



HHS Public Access

Author manuscript

Behav Brain Res. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

Behav Brain Res. 2014 October 1; 272: 46–54. doi:10.1016/j.bbr.2014.06.017.

Experimenter effects on behavioral test scores of eight inbred mouse strains under the influence of ethanol

Martin Bohlen^{a,b}, Erika R. Hayes^a, Benjamin Bohlen^a, Jeremy Bailoo^a, John C. Crabbe^c, and Douglas Wahlsten^{a,d,*}

^a Department of Psychology, University of North Carolina Greensboro, 275 Eberhart Bldg, Greensboro, NC 27412, USA

^b Program in Neuroscience, The University of Mississippi Medical Center 2500 North State St. Jackson, MS 39216, USA, MBohlen@umc.edu

^c Portland Alcohol Research Center, VA Medical Center and Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, OR 97239, USA, crabbe@ohsu.edu

^d Department of Psychology, University of Alberta, Edmonton, AB T6G 2E9

Abstract

Eight standard inbred mouse strains were evaluated for ethanol effects on a refined battery of behavioral tests in a study that was originally designed to assess the influence of rat odors in the colony on mouse behaviors. As part of the design of the study, two experimenters conducted the tests, and the study was carefully balanced so that equal numbers of mice in all groups and times of day were tested by each experimenter. A defect in airflow in the facility compromised the odor manipulation, and in fact the different odor exposure groups did not differ in their behaviors. The two experimenters, however, obtained markedly different results for three of the tests. Certain of the experimenter effects arose from the way they judged behaviors that were not automated and had to be rated by the experimenter, such as slips on the balance beam. Others were not evident prior to ethanol injection but had a major influence after the injection. For several measures, the experimenter effects were notably different for different inbred strains. Methods to evaluate and reduce the impact of experimenter effects in future research are discussed.

Keywords

Alcohol intoxication; anxiety-like behavior; motor coordination; elevated plus maze; balance beam; open field activity; accelerating rotarod; grip strength; laboratory environment; heredity-environment interaction

* Corresponding author. Phone: 1-780-977-9086 Fax: 1-780-492-1768 wahlsten@ualberta.ca. Mailing address: Department of Psychology, P217 Bio Sci Building, University of Alberta, Edmonton, AB T6G 2E9, Canada. erika.r.hayes@gmail.com

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The laboratory mouse now plays a central role in research on animal models of human behavioral disorders [1], and numerous laboratories worldwide work with the same genetically defined mouse strains and mutations to answer complex questions about behavior. Within a laboratory, multiple experimenters often work together in order to increase the amount and rate of data collection, while different labs almost always utilize the services of different experimenters. In terms of research design for genetic studies, the experimenter is part of the laboratory environment and constitutes a control variable rather than a systematically manipulated independent variable in many studies. It is recognized that the laboratory environment can have a noteworthy impact on the results of mouse behavioral tests and can interact with genotype of the research animals [2,3]. Within a lab, the experimenter who administers a test can also be an important influence [4-7]. The specific experimenter who conducts a test is a difference between labs that cannot be eliminated. Within a lab, however, the study design can be carefully balanced and randomized so that the experimenter does not bias treatment effects. The principles behind this kind of balancing are well understood [8], but the interpretation of even a perfectly balanced study can be difficult if any experimenter effects interact with the treatment effects of principal interest. It is therefore important for behavioral neuroscience research that we gain a better appreciation of the prevalence and magnitude of experimenter effects.

The problem of experimenter effects [9] and experimenter bias [10] in studies of rodent behavior has been acknowledged, and a few reports of experimenter effects on mouse and rat behavioral tests have appeared. Results for elevated plus maze behaviors of rats differed markedly between two experimenters [11], while rats in an elevated plus maze showed greater variation between six experimenters in anxiety scores when the experimenters were unfamiliar to them [12]. Although many studies of genetic variation in mouse behavior employ more than one experimenter, it is rare to see this factor included in the report and data analysis.

In the present study where the treatment effect of central interest turned out to be very small, the experimenter effect was the largest effect in the entire study, even though experimenter was included in the design as a control variable. The study was originally conceived after a surprising result obtained at the University of Alberta [13]. In 1998, eight genotypes were tested for several behaviors in three labs[2]. Then in 2002, 20 inbred strains of mice were tested on several behaviors following ethanol or cocaine injection as part of the Mouse Phenome Project, and the experiment was replicated with identical apparatus and protocols at the same time at Oregon Health & Science University [3,14-16]. Between 1998 and 2002, the Alberta lab had to be moved out of the Department of Psychology space into the central animal quarters in a different wing of the Biological Sciences building. The same test apparatus was used in the new quarters, and results were quite different for certain of the tests, especially the elevated plus maze. It was noticed at the time that odors of many other species of rodents were present in the central facility. Mice were exposed to those odors when being transported down a hallway to the test room, and some of the odors were circulated through the test room as well. The experimenters, however, also differed between

the 1998 and 2002 studies done in Alberta. Either the odors, the experimenters, or other unknown factors could have altered results.

Mice are highly sensitive to different kinds of odors and engage in scent marking for social communication [17]. There is clear evidence that rodents exposed to predator odor (fox and cat odor) show anxiety-like behaviors to the potential threat [18-20]. Additionally, it is clear from nearly 65 years of research (see O'Boyle, 1975 for a historical discussion) that rats are muricidal, a stereotypic behavior defined by the tendency for rats to express predatory behaviors when a mouse is present and accessible [21]. These behaviors include hunting, killing and consuming the mouse [22]. The predatory behaviors and their influence on mouse behavior have been further characterized by the Blanchards and coworkers at the University of Hawaii, who have developed a mouse defense battery to characterize responses of mice confronted with a rat [23,24]. Mice presented with a recently euthanized or anesthetized rat tend to keep a large distance from the rat and will flee if an awake, restrained rat can follow. If escape is not available, the mouse will perform defensive (defensive upright posture, vocalizations) and attack (biting, jump attack) behaviors [25,26]. More recently, mice presented with a restrained rat were shown to have altered facial expressive patterns with increased nose and cheek swells, and the behaviors were very similar to those manifested to cat odor presentation [27]. For mice exposed to rat odor, stress-related hormone levels were altered [28-30]. Rat odor also suppressed appetite and markedly increased latency to approach and consume food rewards [31], decreased sucrose intake and time spent in the open arms of an elevated plus maze [32], increased time spent freezing [33], and amplified startle response and time spent in the dark of a light-dark test [34,35]. Some effects were so robust that Calvo-Torrent et al., 1999 suggested rats and mice should not be housed near one-another.

When D.W. moved his mouse lab to UNCG in 2008, the animal research facility was empty and there were many unused testing and colony rooms. This provided an ideal situation to test the influence of rat odors on mouse behavior. The facility manager stated that all air in the facility was fresh to each room and was not recirculated. During preparation for the study, the smell of rats was never detected by the researchers in any of the testing rooms. A study was then conducted using three groups: (a) mice housed and tested in rooms that only contained mice; (b) mice housed and tested in rooms that contained both mice and rats; (c) mice housed only with mice but tested in a room containing rats. It was expected that mice exposed to rat odors for the first time would express greater anxiety-like behaviors and show greater impairment following an ethanol injections. The study used two experimenters to test the animals during each day. The study was carefully randomized and balanced for experimenter and treatment effects over strain, sex, time (morning versus afternoon), and housing room.

While the study was in progress, it was noticed on several occasions that a distinctive odor of coffee brewing was coming into the mouse testing rooms. Neither mice nor rats were ever fed coffee in this study, and our experimenters never brewed coffee anywhere in the animal facility. It was then determined that the animal care personnel employed by the university were making coffee in their office that was inside the controlled access animal facility. Evidently there was recirculation of air among the various rooms, especially during hot

weather when air conditioning was used. This negated the design of our experiment. We decided to complete the study and look at the data. No or very small rat housing effects were found, but there were several substantial experimenter effects.

2. Methods

2.1. Mice

Equal numbers of males and females of eight strains from the Mouse Phenome Database (MPD) Priority list 1 were studied (129S1/SvImJ, A/J, BALB/cByJ, C3H/HeJ, C57BL/6J, DBA/2J, FVB/NJ, SJL/J). All animals were obtained at 6 weeks of age from the Jackson Laboratory, Bar Harbor, Maine, USA. The rats were six week old Harlan Sprague-Dawley imported directly into the lab and never used in any previous study. At the start of the study they were about 10 weeks old.

2.2. Husbandry

On arrival, mice were randomly assigned, two male and two female mice per strain, to one of three rooms: (a) no rats in colony room or behavioral testing room, (b) no rats in colony room but rats in test room, and (c) rats in both colony room and test room. Mice were habituated in their assigned housing condition for two weeks before behavioral experiments commenced. Animals were housed two of the same sex per cage in standard shoebox cages with open wire frame tops and had free access to Purina 5001 mouse chow and Greensboro tap water.

2.3. General procedures

Animals in all treatment conditions except housing room were balanced within a room for shelving position, sex and strain. There were 6 replicates with three shipments of 96 mice each, for a total of 288 mice. Each replicate had each sex-strain-housing condition represented. One of the two animals in each cage was randomly tail marked the day before a replication started. Whichever animal was selected for the first replicate, the other animal was then used for the subsequent replicate. A solution of 20% v/v 200 proof ethanol from Commercial Alcohols (Chatham, ON) in buffered physiological saline was injected intraperitoneally using sterile 27 gauge hypodermic needles. Different dosages were used on different days, as discussed below. The detailed protocol for injections was provided by the lab of J.C.C. at Oregon Health & Science University.

2.4. Procedures for rat odor

Before the study began, all colony and testing rooms and laboratory equipment were washed down first with bleach then with the disinfectant solution Sporidicin® (Sporidicin International, Rockville, MD). Animals were housed in identical colony rooms. In the one colony room where rats were also present, mice were housed on an adjacent, separate rack in cages that were occluded so that the mice could not see any rats. The same was true for the behavioral test rooms; rat cages were occluded so that the mice could not see any rats. Airflow was measured in each room and resident animals (rats or mice) were always placed up-wind of test mice. Rather than have the resident animal cages changed on the same schedule as the experimental animals, we changed the cages so that half the resident cages

were changed at a time to control for changes in odorant throughout the week. Identical test equipment was used in both testing rooms. Experimenters were randomized between morning and afternoon and rat or non-rat testing room, and they were also balanced across sex and strain. To prevent odor contamination, facility staff always changed non-rat colony rooms first. On experiment days the non-rat experimenter transported mice to the rat condition behavioral test room. After morning experiments were completed, both experimenters showered and changed clothes to ensure no cross odorant contamination. Movement of experimenters between the housing and test rooms was carefully controlled and scheduled, as summarized in Figure 1, so that the odor of rats would never be brought into a room with mice by an experimenter. A strict boundary between zones with rats and mice was established, and cages with animals were passed across the boundary but the experimenters never went from one zone to the other during a day unless they had showered and changed clothing.

2.5 Experimenter characteristics

One experimenter, a 30-year old woman, was employed full time as a technician in the lab and had about five years of experience working with rodents at other institutions in North Carolina. She had a B.S degree in Laboratory Animal Science and an M.S. degree in Animal Health Science. She had taken courses involving mouse handling and husbandry, but she did not have prior training in behavioral testing per se. The other experimenter was a 23-year old male graduate student doing his thesis research in the lab. He had a B.Sc. degree in psychology and was skilled at data analysis and video tracking of mice. He had not given intraperitoneal injections until beginning work in the present lab, and he was trained to do this by the lab director (D.W.). The two experimenters reviewed all protocols together and worked closely during the planning phase of the study when pilot experiments were conducted, and they coordinated their activities each day during the study. Both had completed training modules on Laboratory Animals and the Laboratory Mouse as required by the UNCG Institutional Animal Care and Use Committee.

2.5. Test battery

General methods for motor tasks and the battery have been published in detail [36] and are briefly described here. The timing of tests and injections are shown schematically in Figure 2.

Day 1—Animals were first tested on the elevated plus maze for 300 s using ANY-maze video tracking software (Version 4.3; Stoelting Co., Wood Dale, IL) [37], then were pre-trained on both the balance beam (two traverses from both directions) and accelerating rotarod (10 consecutive trials).

Day 2—All animals were weighed first in the morning and syringes for each animal were prepared. Thirty minutes before they were to be run, animals were removed from their home cage and placed into a clean shoebox cage with fresh bedding and then taken to the test room for 30 min of habituation. Animals were timed for latency to cross from one end to the other on the balance beam, and number of foot slips was counted by the experimenter. Following the balance beam, animals received three trials on the grip strength test. Animals

then received a 1.25 g/kg ethanol IP injection and were returned to their habituation cages for 30 mins. Using the same methods, the animals were tested again post-injection on the balance beam (one trial) and grip strength (three trials).

Day 5—Animals were weighed and syringes were prepared according to the animal's weight. Animals were placed into a new cage with clean bedding, taken to the test room, and habituated for 30 min before any behavioral experiment was run. Animals were first run in the open field activity chamber for 300 sec with pink butcher paper on the floor and infrared backlighting (see Bailoo et al. 2010), and numbers of rears and leans were recorded by watching the video in real-time. After the open-field test, animals were given three trials on the accelerating rotarod. Each animal then received a 1.75 g/kg ethanol IP injection and was returned to its habituation cage for 30 min. Then, using the same methods, the animals were run post-injection in the open field (300s) and accelerating rotarod (three trials).

3. Results

3.1. Preliminary analyses

There were four between-subject factors in this study (strain, sex, housing, experimenter) and one within-subject factor (ethanol injection). The first step in the analysis was to examine factors that might have little or no effect and could be pooled for further analyses. It was immediately evident that housing or testing with rats present had no noteworthy effect on any behavioral test. In an analysis of variance with housing included, 36 measures were examined for the five behavioral tests. Ethanol effects were assessed with separate pre-post difference measures. Many effects of strain and ethanol were clearly significant ($P < 0.0001$). Because so many significance tests were evaluated in this and other analyses, $P < 0.001$ was regarded as a reasonable criterion for statistical significance in the preliminary analysis of all measures in all tests, while the criterion was set at $P < .005$ for effects of principal interest, especially experimenter effects, in the more refined analysis. Housing had a significant effect only on time spent near the wall in the open field, sometimes viewed as an indicator of anxiety-like behavior, and this effect was not evident for the change attributable to ethanol injection. No other measure in the open field showed a housing effect. Neither did time in the open arms of the elevated plus maze, another indicator of anxiety-like behavior. It was concluded that housing and testing with rats present or absent had no perceptible influence on behavior in this study. Accordingly, data were pooled across housing/testing condition for further analyses.

A similar analysis was done with sex of the mouse in the analysis. The only main effect showing a sex difference was for grip strength, where the generally larger males had stronger grips ($P < 0.0001$). There were no significant interactions of sex with experimenter, and only one interaction of sex with strain (ethanol effect on vertical movements in the open field; $P < 0.0001$). The data were therefore pooled over sex for further analyses.

3.2. Pattern of significant effects for all measures

All measures of the five behavioral tests were subjected to analysis of variance, the results of which are summarized in Table 1 as significance (P) values for the most important, non-

redundant measures. Except for elevated plus maze where no ethanol was administered, the ethanol effect was analyzed as a repeated measure for pre- versus post-injection. The table shows that strain differences on all measures were clearly significant, a finding that was expected because the strains were chosen on the basis of large differences on the same tests in previous studies [13,36]. Most measures also showed a large ethanol effect, as expected from previous studies. On the accelerating rotarod and open field tests, there were remarkable experimenter effects. For the elevated plus maze, there was a clearly significant interaction between strain and experimenter on percent time in open arms ($P < 0.0001$) but only a modest experimenter main effect, which indicated that some strains were affected in opposite ways by the experimenter who conducted the test. Thus, there were noteworthy effects of the experimenter on three of five behavioral tests in a study that was carefully balanced for experimenter.

3.3. Elevated plus maze

Most mice made more than 10 arm entries and experienced both open and closed arms, while six of 288 mice entered only one arm and froze there the entire 300 sec. Strains differed greatly in all measures of primary interest. The level of exploration of open arms was relatively high in comparison with some studies, and several strains spent more time in the open than the closed arms (Fig. 3a). The preference for the open arm versus the enclosed arm was particularly striking for the two arms at the front of the maze that faced the center of the room and were farthest away from a wall (Fig. 3b). For strain A/J, open arm exploration was strongly influenced by the experimenter. Close inspection of the data revealed that time in arms was highly variable for that strain because of freezing in one arm by several hypoactive A/J animals. Our protocol required that the mouse be placed at the center of the maze and its tail released when it was facing the open arm on the front side of the maze. That had the greatest influence on scores of the A/J strain because of its tendency to freeze not long after being released. Very small differences between experimenters in how this release was done could have resulted in a difference in where an A/J mouse froze. One other measure showed a clear experimenter effect that evidently arose from different criteria used to identify a head dip (Fig. 3c); Experimenter 2 was much more likely to record a head dip across all strains, while the two people agreed closely on the number of rearing behaviors. Stretch-attend behaviors were infrequent for both experimenters.

3.4. Balance beam

Ethanol reduced time required to traverse the beam by a small amount (Fig. 4a) but greatly increased the numbers of slips of a foot or the body off the beam (Fig. 4b) for all strains. Hypoactive strains 129S1/SvImJ and A/J required the most time to traverse the beam, but numbers of slips were not related to the general level of activity on the beam. Only eight of 271 mice fell of the beam before injection and just one fell after the injection. There were no noteworthy effects of experimenter.

3.5 Accelerating rotarod

Ethanol effects on fall latency were remarkably small in this study (Fig. 4c). A reduction in fall latency after ethanol injection was indeed seen for Experimenter 1 but not 2 (Fig. 5). The experimenter effect was significant at $P = .005$ and the interaction between the ethanol and

experimenter effects was of borderline significance ($P = .007$). The experimenter interaction effect was not sufficiently large to obscure the robust strain difference in which A/J was among the first to fall and C57BL/6J remained longest on the rod.

3.6. Open field

Strain differences were very large (Fig. 4d) and in accord with previous observation of hypoactivity in strains 129S1 and A/J in contrast to very high activity in C57BL/6 mice. Pronounced activation by ethanol was seen in strains A/J and DBA, whereas ethanol markedly reduced motor activity in C57BL/6. Ethanol greatly reduced rearing and leaning behaviors in all strains that showed appreciable amounts of these behaviors before ethanol (Fig. 4e), and the reduction was proportional to the baseline level of rearing and leaning. The highly significant strain by ethanol interaction arose primarily from the lack of any perceptible ethanol effect on the 129S1 strain that showed very little rearing or leaning before ethanol. Percentage of time near a wall was altered by ethanol in a strain-dependent manner, such that it increased substantially for BALB and FVB, declined appreciably for C57BL/6 and changed little for the other strains. Nevertheless, wall time showed a fairly narrow range from 70% to 90% across all strains and conditions (data not shown).

A large experimenter effect was apparent for open field activity, and the magnitude of the ethanol effect depended strongly on the specific experimenter. As shown in Fig. 6a, the pattern of activity across all eight strains was remarkably similar for the two experimenters prior to the ethanol injection, which is not at all surprising because the open field test is done with computer-based video tracking involving minimal interaction with an experimenter. After the injection, however, the difference between experimenters was very large (Fig. 6b) except for strain 129S1. Furthermore, the magnitude of the injection/ethanol effect for specific strains depended on the experimenter giving the injection (Fig. 6c, d). With experimenter 2 there was a pronounced activation effect from ethanol for all but two strains (Fig. 6d), whereas for experimenter 1 there was little change after the injection for five of the eight strains and a marked lowering of activity for the other three. The interaction effect was so large that rank orders of strains changed substantially before and after injection for the two experimenters.

3.7. Grip strength

Strain differences were highly significant and the ethanol effect was large and obvious for every strain. Nevertheless, certain strains (129S1, DBA) showed a substantially smaller degree of impairment, whereas others (BALB, C57BL/6) showed a larger impairment (Fig. 4f). There were no noteworthy experimenter effects on this test, despite the extensive handling of mice required during the test.

4. Discussion

4.1 Size and importance of experimenter effects

In a situation where there are two experimenters, the size of the experimenter effect can be expressed as the coefficient d , the number of standard deviations by which group means differ. Using a convenient utility *Effect size from article P.xls* for Excel provided by [8], the

value of d can be found from values of degrees of freedom and the F or t ratio for the significance test. In the present data, the experimenter effect was about 0.3 for percent time in open arms of the elevated plus maze and rotarod fall latency. These amount to small effects that are statistically significant in a situation where degrees of freedom within groups were relatively large, more than 260. For distance in the open field, $d = 0.8$, a large effect, and for head dips in the elevated plus maze $d = 1.22$, also a large effect. Thus, for two of the five tests, the experimenter effect was quite large. When an effect can be this large, it needs to be taken into account in the design of any study that involves more than one experimenter. The method for achieving proper balance of a study with two experimenters is described in detail in Chapter 7 in Wahlsten (2011). If this balancing is not done correctly, there is a serious risk that results for the treatment effects of principal interest in a study might be confounded with or biased by an experimenter effect. If there is an experimenter effect in the data, the power to detect treatment effects should not be greatly reduced, provided the experimenter is included in the data analysis as a factor. If there is an experimenter effect and data are pooled over experimenter, this will inflate error variance and reduce power to detect treatment effects. The pattern of results may be more complicated if experimenter interacts with treatments, but at least proper balancing will make the test of interactions meaningful.

4.2 Sources of experimenter effects

When a study involves human judgment about whether a behavior occurs or not, as in the case of head dips on the elevated plus maze, different criteria applied by the experimenters can have a major impact on results. This kind of effect might be eliminated or greatly reduced by more extensive training of the observers prior to the start of a study. When the difference of opinion is large, this can be detected in preliminary testing. Automated response detection with photocells or video tracking might avoid biased judgments, but some configurations of apparatus are just too complex to permit this. Head dips, for example, can occur on either side of each of the two open arms of the elevated plus maze, such that video tracking from above will not help. A system of four photocell beams might be effective, but it will be challenged by occasions where the foot slips off the maze and the mouse struggles to get it back onto the top of the arm, a situation where the foot might pass in front of a photocell several times very quickly.

Behavioral tests that involve extensive handling may be more prone to experimenter effects. This was not the case in the present study for balance beam or grip strength. In the open field test, experimenter had no influence prior to injection, but things changed in complex ways after the injection. Experimenter 1 had experience working with rodents and had given many injections, while DW trained Experimenter 2 in our lab. There must have been something different about the ways in which the injections were given. To our eyes, the injections looked quite similar, but the behavior of the mice informed us otherwise. This effect might also be reduced by more extensive training prior to the experiment and closer scrutiny of the fine details of an injection. The balance beam and grip strength tests also involved injections, and there were large ethanol effects but no experimenter effects. Thus, it might be helpful to extend the evaluation of experimenter effects to a wider variety of behavioral tests.

A recent report [7] provided extensive evidence that the sex of the human experimenter can substantially alter ratings of pain exhibited by mice. The effect arose from male-related odors and could be produced simply by a T shirt worn by a man the previous day or specific chemicals known to be excreted by men (3-methyl-2-hexenoic acid, androstenone, androstadienone). In our study, Experimenter 1 was male and 2 was female, and any sex-related effects were completely confounded with other kinds of differences between them. The magnitude of experimenter sex effects in that report [7] were so large that it would be wise to employ experimenters of both sexes routinely in future work in order to reduce confounding of experimenter sex with treatment effects of study factors. It also would be advisable to provide more information about the experimenters, including their sexes, in the methods section of a report, a practice that is not common at the present time in behavioral neuroscience.

4.3 Experimenter effects and internal validity of treatment effects

It is possible to observe a large experimenter effect that has no influence at all on the conclusions about treatments of principal interest in the study, provided experimenter does not interact with treatments. We were primarily interested in ethanol effects on behavior and possible differences between strains in the size of ethanol effects. For the accelerating rotarod where the experimenter main effect was small, the interaction of experimenter and ethanol posed a serious challenge to interpretation of results because only one of the experimenters obtained a clear pattern of reduced fall latency after ethanol. For open field distance after ethanol injection, the strains changed in a complex way, depending on experimenter. Open field activity increased substantially following ethanol for several strains for Experimenter 2 but decreased for several strains for Experimenter 1. There was a major difference between experimenters in which strains changed the most or least. Thus, using two experimenters provided a much more complex pattern of results than if there had been just one.

4.4 Experimenter effects versus experimenter bias

Experimenter bias can obscure or augment apparent treatment effects, especially when an experimenter has ideas about what kind of outcome is desirable [10]. This can be important in situations where the experimenter rates behaviors from visual impressions or handles the animals extensively. Having a study carefully balanced for multiple experimenters will not necessarily eliminate such effects. Instead, it may be necessary to use blind coding of treatments such as ethanol injection, so that the experimenter does not know whether a specific animal receives an ethanol or saline injection. This becomes more difficult in a situation such as ours where behaviors are rated before and after injection. An experimenter may expect a mouse to have poorer motor coordination after ethanol. This could be addressed by having different people give the injection and do the rating of behavior, at the cost of a large increase in the complexity of logistics during the test day.

4.5. Automated testing

The less contact a mouse has with an experimenter, the smaller should be experimenter effects. Some laboratories employ sophisticated electronic apparatus to collect data throughout the day and night when humans are far from the test area. Added cost of the

equipment may be justified by its capacity to minimize experimenter effects. Tracking mice with radio frequency chips or infrared video [37] can detect patterns that would likely elude a human observer.

4.6. Experimenter effects and differences between laboratories

Several previous studies have found that results of mouse behavioral tests can differ substantially between laboratories [2,3,8,13-16]. In every case, the experimenters differed between labs, so it is possible that lab differences arose from experimenter effects. This source of lab differences cannot be eliminated through any practical measure. It will pose the greatest threat to generality of results if a lab uses only one experimenter to conduct the behavioral tests. If two or more experimenters are employed and the study is properly balanced for experimenter, generality will be enhanced to some extent. The principal benefit from using two or more experimenters is that this can help to assess the robustness of results within a single lab in the presence of factors that often differ between labs. It is feasible that tests and measures that show the greatest experimenter effects within a lab may also be most likely to indicate differences between labs. If the experimenter effect in a specific study is not significant, this suggests that similar results will be likely in other labs, but it cannot prove this. There could be no experimenter effect while other lab environment differences influence results substantially.

4.7. Animal and human odors

Our experience in this study shows the limitations of many conventional animal facilities for controlling odors. Whenever there is any kind of central ventilation system, there is a possibility of exposure of mice to alien odors, especially if this will reduce heating or cooling costs. Human noses are not dependable judges of the presence of rat odors. Further work on this issue warrants housing and test rooms specially designed to control the flow of airborne odors. A closed vent system for an entire rack of cages can effectively prevent the ingress of animal odors into the general lab environment (e.g. Animal Care Systems M.I.C.E. caging) but does not isolate the mice from odors in the lab. Reversing the flow of air might achieve this. At present, there are reasons to be concerned about effects on mice of rats housed in the same animal quarters, but it is not clear if these effects will be large enough to justify keeping rats out of the facility altogether.

4.8. Conclusions

At the present time, no generalizations can be drawn about the prevalence of experimenter effects in different labs because there have been so few formal reports where they were rigorously assessed. It is possible that many labs have used more than one experimenter but did not report the effects because they were negligible. The present study does not show that experimenter effects are to be expected, but it does show that their influences should be taken seriously. In many labs there will be difficulty replicating specific experimenter effects because of a frequent turnover in personnel. In our study, for example, the two experimenters have already moved to new labs.

We conclude with total confidence that, whenever two or more experimenters are employed to collect data in the same experiment, the design should be carefully balanced for

experimenter. It is common practice and an entirely acceptable practice to carefully balance a study for the possible effects of many control factors such as cage of mice, position on shelves of the rack in the housing room and time of day of testing, but not conduct a thorough analysis of the effects of several control factors. Such a practice pools any effects of control factors, including their interactions with treatment effects, into the variance within a treatment group. Proper balancing insures treatment effects will not be biased by control factors, but it does not guarantee that control factors will have no influence.

Acknowledgments

Research reported in this paper was supported by grant AA12714 from the National Institute of Alcoholism and Alcohol Abuse. JC also received support from grants AA13519, AA10760, and a grant from the US Department of Veterans Affairs. The sponsoring agencies were not involved in the planning, execution or analysis of the experiments and did not contribute to writing the report.

Reference List

1. Crawley, J. What's Wrong With My Mouse: Behavioral Phenotyping of Transgenic and Knockout Mice. 2nd edn.. Wiley-Liss; Hoboken, NJ: 2007.
2. Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. *Science*. 1999; 284:1670–1672. [PubMed: 10356397]
3. Wahlsten D, Metten P, Phillips TJ, Boehm SL, Burkhart-Kasch S, Dorow J, Doerksen S, Downing C, Fogarty J, Rodd-Henricks K, et al. Different data from different labs: lessons from studies of gene-environment interaction. *J Neurobiol*. 2003; 54:283–311. [PubMed: 12486710]
4. Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS. Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci. Biobehav. Rev*. 2002; 26:907–923. [PubMed: 12667496]
5. Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS. Influences of laboratory environment on behavior. *Nat. Neurosci*. 2002; 5:1101–1102. [PubMed: 12403996]
6. Mogil JS. Animal models of pain: progress and challenges. *Nat. Rev. Neurosci*. 2009; 10:283–294. [PubMed: 19259101]
7. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat.Methods*. Apr 28.2014 doi: 10.1038/nmeth.2935 [Epub ahead of print].
8. Wahlsten, D. *Mouse Behavioral Testing: How to Use Mice in Behavioral Neuroscience*. 1st edn.. Academic Press; San Diego, CA: 2011.
9. Sousa N, Almeida OF, Wotjak CT. A hitchhiker's guide to behavioral analysis in laboratory rodents. *Genes Brain Behav*. 5 Suppl. 2006; 2:5–24.
10. Eisenach JC, Lindner MD. Did experimenter bias conceal the efficacy of spinal opioids in previous studies with the spinal nerve ligation model of neuropathic pain? *Anesthesiology*. 2004; 100:765–767. [PubMed: 15087608]
11. Lopez-Aumatell R, Martinez-Membrives E, Vicens-Costa E, Canete T, Blazquez G, Mont-Cardona C, Johannesson M, Flint J, Tobena A, Fernandez-Teruel A. Effects of environmental and physiological covariates on sex differences in unconditioned and conditioned anxiety and fear in a large sample of genetically heterogeneous (N/Nih-HS) rats. *Behav. Brain Funct*. 2011; 7:48. [PubMed: 22118015]
12. van Driel KS, Talling JC. Familiarity increases consistency in animal tests. *Behav. Brain Res*. 2005; 159:243–245. [PubMed: 15817187]
13. Wahlsten D, Bachmanov A, Finn DA, Crabbe JC. Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc. Natl. Acad. Sci. U. S. A*. 2006; 103:16364–16369. [PubMed: 17053075]

14. Crabbe JC, Wahlsten D. Of mice and their environments. *Science*. 2003; 299:1313–1314. [PubMed: 12610278]
15. Wahlsten D, Metten P, Phillips TJ, Boehm SL, Burkhart-Kasch S, Dorow J, Doerksen S, Downing C, Fogarty J, Rodd-Henricks K, et al. Different data from different labs: lessons from studies of gene-environment interaction. *J. Neurobiol.* 2003; 54:283–311. [PubMed: 12486710]
16. Wahlsten D, Metten P, Crabbe JC. A rating scale for wildness and ease of handling laboratory mice: results for 21 inbred strains tested in two laboratories. *Genes Brain Behav.* 2003; 2:71–79. [PubMed: 12884964]
17. Hurst JL, Thom MD, Nevison CM, Humphries RE, Beynon RJ. MHC odours are not required or sufficient for recognition of individual scent owners. *Proc. Biol. Sci.* 2005; 272:715–724. [PubMed: 15906464]
18. Adamec R, Walling S, Burton P. Long-lasting, selective, anxiogenic effects of feline predator stress in mice. *Physiol Behav.* 2004; 83:401–410. [PubMed: 15581662]
19. Berton F, Vogel E, Belzung C. Modulation of mice anxiety in response to cat odor as a consequence of predators diet. *Physiol Behav.* 1998; 65:247–254. [PubMed: 9855473]
20. Hacquemand R, Choffat N, Jacquot L, Brand G. Comparison between low doses of TMT and cat odor exposure in anxiety- and fear-related behaviors in mice. *Behav. Brain Res.* 2013; 238:227–231. [PubMed: 23089645]
21. O'Boyle M. The rat as a predator. *Psychological Bulletin.* 1975; 82:460–462.
22. Albert DJ, Walsh ML, Ryan J, Siemens Y. Mouse killing in rats: a comparison of spontaneous killers and rats with lesions of the medial hypothalamus or the medial accumbens nucleus. *Physiol Behav.* 1982; 29:989–994. [PubMed: 6891798]
23. Blanchard DC, Griebel G, Blanchard RJ. The Mouse Defense Test Battery: pharmacological and behavioral assays for anxiety and panic. *Eur. J Pharmacol.* 2003; 463:97–116. [PubMed: 12600704]
24. Blanchard DC, Griebel G, Blanchard RJ. Mouse defensive behaviors: pharmacological and behavioral assays for anxiety and panic. *Neurosci. Biobehav. Rev.* 2001; 25:205–218. [PubMed: 11378177]
25. Blanchard RJ, Hebert MA, Ferrari PF, Palanza P, Figueira R, Blanchard DC, Parmigiani S. Defensive behaviors in wild and laboratory (Swiss) mice: the mouse defense test battery. *Physiol Behav.* 1998; 65:201–209. [PubMed: 9855467]
26. Griebel G, Blanchard DC, Jung A, Blanchard RJ. A model of 'antipredator' defense in Swiss-Webster mice: effects of benzodiazepine receptor ligands with different intrinsic activities. *Behav. Pharmacol.* 1995; 6:732–745. [PubMed: 11224376]
27. Defensor EB, Corley MJ, Blanchard RJ, Blanchard DC. Facial expressions of mice in aggressive and fearful contexts. *Physiol Behav.* 2012; 107:680–685. [PubMed: 22484562]
28. Morrow BA, Elsworth JD, Roth RH. Fear-like biochemical and behavioral responses in rats to the predator odor, TMT, are dependent on the exposure environment. *Synapse.* 2002; 46:11–18. [PubMed: 12211094]
29. Munoz-Abellan C, Andero R, Nadal R, Armario A. Marked dissociation between hypothalamic-pituitary-adrenal activation and long-term behavioral effects in rats exposed to immobilization or cat odor. *Psychoneuroendocrinology.* 2008; 33:1139–1150. [PubMed: 18644680]
30. Vendruscolo LF, Vendruscolo JC, Terenina-Rigaldie E, Raba F, Ramos A, Takahashi RN, Mormede P. Genetic influences on behavioral and neuroendocrine responses to predator-odor stress in rats. *Neurosci. Lett.* 2006; 409:89–94. [PubMed: 17052845]
31. Merali Z, Levac C, Anisman H. Validation of a simple, ethologically relevant paradigm for assessing anxiety in mice. *Biol. Psychiatry.* 2003; 54:552–565. [PubMed: 12946884]
32. Calvo-Torrent A, Brain PF, Martinez M. Effect of predatory stress on sucrose intake and behavior on the plus-maze in male mice. *Physiol Behav.* 1999; 67:189–196. [PubMed: 10477049]
33. Buron G, Hacquemand R, Pourie G, Lucarz A, Jacquot L, Brand G. Comparative behavioral effects between synthetic 2,4,5-trimethylthiazoline (TMT) and the odor of natural fox (*Vulpes vulpes*) feces in mice. *Behav. Neurosci.* 2007; 121:1063–1072. [PubMed: 17907837]

34. Hebb AL, Zacharko RM, Gauthier M, Drolet G. Exposure of mice to a predator odor increases acoustic startle but does not disrupt the rewarding properties of VTA intracranial self-stimulation. *Brain Res.* 2003; 982:195–210. [PubMed: 12915255]
35. Hebb AL, Zacharko RM, Dominguez H, Laforest S, Gauthier M, Levac C, Drolet G. Changes in brain cholecystokinin and anxiety-like behavior following exposure of mice to predator odor. *Neuroscience.* 2003; 116:539–551. [PubMed: 12559109]
36. Munn E, Bunning M, Prada S, Bohlen M, Crabbe JC, Wahlsten D. Reversed light-dark cycle and cage enrichment effects on ethanol-induced deficits in motor coordination assessed in inbred mouse strains with a compact battery of refined tests. *Behav. Brain Res.* 2011; 224:259–271. [PubMed: 21664382]
37. Bailoo JD, Bohlen MO, Wahlsten D. The precision of video and photocell tracking systems and the elimination of tracking errors with infrared backlighting. *J Neurosci. Methods.* 2010; 188:45–52. [PubMed: 20138914]

Highlights

- A study of ethanol effects on 8 strains of mice was balanced for experimenters.
- Strain differences and ethanol effects were clearly significant ($P < .00001$).
- Open field activity differed for experimenters ($d = 0.8$) after ethanol injection.
- Experimenters rated head dips on the elevated plus maze differently ($d = 1.2$).
- Fall latency on the accelerating rotarod differed between experimenters ($P = .005$).

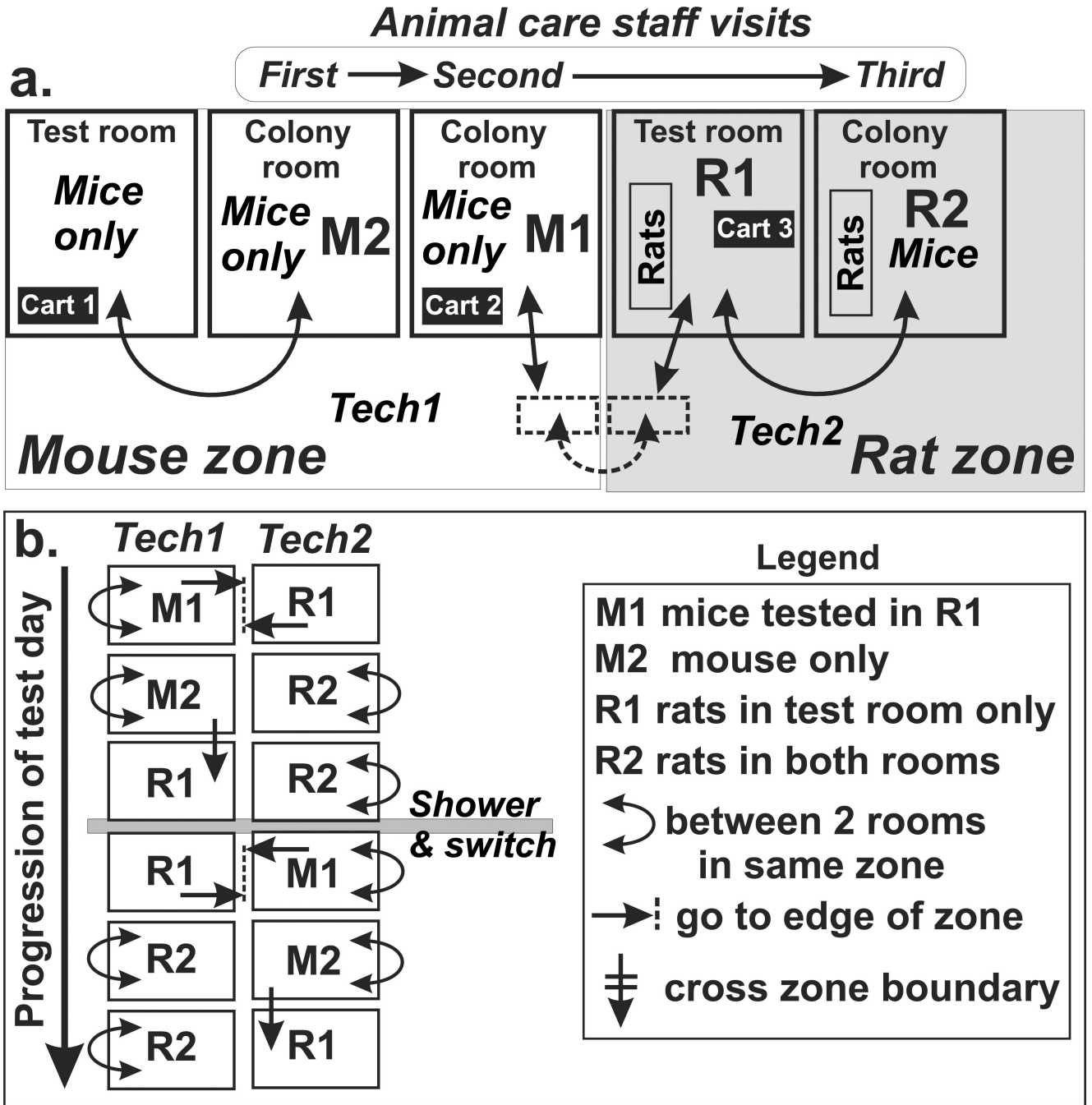


Fig. 1.
 A. Arrangement of rooms used for housing and testing mice and rats. The rooms and hallway in the mouse zone never had rats housed or transported there, while the rat zone provided housing for rats and testing for mice that were to be exposed to rat odors. Mice to be tested in a room with rats were transferred from one cage to another on carts at the intersection of the zones. Colony rooms where animals were housed overnight did not contain test apparatus. B. Work flow for the two technicians during a day. The experimenters

showered and put on clean clothing at mid-day before one who had been working near rats moved to rooms that contained mice not previously exposed to rats.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

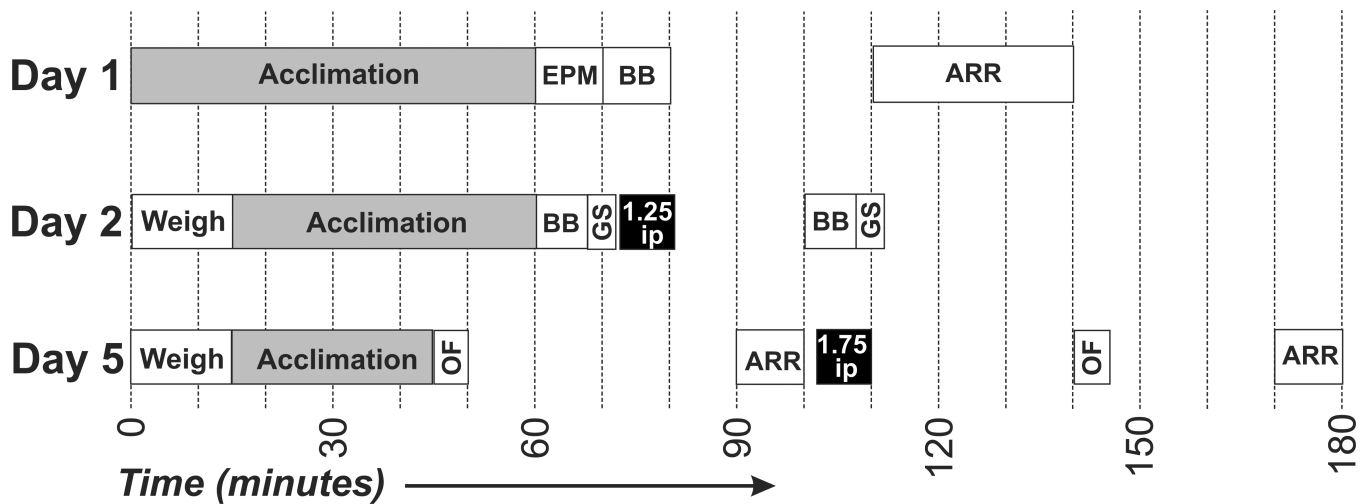


Fig. 2.

Schedule of testing a squad of mice within each of three test days. No alcohol was given on Day 1. On Day 2, mice were tested shortly before intraperitoneal (ip) injection of a 1.25g/Kg dose of alcohol and then again about 30 min after injection. On Day 3, the times between open field and accelerating rotarod are averages because there were always four mice in one squad tested at the same time on the rotarod, whereas mice were tested one at a time in the open field apparatus in a staggered pattern. Abbreviations: ARR, accelerating rotarod; BB, balance beam; EPM, elevated plus maze; GS, grip strength; OF, open field.

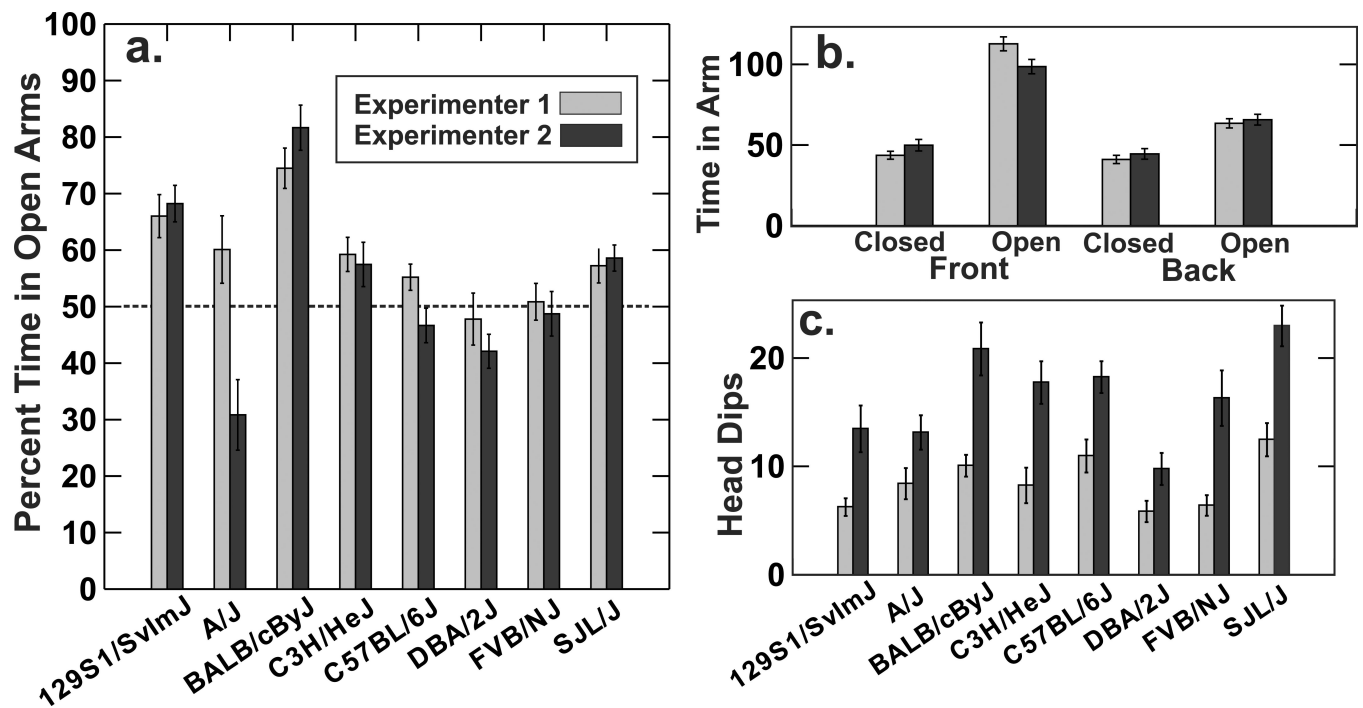


Fig. 3.

Results (means and standard errors) for eight strains tested by two experimenters on the elevated plus maze. **A.** There was a marked difference between experimenters only for strain A/J. **b.** Averaged over all strains and conditions, mice spent considerably more time in the open arms, especially the arm away from the nearest wall. **C.** The two experimenters made substantially different judgments about the frequency of head dips over the edges of the open arms for all strains.

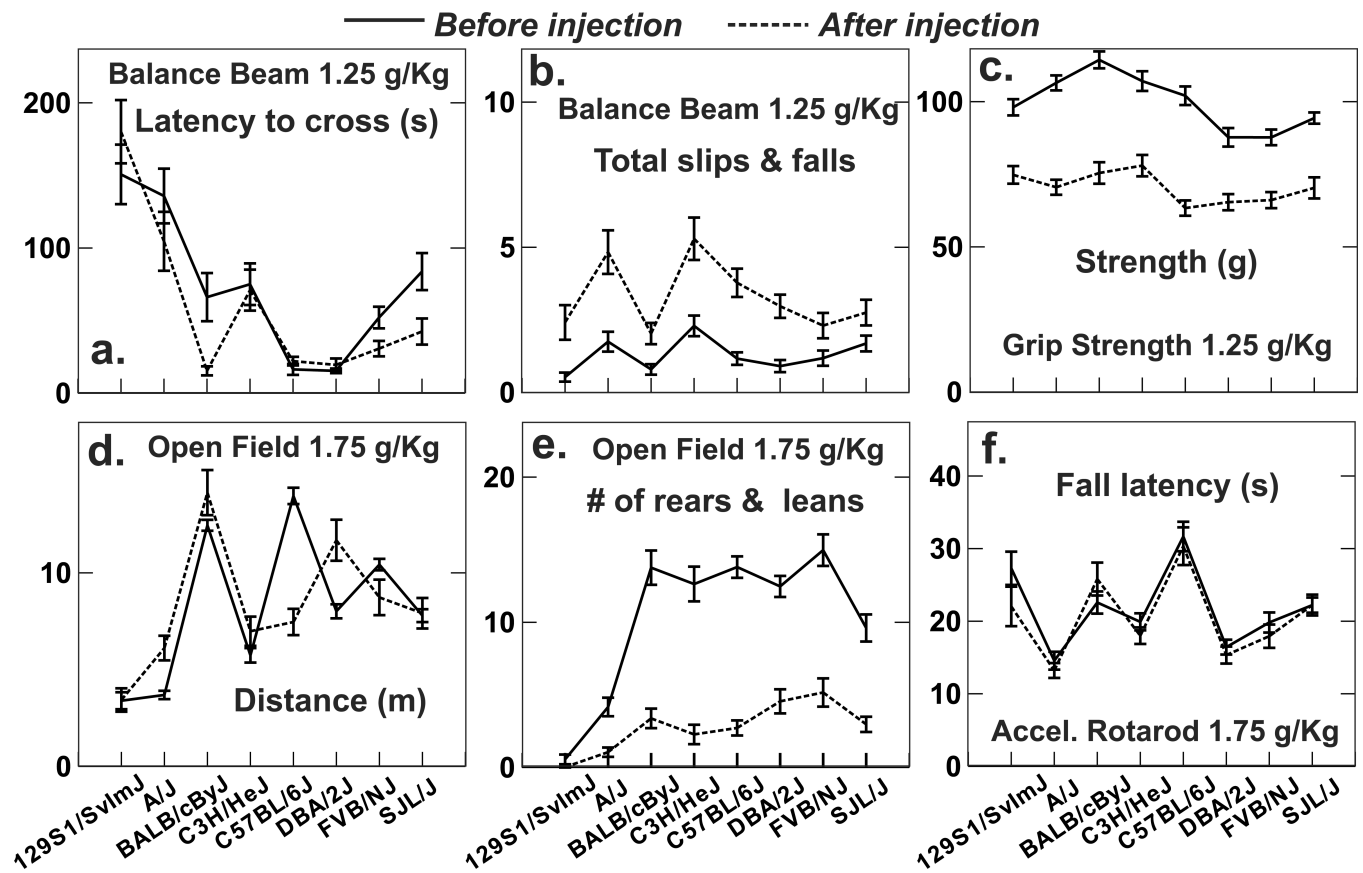


Fig. 4.

Strain mean scores (with standard error bars) before and after alcohol injection for four behavioral tests. A. Time required to traverse the balance beam was considerably lower after alcohol injection for strains BALB/cByJ and SJL/J but changed little for the other strains. B. Slips of a leg or the body from the beam were substantially more frequent after alcohol for all strains. Falls were included in the summary but were quite rare. C. Grip strength was markedly weaker after alcohol for all strains. D. Open field distance was unchanged for most strains but was influenced by alcohol in opposite directions for strains C57BL/6J and DBA/2J. e. Rearing and leaning against a wall were greatly reduced after the alcohol injection, except for strain 129S1/SvImJ that showed very little of these behaviors before injection. F. Fall latency from the accelerating rotarod was virtually unchanged after alcohol injection, although the pattern differed between experimenters (Fig. 5).

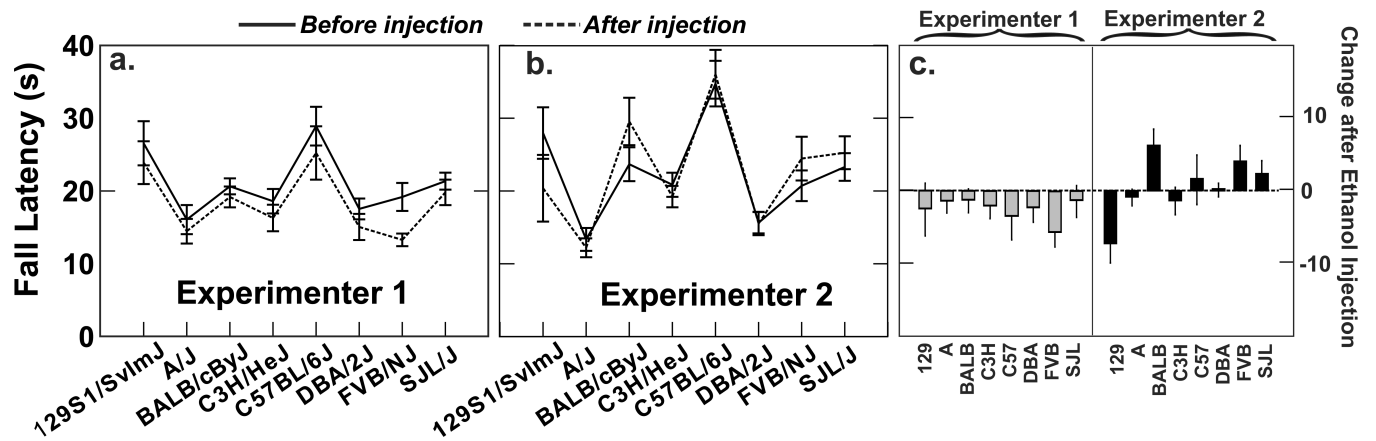


Fig. 5. Effects of alcohol differed markedly between the two experimenters. A. For experimenter 1, fall latency was consistently decreased by a moderate amount. B. For experimenter 2, there was no alcohol effect at all for some strains, while others changed in opposite directions. C. The changes following alcohol were generally small, as indicated by the standard error bars. Detailed statistical analyses are presented in Table 1 and in the text.

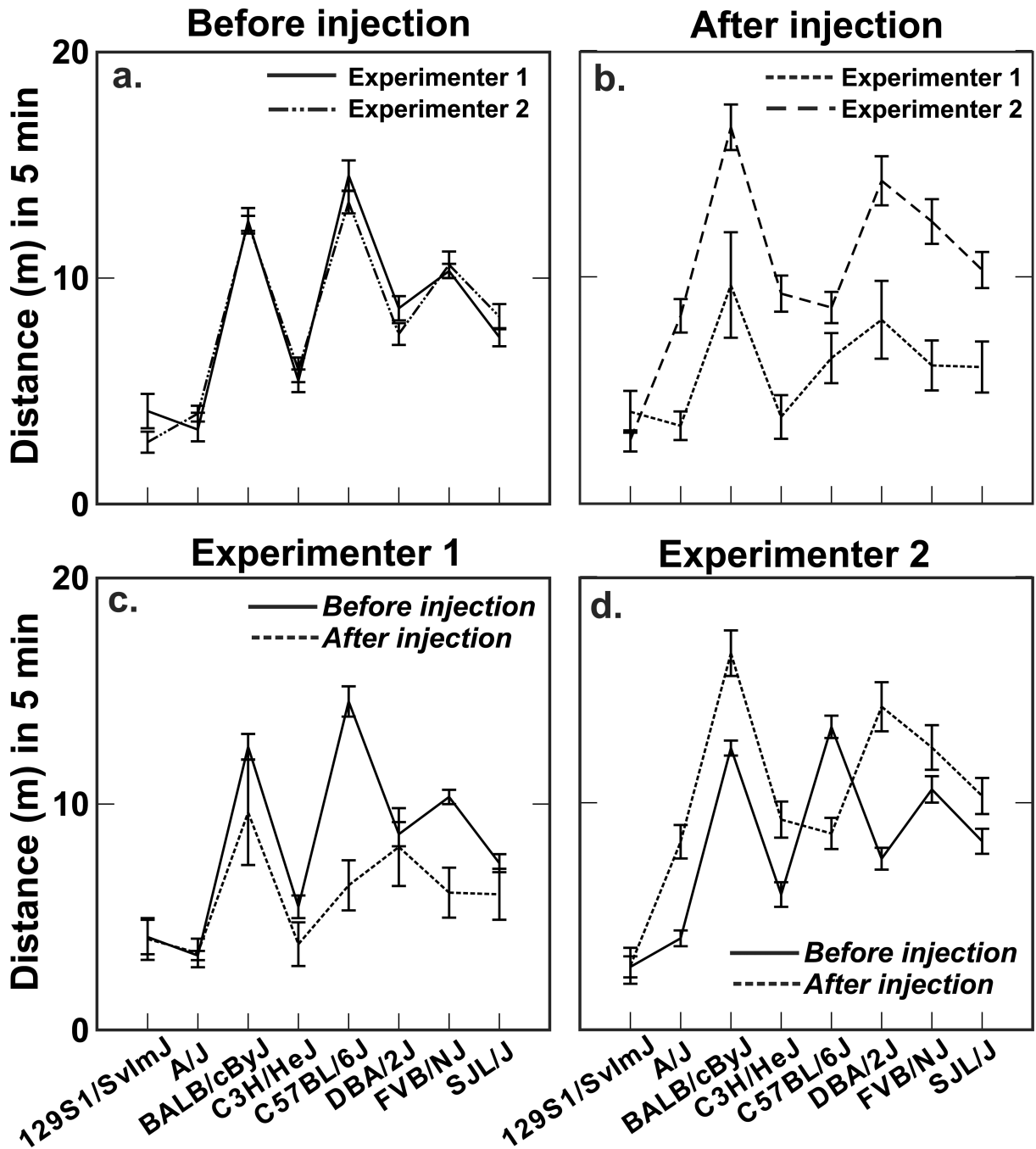


Fig. 6.

A. The pattern of strain differences in activity in the open field, a measure obtained with video tracking, was almost identical for the two experimenters before alcohol injection. B. The pattern of strains differences was similar for the two experimenters after injection, but activity generally decreased for experimenter 1 and increased for experimenter 2. C. The decrease for experimenter 1 is more clearly apparent when shown as before and after

injection. D. Likewise the increase in activity after alcohol for experimenter 2 is evident for all but two strains.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1Results of analysis of variance (P values, two-tailed)^a

Task	Measure	Strain df = 7	Exp df = 1	Str × Exp df = 7	Ethanol df = 1	Eth × Str df = 7	Eth × Exp df = 1	Eth × Str × Exp
EPM	Distance	< .00001	.31	.48	NA	NA	NA	NA
EPM	% time in open arms	< .00001	.015	.0001	NA	NA	NA	NA
EPM	Head dips	< .00001	< .00001	.22	NA	NA	NA	NA
EPM	Rearing	< .00001	.15	.16	NA	NA	NA	NA
BB	Time to cross	< .00001	.05	.33	.01	.002	.36	.92
BB	Total slips	< .00001	.25	.06	< .00001	.03	.68	.40
ARR	Pretrain lat.	< .00001	.07	.17	NA	NA	NA	NA
ARR	Fall latency	< .00001	.005	.07	.03	.11	.007	.04
OF	Distance	< .00001	< .00001	.005	.70	< .00001	< .00001	.007
OF	% wall	< .00001	.92	.82	.02	< .00001	.42	.42
OF	Rear, lean	< .00001	.01	.20	< .00001	< .00001	.13	.64
GRIP	Strength	< .00001	.98	.28	< .00001	.00002	.02	.54

Abbreviations: EPM, elevated plus maze; BB, balance beam; ARR, accelerating rotarod; OF, open field; GRIP, grip strength; Exp, experimenter; Str, strain; Eth, ethanol; NA, not applicable, no alcohol given.

^aDegrees of freedom within groups ranged from 256 to 270 for all tests.