JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2000, **73,** 305–318 NUMBER 3 (MAY)

AN OLFACTORY DISCRIMINATION PROCEDURE FOR MICE

SHEILA M. MIHALICK, JASON C. LANGLOIS, JASON D. KRIENKE, AND WILLIAM V. DUBE

E. K. SHRIVER CENTER FOR MENTAL RETARDATION

This paper describes an olfactory discrimination procedure for mice that is inexpensively implemented and leads to rapid discrimination learning. Mice were first trained to dig in small containers of sand to retrieve bits of buried chocolate. For discrimination training, two containers were presented simultaneously for eight trials per session. One container held sand mixed with cinnamon, and the other held sand mixed with nutmeg. Both containers were baited with chocolate buried in the sand. One odor was designated S+, and mice were allowed to dig and retrieve the chocolate from this container. The other odor was S−, and both containers were removed immediately if subjects began to dig in an S− container. After meeting a two-session acquisition criterion, subjects were given a series of discrimination reversals. In Experiment 1, 12 Swiss-Webster mice (6 male and 6 female) acquired the olfactory discrimination in three to five sessions and completed 3 to 10 successive discrimination reversals within a 50-session testing limit. In Experiment 2, subjects were 14 Pah^{enu2} mice, the mouse mutant for phenylketonuria; 7 were homozygotes in which the disorder was expressed (PKU), and 7 were heterozygotes with normal metabolism (non-PKU). Thirteen mice completed pretraining in four to seven sessions, acquisition required 3 to 12 sessions, and all mice completed at least three reversals. Learning rates were similar in PKU and non-PKU mice. We discuss issues related to implementation and several potentially useful procedural variations.

Key words: olfactory discrimination, discrimination reversal, phenylketonuria, digging, Pah^{enu2} mice, Swiss-Webster mice, mice

Recent developments indicate a growing potential for behavioral research with mice. Nonhuman mammalian genome mapping is currently focused on mice, and advances in bioengineering techniques continually produce new mutant, transgenic, and knockout strains for biomedical research (e.g., Lewis, 1999; Nelson, 1997). This situation offers new opportunities for behavior analysis. Efforts to characterize the behavioral phenotype of mouse strains are incomplete without measures of learning and memory (Crawley & Paylor, 1997; Rogers et al., 1997). Also, appropriate behavioral tests are necessary for valid animal models of conditions that are defined at least in part by behavioral effects (e.g., mental retardation; McIlvane & Cataldo, 1996).

The research literature on learning and memory in mice includes many studies using punishment (e.g., contextual fear conditioning), active or passive avoidance, or escape contingencies (e.g., Morris water maze), but tests for mice that use positive reinforcement contingencies are much less common (e.g., Wehner & Silva, 1996). Of those behavioral test procedures that do use positive reinforcement contingencies, the majority are tasks in which the controlling stimuli are spatial. As an example of current trends in behavioral testing with mice, Crawley and Paylor (1997) recently recommended 55 research papers as resources for behavioral neuroscientists interested in investigating learning and memory in transgenic and knockout mice: 16 examples of the Morris water task, 12 of active or passive avoidance, 7 of response-independent conditioning, 17 of spatial learning tasks in mazes, and 3 of delayed nonmatching to position. There were no recommendations for positively reinforced nonspatial discrimination learning tasks. A search of abstracts from the *Journal of the Experimental Analysis of Be-*

Data collection and manuscript preparation were supported by NICHD P30 HD04147 (W. J. McIlvane, Principal Investigator). We thank Joanne Berger-Sweeney and Jill Arters for allowing us to visit their laboratory and observe experimental sessions, Don Druin for apparatus design and construction, Don and Gerson Tomanari for assistance during preliminary studies, Adele Diamond for contributing mice and advice during the preliminary studies, Charissa Dyer for supplying the Pah^{enu2} mice used in Experiment 2, Bill McIlvane for comments on the manuscript, and Lynn Doucette for photography.

Address correspondence to Sheila Mihalick or William Dube, Psychological Sciences Division, E. K. Shriver Center, 200 Trapelo Road, Waltham, Massachusetts 02452 (E-mail: smihalick@shriver.org or wdube@shriver.org).

havior returned three research papers in which mice served as experimental subjects: One used shock avoidance (Powell & Peck, 1969, Experiment 2), one used a contingency in which responses in a cold (3 °C) environment produced a warm air stream (arguably either escape from cold or positive reinforcement by heat; Sakagami, Hursh, Christensen, & Silberberg, 1989), and the third used pup retrieval as a positive reinforcer for female mice (Van Hemel, 1973).

This paper describes a positively reinforced nonspatial olfactory discrimination learning procedure for mice that is appropriate for either group designs or single-subject research. The procedure offers several practical advantages: inexpensive apparatus, ease of implementation, and rapid acquisition. It is an adaptation of one for rats reported by Bunsey and Eichenbaum (1996) and Dusek and Eichenbaum (1997). In these studies, the stimuli were containers of sand mixed with household spices (basil, nutmeg, etc.). When two containers were presented simultaneously, the rats responded by selecting one on the basis of olfactory stimuli and then digging in the sand for buried food reinforcers.

Berger-Sweeney, Libbey, Arters, Junagadhwalla, and Hohmann (1998) adapted Eichenbaum's methods for use with mice, and the procedure reported here is a substantially modified version of theirs. One modification was to the apparatus. In Berger-Sweeney et al.'s procedure, subjects were moved back and forth between a holding cage and a test cage for every trial. To reduce the potential for stress effects from handling, we developed an apparatus for testing mice in their home cages.

Also, we made several changes in the discrimination procedures after pilot testing suggested the possibility of alternate sources of stimulus control. The pilot tests were conducted with 5 mice that had had substantial training (40 to 45 sessions) with the Berger-Sweeney et al. (1998) procedure. For these pilot tests, the experimenter-specified olfactory cues (cinnamon and curry powder) were eliminated, and the containers designated by the experimenter as correct or incorrect were both filled with plain sand. Several procedural variables were manipulated during a series of sessions. The results of these tests suggested that in some subjects the controlling stimuli could include (a) containers that were reused consistently as the correct and incorrect alternatives for a series of trials and (b) the presence versus absence of bits of chocolate buried in sand. These results were subsequently confirmed by Zagreda, Goodman, Druin, McDonald, and Diamond (1999) in a systematic replication with naive mice and with scented sand.

As a consequence of these pilot tests, the procedures described in the present paper differed from Berger-Sweeney et al.'s (1998) procedures in several ways: (a) We used different, clean containers for every trial, rather than reusing containers within sessions. (b) We baited containers in a way that avoided contact between the chocolate and the sand at the surface of the container, rather than using forceps to insert a piece of chocolate beneath the sand in previously filled containers. (c) We baited both the correct and incorrect containers on every trial, rather than baiting only the correct container. (d) Because the incorrect container was baited, we changed the consequence for an incorrect response from within-trial correction (continuing the trial until the mouse switched from the incorrect container to the correct container and retrieved the chocolate) to immediate removal of both containers.

Finally, we increased the duration of preliminary training from two to six sessions. We planned to conduct a series of discrimination reversals, and a relatively large number of consecutive errors often occurs immediately after a reversal, as the subject continues to select the stimulus that was previously correct. Because there was no within-trial correction procedure, trials with errors ended immediately with no reinforcer. To reduce the likelihood of extinguishing responding with the first reversal, the number of preliminary training sessions was increased to provide a more extensive reinforcement history for digging in the sand containers.

The purpose of this paper is to describe the modified procedures in detail and show their application to a common strain of mouse, the Swiss-Webster. Also, we report a study of discrimination reversal learning in a mutant strain that models the metabolic disorder phenylketonuria.

EXPERIMENT 1

Experiment 1 describes an implementation of the olfactory discrimination procedure in outbred Swiss-Webster mice, which are widely used in psychological and medical research. Specific examples of the range of studies using Swiss-Websters include models of panic attacks (Griebel, Blanchard, & Blanchard, 1996), drug dependence (Gallaher, Jacques, & Hollister, 1987), immunosuppression (Dunn, 1988), and the pharmacokinetics of ethanol bioavailability (Pastino, Sultatos, & Flynn, 1996). Because this mouse stock is a standard laboratory tool with a broad and well-established research literature, it seemed to be a reasonable choice for a benchmark assessment of the procedure.

METHOD

Subjects

Subjects were 12 Swiss-Webster mice (Charles River Laboratories), 6 males and 6 females, 60 days old at the start of testing. All animals were experimentally naive. They were reduced to 80% to 90% of free-feeding body weight, the mean weight for 3 days prior to food restriction. This range, which is somewhat larger than that typical for rats, was adopted because of relatively lower body weight. Mean 85% weight was 22 g, so the allowable fluctuation was approximately ±1 g. Throughout the study, weights were monitored daily and were maintained by supplemental feedings of 3 to 7 g standard rodent chow at least 1 hr after the last experimental session of the day. The mice were housed individually, maintained on a 12:12 hr light/dark cycle, and were provided with water ad lib.

Apparatus

The olfactory discrimination apparatus was a divided plastic platform upon which small containers could be attached with Velcro® patches (see Figure 1A). The divider prevented subjects from making simultaneous contact with both containers. The containers were plastic bottle caps (3 cm diameter, 1.3 cm deep). The sand was Quikrete® washed and screened playground sand, sifted to remove small bits of rock. Playground sand was used because construction-grade sand may cause nasal lesions (H. Eichenbaum, personal

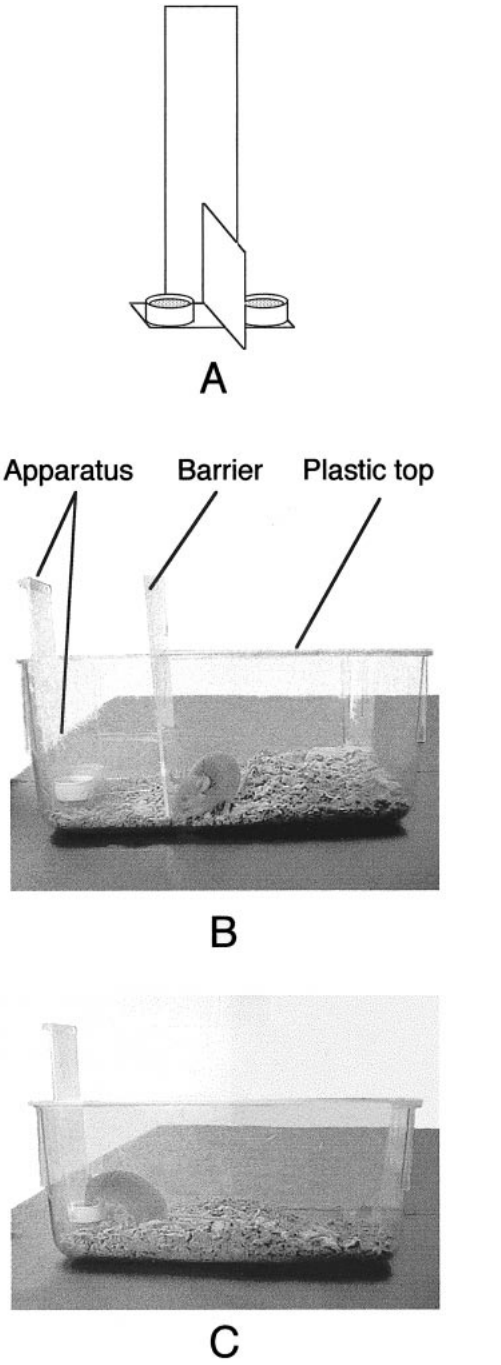

Fig. 1. A: stimulus presentation apparatus. Containers were small bottle caps attached to the apparatus by Velcro® patches. B: experimental arrangement during an intertrial interval with barrier and apparatus in place. C: a subject responding to S+ and digging in the sand to retrieve a small piece of chocolate.

communication, December 1996). Containers were baited by partially filling them with sand to a depth of approximately 0.5 cm, placing a small bit of chocolate (approximately 20 mg of Hershey's Semisweet Morsels) in the center, and then adding sand to a total depth of approximately 1 cm. Within each experimental session, different containers were used on every trial. Containers were washed by hand between sessions. Other apparatus included a sheet of clear plastic large enough to cover a home cage, a second piece of plastic cut to the width of a cage (used as a barrier, see below), and a timer.

Procedure

Experimental sessions were conducted 6 days per week, one session per day. Subjects were tested in their home cages (clear polyurethane bins 29 cm long by 19 cm wide by 13 cm high). Before sessions, the cage was placed on a tabletop, and the wire cage top was removed and replaced by a clear plastic top. During the intertrial interval (ITI), the experimenter slid the plastic top back and inserted a clear plastic barrier vertically into the cage. This barrier was positioned so that the apparatus could be inserted on one side of it while the subject remained on the other side (see Figure 1B). When the 30-s ITI was completed, the barrier was removed, regardless of the subject's position in the cage, to allow the subject access to the apparatus (Figure 1C).

Preliminary training. Subjects were taught to dig in containers of unscented sand to retrieve small pieces of chocolate. During pretraining, a single container was presented on each trial in the left or right location equally often. In Pretraining Step 1, the container was baited with three pieces of chocolate: one buried just beneath the surface of the sand, a second partially buried, and a third placed on top of the sand. Subjects received four trials per session. If both the exposed and partially buried pieces were not retrieved within 15 min, the trial and session were terminated. If the exposed and partially buried pieces, but not the completely buried piece, were retrieved within 15 min, the trial was terminated but the session continued. Step 1 was completed following two consecutive sessions in which all pieces were retrieved on at least one trial.

In Pretraining Step 2, the containers were presented as in Step 1, but with partially and completely buried chocolate only. The buried pieces were placed more deeply on successive trials. As in Step 1, each session consisted of four trials. If the subject failed to retrieve the partially buried piece within 15 min, both the trial and the session were terminated. Step 2 was completed following two consecutive sessions in which both pieces were retrieved on at least seven of eight trials.

In Pretraining Step 3, the containers were presented with one piece of chocolate, buried at a depth of approximately 0.3 to 0.5 cm in the first session and 0.5 cm thereafter. Sessions consisted of two blocks of four trials each, with an ITI of 30 s and 5 to 10 min between blocks. Failure to retrieve the chocolate within 15 min resulted in the termination of the trial and session. Step 3 was completed following two sessions in which the chocolate was retrieved on every trial.

Discrimination acquisition. Discrimination training began in the session following completion of Pretraining Step 3. Sessions consisted of two blocks of four trials each, with a 30-s ITI and 5 to 10 min between blocks. Two containers were presented on each trial. One contained sand scented with cinnamon and the other contained sand scented with nutmeg (0.6 g spice per 100 g sand). For each subject, one odor was randomly designated correct (S+) and the other incorrect (S−). Both the S+ and S− containers were baited on every trial. The left-right position of the correct stimulus was determined randomly with the restrictions that both positions were correct twice in each block and the same position could not be correct more than three times consecutively in one session.

A response was defined as paws or nose in contact with the sand plus digging motions with the paws or nose. If the first response was to S+, the mouse was allowed to dig until it retrieved the chocolate, the apparatus was removed, and a correct response was recorded. Occasionally, the first response was to S+, but the mouse did not retrieve the chocolate (it moved away before finding the chocolate, flipped the chocolate out onto the bedding while digging, etc.). In such cases, to ensure that all correct responses would be followed by reinforcers, the apparatus was raised, the S− cap quickly removed, and the apparatus was re-presented until the subject retrieved

Table 1
Sessions completed by Swiss-Webster mice in Experiment 1.

	P	A	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
Females												
SWF1	6	3	6	9	10	10	6	*6*				
SWF2	6	3	7	6	11	9	*14*					
SWF3	6	3	6	7	12	14	*8*					
SWF4	6	4	9	7	6	8	9	*7*				
SWF5	6	3	6	11	14	10	*7*					
SWF6	6	3	8	9	9	19	*2*					
M	6.0	3.2	7.0	8.2	10.3	11.7						
Males												
SWM1	6	3	7	9	4	4	5	3	4	6	3	3
SWM2	6	3	4	8	12	9	8	*6*				
SWM3	6	3[a]	7[a]	6	3	5	11	16				
SWM4	6	5	13	6	15	*11*						
SWM5	6	3	5	5	7	12	7	2	3	*6*		
SWM6	6	3	6	10	5	6	14	*6*				
M	6.0	3.3	7.0	7.3	7.7							

Note. P = pretraining; A = olfactory discrimination acquisition; R1, R2, etc., = discrimination reversals; italics indicate that the learning criterion was not met.

[a] Reversal 1 was initiated before SWM3 had met the learning criterion because of experimenter error; data for SWM3 were not included in the calculation of group means for acquisition or the first reversal.

the chocolate. If necessary, a piece of chocolate was placed on top of the sand in the S+ container before it was re-presented.

If the first response was to S−, the apparatus was removed immediately, before the mouse could retrieve the chocolate, and an incorrect response was recorded. If there was no response within 2 min, the apparatus was removed, and ''no response'' was recorded for the trial.

The only exception to the S− procedure occurred on the first block of trials in the first discrimination training session (i.e., the first four trials after pretraining). A correction procedure was used to ensure delivery of reinforcers on initial discrimination trials. If the mouse began to dig in the S− container, the apparatus was raised (but not removed) until the mouse moved away, and then the apparatus was lowered immediately. The correction procedure was repeated as necessary with no time limit until the mouse retrieved the chocolate from the S+ container. Correction trials were scored as errors.

The acquisition criterion was at least seven of eight correct responses for two consecutive sessions, with the exception that the first session of discrimination training did not count toward this learning criterion (because of the correction procedure). Thus, a minimum of three sessions was required for acquisition of the first discrimination.

Repeated discrimination reversals. After a subject met the acquisition criterion, the odors designated as S+ and S− were reversed in the following session; the previous S+ became the S−, and vice versa. Training continued to a reversal learning criterion of seven of eight correct for two consecutive sessions. Thereafter, each time a subject met the reversal learning criterion, S+ and S− stimuli were reversed again in the following session, for a series of repeated reversals. Testing ended after 50 sessions; one additional session was scheduled if Session 50 was the first session meeting a learning criterion.

Results

Table 1 shows the number of sessions for each subject in pretraining, discrimination acquisition, and subsequent reversals. All mice completed pretraining in the minimum of six sessions. Most mice also met the learning criterion for the initial odor discrimination in the minimum of three sessions. All females completed at least four reversals, and mean sessions per reversal for the first three reversals increased from 7.0 for the first reversal to 10.3 for the third. All males completed at least three reversals, and the mean

number of sessions per reversal increased only slightly, from 7.0 to 7.7.

Figures 2 and 3 show, for females and males, respectively, session-by-session accuracy for acquisition and all reversals in which the learning criterion was met. Sessions for incomplete reversals, shown by italics in Table 1, are not included. Half of the subjects (e.g., SWF1) had criterion-level accuracy of at least seven correct responses in the first session of discrimination training with the cinnamon and nutmeg olfactory cues. One possible concern with such results is artifactual control by stimuli other than the cinnamon and nutmeg odors. This possibility seems to be ruled out by the results of the initial sessions of the first discrimination reversal. Most subjects had zero correct responses in the first reversal session, an indication of stimulus control by the odor that was formerly S+. This pattern of results continued throughout training, with accuracy at or below chance levels in initial reversal sessions; the only exception was SWM5 in Sessions 14 and 40. Failures to respond within 2 min were rare (a total of six trials for 4 subjects; open points in Figure 2). Taken together, the data in Figures 2 and 3 show that the procedures produced reliable responding, rapid acquisition of discriminative control by the olfactory stimuli, and orderly reversal learning.

EXPERIMENT 2

In Experiment 2, we tested Pah^{enu2} mice, a mutant in the BTBR strain that is a genetic mouse model for the disorder phenylketonuria (PKU; McDonald, Bode, Dove, & Shedlovsky, 1990; Shedlovsky, McDonald, Symula, & Dove, 1993). The homozygous animal, in which the metabolic defect is expressed, has a deficiency in phenylalanine hydroxylase, the enzyme that catalyzes the conversion of dietary phenylalanine to tyrosine. Homozygotes exhibit many of the biological characteristics of human PKU, including hypomyelination and gliosis (Dyer et al., 1996). Phenylalanine metabolism in the heterozygous animal is within normal limits. These mice are somewhat fragile and difficult to breed. Pah^{enu2} mice were tested with the olfactory discrimination procedure as part of an effort to develop a biobehavioral animal model of myelin-related disorders.

Method

Subjects

Subjects were 7 homozygous Pah^{enu2} males (PKU) and 7 heterozygous males (non-PKU) obtained from the breeding colony of Charissa Dyer in the Neurology Department at Children's Hospital of Philadelphia and the University of Pennsylvania Medical School. The mice were drawn from four litters produced by mating homozygous (−/−) males with heterozygous (+/−) females. Mean age was 2.3 months when they were shipped to our laboratory and 7.2 months at the start of testing. All animals were experimentally naive. Food restriction and weight maintenance procedures were as described in Experiment 1. Because they were fed standard rodent chow that contained phenylalanine, the homozygous mice exhibited uncontrolled PKU. High-performance liquid chromatographic analysis of blood samples taken from the 13 subjects that completed behavioral testing showed that the PKU mice had significantly elevated concentrations of blood phenylalanine compared to the non-PKU mice [group means of 9.23 mg/dl and 1.00 mg/dl, respectively; $t(11) = 8.30$, $p < .001$]. The PKU mice were hypopigmented and were easily distinguished from non-PKU mice by their lighter coat color.

Apparatus and Procedure

The apparatus, preliminary training, and discrimination procedures were as described for Experiment 1 with the following exceptions: Reinforcers were small bits of phenylalanine-free chocolate (Ambrosia chocolate-flavored bark coating), used in anticipation of follow-up experiments in which dietary phenylalanine would be controlled. The criterion for completing Pretraining Step 3 was one session in which the chocolate was retrieved on every trial. The minimum number of training sessions was 54, not including pretraining. The experiment ended after 54 sessions if subjects had completed three reversals; if not, sessions continued until three reversals were completed.

Results and Discussion

Pretraining. One non-PKU mouse was eliminated during Pretraining Step 1 when it retrieved only one piece of chocolate in six ses-

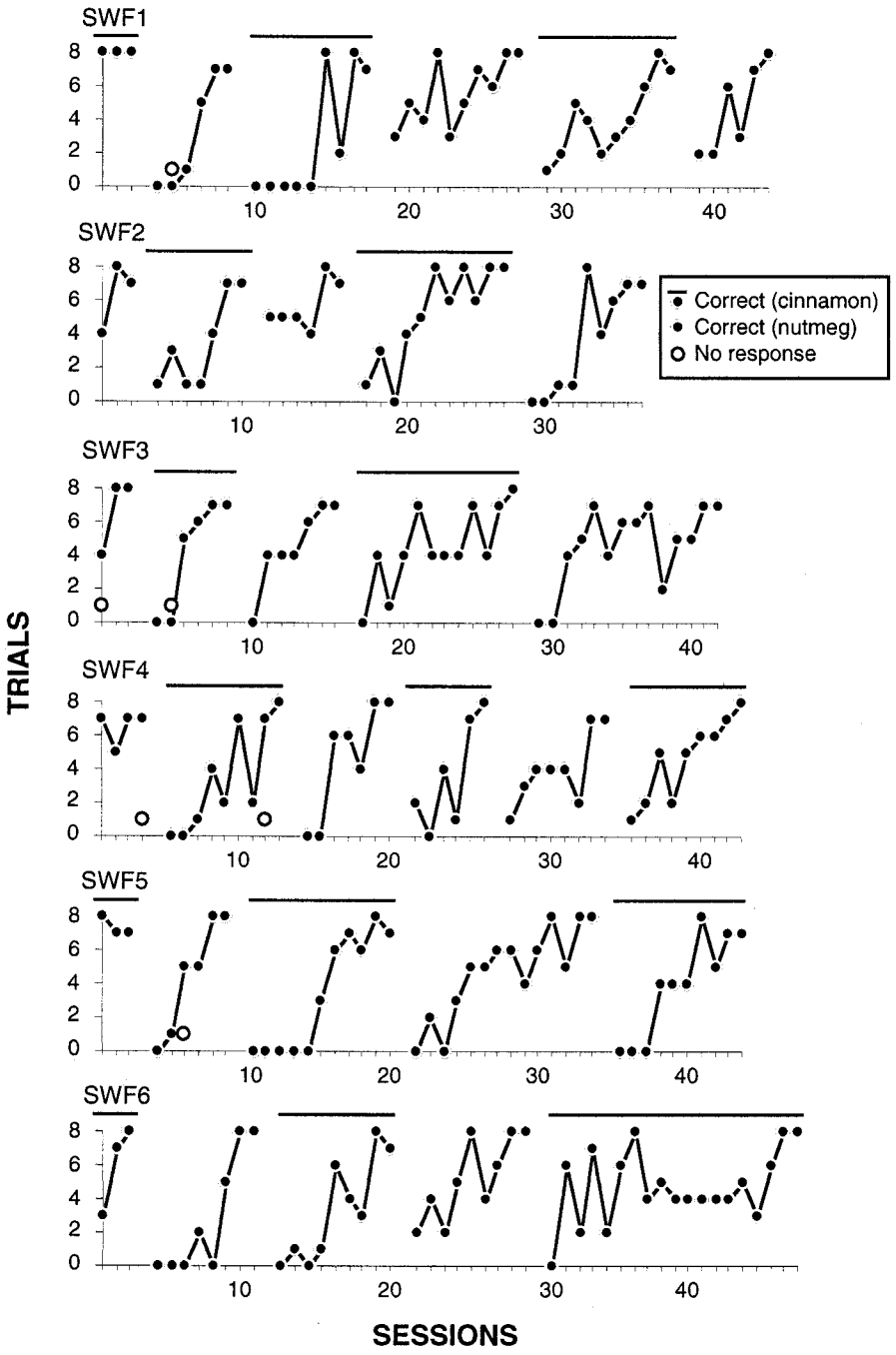

Fig. 2. Individual session scores for female Swiss-Webster mice in Experiment 1. Each session consisted of eight trials. Filled points show the number of correct responses, and open points show the number of trials terminated after 2 min without a response (zeros omitted). Cinnamon was the correct odor in conditions in which a horizontal bar appears above the data; nutmeg was correct in conditions with no bar.

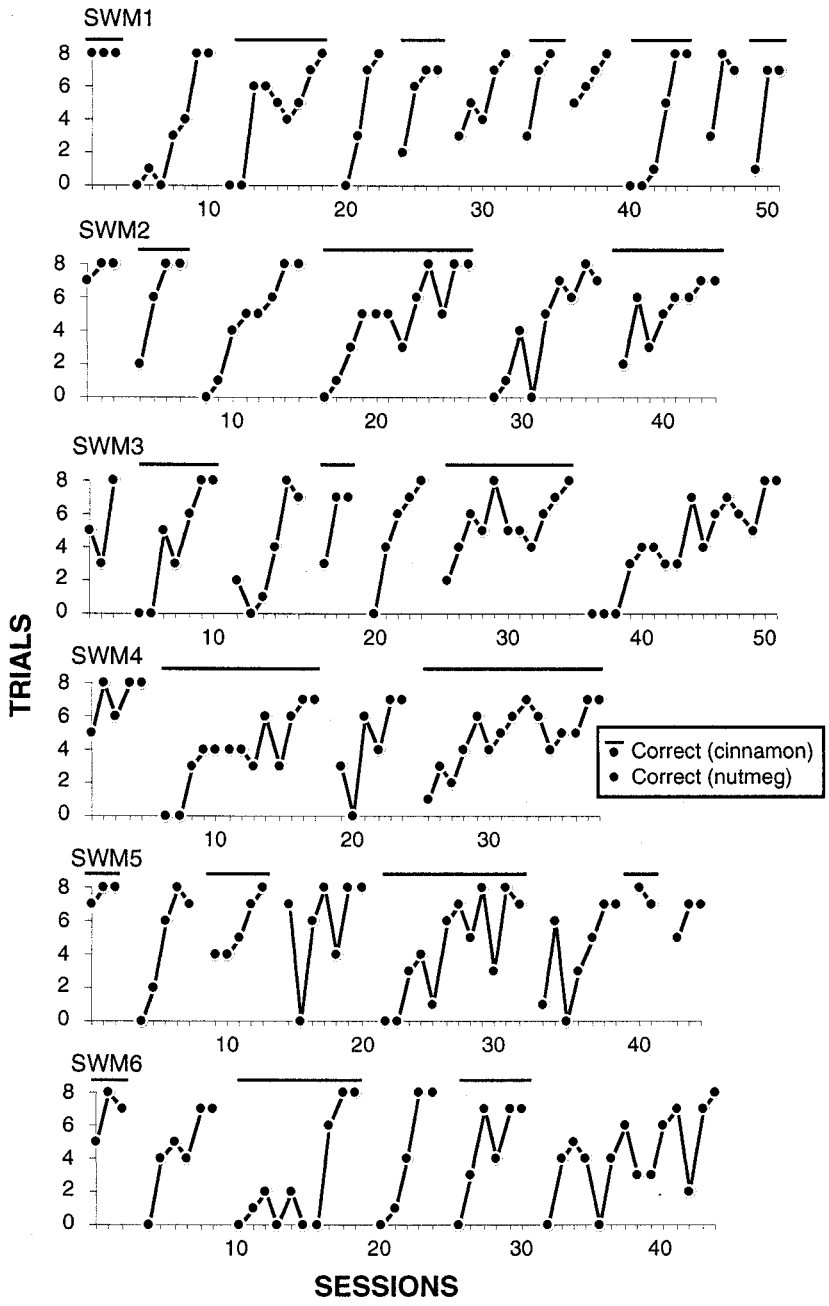

Fig. 3. Individual session scores for male Swiss-Webster mice in Experiment 1. See Figure 2 caption for details.

sions. Table 2 shows that all but 1 of the remaining animals required either five (the minimum) or six sessions to complete pretraining.

Discrimination acquisition and reversals. Table 2 shows that the mice required 3 to 12 sessions to meet the criterion for discrimination acquisition. Figures 4 and 5 show the data for individual sessions for non-PKU and PKU mice, respectively. The figures show only reversals in which the learning criterion was met; sessions for incomplete reversals, shown by italics in Table 2, are not included.

Data from the first discrimination session

Table 2
Sessions completed by Pah^{enu2} mice in Experiment 2.

	P	A	R1	R2	R3	R4	R5	R6	R7
Non-PKU									
N104	4[a]	8	9	3	6	8	9	5	*6*
N105	6	3	11	8	11	8	*13*		
N106	5	4	11	18	27				
N205	5	7	11	9	15	*12*			
N206	5	6	21	11	13	*3*			
N208	6	7	10	19	23				
M	5.4	5.8	12.2	11.3	15.8				
PKU									
P101	6	4	11	13	16	*10*			
P102	6	9	9	17	6	*13*			
P103	5	12	21	22	20				
P201	6	6	6	12	7	23			
P202	5	8	25	11	18				
P203	5	10	18	12	11	*3*			
P204	7	5	13	13	19	*4*			
M	5.7	7.7	14.7	14.3	13.9				

Note. P = pretraining; A = olfactory discrimination acquisition; R1, R2, etc., = discrimination reversals; italics indicate that the learning criterion was not met.

[a] Pretraining Step 2 ended after one session because of experimenter error; data for N104 were not included in the calculation of the pretraining group mean.

suggest a stimulus preference. Four of the 6 animals with cinnamon as S+ made correct choices on seven of eight trials, and 5 of the 7 animals with nutmeg as S+ made correct choices on only two of eight trials. The effects of any preference for cinnamon were apparently brief, and the mean number of sessions to acquisition was similar for mice with cinnamon and nutmeg as S+, 6.3 and 7.3 sessions, respectively. No preference was seen with the Swiss-Webster mice in Experiment 1, and this difference suggests that stimulus preference pretests may be useful when implementing the procedure with different mouse strains.

As in Experiment 1, accuracy in the initial sessions following discrimination reversals was always at or below chance level. Reversal learning was generally slower than in Experiment 1, and 4 mice required testing beyond the 54-session limit to complete three reversals (N106, N208, P103, and P202). Figures 4 and 5 also show that all non-PKU and PKU mice occasionally failed to respond within the 2-min trial limit (open points in the figures). Often such failures occurred during the initial sessions of a reversal, when accuracy was low for several consecutive sessions. The frequency of these no-response trials subsequently declined as accuracy improved. In contrast, the Swiss-Webster mice in Experiment 1 rarely failed to respond within the trial limit of 2 min, suggesting a possible strain- or perhaps age-related difference in response to sudden changes in reinforcement contingencies.

Zagreda et al. (1999) recently reported significant differences between groups of PKU and non-PKU mice on rate of learning olfactory discrimination reversals. For the present data, a repeated measures analysis of variance (ANOVA) on the number of sessions to meet criterion for acquisition and the first three reversals did not reveal an effect of genotype, $F(1, 11) = 0.47$, $p > .05$. Thus, by the metric of a null hypothesis significance test, we failed to replicate the main finding in Zagreda et al., although the procedures were similar in many respects. Crabbe, Wahlsten, and Dudek (1999) recently reported significant and systematic differences among laboratories in the results of behavioral testing with mutant mice, even with rigorous efforts to standardize procedures. The protocols for the present study and that of Zagreda et al. were similar because of collaboration during procedural development, but there was no effort to standardize the procedures, and our two laboratories conducted completely independent studies of PKU mice. There were a number

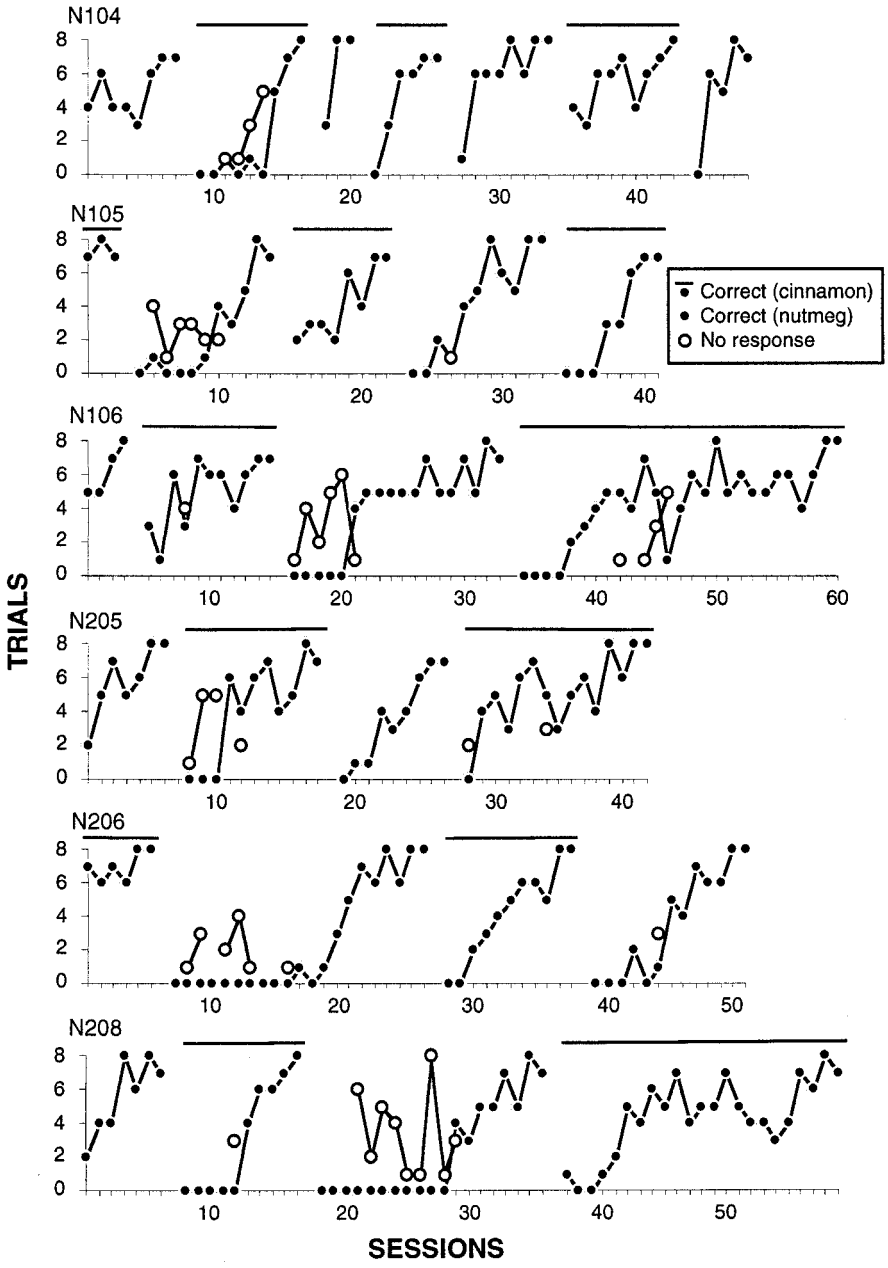

Fig. 4. Individual session scores for non-PKU mice in Experiment 2. See Figure 2 caption for details.

of differences between the two studies, and some that may be relevant to the outcomes are discussed below.

The studies differed in some of the characteristics of groups and subjects. Zagreda et al. (1999) reported olfactory discrimination data for PKU and two non-PKU groups, heterozygotes and wild-type BTBR mice. Although they found no significant differences between heterozygous and wild-type control groups on any measures and a significant difference between the PKU group and both control groups for the second reversal, they found significant differences only between PKU and wild-type control groups for the third and fourth reversals. Their analyses in-

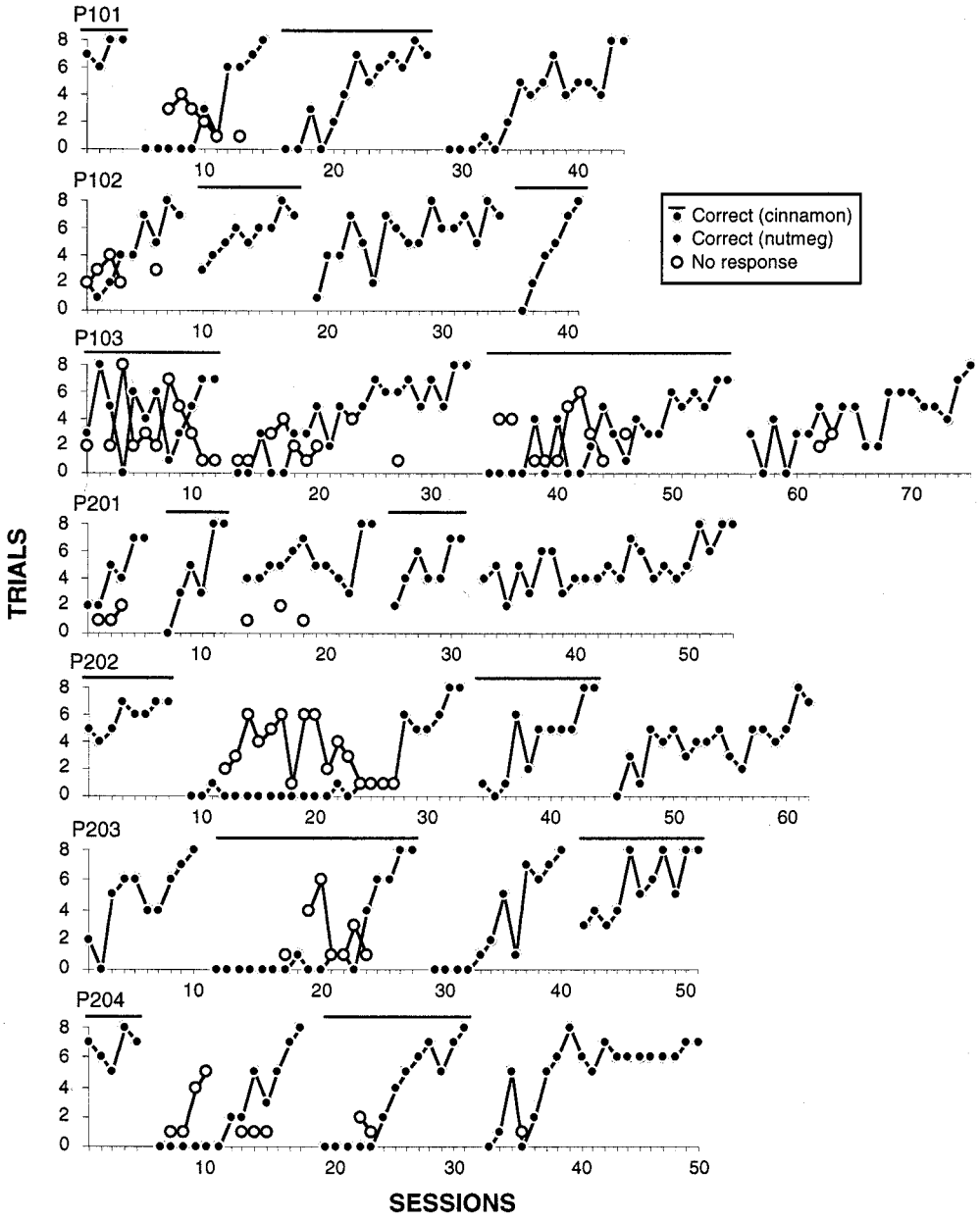

Fig. 5. Individual session scores for PKU mice in Experiment 2. See Figure 2 caption for details.

cluded totals ranging from 28 to 39 animals in different reversals, with independent one-way ANOVAs conducted on each reversal. The present study included a total of 13 animals in all conditions, and the data were analyzed by a 2 × 4 mixed-model ANOVA, with repeated measures across learning conditions (acquisition plus three reversals). Mean age of the heterozygous mice in Zagreda et al. was similar to that of the non-PKU mice in the present study (6.3 and 6.9 months, respectively). The PKU mice in the present study, however, were somewhat older (7.4 months) than those in Zagreda et al.'s study (4.1 months). Our study included male subjects only, and Zagreda et al. included both males and females but found no significant sex differences or interactions.

Blood phenylalanine levels for the PKU mice were also different, 9.23 mg/dl in the present study and 18.60 mg/dl in Zagreda et al. (1999). We obtained the blood samples after behavioral testing was completed, at approximately the same time of day as test sessions, and while the food restriction procedures were still in effect. Thus, the animals had no source of dietary phenylalanine for approximately 20 hr before the blood samples were taken (this was also the case for behavioral test sessions), and the lower phenylalanine levels may be related to the food restriction schedule. The levels reported by Zagreda et al. are consistent with those we and others have found in PKU mice with ad lib access to chow containing phenylalanine (e.g., Dyer et al., 1996; Shedlovsky et al., 1993).

The lower phenylalanine levels in the present study, however, were not accompanied by faster learning rates. A comparison of Table 2 in the present paper with Figure 2A in Zagreda et al. (1999) shows that the PKU mice in our study were slower in acquisition and the first reversal, and about the same in the second and third reversals. Learning rates for non-PKU mice in our study were slower than those in Zagreda et al., and this difference was greater over successive reversals.

Some of the difference in acquisition rates may be due to our use of a correction procedure during the first discrimination training session. Because that first session did not count toward the initial learning criterion, the minimum number of sessions for acquisition in our study was three. Zagreda et al. (1999) did not use a correction procedure. Some of the overall difference in learning rates may be related to different learning criteria, at least seven of eight correct for two consecutive sessions in the present study and 14 of 16 correct for two combined sessions in Zagreda et al. Thus, in Zagreda et al., successive scores of eight of eight and six of eight correct or vice versa could meet a learning criterion. Figures 4 and 5 show that application of the Zagreda et al. criterion to the present data would have resulted in fewer sessions to criterion in some instances (e.g., Figure 4, N105, third reversal). Our data set cannot be reanalyzed with their criterion, however, because the effects of changing the amount of prereversal training on learning rates for subsequent reversals cannot be estimated.

Finally, there were differences in experimental design and data characterization. We tested each mouse until it met the learning criterion for each reversal, and there was considerable variability in the number of sessions per reversal (see Table 2). Zagreda et al. (1999) imposed a 21-day testing limit for each reversal, and animals that did not meet the learning criterion within that limit were all assigned scores of 22 days for the days-to-criterion analysis (their Figure 2A). Although an animal that failed to meet a reversal criterion was not tested further, it was considered to have failed all subsequent reversals and was included in the numbers for those reversals when failure rates were calculated (their Figure 2B; A. Diamond, personal communication, February 23, 2000). In contrast, we tested all animals on all three reversals. Table 2 shows that 2 of 7 PKU mice (Subjects P103 and P202) required more than 21 sessions to meet a reversal learning criterion but nevertheless went on to complete the next reversal in fewer than 21 sessions. These data illustrate one way in which assumptions made for statistical control may fail to describe the behavior of individual subjects.

GENERAL DISCUSSION

The olfactory discrimination procedure produced rapid discrimination learning in mice. In Experiment 1, most Swiss-Webster mice met a fairly stringent criterion for acquisition of the initial discrimination within three eight-trial sessions, and there was evidence of stimulus control by odor in the initial session for many mice. Learning rates were comparable to those reported by Berger-Sweeney et al. (1998) with BALB/cByJ mice. Acquisition for the Pah^{enu2} mice in Experiment 2 was generally slower than for the Swiss-Websters, and was also somewhat slower than that reported by Zagreda et al. (1999). We have found no other reports of two-choice olfactory discrimination learning in mice.

In both experiments, reversal learning generally required about twice as many sessions as initial acquisition, although there were a few exceptions (e.g., SWM2, P102; see Tables 1 and 2). Responding in initial reversal sessions was virtually always at or below chance

levels, and accuracy scores were often zero correct for one or more sessions. These data indicate selective and reliable stimulus control by the olfactory stimuli during repeated discrimination reversals.

Because the procedure was not automated, the experimenter observed the subject and judged whether each approach to a container was merely sniffing (analogous to observing) or included the digging motions that defined a response. Such procedures require attention to the issues of experimenter bias and consistency in applying response criteria. When possible, an experimental blind would eliminate selective bias (a blind was not possible in Experiment 2 because the experimenter could easily identify each Pah^{enu2} genotype by coat color). Consistency can be evaluated by interobserver agreement. For a rough evaluation of interobserver agreement in Experiment 2, a second observer sat slightly behind and to one side of the experimenter and recorded responses for approximately 10% of the sessions for each subject. The two observers agreed on 537 of 544 trials. We present these data in the context of this discussion, rather than as part of the results, because the two observers were not independent. Although the second observer attempted to make judgments based solely on the subject's behavior, complete independence was impossible because the experimenter revealed his or her own judgment when providing the differential consequences for responses. Interobserver reliability evaluations could be improved in future studies if the second observer scored videotapes edited to show subjects' responses but not the consequences that followed.

Two other issues that may arise when considering the procedures reported here are those of time and sensitivity. The studies of serial reversal learning in Experiments 1 and 2 each required about 10 weeks to complete. Clearly, a protocol of this type and duration is not meant to function as a rapid behavioral screen, but rather as a longer term assessment of discrimination learning. The complex issue of sensitivity arises because we failed to find a statistically significant group difference in Experiment 2. Insensitivity can result if a procedure's behavioral requirements are not relevant to the behavioral repertoire one wishes to measure. We studied discrimination reversal because reversal learning deficits have been found in humans with mental retardation (e.g., Heal, Ross, & Sanders, 1966), and reversal learning has been used to index learning capability in nonhuman animals (e.g., Bitterman, 1965; Rumbaugh & Pate, 1984). Insensitivity could also result from inadequate preparation for testing or an inappropriate level of difficulty (e.g., floor or ceiling effects). For the procedures reported in the present paper, systematic replications with variations in parameters such as duration of pretraining, correction procedures, and so forth may clarify this issue.

We implemented a simple simultaneous discrimination procedure. Several potentially interesting variations may be practical, including many standard neuropsychological tests such as single alternation with either location or odor as relevant cues, delayed alternation, and delayed nonmatching to sample with trial-unique stimuli (Berger-Sweeney et al., 1998). A learning-set paradigm could be implemented by training a large number of discriminations with different stimuli. A left-right conditional discrimination could present the same olfactory cue in both locations (e.g., left side correct if both cinnamon, right side correct if both nutmeg).

This procedure also has the potential to be elaborated for studies of functional stimulus classification (Dube, McIlvane, Callahan, & Stoddard, 1993; Vaughan, 1988). One question for further research is whether rodents have capabilities for olfactory discriminations with large stimulus sets that parallel pigeons' capabilities for visual discriminations (Vaughan, 1988). Finally, more complex variations and elaborations of this procedure may be appropriate for any species in which olfactory cues are particularly salient. For example, Bunsey and Eichenbaum (1996) implemented a matching-to-sample version of the procedure with rats.

REFERENCES

Berger-Sweeney, J., Libbey, M., Arters, J., Junagadhwalla, M., & Hohmann, C. E. (1998). Neonatal monoaminergic depletion in mice (*Mus musculus*) improves performance of a novel odor discrimination task. *Behavioral Neuroscience, 112,* 1318–1326.

Bitterman, M. E. (1965). Phyletic differences in learning. *American Psychologist, 20,* 396–410.

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

Crabbe, J. C., Wahlsten, D., & Dudek, B. C. (1999). Genetics of mouse behavior: Interactions with laboratory environment. *Science, 284,* 1670–1672.

Crawley, J. N., & Paylor, R. (1997). A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Hormones and Behavior, 31,* 197–211.

Dube, W. V., McIlvane, W. J., Callahan, T. D., & Stoddard, L. T. (1993). The search for stimulus equivalence in nonverbal organisms. *The Psychological Record, 43,* 761–778.

Dunn, D. L. (1988). Antibody immunotherapy of gram-negative bacterial sepsis in an immunosuppressed animal model. *Transplantation, 45,* 424–429.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Science USA, 94,* 7109–7114.

Dyer, C. A., Kendler, A., Philibotte, T., Gardiner, P., Cruz, J., & Levy, H. L. (1996). Evidence for central nervous system glial cell plasticity in phenylketonuria. *Journal of Neuropathology and Experimental Neurology, 55,* 795–814.

Gallaher, E. J., Jacques, C. J., & Hollister, L. E. (1987). Alprazolam dependence in mice. *Alcohol and Drug Research, 7,* 503–510.

Griebel, G., Blanchard, D. C., & Blanchard, R. J. (1996). Predator-elicited flight responses in Swiss-Webster mice: An experimental model of panic attacks. *Progress in Neuropsychopharmacology and Biological Psychiatry, 20,* 185–205.

Heal, L. W., Ross, L. E., & Sanders, B. (1966). Reversal and partial reversal in mental defectives and normal children of a comparable mental age. *American Journal of Mental Deficiency, 71,* 411–416.

Lewis, R. (1999). The rodent wars: Is a rat just a big mouse? *The Scientist, 13*(14), 1–7.

McDonald, J. D., Bode, V. C., Dove, W. F., & Shedlovsky, A. (1990). $Pah^{hph\text{-}5}$: A mouse mutant deficient in phenylalanine hydroxylase. *Proceedings of the National Academy of Science, 87,* 1965–1967.

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.

Nelson, R. J. (1997). The use of genetic ''knockout'' mice in behavioral endocrinology research. *Hormones and Behavior, 31,* 188–196.

Pastino, G. M., Sultatos, L. G., & Flynn, E. J. (1996). Development and application of a physiologically based pharmacokinetic model for ethanol in the mouse. *Alcohol and Alcoholism, 31,* 365–374.

Powell, R. W., & Peck, S. (1969). Running-wheel activity and avoidance in the Mongolian gerbil. *Journal of the Experimental Analysis of Behavior, 12,* 779–787.

Rogers, D. C., Fisher, E. M. C., Brown, S. D. M., Peters, J., Hunter, A. J., & Martin, J. E. (1997). Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. *Mammalian Genome, 8,* 711–713.

Rumbaugh, D. M., & Pate, J. L. (1984). Primates' learning by levels. In G. Greenberg & E. Tobach (Eds.), *Behavioral evolution and integrative effects* (pp. 221–240). Hillsdale, NJ: Erlbaum.

Sakagami, T., Hursh, S. R., Christensen, J., & Silberberg, A. (1989). Income maximizing in concurrent interval-ratio schedules. *Journal of the Experimental Analysis of Behavior, 52,* 41–46.

Shedlovsky, A., McDonald, J. D., Symula, D., & Dove, W. F. (1993). Mouse models of human phenylketonuria. *Genetics, 143,* 1205–1210.

Van Hemel, S. B. (1973). Pup retrieving as a reinforcer in nulliparous mice. *Journal of the Experimental Analysis of Behavior, 19,* 233–238.

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

Wehner, J. M., & Silva, A. (1996). Importance of strain differences in evaluations of learning and memory processes in null mutants. *Mental Retardation and Developmental Disabilities Research Reviews, 2,* 243–248.

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. *Journal of Neuroscience, 19,* 6175–6182.

Received March 29, 1999
Final acceptance February 4, 2000