Learning curve assessment of rule use provides evidence for spared implicit sequence learning in a mouse model of mental retardation

Cindy Cai, Jeremy Bauchwitz, Janaile Spence, Jeff Chang, Teresa Concha, Sarah Reisberg, Ga Young Lee, Kseniya Barkova, Yvette Wojciechowski, Christina Poopatana, Brian Maitland, Daisy Duan, Diane Mei, Yue Michael Ma, Karina Illescas, Christopher Sikorski, Orellana del Fierro, QiJiang Yan, Ann M. Rogers, and Robert P. Bauchwitz [1]*

[1]St. Luke's-Roosevelt Institute for Health Sciences, Columbia University, New York, NY 10019, USA

*Corresponding author. Tel.: +1 212 523 8869; fax: +1 212 523 7623

Email addresses:

CC: cxc2103@columbia.edu

JB: jbauchwitz@gmail.com

JS: xycopixie@aol.com

JC: yc2050@columbia.edu

TC: teresa.concha@yale.edu

SR: Sreisber@fandm.edu

GYL: ssulunghae@hanmail.net

KB: Ksenia1987@aol.com

YW: yvette.Wojciechowski@fairwindspartners.com

CP: thaiflip92@yahoo.com

BM: bmaitland17@yahoo.com

DD: day_zee0625@yahoo.com

DM: Diane7190@aol.com

YMM: yuema@bergen.org

KI: kmi201@nyu.edu

CS: shishkax@yahoo.com

ODF: odelfierr0@yahoo.com

QJY: qijiangyan@yahoo.com

AMR: amrogers@luxsci.net

RPB: rpb3@columbia.edu

Keywords:

Learning curve, learning set, sequence learning, serial learning, mental retardation, Fragile X syndrome, mouse models, explicit cognition, implicit cognition,

Abstract

Humans with Fragile X Syndrome (FXS) have a mental retardation of which a notable characteristic is a weakness in recalling sequences of information. A mouse model of the disorder exists which exhibits behavioral and neurologic changes, but cognitive testing has not revealed learning deficits seemingly comparable in magnitude to that seen in the human condition. A working memory task for olfactory sequences was employed to test learning set acquisition in mice, half of which had a disruption of the gene responsible for FXS in humans. The task protected against reward detection artifact and demonstrated stringency-dependent task acquisition. A comparable image-based sequence learning set task was used to test humans. The performances of human subjects who did and did not report consciously acquiring the task rules were used as positive and negative controls to assess the mouse learning curves. Learning curve plateau error fluctuation for individual mice was comparable to that of human subjects who never acquired an explicit rule to perform the task, but different from those of human subjects who could state a rule to solve the problem. Sliding window error plots and nonparametric statistical analysis discriminated between the consciously rule-based human performances and that of the mice and humans who did not explicitly obtain the rule. Based on comparison to the human results, wild-type and FX mouse learning curves with a continuingly variable terminal plateau error rate in sliding epochs were classified as “implicit”. Although a moderately large difference in performance of the olfactory task was observed among mouse strains, there was no significant effect of FX genotype. The wild-type performance of the FX mice in this sequence task suggests that implicit learning may be relatively spared in FXS.

Introduction

The ability to "learn how to learn efficiently in the situations an animal frequently encounters", i.e. to obtain a learning set (Harlow, 1949) is believed to be a sign of higher order cognition. Specifically, Harlow noted, "This learning to learn transforms the organism from a creature that adapts to a changing environment by trial and error to one that adapts by seeming hypothesis and insight." (Harlow, 1949). In effect, the animal goes beyond learning simple associations to learning rules, i.e. cognitive methods by which behavior can be generalized. For example, in serial (sequence-based) learning set tasks, use of a rule could provide a heuristic guide or exact algorithmic procedure to make a correct choice. Rule use can be conscious, and even involve language, but it does not have to (Kellogg and Bourne, 1989; Levine, 1959; Neal and Hesketh, 1997; Reber, 1989b; Seger, 1994). Hence both people and non-human animals can give evidence of using rules without conscious awareness of the underlying concept (category commonality) inherent to a rule.

Learning set acquisition has primarily been assessed by repeated presentation of a rewarded/non-rewarded object pair for a block of multiple trials, after which a new object pair was used for each subsequent block (see Supplemental Information EN1 for additional detail). Learning set tasks can also be performed with stimuli changing with each trial as a means to examine the dependence of learning and intelligence on working memory, as well as on acquisition of a rule. It is working memory (active maintenance of information in short-term memory for manipulation by a variety of cognitive processes (Baddeley, 1998; Engle et al., 1999; Numminen et al., 2000)), but not short-term memory, that shows significant correlations with important measures of IQ (Engle et al., 1999).

The olfactory learning set task presented here is a modification of one designed for rats (Fortin et al., 2002; Katz et al., 2003), which has constantly changing odors in each trial. Therefore, each trial can be considered a test of learning set achievement, i.e. as if it were a "trial 2" of a learning set. Once the task is learned, performance gives an indication of working memory. Furthermore, learning the task itself also employs working memory in several possible ways, for example, to make associations between operant responses and consequences, or even to comprehend the underlying task rule.

In this study, we extend previous experiments (Katz et al., 2003) which showed that mice with a disruption in the gene which causes FXS in humans (Bakker and Consortium, 1994) were able to achieve learning sets of such continually changing sequences of odorant stimuli. Furthermore, their performance was remarkably similar to that of wild-type (wt) control mice. To further assess why humans with FXS do appear to show significant working memory deficits in sequence tasks with delays of the same duration as those employed here (Cornish et al., 2001; Hodapp et al., 1992; Kwon et al., 2001; Maes et al., 1994), a comparable learning set task for human subjects was produced which employed a computerized, image-based 2-sequence task.

Significant differences in the learning curves between mice and normal humans were noted, with those of the mice being initially gradually increasing and considerably more error-prone near the terminal plateau region, while that of the humans often showed a step-function in learning followed by a greatly reduced error frequency. In particular, graphical examination of terminal error fluctuation among individual human subjects who did and did not acquire the test rule explicitly provided a very useful set of controls with which to assess the underlying cognitive basis for the performance of the mouse subjects. Post-test comments by the human subjects indicated that the unique features of the human learning curves (among those subjects who had acquired the underlying rule in an explicit manner) were the result of conscious rule acquisition and subvocal memory strategies. Based on the differences in species learning curves, it was concluded that the mice did not acquire whatever rule-based generalization they may have employed to achieve the learning set in a manner comparable to that of some of the human subjects. Therefore, the quality of rule acquisition and use could be discriminated by learning curve plateau error analysis.

Differences in the quality of rule acquisition may have significant implications on the use of animal models to investigate the cognitive deficits of human mental retardations. Nonetheless, the FX mice in this study did achieve learning set efficiently, suggesting that alternate learning and memory systems might be relatively spared, and that such systems could be more directly exploited in education of those with Fragile X Syndrome.

Methods

Subjects

FVB/NJ, C57BL/6J and C57BL/6J Fragile X (*fmr1*^{tm1Cgr}) mice were procured, housed, and genotyped as noted in a prior series of related experiments (Katz et al., 2003; Yan et al., 2004). Briefly, male C57BL/6J *fmr1-tm1Cgr* mice ("ko", "*fmr1*", or "FX") were bred to wild-type ("wt") female C57BL/6J mice to produce females heterozygous for the *fmr1* mutant allele. Male FVB/NJ mice were bred to heterozygous C57BL/6J *fmr1*^{tm1Cgr} /+ female mice to produce F1 hybrid ("HYB") litters with approximately half wt and half *fmr1* mutant males. Adult HYB males of 3 – 5 months of age were taken from a barrier facility to a room for cognitive testing, at which time they were housed individually. Light cycle began at 7AM for 12 hours and all testing occurred during this time. Food and water were supplied *ad libitum* until one week prior to testing, at which time food restriction was initiated (see below). Treatment of the mouse subjects was approved by the St. Luke's-Roosevelt Institute for Health Sciences IACUC and was in accordance with APA ethical standards. Human subjects were volunteers of apparently normal intelligence who agreed to perform the computer-based task daily as described below. Personal identifying information was removed as per SLRIHS IRB protocol.

Two-sequence task for mice

Dieting and testing occurred in the home cage using apparatus essentially as described previously (Katz et al., 2003). Briefly, 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). Mice were then shaped to dig (days 4-8) for one-sixteenth piece of cereal (FrootLoops, Kellogg Co., Battle Creek, MI; approximately 5 mm x 5 mm x 2-3 mm, with an average weight of 11 mg) in non-toxic aquarium gravel (Estes' UltraReef Marine Sand, Totowa, NJ; approximately 23 g and 1 cm deep; see also Fig. 1 of (Katz et al., 2003)). On day 11 (generally the second Monday, if calorie restriction began on a Friday), testing was initiated. The two-sequence task was then conducted as described in (Katz et al., 2003) and Figure 1, or with the variations elaborated upon in the Results section of this work.

Two-sequence task for human subjects

A computer-based version of the 2-sequence task was produced using images in place of odorants. Images were chosen from various font symbols (generally Monotype Sorts, 96 point; see Supplemental Information file), which were shuffled as for the odorants. Five slide files corresponding to five shuffled image-blocks were produced. The primary difference from the odorant task was that the visual task had 16 trials per day instead of 12, and there were no example trials during the first blocks. Prior to testing, the tester told the subject the following: “You are going to see some symbols in a slide presentation. You will have to make a choice of one symbol when you see more than one. I will respond correct or incorrect. You try to make as many correct choices as you can. You are to say nothing during the test. Simply point to a symbol. After we are done today, I may ask you some questions.” The exposure phase (sequence of two single images) was separated by a blank slide from the test phase (three images, followed by two, if a correct answer was given). Blank time-out slides followed incorrect answers. The duration that each blank was shown is given on the scoring sheet (see Supplemental Information); the default between exposure and test phases was 10 seconds. The time during which the subject could view the image slides was also fixed (generally two or three seconds). If the subject did not make a choice in the time allowed, the tester said nothing more than “Please choose a symbol.” At no time did the tester provide any further information or comment, so as to make the experience as comparable to that of the mice as possible. After the completion of the day’s block of trials, the tester stated, “Did you notice anything about the test?” (Willingham et al., 1989). The subject’s comments were transcribed by the tester. Once subjects had essentially perfect responses for more than 20 trials, and had indicated that they were using an internal rehearsal strategy, delays of 30 seconds and 90 seconds were added between the exposure and test phases. The eight normal subjects were 10, 11, 13, 15, 19, 41, and two of 45 years of age. The sex of each subject is listed above the individual learning curve (Fig. 7 and 7S).

Statistics and Animal Numbers

Learning curves for each of the possible four choices ("1 then 2", "1 then 3", 2 first" and "3 first") were plotted, and probable regions in which performance was relatively unchanging (graphical plateaus or approach to seeming linear asymptote) were estimated. Data were combined in blocks of three days to simplify and enhance statistical analysis. A two-factor, repeated measures ANOVA of the days of interest (e.g. days 15-17) by treatment (test and genoytpe, e.g. 2NC+/wt, 2NC+/ko, etc.) was then performed. In all but one case (2NC+ days 21-23, for choice 1 then 3), it was found that there was no interaction of days x treatment (for plateau regions), nor were there any main effects of the days. In the one exception noted, 2NC+ days 20-22 for choice 1 then 3 was used. In no case in any test for any choice was a statistically significant difference by genotype seen in the two-factor ANOVA (no main effect of genotype), nor were such effects seen in t-tests between the two genotypes for a given test and choice (not shown). Therefore, in all cases, wt and ko genotype results for a given test and choice were combined ("combo-geno"). The combined genotype data at plateau points, generally groups of three days for the correct choice "1 then 2", were compared to similar results for days 1-3 to determine whether statistically significant changes in performance had occurred from the start of the test to the first plateau (days 15-17). These comparisons were made by t-tests (e.g. 2NC+combo-geno 1 then 2 days 1-3 vs. same for days 15-17). Also, days 15-17 were compared to subsequent periods in the extended learning tests. One-factor ANOVA was used to compare the three tests for a given choice and range of days. All tests used a level of 0.05 for significance. All t-tests were two-tailed unless indicated otherwise. In charts, probabilities are designated by asterisks as follows: $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***. Means ± standard error of the mean or standard deviation are given as indicated, e.g. (M ± SEM). All error bars in charts indicate SEM. Only results with statistically significant probabilities are given in the text. Comprehensive statistical results for the mouse learning curves are listed in supplemental data Table 1S and 2S.

Repeated measures ANOVA was not employed for the time-series data because of need to accommodate differing numbers of mice in each treatment group, and to avoid concerns of circularity assumption violations.

The number of mice (subjects) involved in a test can be determined from the statistical presentation of degrees of freedom as follows: F(groups - 1, subjects - groups) and t(subjects - 2). The number of groups (for genotype-test) can also be determined from the same information. The exception to this is that in the day x test-genotype two-factor ANOVAs, the total number of "subjects" are the days and the groups are test-genotype (e.g. 2NC+/wt); therefore, the data were subsequently rearranged and pair-wise t-tests (wt vs. ko for a given task) are also presented in the supplementary data table. The exact number of mice for each genotype can also be ascertained from the left panels of Figure 2S in which each dot represents the performance of a single mouse.

All mice are F1 hybrid adult males, unless otherwise specifically noted. Data used for F1 hybrid 2-odor-sequence learning test analysis are independent of those in (Katz et al., 2003). Statistical tests were primarily performed using the Prism program version 4.0c for Macintosh (GraphPad, San Diego). In some cases in which Tukey-Kramer CD values for a two-factor ANOVA or a three-factor ANOVA were desired, the GB-STAT program (Dynamic Microsystems) was used. All odor tests were conducted by experimenters blind to the genotype of the mice. Each investigator tested his or her own group of mice.

Results

Two-odor sequence task

A 5-sequence protocol for rodent working memory (Figure 1S) was modified to reduce risk of artifact in tests of working memory for sequences and learning set formation, as described previously (Katz et al., 2003). The basic 2-sequence (two-sequence) task is shown schematically in Figure 1. In this study, variations of the 2-sequence task were performed to assess task stringency and ability to discriminate wt from FX mouse performances. In addition, quantitative assessments of the effects of punishment, motor sequence learning, and task extension were made.

In the first variant assessed ("2NC-" for "two not constant minus [emphasis]"), mice were exposed once each to odors 1 and 2 in sequence, and odor 2 changed with each trial. After two weeks of training and testing, mice were choosing the correct sequence of odor 1 then odor 2 just over 50% of the time (with random choices expected to produce 16.5% "1 then 2" responses). To determine if the correct response rate could be elevated in order to reduce the total number of trial-days, the first odor was emphasized by presenting it twice (2NC+).

In the 2NC- and 2NC+ task variants, odor 2 could be rewarded if chosen after odor one, or punished if chosen first. (Choice of odor 1 was always rewarded and odor 3 always punished.) It seemed possible that the variable valuation of odor 2 might introduce an element of significant complexity to learning the rules of the task. Therefore, to investigate whether plateau performance might be a sign of the rigor of the task, a third task variant was simplified by holding odor two constant (2C+), while still emphasizing the first odor (see also (Katz et al., 2003)). It was predicted that task difficulty would increase in order: 2C+, 2NC+, 2NC-.

Genotype and Strain Comparisons

In each task, half the F1 hybrid mice carried a mutation in the X-linked *Fmr1* gene, inactivation of which causes Fragile X mental retardation in humans (Bakker and Consortium, 1994). Therefore, the first analysis performed for each of the four possible choices in the task ("1 then 2", "1 then 3". "2 first", "3 first") was a two-factor ANOVA to assess the effect of genotype. The comparisons were initially made for the average performance over days 1 through 3 and days 15 – 17 (Table 1S). In each case, a t-test directly comparing the genotypes in a particular task was subsequently performed; those results, as well as mean ± SEM are also presented in Table 1S. In no case was a significant difference by genotype found for any choice (see Fig. 2 and left panels in Figures 2S(a) and 2S(b)).

A similar comparison of wt and FX genotypes had been made in an inbred background (FVB/NJ) for the 2NC+ task (Katz et al., 2003) with no difference by genotype detected. Despite the absence of significant difference in performance between wt and FX mice, a very significant difference between the F1 hybrid and FVB/NJ background plateaus (last three days of training) had been observed in the earlier study. The F1 hybrid results presented here are consistent with the prior results.

Power analysis for the wt and FX 2NC+ data at days 15-17 (first part of the plateau) indicated that there would have been a less than 20% chance of a false negative result (a standard level for assessment) if the means had a difference of 15.74 or greater (both Ns = 11; SD_{wt} = 15.92; SD_{FX} = 7.67). A much larger effect size for a true mental retardation gene on human higher order cognitive performance would be expected. Consistent with this, mice of the FVB/NJ strain background, which are not known to have major cognitive deficits, nonetheless had a mean performance almost 20 points (percentage correct) lower than the F1 hybrid (C57BL/6J x FVB/NJ) strain. (An assessment of genotypic effect is also given for individual mice in the 2C test as compared to human subjects beginning in the section titled, "comparative learning curves".)

Odor choice analysis

Upon finding that there was no significant difference by genotype, the results for the wt and FX genotypes were combined for subsequent analyses. t-tests indicated that there was a very significant improvement in performance (choice of correct response odor 1 then 2) from days 1 – 3 to days 15 – 17 (Fig. 3 and Table 1S). These results indicated that the mice had successfully learned the task, i.e. to choose the exposure odors in the order in which they were presented.

When one-factor ANOVA was used to compare the three tasks at days 1-3 and separately at 15-17, few significant differences were observed for any choice. The exceptions were 1) a modest elevation of choice 2 first for 2NC- over 2NC+ on days 1-3 and also for 2 NC- over 2C+ on days 15-17, and 2) elevated choice of 3 first for 2NC- over both 2NC+ and 2C+ on days 15-16 (Table 1S and right panels of Fig. 2S(a) and 2S(b)).

These results are supportive of the prediction that 2NC- would be the most difficult task. However, it is also clear that by days 15-17, the mice were learning to choose the correct choice, 1 then 2, at approximately the same rate for all the tasks. Therefore, performance in making the correct choice of sequence for days 4 –15, was assessed, again in blocks of three days. It was apparent from these data that the mice performed the 2C+ task significantly better than 2NC+ for days 4-12, and the 2NC- task significantly less well between days 7 through 15 (Fig. 4 and Table 1S). These data show that altering task rigor can affect the rate at which the task is acquired, though not necessarily having as great an effect on the near-asymptote memory performance, especially in tasks in which one of the odors was emphasized during exposure.

Motor sequence controls

In the 2-sequence tasks (2NC+, 2NC-, 2C+) discussed thus far, the odor cups were repositioned on each trial in a predetermined order, i.e. as part of a single long sequence. In order to assess whether the mice might have learned to improve their performance by learning the complex motor sequence, six of the mice were further tested for nine additional days after completing 28 days of the 2C+ task. The new test ("dance" control) remained the same in terms of the 2C odors, but the positions of the cups were changed into a new sequence. If the mice had been relying on the motor learning, their performance should have deteriorated. However, they were able to maintain their high level of performance, indicating that they were using odor cues, not motor sequence learning, to choose the correct odors. There was no statistical difference in performance between wt and FX in the first three days of the motor sequence control ($t(4) = 0.62$, p =

0.57), nor was there any difference between a combination of the wt and ko mice on those days with performance on the last three days of the 2C test using the standard cup positions (t(14) = 0.09, p = 0.93.)

Aversive stimulus impact

In a prior study we noted that punishment (use of an aversive stimulus after an incorrect response) in this learning paradigm appeared to be effective in maintaining performance (Katz et al., 2003). This result was consistent with findings that punishment can be an effective means of redirecting behavior during learning to a rewarded outcome when a positively rewarded option is always present (Bernstein et al., 2003). In order to get a more quantitative measure of the effect of punishment, after 17 days of training to plateau performance mice were tested for another 5 days in its absence. Once the mice had learned the task, dropping punishment produced a significant decline in performance, regardless of whether the task used emphasis of an exposed odor (Fig. 5). Furthermore, a comparison of Figure 4 with Figure 5 shows that the most difficult task had a plateau that exceeded 45% correct with punishment, while the same task without punishment had a correct choice rate at least 10% below that level. Therefore, punishment was valuable for maintenance of enhanced performance.

Extended Training

In order to assess whether extending the training period would allow further improvement in performance, the two less rigorous tasks with emphasis of an odor during the exposure phase (2NC+ and 2C+), were performed for 23 and 28 days, respectively (the length of extension was based on the effect observed during the first six days added for both tests; see Fig. 6). A t-test was used to compare the mean scores obtained during the last three days of testing in the normal testing (days 15-17) with the terminal three-day period during extended testing (days 21-23 for 2NC+ and days 21-23 and 26-28 for 2C+; Table 2S and Fig. 6C). The analysis revealed no difference by genotype, but a significant increase in the average correct responses from days 15-17 to days 26-28 for the 2C+ task, with an increasing mean throughout the 11 day extension. Therefore, continued training, in this case for an additional six to eleven days, allowed improvement in performance in the easiest task (2C+), indicating that task stringency effects are not limited to differences in initial rate of task acquisition (Fig. 4). Nonetheless, despite these improvements, individual mice still did not achieve consistent near-errorless performance (see below).

Human visual sequence task

It was of note that humans with FXS are seriously impaired in tasks requiring working memory of two images in sequence (see Introduction), yet as demonstrated above, FX mice were not noticeably deficient in a comparable task involving odors. To begin to assess why this might be, a computer-image-based sequence learning set task was produced in which human subjects would perform the same process as the mice. Eight subjects of normal intelligence ranging in age from 10 years to 45 years were given 16 trials a day without any example or explanation. They were shown two individual images in sequence, and then, after a delay, shown a set of three images containing the two original images. They were instructed to point to a choice and were then told “correct” or “incorrect”.

The subjects were given no information as to the basis of the task, nor even the demarcations of a trial. In this way, it was believed that the subjects would not start with any advantage compared to the mice. On the contrary, with the original intent to reduce the random extent to which mice might happen upon a "rule" (were they actually to use such – see Discussion), the mice were given a few example trials among the initial tests, which were not given the human subjects.

At the conclusion of each day's testing, the subjects' comments were noted. Although they were never told whether a pattern was present, it was evident that all had the inclination to try to discern such. Those who did acquire the sequence rule did so by consciously testing various rules expressed subvocally, e.g. one subject reported, "For the row of three, it would always be the first symbol that you had seen. And for the row of two, it would always be the second one you had seen." (See Table 3S in Supplemental Information for post-test comments). They also kept the images in memory by naming them, e.g. "I gave them all names, like 'target', 'hubcap', 'check', 'square-up', 'square-down'. I said them as I went along in my mind" (Table 3S).

The speed with which the correct pattern was discovered varied significantly among the participants. It should be noted that three normal human subjects (15 years old, and two 45 years old) in this study did not acquire the sequence pattern (rule) within 100 trials, nor did they show substantial improvement over the five to six days they were tested (Fig. 7 and 7S) – a period in which the mice showed significant performance increases. (The 15 year old female who never acquired the complete rule, did realize by the last day of testing that the first object seen was to be chosen first when all three were shown together, i.e. she had learned part of the rule; her performance modestly trended upward over the last days of testing.) In all three cases, the subjects who did not acquire the complete rule lost interest in the task after the first two days, and reported not only little or no further attempts to understand the sequence, but also forgetfulness during trials due to distraction by other thoughts. It was concluded that this sequence task is not necessarily trivial for human subjects who are given no information about its nature.

Comparative learning curves

Five of the eight human subjects acquired the sequence rule within the 100 trial testing period. The individual human learning curves showed dramatic differences between those of subjects after they acquired the rule for choosing the images in sequence and those who did not. In particular, subjects who had consciously acquired the rule were able to maintain their essentially perfect performance over a large number of subsequent trials, including those blocks for which delays between image exposure and test phases were increased from 30 to 90 seconds (Fig. 7 and Supplemental Fig. 7S). Furthermore, acquisition of the rule led to an abrupt increase in performance and decline in error rate (see below for statistical analysis).

By contrast, individual learning curves for mice performing the easiest task (2C+) over an extended period (almost 350 trials) showed no comparable sign of an abrupt increase in performance and decline in error rate as seen for the humans (Figure 8 and Supplemental Fig. 8S; see also statistical analysis below). Instead, the individual mouse learning curves showed gradual improvement. It was also interesting to note that when the human scores were combined and compared to the combined mouse scores (Fig. 8S(b)), the evidence of human explicit rule acquisition was obscured and the overall learning among humans was inferior in rate of improvement to that of the mice.

Species error patterns

To get a better sense of how the human and mouse performances compared, errors at terminal plateau were examined in two ways. First, a comparison was made of the number of trial-ending errors in the last two blocks of twelve trials (313-336) for the mice, and the same number of terminal trials for the humans. Second, an error score, defined as the actual trial score minus the maximum possible score, was graphed for each trial.

With respect to analysis of trial-ending errors, when group performance was compared, mice had lower mean error rates than the human subjects in both the penultimate (15.8% mouse vs. 29% human) and the final group of twelve trials (21% mouse vs. 27% human). The only statistically significant differences among mean errors were observed when the human performances were stratified into those who could verbally state the correct rule and those who could not (Fig. 9). The mean error number made by those humans who had acquired the rule was significantly lower in both sets of trials (3.3% and 1.7%) than for those humans who had not acquired the rule (72% and 69%), and also in the last twelve trials compared to the mice as a group in their terminal twelve trials (21%). Since the mice could not directly report whether they had acquired the rule (but see Discussion for indirect assessment), they were arbitrarily divided into two groups: those that made 1 or no errors on the penultimate twelve trials, and those that made two or more errors. In each case, there was a trend of increased means from the penultimate to the final twelve trials for errors made by the mice (Fig. 9). None of the five human subjects who had explicitly acquired the rule had an increase in error number in the last 12 trials compared to the penultimate twelve, while two of the five mice in the group with 1 or fewer errors in the penultimate block did show an increase. The number of subjects was too small to determine whether the odds of an increase were actually elevated for these arbitrarily selected group of better performing mice.

Given the observed trends just noted, it was of interest to determine whether errors could be compared in less arbitrary blocks of trials and thereby qualitatively distinguished at the individual level. The average error score for every 10 trials was computed as a sliding window of scores. The resulting sliding 10-trial error scores were plotted, producing a pattern of performance with fingerprint-like detail and uniqueness (Fig. 10 and Supplemental Fig. 10S). The step-function increase to near-errorless performance for the human subjects who were able to state the correct rule was still quite apparent (Fig. 10, three right panels). In addition, it became clear that prior to consciously achieving the rule, those subjects, as well as those who never acquired the rule consciously (explicitly) had a highly variable error rate, with the error sliding window frequencies showing peaks followed by troughs. We term this pattern a "streak and slump" performance, for reasons outlined in the Discussion.

When the error performance of the mice was examined in a similar manner, the variable nature of error pattern, and its similarity to that of the human subjects prior to acquiring the rule explicitly, became evident (Fig. 11 and Supplemental Fig. 10S). Most notably, the charts make clear that even when the mice achieved near errorless, or even brief errorless, performance (Fig. 9, uppermost, maximal two-point value black diamonds after 275 trials), this was never sustained, but rather followed by significant increases in error frequency. The latter is seen as a pattern of continual fluctuation and sharp peaks and valleys within the sliding error rate lines (Fig. 11). By contrast, the error lines for those human subjects who explicitly acquired the underlying sequence rule became essentially flat and perfectly errorless (Fig.s 7 and 10). Nonetheless, most mice continued to improve their overall average performance (in the 2C+ task), while there was no comparable strong sign of learning among those human subjects who did not acquire the rule explicitly, even when comparing only the first 100 trials among the species.

Nonparametric analysis of individual learning curves

Although the qualitative, graphical differences between the ruled-based and non-rule-based human learning are quite apparent, a more mathematical, predictive method of characterizing such learning curves would be of substantial value. Two characteristics of the learning curves presented here are of potential interest: the "abruptness" of increase in performance (the moment of insight for some of the human subjects) and the "plateau" error fluctuation (the change in behavior after rule insight). As noted in the results presented above, humans who acquired and were able to state the correct underlying rule to the problem appeared to show a dramatic and persistent decline in subsequent error rate. The error scores that we use here are categorical, i.e. for a given trial, a subject can only get either a score of 0 (no error), -1 (second choice was in error), or -2 (both choices were in error). These nonparametric data, though noncontinuous, can be examined by various "rank-sum" methods. For the subsequent analysis, the error scores were divided into groups of 10 (initial) or 11 (thereafter to allow overlap) consecutive trials, except for the last block which most often was less than 10 trials (since the human and mouse total trials were generally not multiples of 10 or 11).

A Kruskal-Wallis test was initially used to assess whether significant differences in errors were present in the blocks of error scores for an individual subject. In this test, every error score was assigned a rank; ties were given a common, intermediate value. The groups were then distinguished by the sum of their ranks. The Kruskal-Wallis test determined whether a significantly different error score rank-sum was present. Then, to determine which specific group was significantly different, a Dunn's Multiple Comparison test was performed. The Dunn's post-test takes into account the number of groups (as well as group size); however, it was found that in some cases greater sensitivity was achieved by limiting the total number of blocks being assessed to 10 - 11, similar to the number of blocks for many of the human subjects. (The mice had 34 blocks as they had undergone more trials in the extended 2C tests than the human subjects.)

The Kruskal-Wallis/Dunn's tests were chosen to allow discrimination between the error changes in human subjects who obtained the rule (positive controls) and those who did not (negative controls). The results of the Kruskal-Wallis/Dunn's test analysis is shown in Table 4S. The results clearly demonstrate that four of the five human subjects who had acquired the rule had a statistically significant decline in error scores, but that those subjects who had not acquired the rule did not. In the case of the 19 year old female who did acquire the rule but for whom no statistically significant change in error was observed, it is notable that she did so in the first block of 10 trials. This led to insignificant changes in error rates from the first to subsequent blocks. Therefore, one caveat to such methods is that they may miss significant changes if they occur in the first or last block of trials. Generally, when a significant reduction in error rate was observed in consecutive trials, that same change persisted for the remaining trials (Table 4S, column F). However, the 41 year old female who acquired the rule did not show a significant difference until two blocks after she reported having deduced the pattern (Table 4S). Thus, "abruptness" or change in error rate will not always be seen in a consecutive block pair, so we examined sequences of three blocks for significant change, e.g. blocks 5 and 6 following block 4.

For the human control subjects, all blocks except the last would continue to be significantly reduced in error rate if the rule had been acquired in this test. Thus, the curve characteristic of interest associated with acquiring a rule is not only an abrupt change in error rate, but also a persistent “near-errorless” performance thereafter. To assess the persistence of near-errorless learning statistically, the Wilcoxon signed-rank test was used to compare the signed rank of the error scores in a block to a theoretical perfect performance of zero errors in a block. Significantly, the Wilcoxon test showed that none of the humans who failed to acquire the rule ever had a block of errors statistically near errorless, while the human subjects that did acquire the rule never lost the near-errorless performance after acquiring the rule (Table 4S, column E). The Wilcoxon test made it obvious that the 19 year old female had been performing in a near-errorless fashion from the first block of trials.

The number of perfect blocks was also examined. It was found that all of the human subjects who acquired the rule had three or more perfect blocks of ten trials after they acquired the rule, versus no perfect trials for the human subjects who did not acquire the rule (Table 4S, column F). Of the trials following rule acquisition, three subjects never made an error (11 year old male, 10 year old male, and 41 year old female), while two others had correct choices of 93% and 98% of the time (19 year old female and 13 year old male, respectively). The percentage of correct choices for those who did not acquire the rule was 43%, 30%, and 7% (15 year old female, 45 year old female, and 45 year old male, respectively). These data further illustrated the substantial difference in performance dependent on whether a rule was acquired. Therefore, nonparametric analysis of the individual error data very strongly quantified the differences in performance between the human subjects based on whether they had the “insight” of correctly acquiring the rule to solve the problem.

The same statistical techniques were applied to the mouse error data (Table 4S, bottom). For the mice, blocks 1 – 10 were examined with Kruskal-Wallis and Dunn’s Multiple Comparison post-test among all the blocks using the GraphPad Prism statistical software. The results were scanned for significant changes as follows: block 2 and 3 following block 1, block 3 and 4 following block 2, and so on.

Only two of the ten mice showed any statistical indication of an abrupt change in error scores (wt3 and ko6; Table 4S, column B). An examination of the learning curves in Figures 11S(a) and 11S(b) suggested that such abrupt changes might have been the result of a preceding decline in performance. All of the mice had a significant number of blocks with near-errorless performance as defined statistically by the Wilcoxon test. The near errorless performances began as early as block 5 and as late as block 11, and the length of consecutive near-errorless blocks became longer as the mice became more trained. However, it was notable that the near-errorless performance always waned before resuming. This was very much in contrast to the human subject performance in which there were no prolonged “slump” periods of less than near-errorless, or often perfect, performance. Furthermore, although six of the 10 mice had one or more perfect blocks, the terminal performance on the last 30 trials was intermediate to that of the two groups of humans, with an average of 82% correct (range 57% - 96%). Four mice had terminal correct performances in the 90% range, and three more were over 80%. Of the two mice which had seemingly abrupt improvements in their error scores, one obtained only 77% correct after that change (ko6), while the other obtained 87% correct (wt3). Furthermore, in contrast to the human subjects, near-errorless performance always preceded the seeming abrupt decline in error rate for these two mice. Thus, the mice individually never showed the pattern of error prone performance followed by abrupt change to persistent near-errorless, and often perfect, performance of the human subjects who had acquired the rule. Nevertheless, the mice out-performed the human subjects who did not acquire the rule.

Discussion

We have shown here that despite significant differences in acquisition of olfactory learning set between inbred and F1 hybrid mouse strains (Katz et al., 2003), a disruption of the gene producing FXS in humans does not noticeably diminish performance in either strain background. These results hold true even when test stringency is increased by not enhancing the salience of one of the odors presented in the trial, though such increase in test stringency alters the rate of acquiring the task for all strains and genotypes.

Humans with Fragile X Syndrome have been shown to have deficits in short-term retention of sequences of stimuli (Cornish et al., 2001; Hodapp et al., 1992; Maes et al., 1994). For example, females with FXS performed significantly worse than normals in a 2-back task requiring maintenance of sequences of spatial information in short-term memory (Kwon et al., 2001). Therefore, it was very surprising that a mutation in the mouse *Fmr1* gene (*fmr1*^{tm1Cgr} allele) that dramatically impacts FMRP production would not produce a noticeable deficit in sequence learning and memory in the FX mouse. The biochemical alterations found in the FX mouse do produce neurologic and behavioral differences from wild type, e.g. (Antar et al., 2004; Chen and Toth, 2001; Chuang et al., 2005; El Idrissi et al., 2004; Fisch et al., 1999; Huber et al., 2002; Koekkoek et al., 2005; Li et al., 2002; Mineur et al., 2006; Spencer et al., 2005; Weiler and Greenough, 1993; Yan et al., 2004; Yan et al., 2005). Nonetheless, the findings presented here are consistent with previously reported minimal or nonexistent differences between F1 hybrid FX and wild-type mice in a battery of cognitive tests, including a radial maze task in which mice had to maintain a list of up to four spatial locations in working memory (Yan et al., 2004).

In comparing performance of FX mice and FXS humans in sequence working memory tasks, it may have been some aspect of the tasks other than memory demands which distinguished them. Working memory in these experiments can be seen as procedural memory (the behavior pattern learned during training) acting on the contents of short-term memory (of the odors presented in a given trial). Specifically, the mice had to actively maintain short-term memory of the two most recently exposed single odors in temporal sequence, and then compare those to the three odors presented simultaneously thereafter. Furthermore, the mice had to have correctly learned the basis for making the choices. In the case of the human subjects in this study, it was clear from post-test reports that those who learned the rule underlying the sequence did so by naming the single images to facilitate holding them in short-term memory, and then hypothesis-testing, a form of metacognitive, or higher order, strategy.

Rule use and learning curves

There has been a longstanding debate regarding the ability of non-human animals to employ conscious cognitive strategies comparable to those of humans e.g. (Blumberg and Coppinger, 2005; Wynne, 2004). Recent evidence has been presented to suggest that rats can think in a metacognitive (e.g. seemingly self-reflective) fashion (Foote and Crystal, 2007). It is also clear from Harlow's use of the words "hypothesis", "insight" and "rational" in describing the performance of monkeys in his learning set experiments that he interpreted the quality of thought being assessed to be distinct and of higher order than in other learning tests (Harlow, 1949). For example, Harlow concluded that his data "clearly show that animals can gradually learn insight". Indeed, gradual performance improvements are commonly seen in learning curves produced by animals, including the mice tested here. (Supplemental Information EN2 has additional comment on characteristics of the learning curves). In contrast, the mark of insight and "higher order" rule acquisition on the human learning curves in the present task was quite abrupt, often leading to a step function in the learning curve once the correct rule was deduced (EN3). Nevertheless, what have been termed abrupt increases in learning curves have been observed in conditioning studies of non-human animals (Gallistel et al., 2004), so the specificity of this curve characteristic in signaling the quality of rule use can be problematic. Consistent with this, two of the mice in this study showed what could be interpreted as an "abrupt" improvement in correct choices; however, in both cases the abruptness was produced by a preceding decline in performance.

Of potentially greater diagnostic significance, it was observed in these studies that correct choice fluctuation of the learning curve in its plateau region was much larger for mice than for the human subjects who had acquired and could state the underlying rule of the task. The performance of the latter was essentially perfect, even over extended delays between image presentation and testing.

We term the more variable performance at the relative plateau phase of the learning curve a “streak and slump” pattern (see also (Katz et al., 2003)), since in almost all cases, the error rate at any given trial for even the best performing mice was highly variable, and always would rise several percent after hundreds of trials. These error-frequency effects, and their association with explicit rule use by the human subjects, were made clear by plots of error rates as sliding averages.

Prior indications of “streak and slump” variability in performance for individual rodent subjects are suggested in the learning set literature. (See EN4 for examples.) On average, mice and rats could improve as a group to achieve near-errorless performance, but individual subjects rarely, if ever, do so continuously (see also results of (Reid and Morris, 1993)). Questions have also been raised as to the variable performance of primates in learning set tasks (Slotnick et al., 2000). In Harlow’s error factor analysis of monkey learning set, monkey subjects made “response shift errors” after having made a series of correct choices in a particular object discrimination trial (Harlow, 1950). These errors persisted after more than three hundred 2-object discrimination problems. Thus, the streak and slump performance pattern in learning set studies may be a general one for such non-human species. We argue that resumption of elevated error rates after seeming near-errorless performance characteristic of a learning set actually indicates the absence of higher-order strategy use as defined by human control performance. Extension of testing to quantify learning curve plateau errors demonstrated that differences in error patterns could be discriminated both graphically and statistically. Such a process should provide a mechanism by which to qualify rule use in non-verbal animals and people, including those with mental retardation (EN5).

Fragile X Syndrome and implicit learning and memory systems

As noted above, the degree to which various non-human animals employ conscious problem-solving techniques remains controversial. Cognitive functions responsible for learning and memory can be divided into an implicit, procedural, largely unconscious form, and an explicit, reflective, conscious form (Lewicki et al., 1987; Lewicki et al., 1992; Milner, 2005; Reber, 1989a; Reber, 1992; Seger, 1994). All animals, including humans, use implicit cognitive systems to learn by practice and "feel".

It should be noted that for the purpose of this discussion, we are not concerned with the question of whether mice were actually in some way aware or conscious of what they were learning or remembering. There are several forms of implicit learning in humans, such as implicit expertise, in which there are conscious elements (Holyoak and Spellman, 1993; Perruchet and Pacteau, 1990; Schacter et al., 1993). Furthermore, depending on the definition of explicit learning, non-human animals may exhibit several features in common with humans (Wynne, 1998); indeed, rats apparently can make choices in a metacognitive (seemingly reflective) fashion (Foote and Crystal, 2007). It is also worth noting that not all rules give exact results, as seen when heuristic short-cuts are employed, and it has been well established that non-human animals use such approximation methods.

Therefore, we restrict our definition of "explicit" here to refer to the conscious manipulation of information in working memory in the service of insightful, error-suppressing, rule acquisition. Furthermore, such discrimination of the quality of rule use, e.g. as "explicit", must be limited to tasks in which an algorithimic approach can produce exact results, as was be shown by human controls in this study.

Importantly, we propose that extended perfect performance immediately retrievable upon loss by error provides evidence of algorithmically exact, and therefore very likely explicit, rule use by non-verbal animals (as based upon human subject reports – see Results). We do not require language use, nor, for example, do we exclude the possibility that mice can use internally generated tones (or other neurologic means) to tag or categorize stimuli. What we do require by "explicit" is that the subject appear to have control over the rule, i.e. that if giving evidence of adequate motivation, performance can immediately resume to previously achieved high levels.

The importance of implicit processes with respect to FXS and the performance of FX mice in learning tasks may be based upon the relative robustness of implicit cognition against degradation and loss due to human neurologic diseases such as amnesia and Alzheimer's disease compared to explicit processes (Reber, 1990). Such robustness of implicit thought has been proposed to reflect its evolutionarily older position compared to explicit processes, with retention of greater parallel processing by neural circuitry. If so, this could explain why not much (see also EN6), if any, effect of mutation of the FX gene on cognition in the mouse has been observed, as this species may rely much more, if not exclusively, on implicit rather than explicit processes to learn to solve problems. Conversely, in humans, the effect of FXS would be expected to be much more pronounced, as normal human problem-solving involves a substantial explicit component. This is not to say that implicit processes are completely unaffected by the absence of FMRP, but rather that explicit processes are more generally sensitive and apparently more severely affected in FXS.

We therefore posit that the FXS humans, having disability in verbal and other higher-order strategic functions ordinarily employed in explicit problem solving, must rely much more on implicit procedural learning, as both wt and FX mice apparently do (EN7). It will be interesting to perform extended learning curve analysis on humans with FXS to assess their implicit learning and memory capacities. However, since people with FXS are known to have attentional issues (Munir et al., 2000; Turk, 1998), it might be quite challenging to perform extended sequence testing in this population. Whatever attentional deficits exist as a result of mutation of *Fmr1* in the mouse (Moon et al., 2006) may have been masked here by the use of hunger motivation and punishment threat. (EN8).

Animal modeling of human cognitive deficits

Animal models for cognitive deficits in humans have almost always been tested without any consideration of the problem solving approach taken by humans in a comparable task. For those tasks in which normal human subjects do not employ an explicit, subvocal language-based strategy, it would probably be reasonable to expect similar performances in non-human animals, and therefore genetically modified models of such might be informative as to the nature of the deficit faced by humans who have a related genetic change. In these cases it would be expected that the species learning curves would be similar (assuming that test instruction for the human subjects was limited, as in these studies).

However, in tasks such as those presented here in which it has been demonstrated that at least some significant portion of human subjects do use explicit, language-based hypothesis testing as a strategy, the animal model's learning curve may not correspond to that of the human subjects, and consequently the animal model might have a much more limited deficit, if any, compared to the performance of wild-type mice. (Specific anatomical reasons which might underlie differences in working memory and executive function ability among the species are discussed in EN9.) If the quality of the species learning curves are quite divergent, as was the case in these studies, then the use of the animal model in such tasks to further characterize the human deficit might be somewhat misleading, especially in cases in which there was also a difference between the mouse model and wild-type mice. Therefore, comparative learning curve analysis should be quite informative as to the potential validity of an animal model of a human cognitive disorder. In particular, there should be considerable value in extending the plateau phase of learning curve based tests and comparing the error fluctuations found there with the same from human performing comparable tests.

Abbreviations

FXS: Fragile X Syndrome, a human mental retardation caused by mutation of the X chromosomal *FMR1* gene; FX: Fragile X, generally referring to mice with an *Fmr1* gene disruption; 2-sequence (two-sequence): learning and memory task in which for each trial a subject is serially exposed to two stimuli (odorants for mice and images for humans in this study) at least one of which is distinct from stimuli of the prior trial, after which, in the basic variation, the subject must learn to choose those stimuli in the same order during a test phase of the same trial; 2NC+: [odorant] two not constant [changing in each trial, as is always odorant one], plus [emphasis of odorant one in the exposure phase]; 2NC-: [odorant] two not constant, without [emphasis]; 2C+: [odorant] two constant, plus [emphasis]; LC: learning curve.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CC collected her own data set, trained and supervised others in experimental procedure and the hardcopy and electronic recording of data, contributed significantly to the data analysis, and reviewed the manuscript. JS, JC, TC, SR, GYL, KB, YW, CP, BM, DD, DM, YMM, KI, CS, OF, JB, and QY collected their own data sets and participated in data analysis, AR helped to analyze data and draft the manuscript. JB originally performed the olfactory 8-arm radial maze tasks for the 2009 Pennsylvania Junior Academy of Sciences and CASEF science and engineering fairs. Experimental design, statistical analysis, and primary writing of the manuscript was performed by RB.

Acknowledgments

This work was supported by grants from the FRAXA Research Foundation and the National Institute of Mental Health (R21MH067237) to R.P.B.

References

Baddeley, A., 1998. Working memory. Comptes Rendus de l Academie des Sciences. Serie III, Sciences de la Vie 321, 167-173.

Bakker, C. E., Consortium, D.-B. F. X., 1994. Fmr1 knockout mice: a model to study fragile X mental retardation. Cell 78, 23-33.

Bernstein, D. A., Penner, L. A., Clarke-Stewart, A., Roy, E. j., 2003. Psychology. Houghton Mifflin.

Blumberg, B., Coppinger, R., 2005. Can Dogs Think? Natural History, 48-51.

Cornish, K. M., Munir, F., Cross, G., 2001. Differential impact of the FMR-1 full mutation on memory and attention functioning : a neuropsychological perspective. Journal of Cognitive Neuroscience 13, 144-150.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., Conway, A. R., 1999. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. Journal of Experimental Psychology: General 128, 309-331.

Foote, A. L., Crystal, J. D., 2007. Metacognition in the rat. Current Biology 17, 551-555.

Fortin, N. J., Agster, K. L., Eichenbaum, H. B., 2002. Critical role of the hippocampus in memory for sequences of events. Nature Neuroscience 5, 458-462.

Gallistel, C. R., Fairhurst, S., Balsam, P., 2004. The learning curve: implications of a quantitative analysis. Proceedings of the National Academy of Sciences of the United States of America 101, 13124-13131.

Harlow, H. F., 1949. The formation of learning sets. Psychological Review 56, 51-65.

Harlow, H. F., 1950. Analysis of discrimination learning by monkeys. . Journal of Experimental Psychology 40, 26-39.

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. American Journal of Mental Retardation 97, 39-46.

Holyoak, K. J., Spellman, B. A., 1993. Thinking. Annual Review of Psychology. Vol 44, 265-315.

Katz, E., Rothschild, O., Herrera, A., Huang, S., A., W., Wojciechowski, Y., Gil, A., Yan, Q., Bauchwitz, R., 2003. Odor Based Behavioral Tasks Confounded by Distance Dependent Detection: Modfication of a Murine Digging Paradigm. CogPrints cogprints.ecs.soton.ac.uk/archive/00003316.

Kellogg, R. T., Bourne, L. E., Jr., 1989. Nonanalytic-automatic abstraction of concepts. Sidowski, Joseph B, 89-111.

Kwon, H., Menon, V., Eliez, S., Warsofsky, I. S., White, C. D., Dyer-Friedman, J., Taylor, A. K., Glover, G. H., Reiss, A. L., 2001. Functional neuroanatomy of visuospatial working memory in fragile X syndrome: relation to behavioral and molecular measures. American Journal of Psychiatry 158, 1040-1051.

Levine, M., 1959. A model of hypothesis behavior in discrimination learning set. Psychological Review 66, 353-366.

Lewicki, P., Czyzewska, M., Hoffman, H., 1987. Unconscious acquisition of complex procedural knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition Vol 13(4) Oct 1987, 523-530.

Lewicki, P., Hill, T., Czyzewska, M., 1992. Nonconscious acquisition of information. American Psychologist 47, 796-801.

Maes, B., Fryns, J. P., Van Walleghem, M., Van den Berghe, H., 1994. Cognitive functioning and information processing of adult mentally retarded men with fragile-X syndrome. American Journal of Medical Genetics 50, 190-200.

Milner, 2005. The Medial Temporal-Lobe Amnesic Syndrome. Psychiatric Clinics of North America 28, 599-611.

Moon, J., Beaudin, A. E., Verosky, S., Driscoll, L. L., Weiskopf, M., Levitsky, D. A., Crnic, L. S., Strupp, B. J., 2006. Attentional dysfunction, impulsivity, and resistance to change in a mouse model of fragile X syndrome. Behavioral Neuroscience 120, 1367-1379.

Munir, F., Cornish, K. M., Wilding, J., 2000. Nature of the working memory deficit in fragile-X syndrome. Brain and Cognition 44, 387-401.

Neal, A., Hesketh, B., 1997. Episodic knowledge and implicit learning. Psychonomic Bulletin & Review. Vol 4, 24-37.

Numminen, H., Service, E., Ahonen, T., Korhonen, T., Tolvanen, A., Patja, K., Ruoppila, I., 2000. Working memory structure and intellectual disability. Journal of Intellectual Disability Research 44 (Pt 5), 579-590.

Perruchet, P., Pacteau, C., 1990. Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? Journal of Experimental Psychology: General. Vol 119, 264-275.

Reber, A. S., 1989a. Implicit Learning and Tacit Knowledge. Journal of Experimental Psychology: General 118, 219-235.

Reber, A. S., 1989b. More thoughts on the unconscious: reply to Brody and to Lewicki and Hill.[comment]. Journal of Experimental Psychology: General 118, 242-244.

Reber, A. S., 1990. On the primacy of the implicit: Comment on Perruchet and Pacteau. Journal of Experimental Psychology: General. Vol 119, 340-342.

Reber, A. S., 1992. The cognitive unconscious: An evolutionary perspective. Consciousness and Cognition 1, 93-133.

Reid, I. C., Morris, R. G., 1993. The enigma of olfactory learning. Trends in Neurosciences 16, 17-20.

Schacter, D. L., Chiu, C. Y., Ochsner, K. N., 1993. Implicit memory: a selective review. Annual Review of Neuroscience 16, 159-182.

Seger, C. A., 1994. Implicit learning. Psychological Bulletin 115, 163-196.

Slotnick, B., Hanford, L., Hodos, W., 2000. Can rats acquire an olfactory learning set? Journal of Experimental Psychology: Animal Behavior Processes 26, 399-415.

Turk, J., 1998. Fragile X syndrome and attentional deficits. Journal of Applied Research in Intellectual Disabilities. Vol 11, 175-191.

Willingham, D. B., Nissen, M. J., Bullemer, P., 1989. On the development of procedural knowledge. Journal of Experimental Psychology. Learning, Memory, and Cognition 15, 1047-1060.

Wynne, C. D. L., 1998. A Natural History of Explicit Learning and Memory. In: Kirsner, K., Speelman, C., Maybery, M., O'Brien-Malone, A., Anderson, M., MacLeod, C., (Eds), Implicit and Explicit Mental Processes. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 255-269.

Wynne, C. D. L., 2004. The perils of anthropomorphism. Nature 428, 606.

Yan, Q. J., Asafo-Adjei, P. K., Arnold, H. M., Brown, R. E., Bauchwitz, R. P., 2004. A phenotypic and molecular characterization of the fmr1-tm1Cgr Fragile X mouse. Genes, Brain and Behavior 3, 337-359.

Figure Legends

Fig. 1. Schematic of the 2-sequence protocol. (A) Exposure and test phases. (B) Odor cup array and cup shuffling. A similar cup array was used for the 5-sequence test, except that the cups were used in a series of five per trial and an additional 26th cup was added in order to shift the cups out of phase after the first 5 trials.

Fig. 2. Effect of genotype and strain on 2-sequence plateau performance after 17 days of training in the 2NC+ task. (A) Wild-type and FX (ko) F1 hybrid adult male mice achieved the same level of performance by the last four days of testing. (B) FVB/NJ inbred mice of either genotype [11] did not match the F1 hybrid performance.

Fig. 3. Change in 2-sequence response choices from days 1-3 to days 15-17. Combined data for wild-type and FX F1 hybrid adult male mice showing changes in the four possible responses for each of three tasks (2NC+, 2NC-, 2C+). “2NC+”: odor 1 was emphasized by presenting and rewarding it twice while odor 2 was not held constant (changed on every trial); “2NC-”: odor 1 was not emphasized and odor 2 was not held constant; “2C+”: odor 1 was emphasized and odor 2 was held constant.

Fig. 4. Learning curves for three task variants. Performance for F1 hybrid adult male mice in each of three tasks was compared using one-factor ANOVA with Tukey-Kramer post-hoc analysis. “% 1 then 2”: correct responses. Statistical significance is indicated as follows: * are for 2NC- vs. 2C+, ‡ is for 2NC- vs. 2NC+, ◇ is for 2NC+ vs. 2C+.

Fig. 5. Effect of dropping punishment. After 17 days of training, punishment was dropped in two experiments for five days (“wPdrop” means “with punishment drop”). Regardless of emphasis of odor one, a significant decline in performance by the F1 hybrid males was observed.

Fig. 6. Effect of extending the test period. (A) “2NC+”: standard task in which wt and ko (FX) F1 hybrid mice were trained to 23 days. (B) “2C+”: simpler task in which wt and ko (FX) F1 hybrid mice were trained to 28 days. (C) Improvement in performance from original test end at day 17 compared to extended periods was assessed by one-factor ANOVA with Tukey-Kramer post-hoc analysis. Task abbreviations are as described in Figure 3.

Fig. 7. Human 2-sequence learning curves. For each trial: 2 points for choice of image 1 then 2 (correct), 1 point for 1 first, 0 points for 2 first or 3 first. “yo” means “years old”. Trial performance - black diamonds. Day scores are the sum of trial scores - white circles; 32 points maximum. Trial on which explicit rule acquired - open triangle; 15 yo female did not acquire the rule, so no triangle is present. Delays between exposure and test phases were 10 seconds, except for days showing “+30s” and “+90s” which had delays of 30 and 90 seconds, respectively.

Fig. 8. Mouse 2-sequence learning curves. Data for individual mice in the 2C+ extended test (odor two constant, least stringent task) were converted to point scores as for the individual human learning curves (Fig.7 and 7S). Maximum of 24 points per day (less than for human subjects due to fewer trials per day for the mice). Delays between exposure and test phases were 10 - 15 seconds. Mouse identification numbers follow the wild type (wt) and knockout (ko) designations.

Fig. 9. Terminal error rate changes. Terminal errors for last (1-12) and second to last (13-24) block of twelve trials in the 2-sequence task. For mice, the task version was the least stringent, 2C+: odor two constant with emphasis on the first odor. "Hu explicit": human subject who could consciously state the correct rule; "Hu non-explicit" never stated the correct rule; "Mu 0-1 start": mouse with 0 or 1 error in trials 24 to 13 prior to the task end; "Mu 2+ start": mouse with 2 or more errors in trials 24 to 13 prior to the task end. Statistical significance: Kruskal-Wallis (29.71 ***, $p < 0.001$) followed by Dunn's Multiple Comparison Test (* indicates $p < 0.05$, as indicated on the chart). Error bars are SEM.

Fig. 10. Human sliding error score plots. Error score = trial points obtained minus maximum trial points. Each data point represents the average error score of the current trial and the prior nine trials. Trial points are calculated as indicated in Fig.s 7 and 8. 15 yo female, 45 yo male, and 45 yo female did not acquire the rule. All other human

subjects in this figure and Fig. 10S did so: 10 yo male on trial 56, 13 yo male on trial 14, and 11 yo male on trial 61.

Fig. 11. Mouse sliding error score plots. Error score = trial points obtained minus maximum trial points. Each data point represents the average error score of the current trial and the prior nine trials. Trial points are calculated as indicated in Fig. 8.

Additional Data Files

File name: Cai_et._al._Supplemental_Information_0909.pdf

File format: pdf

Title: Supplemental Information

Description: Supplementary tables of statistical results, figures, text, and testing materials as referenced in the article.