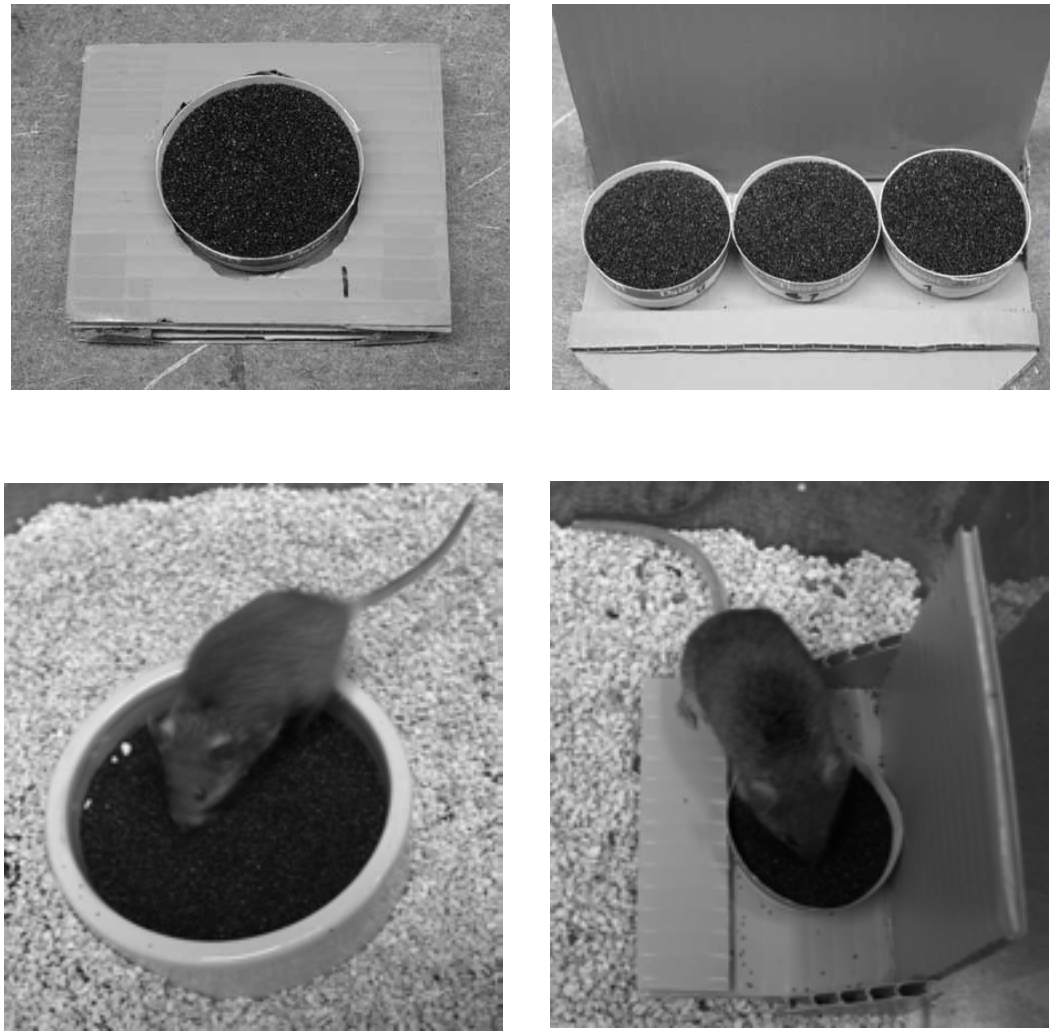

Figure 1. Olfactory digging task cups and trays. Top, left: Single odor exposure cup and tray used for 5-sequence and 2-sequence tests. Top, right: Three cup tray used for 2-sequence learning and memory task. The 5-sequence task used a similar tray wide enough to hold two cups. Bottom, left: Mouse on rat-sized cup with 3.0 cm scented gravel. Bottom, right: Mouse on mouse-sized cup with scented gravel.

Five odor sequence task. The method was that of Fortin et. al. (Fortin et al., 2002) with some modifications. The task consists of two phases: an exposure phase in which the mouse was sequentially exposed to 5 odorants mixed into gravel, followed by a test (probe) phase in which the mouse was presented with pairs of the odors from the exposure phase and required to chose the earlier odor to obtain a food reward buried in the gravel. Each day a mouse was presented with six sequences of five odors (a trial); each sequence had a distinct order to the odors, and 26 of the 30 odors used in each trial were unique; the last sequence reemployed four of the odors from the first sequence, but in a different order. The odors were shuffled every day as described below. A fixed delay between exposure and test phases was used for a given experiment, but was varied in different experiments (e.g. 15 seconds, 3 minutes, or 30 minutes). Each experiment continued until a stable average performance score was obtained (see below).

To begin an experiment, free fed mice were food restricted (3 days without food followed by 2 grams/day of PicoLab Mouse Diet 20, (PMI Nutrition International #5058; www.labdiet.com) in order to achieve approximately 85% of free feeding weight. Mice were then shaped to dig for one-sixteenth piece of cereal (FrootLoops, Kellogg Co., Battle Creek, MI; approximately 5mm x 5mm x 2-3mm, average weight 11 mg) by placing the cereal increasingly deep in non-toxic aquarium gravel (Estes' UltraReef Marine Sand, Totowa, NJ; approximately

23 g and 1 cm deep; Figure 1). Next, the mice were taken through 4 sequences (of 5 odors each, i.e. a trial) with the food visible in the correct cup (a “shape”), followed by one buried reward sequence, and concluded with a visible reward sequence. On the next day, the mice were given a visible reward sequence followed by an invisible one, which concluded training; following this, testing with only buried rewards began.

Animals were exposed to five odors in sequence (denoted A-B-C-D-E) in the single odor tray using baited cups. Cereal rewards for the exposures were half that of the test probes (1/32 lengthwise slivers). Each trial consisted of exposure to six distinct odorants followed by 6 pair-wise probe tests for working memory o the sequence of odors: B vs. E, A vs. D, A vs. C, C vs. E, B vs. D, and A vs. E. If an animal obtained more than 65% correct in the hidden reward trials (2 or fewer incorrect out of 6 sequence probes), then another invisible trial was performed. If however, an animal got less than 65% correct, then a visible trial would commence. One point was given for each correct answer; each day’s session had up to 36 points (6 trials of 6 discriminations). This cycle was continued until an animal achieved criterion (>80% correct). Animals were then followed for at least four days upon achieving criterion to establish that they could consistently recall the odors (plateau performance).

Odor cups were arranged on the bench in a 5 x 5 grid, (with an extra, 26th odor cup to the side). Columns were labeled A - E (as in Fortin et al., 2002). Five

trials of five odors were conducted from top row to bottom, left to right. The sixth trial was conducted by using the 26th odor as cup A1, and then continuing with the grid as previously; this frame shifted the odor reward values (for cup position versus potential reward) by one. The next day, the columns were shuffled from the 5 x 5 grid present at the end of the prior day (i.e. with cup 25 to the side after day one). Each day, some of the shuffled columns were also inverted (denoted by an "f" for flip) as follows: ABCDE -> Bf, C, Ef, C, Af. In the test tray, a pseudorandom sequence of positions for the rewarded cup were used (left/L or right/R): odd trials: L-R-R-L-L-R and even trials: R-L-R-L-L-R.

Analysis. All odor tests were conducted by experimenters who were blind to the genotype of the mice. As no statistically significant difference in performance between wild type and *fmr1* mutant mice was observed, the data were grouped for some of the analyses presented here. In all statistical tests, specificity was set with significance level of $\alpha = 0.05$. All error bars in the figures represent one standard error of the mean (S.E.M.). An asterisk (*) in the figures indicates the following statistical probability: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Results and Discussion

A five odor sequence working memory task (the "5-sequence" task) was performed with FVB/NJ x C57BL/6J F1 hybrid mice essentially as described for

rats (Fortin et al., 2002). Half of the mice were littermates carrying the *fmr1-tm1Cgr* insertional mutation ("ko") which disrupts *Fmr1*, a gene responsible for Fragile X Syndrome (FXS) in humans when inactivated. As a pre-training control experiment, the mice were presented with baited cups containing gravel with no scent, one common scent (sage), or one scented and one unscented cup (Figure 2). One sample t-tests comparing the means to 50% for random choices among pairs showed no significant difference: no odor ($M = 54.11$, $SD = 13.67$), $t(9) = 0.95$, $p = 0.37$, sage ($M = 50.00$, $SD = 17.48$), $t(9) = 0$, $p = 1.0$, combined ($M = 52.01$, $SD = 15.12$), $t(15) = 0.54$, $p = 0.59$. Thus, the results were random choices, as expected if the animals were not using the scent of the buried food reward to choose the cup in which they dug.

After the pre-training control test, the 5-sequence task was performed (Experiment 1, Methods). Criterion of greater than 80% correct was reached within two sessions (days). Plateau performance over the final 9 days was very similar to that previously reported with rats (Fortin et al., 2002) for both wildtype (wt) and *fmr1-tm1Cgr* mice, i.e. near errorless performance (Figure 2). There was no significant difference in working memory performance by genotype in this task as assessed with a two-tailed t-test ($\alpha = 0.05$): mean correct response for wt ($M = 88.5\%$, $SD = 4.9$), ko ($M= 88.9\%$, $SD = 5.6$), $t(13) = -0.53$, $p = 0.60$, $\eta^2 = 0.02$.

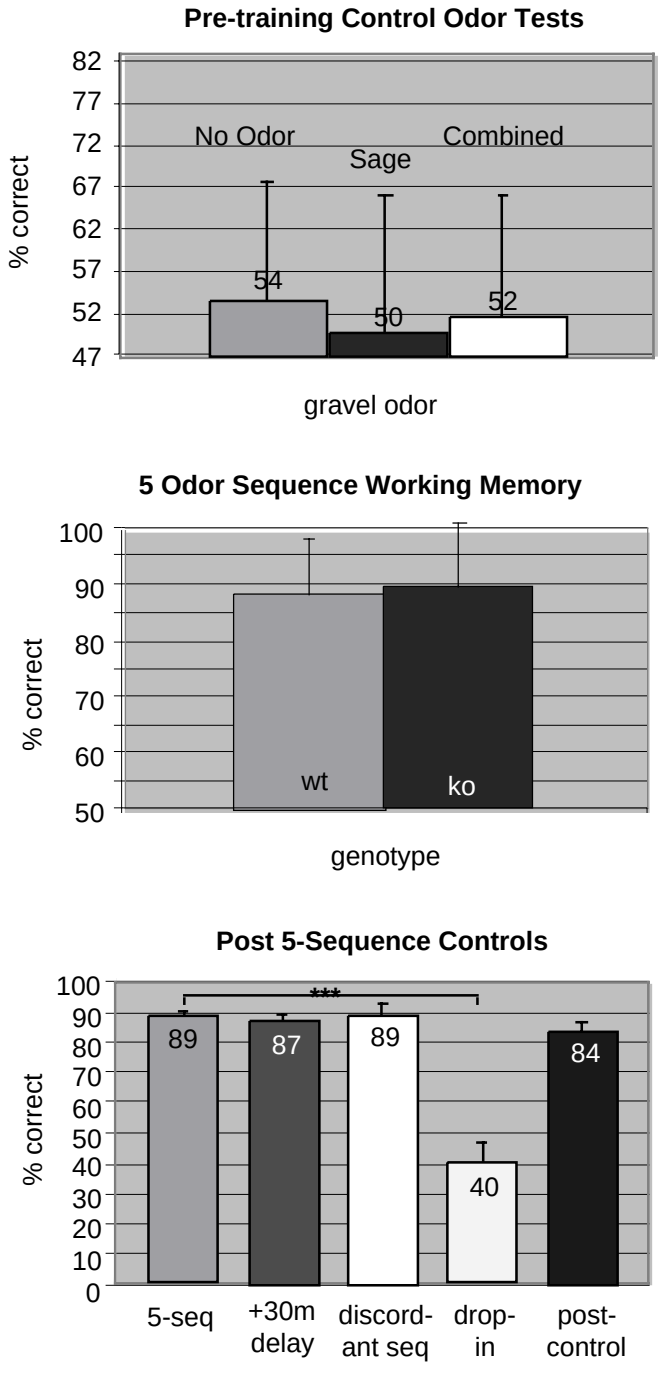

Figure 2. Experiment 1. Five odor sequence recall, a test of working memory for sequences of olfactory information (adapted from Fortin et al., 2002). Subjects were sixteen F1 hybrid mice of two inbred strains (FVB/NJ x C57BL/6J) for which half the littermates ("ko") carried a mutation in the *Fmr1* gene involved in Fragile X mental retardation in humans. **"**Wt" indicates the wild type littermates. Upper panel: Pre-training control odor tests for the 5-sequence task. Six naïve mice (three of each genotype) trained only to dig for cereal rewards were presented with a pair of cups containing gravel, one of which also contained a buried piece of cereal. Cup pairs either had no scent, a common scent (sage), or a combination of one scented and one unscented cup. No significant difference from random choices (50% correct). Middle panel: Plateau performance (final 9 days) of the percentage of correct responses for each genotype is plotted for the 5-sequence task. No significant difference by genotype (see text for statistics). Lower Panel: Post-training controls for the 5-sequence task. The standard 5-sequence result (far left) is followed by (left to right): 1) performance after a 30 minute delay. (Subsequent controls had no delay.) No significant decline in performance relative to the task without delay was detected; 2) performance when the exposure odors were different from the test odors. The expected random performance is 50%; 3) 5-sequence performance when cups were not baited with rewards, which instead were provided post-choice; 4) results from the standard, buried reward 5-sequence task performed after the preceding controls. No statistical differences by genotype were observed.

After achieving a stable, near errorless plateau performance, additional experiment were conducted in which delays of 3 minutes, 10 minutes, and 30 minutes were added between the exposure and the test phases. Remarkably, performance did not deteriorate with a 30 minute delay between exposure and testing: 15

second delay (M = 89.24%, SD = 5.13), 30 minute delay (M = 87.73%, SD = 5.19), $t(29) = 0.81$, $p = 0.42$, $\eta^2 = 0.02$, nor was there any statistically significant difference by genoytpe: wt (M = 88.43, SD = 5.0), ko (M = 87.04, SD = 5.6), $t(14) = 0.52$, $p = 0.61$, $n2 = 0.02$ (Figure 2).

As an additional control experiment, a series of five odors was presented in the prescribed manner. The mice were then tested for recall with minimum delay (15 seconds), except that instead of the odors to which they had just been exposed in the individual cups (A-E), they were presented with a new, random assortment of odors, with one cup of each pair still baited. Surprisingly, the mice in this discordant odor control test performed as well as when they did when the odors matched those presented in the exposure phase (Figure 2): matching sequence (M = 89.24%, SD = 5.13), discordant sequence (M = 89.56%, SD = 11.99), $t(21) = -0.11$, $p = 0.92$, $\eta^2 = 0.0005$. Thus, the mice apparently dug in the cups containing food rewards, no matter what odors were mixed into the sand containing them. There was also no difference by genotype detected in the

discordant sequence control: wt (M = 91.67%, SD = 8.91), ko (M = 87.50, SD = 14.77), $t(14) = 0.68$, $p = 0.51$, $\eta^2 = 0.03$.

The 5-sequence task was then repeated without buried rewards; cereal was dropped into the cups only after observing the mice digging in the correct cup. This produced a dramatic deterioration in performance, with mice choosing the correct sequence of odors approximately 50% of the time on average: 5-sequence drop-in one-sample t-test compared to a mean of 50% (M = 58.89%, SD = 19.79), $t(14) = 1.74$, $p = 0.10$. The difference in mean between the original buried reward results and those using reward drop-in were significant: buried (M = 89.24, SD = 5.13), $t(16) = 5.75$, $p < 0.0001$, $\eta^2 = 0.75$ (Figure 2). There was no statistically significant difference by genotype: wt (M = 66.67, SD = 16.52), $t(13) = 1.48$, $p = 0.16$, $\eta^2 = 0.15$.

To assess whether the mice had become confused by the discordant sequence controls, the original task, with the five original odors corresponding to the test odors, was repeated with buried cereal. The resulting scores were as high as the original plateau performance of these animals: original 5 sequence without delay (M = 89.24%, SD = 5.13), post-control 5-sequence test (M = 84.37%, SD = 9,56), $t(24.4) = 1.78$, $p = 0.09$, $\eta^2 = 0.11$ (Figure 2). One explanation for the seeming discrepancy between performance in the pre-training and post-training 5-sequence controls might be that in pre-training tests, the mice may have believed that any cup could contain food, so they did not yet carefully make an olfactory

discrimination based on odor as they would come to do after training. In summary, the 5-sequence control experiments presented here indicated that the mice could learn to smell the cereal rewards buried in the cups and were choosing cups largely based on this information, not the memory of the odor sequence.

EXPERIMENT 2

Method

Subjects for 2-sequence task. F1 hybrid mice ("HYB") of the same number and strain background (FVB/NJ x C57BL/6J) used in the 5-sequence testing (Subjects, Experiment 1), including half with the *fmr1-tm1Cgr* ("ko") allele, were used. Visually impaired FVB/NJ inbred mice which carry the retinal degeneration allele $Pde6^{rd1}/Pde6^{rd1}$ were also used. For training and testing, male mice 3 - 9 months of age were housed individually with free access to water. Dieting and testing occurred in the home cage. Mice were calorically restricted (3 days without food; generally beginning Friday) and then maintained at ~80-85% free feeding weight on 8.5 - 10.2 kilocalories per day (five to six 500 mg rodent diet pellets; #F0171, BioServ, Frenchtown, NJ). Animals were shaped to dig beginning day 4 by placing a cereal sliver on, and then increasingly deep in, a cup of gravel. The animals were fed 4 pellets for the evening of day 4. Shaping to dig and restricted caloric intake were continued days 5 - 8. During this time, stability

of the animal's weight and appearance were assessed. Excessive hunger was evident by animals walking gingerly on their toes with an arched back. Up to two additional pellets of food daily were added as necessary. On day 11 (generally the second Monday, if calorie restriction began on a Friday), testing was initiated.

Apparatus. To make a three cup testing tray (Figure 1), a corrugated polypropylene sheet was cut to 13 x 20 cm. The 20 cm length was sliced on one side at 8 cm from one end (12 cm from the other) with a razor blade. The plastic was then bent to 90 degrees at the slice and taped in place with Manco HP260 Crystal Clear tape. With the 8 cm piece as the base and the 12 cm piece projecting vertically as the back, a 1 x 13 cm strip of plastic was taped on the base, 4.9 cm from the back, to provide a restraint for the three test cups (Figure 1). Odorants were added to the gravel as described for Experiment 1.

Sequence exposure phase. Mice dug in odor cup 1 with a buried reward present; in some experiments (with "emphasis"), odor cup 1 was removed and returned with another piece of buried food. After consuming the food in odor cup(s) 1, the mouse was presented with odor cup 2, which also contained a piece of buried food. Memory for the sequence of odors 1 and 2 was then tested after a delay of 15 seconds (considered no delay), 3 minutes, or 30 minutes; delays were constant throughout a single experiment. Exposure apparatus and reward sizes were as for 5-sequence testing (Experiment 1 and above).

Acquisition Testing Phase. Odor cups 1 and 2, as well as a novel odor in cup 3, were presented in a linear plastic tray. No food reward was present in any of the test phase odor cups. If the mouse dug in odor cup 1, a piece of cereal ($1/16^{th}$ FrootLoop) was dropped onto it for consumption. Odor cup 1 was then removed. The mouse would then choose between odor cups 2 and 3. If he chose odor cup 2, he was again rewarded with cereal, and if odor cup 3, he received an immediate punishment. Punishment was also immediately applied if the animal dug first in either odor cup 2 or 3. The punishment consisted of a moderate pinch (squeeze) of the tail with a curved hemostat clamp (VWR #25601-066) padded with tape, followed by a one-minute time out. After some trials it was observed that many animals needed to merely see, or in the case of visually impaired FVB/NJ mice, be touched with, the clamp to produce a reaction (i.e. retreat from the cups). In such cases, the animals were not chased and pinched. When the animals began to ignore sight or touch of the clamp, punishment was again employed, as well as after the animals made the second of two consecutive errors. After a time out, three new odors were presented for the next trial.

Twelve trials were conducted for each mouse per day, except for the first day, which had a thirteenth trial. One trial comprised an exposure phase followed by a test phase (probe) of working memory for the exposure sequence. During the first day, every fourth trial, beginning with the first, was performed with the food dropped on top of the correct cup prior to choice, in an attempt to train the animal

to the rule, and thereby reduce the variability of learning the rule by guessing it over time. On day 2, the 13th trial was dropped. On day 3, the 10th trial became an ordinary probe of memory, i.e. it was no longer used as a shape for rule acquisition. On day 4, the 7th trial was no longer a shape, and so on until on day 6, the first trial was no longer a shape. Thus, the animals received 5 days in which a steadily diminishing example of the task was provided. Testing was continued until asymptotic or plateau performance was reached for each animal. Plateau performance was calculated as a statistically similar mean performance over at least three consecutive test days.

Scoring. Performance was followed by percentage of each of four possible responses: choice of odor cup 1 followed by odor cup 2, choice of odor cup 1 then cup 3, choice of cup 2 first, and choice of cup 3 first. In addition, the four responses were combined into a single score using a 3 point system. Choice of odor 1 then 2 was given 3 points, odor 1 then 3 was given 2 points, odor 2 first 1 point, and odor 3 first 0 points. This system provided a simple means of roughly assessing bias for choice of odor 1 first and against odor 3 first.

Cup position and shuffling. The position of the odor cups from left to right in the testing tray for successive trials were assigned in pseudorandom order as follows: trial 1 = cup3 cup1 cup2 [i.e., 312], trial 2 = 123, 3 = 231, 4 = 321, 5 = 132, 6 = 213, 7 = 231, 8 = 123, 9 = 312, 10 = 132, 11 = 321, 12 = 213, 13 = 123.

To shuffle the odor cups, they were arranged in the same 5 x 5 matrix used for the 5-sequence testing (Experiment 1). After the first twenty-four cups had been used (8 trials), cup 25 was inserted as the starting cup, with cup 1 as the new second cup. The other cups would then shifted correspondingly, i.e. the new odor set would be 2-3-4, 5-6-7, and so on for the remaining 4 or 5 trials.

Each day, the cups would be shuffled in a predetermined manner as described for the 5-sequence testing (Experiment 1). Entirely random shuffling could have been used, although this would have required recording cup positions each day in order to be able to assess any association between odor and cup position.

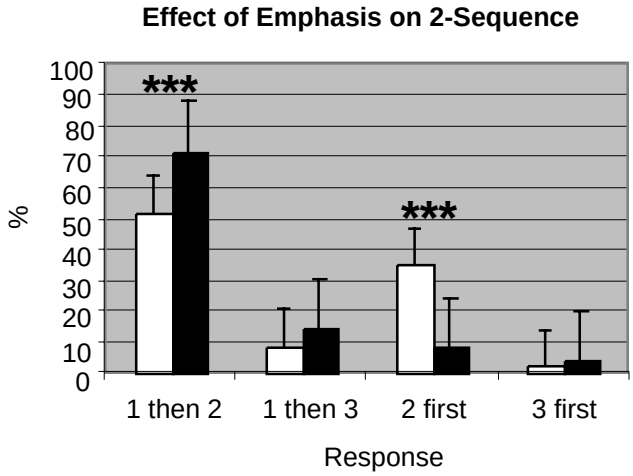

Figure 3. Experiment 2: Effect of emphasis of odor 1 on 2-sequence performance. F1 hybrid mice were tested in the 2-sequence task without emphasis on odor 1 (white bars) or with emphasis by presenting odor 1 twice (black bars) during the exposure phase of each trial. The percentages presented are the average of the final three days of the tests, by which point scores were stable. (*** indicates $p < 0.001$; Methods).

Results and Discussion

The 5-sequence protocol was modified in a further attempt to test working memory for sequences and learning set formation. In addition to potentially smelling buried rewards, it appeared that an animal might be able to perform well in the 5-sequence test by always choosing the first two odors encountered in the exposure phase (A or B) when present in the test phase without regard to sequence. In 4 of 6 discriminations in the 5-sequence task, odors A or B are paired with D or E, and in a 5^{th} discrimination, A is paired with C. In one case, C is tested against E. Therefore, by recalling only odors A and B and choosing those, in theory it would be possible to get 5 of 6 correct answers. The odor value of C can be both correct or incorrect, but even guessing on this discrimination might allow an animal to average 5.5 of 6 correct, or 92% (as observed for rats, and in this study, mice). Consequently, the new task required mice to make a choice with all odors initially present.

As pilot experiments suggested that recall of five odors was likely to be quite difficult to master under such conditions, the test was simplified to require recall of only two odors in sequence (2-sequence task). To expand the range of scores, and thereby assist in statistical analysis, a novel odor not present in the exposure sequence was added to the test panel as a decoy. Thus, the mice were first exposed to odor 1, then odor 2, with both cups baited in order to ensure that the mice dug into and smelled each odor. Next the mice were exposed to a tray

upon which odors 1, 2, and 3 were presented (Figure 1), unbaited, in a fixed, "pseudorandom" order (see Methods). Mice were rewarded for picking odor 1 first, but punished (pinching with tweezers and one minute 'time out'; see Methods) if they chose 2 or 3 first. If they chose odor cup 1, only it would be removed and the mice would then have to choose between odor cups 2 and 3. If odor cup 2 was chosen next, the mice would again be rewarded, but would be punished for picking odor cup 3.

Table 1: Effect of emphasis

	1 then 2	1 then 3	2 first	3 first
- emphasis	279	49	191	15
+ emphasis	401	82	48	26

final 3 day plateau

Preliminary results suggested that mice were often not discriminating the sequence of odor 1 versus 2 during exposure. Therefore, the first odor was emphasized by presenting it twice (Methods). The effect of emphasizing odor 1 during the final 6 days of plateau performance of hybrid mice was determined using a chi-square test of independence. This test was chosen because the behavioral responses were dependent, mutually exclusive categories. Table 1 shows plateau performance over the last three days of testing without and with emphasis on odor 1. The distribution of responses was significantly different as assessed by a chi-square test of independence: χ^2 (3, 1091) = 118.3, $p < 0.0001$. A

comparison of each choice with and without emphasis was also made parametrically by converting responses of each type to daily percentages; these values are likely normatively distributed. Means for each type of response were separately compared by t-test: 1 then 2 no emphasis (M = 51.7%, SD = 1.2), emphasis (M = 72.0%, SD = 2.6), t(4) = -12.2, p = 0.0003, η^2 = 0.97, 1 then 3 no emphasis (M = 9.7%, SD = 4.0), emphasis (M = 14.7, SD = 3.5), t(4) = -1.6, p = 0.18, η^2 = 0.40, 2 first no emphasis (M = 35.3%, SD = 3.1), emphasis (M = 8.7%, SD = 1.5), t(4) = 13.5, p = 0.0002, η^2 = 0.98, 3 first no emphasis (M = 3.3%, SD = 1.5), emphasis (M = 4.7%, SD = 0.6), t(4) = -1.4, p = , η^2 = 0.33 (Figure 3). (Neither multiple t-test comparisons nor ANOVA among the responses were made as these are not independent samples, e.g. choosing 1 then 2 precludes the other choices.) Therefore, the primary effect of emphasis was to increase the correct response “1 then 2” ($p < 0.001$) at the expense of choosing odor 2 first ($p < 0.001$). There was little effect on the already low rate of choosing odor 3 first. Although the animals clearly achieved a greater selection of correct responses (1 then 2), additional tests will be required to rule out the influence of other factors, such as increased time between exposure to odors 1 and 2, to be certain that repetition of odor presentation was responsible for the increase in performance. Nonetheless, for practical purposes, all subsequent tests were conducted with a repeated exposure of odor cup 1 (emphasis).

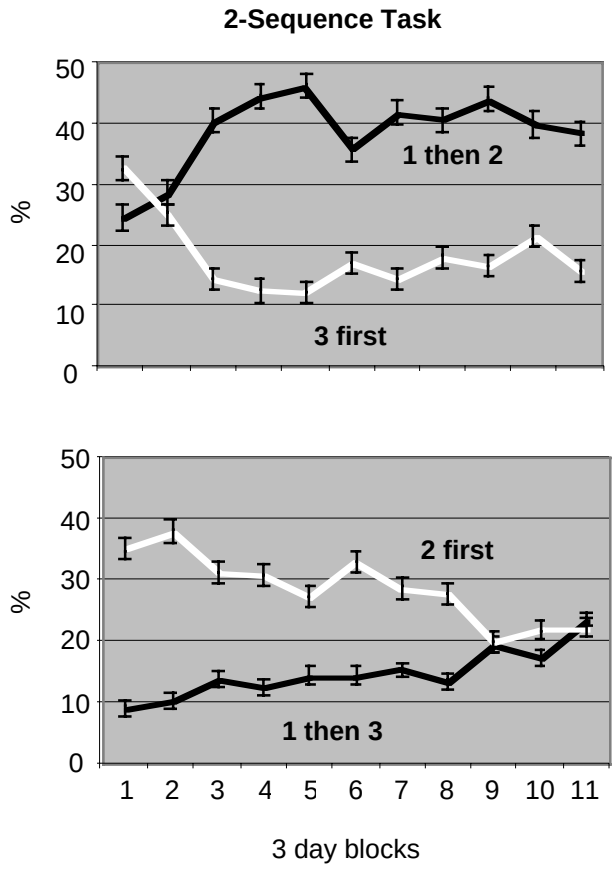

Figure 4. Experiment 2: Plot of all choices made in a 2-sequence task by FVB/NJ mice. Rolling average in three day blocks are shown. The standard error of the mean indicated by bars refers to mean percentage for each response.

The percentage of the four possible 2-sequence responses over time is shown in Figure 4. FVB mice reached a plateau performance of choosing odor 1 then 2 in five days of testing. Consistent with the visual impression from Figure 4, one-way ANOVA of correct choice (1 then 2) x day showed a learning effect over days 1 through 5, $F(12, 65) = 3.98$, $p < 0.001$. The increase in correct choice "1 then 2" doubled from start to plateau for the FVB/NJ mice as depicted in Figure 4 (F1 hybrid mouse performance improved three-fold; see below). During this time, the choice of odor 3 first dropped from random (one-third) to half that rate. The mice learned to avoid picking odors 2 first as well; however, after 195 trials, the FVB mice still sporadically chose odor cups 2 or 3 first.

Table 2: HYB wt vs. ko

	1 then 2	1 then 3	2 first	3 first
wt	401	87	47	20
ko	379	109	67	19

final 6 day plateau

A comparison of hybrid wt and ko littermates in learning to choose odor 1 followed by odor 2 was performed using analysis of variance (ANOVA) with a significance level of $\alpha = 0.05$. Both groups achieved what appeared to be stable performance by the last six days of testing. That the last six days were not significantly differ in mean performance was confirmed by one-way ANOVA of genoytpe x day: for wt, $F(5, 41) = 1.06$, and for ko $F(5, 42) = 0.33$. A two-way ANOVA in which all trials on a given day were treated as replicates found no significant difference by genotype [$F(1, 84) = 3.56$], nor was there any interaction of genotype x day. Performance of wt and ko was also compared across all four categories of choice using nonparametric analysis. Observed responses are reported in Table 2. With a significance level $\alpha = 0.05$, χ^2 (6, N = 1129) = 6.31, $p > 0.05$. Although larger groups of wt and ko mice could provide increased statistical power which might reveal a difference in performance in this task, at the very least we can conclude that no major difference is likely, i.e. one comparable in magnitude to that observed in similar tests between normal and FXS humans.

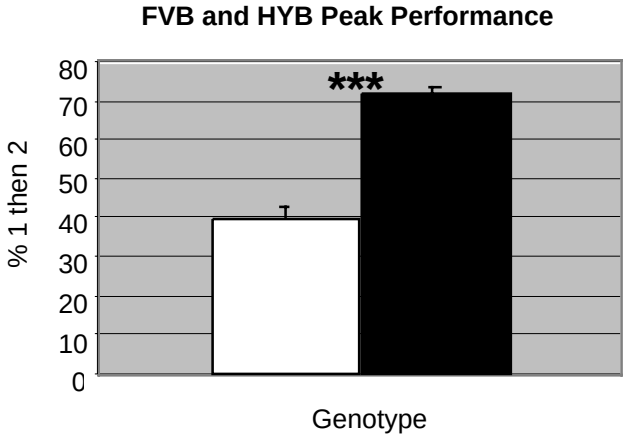

Figure 5. Experiment 2: FVB vs F1 hybrid performance in the 2-sequence task. While both strains seem to reach near asymptotic levels of correct choices (odor 1 then odor 2) after 4-6 days of testing, the hybrid mice (black bar) reached a higher level of performance by the last four days of testing than the visually impaired, inbred FVB/NJ strain (white bar).

It was also of interest to compare performance of the inbred FVB/NJ mice with F1 hybrid mice on the 2-sequence task. FVB mice (of which half also carried the *fmr1-tm1Cgr* FXS mutation) have a genetically determined retinal degeneration allele (*Pde6*^{rd1}/*Pde6*^{rd1}) and are therefore severely visually impaired. While both strains reached plateau levels of performance in four to six days of testing, the hybrids had a statistically superior performance at plateau (Figure 5): HYB (M = 72.0%, SD = 5.5), FVB (M = 40.1%, SD = 6.0), t(14) = -11.1, p = , $\eta^2 = 0.90$. Experiments are underway to try to determine the basis of the difference, but as has been demonstrated for other behavioral tests, strain background can have a significant effect; this despite the fact that a mutation in a gene known to cause mental retardation in humans (*Fmr1*) did not cause any significant difference in performance within the hybrid or FVB backgrounds.

Even among the hybrid mice, few, if any, consistently performed at a level which might be described as near errorless performance (see also General Discussion). Figure 6 shows data by individual F1 hybrid mice for three days of performance during the plateau phase of testing. Although the average performance was quite constant, it is apparent that there was day to day, and even trial to trial variability. The raw data for one animal are illustrated (Figure 6, right panel) to suggest that, even after nearly two hundred trials, rather than near errorless performance, the animals were performing in a type of streak and slump pattern, i.e. with streaks of several correct responses interspersed with slumps of one or more errors. Several possible explanations are apparent, including, 1) that the animals had not yet achieved a true learning set, which might be expected to allow them to solve the continually novel problems accurately, or 2) the level of distractors in such a test (using shuffled odors) relative to the attention levels of the species for these olfactory cues was sufficient to increase the error rate.

d.13	d. 14	d. 15
75	58	92
83	83	67
67	75	58
75	58	83
58	50	42
75	83	92
58	50	58
57	67	58
92	83	67
92	67	50
83	75	83
92	83	83
57	67	83
67	67	75
83	75	67
ND	75	83

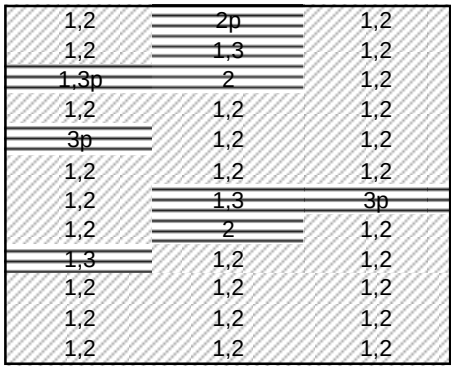

Figure 6. Experiment 2: Performance of individual hybrid mice showing streak and slump pattern in the final three days of the 2-sequence task. Left panel: the percentage of correct sequence choices (choosing odor 1 followed by odor 2, indicated by "1,2") made each day by sixteen mice. Right panel: raw data from one animal during the three days. Horizontal lines in the raw data indicate errors (unsuccessful trials; any response except choosing odor 1 followed by odor 2).

To investigate whether the plateau performance might be a sign of the rigor of the task, it was simplified by holding odor 2 constant. In the original 2-sequence task, odor 2 could be rewarded, if chosen after odor 1, or punished, if chosen first. (Odor 1 was always rewarded and odor 3 always punished if chosen.) We suspected that the variable valuation of odor 2 might introduce an element of significant complexity to learning the rules of the task. The effect of holding odor 2 constant was evaluated using a chi-square test of independence. This test was chosen because the behavioral responses were dependent, mutually exclusive categories. Table 3 presents the total observed frequencies of each category of response (1 then 2, etc.) x treatment (holding odor 2 constant or having odor 2

continuously change): χ^2 (3, N = 1478) = 50.81, $p < 0.05$. To determine which of the categories was responsible for the difference between the two distributions, columns were individually dropped. This analysis revealed that holding odor 2 constant led to significant declines in choice of odor 2 or odor 3 first: χ^2 (1, N = 904) = 3.02, $p > 0.05$ when these two columns were dropped; no other single or paired dropped columns eliminated the difference in the distributions. Although it may not seem unexpected that holding odor 2 constant could reduce the choice of odor 2 first, it is less clear why this also reduced the choice of odor 3 first, yet did not specifically elevate choice of 1 then 2 over 1 then 3. It would appear that holding odor 2 constant allowed the FVB/NJ mice to simply focus on choosing odor 1 first. But there is no evidence here that there was a benefit to the temporal choice of the odors in sequence as would be indicated by a specific increase in choice of 1 then 2. These data do support an interpretation that the original 2-sequence task had indeed been quite challenging for the mice. Additional experiments are under way to determine what elements of the task might further contribute to its complexity, and at what level of difficulty (relative to the number of trials) the mice might achieve near errorless performance in recalling a sequence of two odors in correct order.

Table 3: Odor 2 constant vs. changing

	1 then 2	1 then 3	2 first	3 first
2 constant	280	156	94	72
2 changing	326	142	255	153

It has been shown for rats that increasing the number of odors in a non-matching-to-sample task increased accuracy (Slotnick, 2001). As noted previously (Fortin et al., 2002), the use of larger odorant sets should further diminish the risk of interference from one trial (e.g. choice of A vs. B) to another (choice of A vs. C). Therefore, use of many odors in the 2-sequence task (25 in the current set of experiments) should reduce the likelihood of forming associations between an odor and any particular reward value. Use of a predetermined "pseudorandom" odor sequence allowed determination of whether the mice developed any bias in choice of the odors. All possible odor cup sequences for thirteen days with 12 trials per day were entered into a spreadsheet program. A histogram of odor usage over the first five days of testing was established for each odor value, i.e. the number of times each scent (1-25) had odor values one, two, or three over that period. In the first five days, the greatest disparity in odor values occurred for odors 2 (number of times with odor value 1 = 0 and number of times with odor value 3 = 4) and 22 (number of times with odor value 1 = 5 and number of times with odor value 3 = 1). The selection of odors 2 and 22 during the second five day block of testing showed that odor cup 2 was correctly chosen first (i.e. as odor value 1) on 12 of 30 opportunities (40.0%). Odor cup 2 was chosen incorrectly first on 7 of 18 opportunities (38.8%). Odor cup 22 never had a chance to be chosen as odor value 1 during the second five day block; however, it was chosen incorrectly first 14 of 36 of relevant choices (38.9%). It is quite evident that even

the most extreme cases of odor use during the first five day block had no effect on choice of those odors during the second five day block. Therefore, analysis of choice of these and other scents (e.g. odors 3 and 11 over the entire testing period, for which these had the largest difference in net odor value; not shown) indicated that the mice were not favoring or avoiding any particular odor.

How much the various odors generalize one another is also unknown. Further efforts to define the effective stimulus will be in order, although there is no consistent example over these trials in which one odor consistently lacked discrimination against any others. This is not too surprising as all of the odorants (spices) used are easily discriminated by most humans.

The digging task described here employs differential emphasis during exposure and punishers as well as positive reinforcement during testing in order to achieve an improved rate of acquisition. The use of positive reinforcement in olfactory digging tasks may not be essential. In some studies, mice have been trained to associate one odor of two with a sugar reward, but no reward was present during the test (Schellinck, Forestell, & LoLordo, 2001). Nonetheless, there was a marked decline in correct response over days in olfactory discrimination tests not employing positive reinforcement (Schellinck et al., 2001). Two punishers were also used: 1) immediately removing the cups when a wrong choice was made (“time out”), and 2) a tail pinch for an incorrect choice. Pilot testing for the procedure suggested that an active negative stimulus

(punishment) was important in rapidly training the mice to avoid the incorrect odor and improve performance, especially in getting the mice to attend to the second odor, which changed from being incorrect in the first choice to correct in the second. (A more detailed, quantitative examination of the value of punishment in this test is the subject of another study. We note that even after 50 days of testing, some mice show large declines in performance when punishment was dropped, while others showed little or no effect; Chang et. al., unpublished.) Punishment in learning paradigms is generally considered effective when a positively rewarded option is always present, as is the case for this protocol, i.e. it can be an effective means of redirecting behavior during learning to a rewarded outcome. Punishment in the 2-sequence task did not always require physically touching the subject: after an early pinch, the mere sight of the padded clamp was sufficient to induce a negative response leading the mouse to quickly move to the opposite end of the test box from the odor cups. When the subjects habituated to the sight of the tongs, the operator would apply a new pinch. In the case of the FVB/NJ mice with retinal degeneration, the clamp was touched to the tail to elicit a response in later trials.

EXPERIMENT 3

Method

Detection of baited cups at the ends of the arms of an 8-arm radial maze was tested using an apparatus with 25 cm runways (Crusio, Schwegler, & Brust, 1993; Yan et al., 2003). A T-maze was also made from two 25 cm arms emanating from a central platform. In the olfaction distance tests, one baited and one unbaited cup were placed at varying distances and in a random order in the arms of a T-maze. Calorically restricted mice (Experiments 2 Methods) were allowed to make a single choice on each trial in the 8-arm and T-maze tests. Mice were allowed to consume the reward if the correct cup was chosen. No punishment was applied for an incorrect choice.

Detection of cereal reward by mice in a system conforming to the scale of that used for rats (Fortin et al., 2002; Kesner, Gilbert, & Barua, 2002) was assessed using two rat food dishes 8.25 cm diameter x 4 cm deep. Fortin et. al. (2002) buried 1/4 FrootLoop cereal piece under 100 g sand in a cup 7 cm in diameter x 6 cm height; Kesner et. al. (2002) used a cup of the same diameter. To compensate for the increased surface area of our rat-sized cups, approximately 140 g of sand would be required. 140 ml water filled 3.0 cm of the 8.25 cm diameter rat dish, but the same weight of our gravel had a height of only approximately 1.5 cm. To make the test more stringent, we doubled the amount of

our relatively dense gravel to 280 g (3 cm). Ten adult male naïve F1 hybrid mice were food restricted as described above, shaped to dig in the gravel for cereal rewards, and then tested for ability to detect the buried cereal at a depth of 3.0 cm for 5 days, 6 trials per day, followed by a second 5 day block with cereal buried at 1.5 cm depth (gravel height was reduced to 1.5 cm such that the reward again rested on the bottom of the cup). In each test, the mouse was presented with two rat feeding dishes containing gravel as described above. One of the cups was baited with cereal placed at the bottom of the cup. The subject was allowed to make a single choice of cup as indicated by digging. Cups were baited in a pseudorandom order for each trial such that no baited cup was ever in the same position (left or right) for more than two trials in a row.

Results and Discussion

As one goal of this study was to determine whether buried rewards might create artifactually elevated performances in murine learning tasks, other tasks which have employed buried food rewards were examined. First, calorically restricted, F1 hybrid mice trained in the 2-sequence task (Experiment 2) were placed in an 8-arm radial maze with cups of gravel near the ends of four of the 25 cm arms. One of the cups had a buried cereal reward. Eight mice were tested for 10 days. The total choice of baited arms was 8 out of 76 total arms entered by all the mice (frequency of 0.11). Our hypothesis was that a mouse that could smell

the baited food at the end of an arm from the central platform would choose the baited arm more than the random one in eight chance (probability 0.125). Therefore, a one-tailed binomial probability was calculated for which the outcomes were not equal, comparing the observed frequency of correct arm choice (0.11) with the null hypothesis (0.125), $p > 99.99$, supporting the apparent outcome that the trained mice could not the scent of buried cereal at a distance of 25 cm.

In order to determine the distance at which the mice could detect buried food, the cups were brought increasingly closer to the central platform of a T-maze. As for the radial maze odor detection test (above) our hypothesis was that if the animals could detect the scent of buried cereal, then the choice of baited cups would be increased; therefore, a one-tailed binomial test was used. The mice could detect the cereal at 5 cm (2 inches): $M = 0.78$ per sample over 40 sample sets, which, relative to a random mean of 0.5, produced a probability $p = 0.0003$ that the mice had achieved such choices randomly. In contrast, by 7.5 cm (3 inches) the mice appeared to be guessing which arm contained food: $M = 0.5$, $p = 0.56$. Therefore, only when the cups were brought to within 5 cm (2 inches) from the central start site was a significant increase in choice of baited cups observed (Figure 7). We conclude that ability to smell a buried reward may not affect radial maze performance in which commitment to an arm is made from a greater distance than that which can be used in olfactory digging tasks.

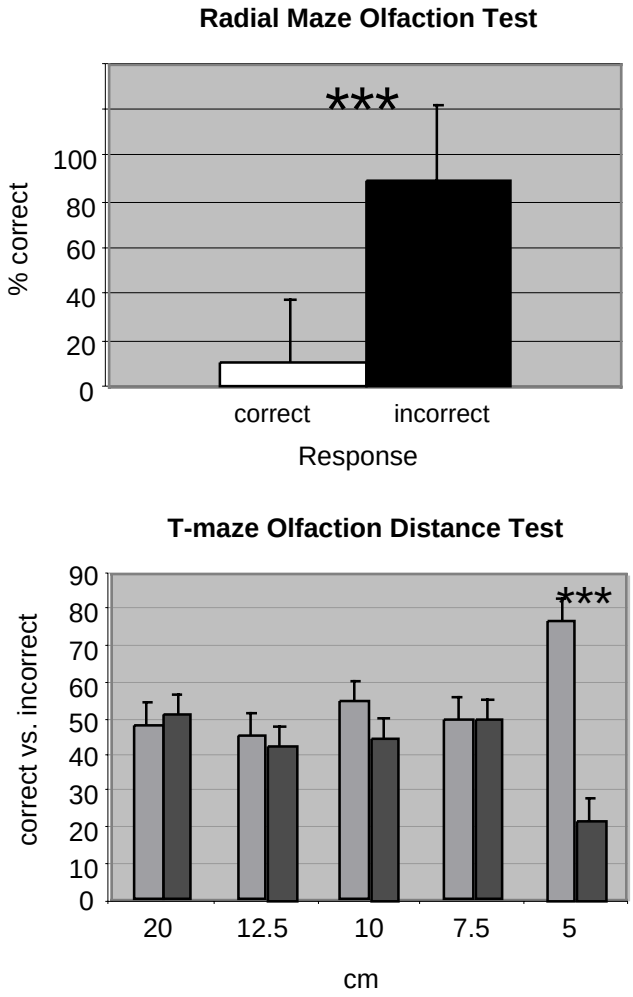

Figure 7. Experiment 3: Top panel: Mice do not smell food placed randomly at the end of an arm in an 8-arm radial maze. Eight mice were tested once a day for 10 days in an 8-arm radial maze with one baited arm; the other arms had unbaited, decoy cups. Each animal was food restricted and had been previously trained in a standard four-baited arm version of the 8-arm radial maze. When placed into this maze, the animal was allowed to enter one arm, and the choice scored as correct or incorrect. Random performance would be expected to produce a correct response rate of 12.5%. Bottom panel: Mice begin to detect buried food at approximately 5 centimeters (left bars reflect correct responses). A cereal reward was buried in one cup at the end of a T-maze, while the second arm contained a cup with no food. Eight food restricted mice were tested with the cups at increasingly close distances to see if the buried food could be detected. Random response would produce a 50% correct rate.

An important question regarding the attempted application of the 5-sequence task to mice is whether scaling down the apparatus was appropriate. Figure 1 shows the relative size of mice to odor cups scaled for the rat and the mouse. It may be that, despite the ten-fold difference in body size, the olfactory sensitivity of mice does not differ much, or may even be superior to, that of rats. To assess this, mice were tested for ability to detect buried cereal of the same

amount (1/4 FrootLoop) in a cup scaled similar to that used in 5-sequence testing of rats (Fortin et al., 2002; Kesner et al., 2002) except that initially twice as much depth of filling material was used (see Methods, Experiment 3). Ten adult male F1 hybrid mice which had not been previously used in testing were food restricted and then trained to dig in the rat scale cups for a cereal reward. As for the T-maze tests, above, our hypothesis was that if the animals could detect the scent of buried cereal, then the choice of baited cups would be increased above average (0.5); therefore, a one-tailed binomial test was used. The mean over 5 days of testing with 6 trials per day ($M = 0.355$) was actually below random guessing (0.5); $p = 0.99$ that it was not different (hypothesis larger) than 0.5. (It appeared that the mice were not merely randomly guessing, but attempting to employ a win-stay, lose-shift strategy which hindered their performance in the pseudorandom sequence used.) Therefore, testing over several days indicated that the mice could not detect cereal buried at 3.0 cm beneath our gravel. The same mice were then retested with cereal buried at 1.5 cm, a depth we calculate as equivalent to that used for the prior rats studies (see Methods, Experiment 3). In this case, $M = 0.658$ for which the one-tailed binomial probability was $p < 0.001$. Therefore, the mice could detect the buried food to some extent. It should be noted that in these experiments it took substantial effort for the mice to dig through such large amounts of gravel to obtain food rewards; it was often difficult to complete trials with some mice under these conditions. Therefore, it does not

appear practical to use such depths of gravel to conduct such olfactory digging tasks. Nonetheless, our results would be consistent with the integrity of the use of such tests for rats. It is quite possible that rats have similar (or less) olfactory sensitivity compared to mice, such that they would not necessarily detect food buried in the configurations described in prior studies, as would be consistent with controls described therein. However, the potential difficulties of attempting to scale such tests for mice are demonstrated in this study.

GENERAL DISCUSSION

Olfactory digging tasks were originally designed for rats (Bunsey & Eichenbaum, 1996; Dusek & Eichenbaum, 1997) and then adapted for mice, e.g. (Berger-Sweeney, Libbey, Arters, Junagadhwalla, & Hohmann, 1998; Bodyak & Slotnick, 1999; Mihalick, Langlois, Krienke, & Dube, 2000; Zagreda, Goodman, Druin, McDonald, & Diamond, 1999). In simple two odor discrimination tests, both rats and mice quickly achieved accurate first choices of greater than 80% (Berger-Sweeney et al., 1998; Slotnick, 2001). In mouse studies which employed buried rewards such as bits of chocolate, comment was made that some mice might have sensed the presence of food buried in the sand (Zagreda et al., 1999). We show here that the situation may require considerable caution, since the mice may smell buried food rewards for some distance. The relative volatility of odorants produced by various rewards would be expected to influence the degree

with which mice may sense rewards. Nonetheless, our cereal reward is a standard one used in rodent behavioral studies and therefore exemplifies the need to account for reward detection.

It remains possible that a stronger sense of smell or a difference related to the size of the test apparatus relative to the size of the mice, may have led to their greater ability to detect buried rewards than may exist for rats (Fortin et al., 2002; Kesner et al., 2002). For example, although mice are 10-fold smaller than rats by weight and the scaled down odor cups here were only 1.8-fold shallower, the olfactory ability of the two species may be comparable enough that buried rewards could be detected in the smaller cups, but not as gravel depth is increased, as was demonstrated in Experiment 3.

Aside from removing the risk of experimental confound from detection of buried rewards, the 2-sequence task as described here should also reduce the possibility that rodents could achieve good scores by remembering less than the full sequence of odorants. In order to avoid this outcome in 5-sequence olfactory memory tasks such as that employed in Experiment 1 (see also Fortin et. al., 2002), all odors would have to be compared, e.g. D vs. E and A vs. B. Otherwise, subjects might simply learn to always choose in the test phase the first two odors presented in the exposure phase (i.e. odors A and B), avoid or ignore the last two odors presented in the exposure phase (D and E), and then guess for discriminations involving odor C (versus E).

It was also demonstrated that emphasis of the first odor during the exposure phase (e.g. by re-exposure) improved performance. Emphasis does not eliminate the need for the animal to make choices in a correct sequence, e.g. choosing the emphasized odor first, followed by choosing the less emphasized odor or by avoiding the novel odor. Similar benefit was obtained when exposure odor 2 was rewarded instead of odor 1 (data not shown). One explanation for the value of emphasis might be that, while olfactory cues may indeed have sensory dominance in the mouse, the species may lack a preparedness to respond to sequences of odors. Thus, two odors of the nature used here, presented in sequence, may not be as memorable to a mouse as two syllables in sequence to a human. If mice are not particularly prepared to recall temporal sequences of odors, this may raise the challenge of the task; it is also at the heart of discourse regarding the special place of olfaction in mouse learning and memory. Perhaps olfactory associations are easily gained, but not necessarily sequences of such. The data presented here suggest that mice can keep odors in temporal sequence, although facility to do so remains uncertain. In the future, it will be of interest to assess whether providing richer cues may increase performance to near errorless levels. Hybrid mice, for example, may have obtained some performance advantage relative to visually impaired FVB/NJ mice from visual cues if they were able to detect changes in appearance of the gravel caused by specific spices.

Another element which may have led to improvement in performance in this task is that increasing experience with some stimuli, such as the odorants used here, would allow the mice to become more familiar with them and recognize, discriminate, group, or recall them better. Conversely, decay of working memory over the minimum delay (of about 15 seconds) between the exposure and testing phases of the 2-sequence task may have also diminished the chance that near errorless performance could have been achieved. Reducing these times may require automation. Conversely, increasing the time between presentation of the two odorants during the exposure phase may decrease any negative effect of interference on memory; thus one possible benefit of emphasis during exposure may have been to alter the time between the two stimuli (odors 1 and 2). Naturally, emphasis may also have led to a stronger memory trace.; even so, that would not obviate the requirement to choose odorants in the correct temporal sequence experienced. Therefore, while the effects of a host of potential variables must be assessed, what is important to establish initially is that, at a minimum, the task does appear to allow an assessment of an ability to respond to stimuli in a temporal manner without risk of contamination artifact or gaming (e.g. succeeding by recalling only a subset of odors).

Although one purpose of this work was to provide data explicitly sounding a note of caution regarding buried rewards in murine olfactory digging tasks, these experiments also touch upon a more fundamental debate regarding the

quality of learning occurring in rodents via olfaction (Reid & Morris, 1993; Slotnick, 1994). As noted above, it was quite surprising to find that mice could recall changing sequences of five odors to which they were quickly exposed with near flawless accuracy even after a delay of 30 minutes. This was a performance level that the human operators felt would take significant effort for them to match, even using visual or verbal cues. Our work demonstrates that this near flawless ability evinced by the original test is likely an artifact, consistent with a "contamination hypothesis" previously discussed (Reid & Morris, 1993; Slotnick, 1994). Clearly, the data presented from the 2-sequence task (Experiment 2) indicate that the mice were able to improve their performances as if they had obtained a sense that they should pick the first odor encountered first and avoid the novel third odor. A more central issue is whether the mice have learned a rule (or acquired a "learning set") to solve the unbaited 2-sequence tasks. The plateau performance of both the hybrid (FVB/NJ x C57Bl/6J) and inbred (FVB/NJ) mice indicate that near errorless performance was not achieved in the task despite reaching a stable plateau. Performance may have improved, in part, because use of emphasis and punishers may have offset some of the "disruptive responses" (Reid & Morris, 1993; Slotnick, 1994) which initially may reduce performance. Two such responses which were obvious were 1) the tendency of the mice to explore, and 2) the willingness of the mice to "game" the task by guessing and if necessary waiting out the turn. Nonetheless, even the hybrid mice did not

explicitly appear to have a firm grasp on the rule “choose one and then two”, or “choose one and then not three”. While the mice were able to produce strings or streaks of perfect performance (“one and then two”), this was easily lost even within a single block for all mice, even well after nearly 200 distinct trials. This sort of performance might be more consistent with gaining a “feel” for the correct response pattern, e.g. a procedural response. Therefore, based on the data we have been able to obtain to date, we cannot state that there is evidence for acquisition of a rule, and certainly not an abstract one, which would produce near errorless performance. Nonetheless, these data do not argue that near errorless performance is unachievable in rodents using simpler olfactory tasks, although these too have been the subject of some controversy (Reid & Morris, 1993; Slotnick, 1994). In fact, we showed that by making our 2-sequence task simpler (by keeping odor 2 constant), performance was further increased. It may also be possible that breakthroughs to near errorless performance might occur from the plateau performance demonstrated in these tasks, e.g. as trial number is further increased. The question of whether the mice in the 2-sequence task acquired a learning set also depends on exactly what is meant by the term. If learning set means a cumulative improvement in performance between problems of a class (Warren, 1973), then the mice did achieve a learning set. However, if a more stringent definition of learning set is used, such that one trial learning with response of greater than 90% on the second trial is required (Reid & Morris, 1993), then the

mice did not achieve this standard. The latter definition is interesting, because it would seem to require a subject to employ some sort of strategy, rule, or other abstraction. But even the use of an abstract rule might be masked by the quality of response availability. For example, a human could swing a stick (or bat) to hit a thrown ball with very strict rules in mind as to how to proceed, yet the speed with which events occur might preclude near errorless performance on grounds other than use of a rule. Therefore, although the data presented here do not provide evidence for use of the most stringent definition of learning set by mice, it also does not rule it out. It is worth noting that even with the most stringent definition of learning set, the term "rule" use can suggest an abstraction that may not be employed by the rodent. Rather, the use of temporal ordering may be all that is required, along with learning to ignore the possibly innate tendency to associate a specific odor with a reward (just as subjects must learn to ignore positional cues.) Thus, the "rule" may be an implicit "THEN" function: odor A THEN odor B (or odor A THEN not odor C), e.g. as in a chained operant response. Therefore, the 2-sequence (or related) task need not be an indicator of abstract rule use or even explicit memory. What we can conclude from these experiments is that 1) baited olfactory digging tasks have an inherent risk which we demonstrate here can have a substantial impact on performance, and 2) when such a confounding effect is removed, mice do not necessarily show an easy acquisition of learning set for recalling sequences of odors.

REFERENCES

Artal, P., Herreros de Tejada, P., Munoz Tedo, C., & Green, D. G. (1998). Retinal image quality in the rodent eye. Vis Neurosci, 15(4), 597-605.

Berger-Sweeney, J., Libbey, M., Arters, J., Junagadhwalla, M., & Hohmann, C. F. (1998). Neonatal monoaminergic depletion in mice (Mus musculus) improves performance of a novel odor discrimination task. Behav Neurosci, 112(6), 1318-1326.

Bodyak, N., & Slotnick, B. (1999). Performance of mice in an automated olfactometer: odor detection, discrimination and odor memory. Chem Senses, 24(6), 637-645.

Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. Nature, 379(6562), 255-257.

Crusio, W. E., Schwegler, H., & Brust, I. (1993). Covariations between hippocampal mossy fibres and working and reference memory in spatial and non-spatial radial maze tasks in mice. Eur J Neurosci, 5(10), 1413-1420.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. Proc Natl Acad Sci U S A, 94(13), 7109-7114.

Dykens, E. M., Hodapp, R. M., & Leckman, J. F. (1987). Strengths and weaknesses in the intellectual functioning of males with fragile X syndrome. Am J Ment Defic, 92(2), 234-236.

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. Nat Neurosci, 5(5), 458-462.

Hodapp, R. M., Leckman, J. F., Dykens, E. M., Sparrow, S. S., Zelinsky, D. G., & Ort, S. I. (1992). K-ABC profiles in children with fragile X syndrome, Down syndrome, and nonspecific mental retardation. Am J Ment Retard, 97(1), 39-46.

Jennings, J. W., & Keefer, L. H. (1969). Olfactory learning set in two varieties of domestic rat. Psychol Rep, 24(1), 3-15.

Kemper, M. B., Hagerman, R. J., & Altshul-Stark, D. (1988). Cognitive profiles of boys with the fragile X syndrome. Am J Med Genet, 30(1-2), 191-200.

Kesner, R. P., Gilbert, P. E., & Barua, L. A. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors. Behav Neurosci, 116(2), 286-290.

McIlvane, W. J., & Cataldo, M. F. (1996). On the clinical relevance of animal models for the study of human mental retardation. Mental Retardation and Developmental Disabilities Research Reviews, 2, 188-196.

Mihalick, S. M., Langlois, J. C., Krienke, J. D., & Dube, W. V. (2000). An olfactory discrimination procedure for mice. J Exp Anal Behav, 73(3), 305-318.

Nigrosh, B. J., Slotnick, B. M., & Nevin, J. A. (1975). Olfactory discrimination, reversal learning, and stimulus control in rats. J Comp Physiol Psychol, 89(4), 285-294.

Reid, I. C., & Morris, R. G. (1993). The enigma of olfactory learning. Trends Neurosci, 16(1), 17-20.

Schellinck, H. M., Forestell, C. A., & LoLordo, V. M. (2001). A simple and reliable test of olfactory learning and memory in mice. Chem Senses, 26(6), 663-672.

Slotnick, B. (2001). Animal cognition and the rat olfactory system. Trends Cogn Sci, 5(5), 216-222.

Slotnick, B., Hanford, L., & Hodos, W. (2000). Can rats acquire an olfactory learning set? J Exp Psychol Anim Behav Process, 26(4), 399-415.

Slotnick, B. M. (1994). The enigma of olfactory learning revisited. Neuroscience, 58(1), 1-12.

Strupp, B. J., & Levitsky, D. A. (1990). An animal model of retarded cognitive development. Adv Infancy Res, 6, 149-185.

Warren, J. M. (1973). Learning in Vertebrates. In D. A. Rethlingshafer (Ed.), Comparative psychology: a modern survey (pp. 471-509). New York: McGraw-Hill.

Yan, Q., Asafo-Adjei, P. K., Arnold, H. M., Brown, R. E., & Bauchwitz, R. (2003). A Phenotypic and Molecular Characterization of the fmr1-tm1Cgr Fragile X Mouse. Genes Brain Behav, in revision.

Zagreda, L., Goodman, J., Druin, D. P., McDonald, D., & Diamond, A. (1999). Cognitive deficits in a genetic mouse model of the most common biochemical cause of human mental retardation. J Neurosci, 19(14), 6175-6182.

ACKNOWLEDGMENTS

This research was supported by a grant from the FRAXA Research Foundation (to RPB) as well as American Chemical Society Project Seed grants (to AH and AG). We would like to thank Howard Eichenbaum for introduction to the 5-sequence task and Rebecca Corwin for review of the manuscript. Thanks also to support from the New York City JUMP program (SH and AW) and Susan Fenton of St. Luke's-Roosevelt Hospital Center.