Structured evaluation of rodent behavioral tests used in drug discovery research

*Anders Hånell[1,2] * and Niklas Marklund[1]*

[1] Department of Neuroscience, Section for Neurosurgery, Uppsala University, Uppsala, Sweden
[2] Department of Anatomy and Neurobiology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

A large variety of rodent behavioral tests are currently being used to evaluate traits such as sensory-motor function, social interactions, anxiety-like and depressive-like behavior, substance dependence and various forms of cognitive function. Most behavioral tests have an inherent complexity, and their use requires consideration of several aspects such as the source of motivation in the test, the interaction between experimenter and animal, sources of variability, the sensory modality required by the animal to solve the task as well as costs and required work effort. Of particular importance is a test's validity because of its influence on the chance of successful translation of preclinical results to clinical settings. High validity may, however, have to be balanced against practical constraints and there are no behavioral tests with optimal characteristics. The design and development of new behavioral tests is therefore an ongoing effort and there are now well over one hundred tests described in the contemporary literature. Some of them are well established following extensive use, while others are novel and still unproven. The task of choosing a behavioral test for a particular project may therefore be daunting and the aim of the present review is to provide a structured way to evaluate rodent behavioral tests aimed at drug discovery research.

Keywords: animal behavior, translational medicine, phenotyping, mice, rats

INTRODUCTION

Charles Darwin may be considered to be the founder of behavioral research (Thierry, 2010). Since then, behavioral testing has been extensively used to gain a better understanding of the central nervous system (CNS) and to find treatments for its diseases. Early experimental work on animal behavior includes Ivan Pavlov's work on conditional reflexes in dogs which began at the end of the 19th century (Samoilov, 2007). Continued interest in animal behavior gave rise to the field of ethology which resulted in the Nobel Prize in Physiology and Medicine in 1973 shared by Karl von Frisch, Konrad Lorenz and Nikolaas Tinbergen. They studied animals in their natural habitat, which made controlled experiments difficult. This problem was addressed by introducing behavioral testing in a laboratory setting in the early twentieth century, which evolved into the field of comparative psychology in a process facilitated by the important contributions made by B F Skinner (Gray, 1973; Dews et al., 1994).

Historically, a large variety of species has been used for behavioral testing but rodents have always been widely used, likely since they are mammals and easy to house and breed. In contrast to common pets such as cats and dogs, there may also be a higher acceptance in the general public for the use of rodents in medical research. Although hamsters, guinea pigs and Mongolian gerbils have been subjected to behavioral testing, mice and rats are far more popular and firmly established as model organisms with several outbred stocks and inbred strains available for experiments. Unlike other rodents used for research, mice and rats

belong to the subfamily Murinaea and are sometimes referred to as murine models. Early examples of rodent behavioral testing include Karl Lashley's work on learning and memory using mazes in the early twentieth century. Initially, wild-caught rodents were used for experiments (McCoy, 1909) but this practice changed with the introduction of strains bred by mouse and rat fanciers (Steensma et al., 2010). Following approximately a hundred years of breeding, contemporary laboratory animals are now considerably more docile than their wild counterparts (Wahlsten et al., 2003a). Over time, there has been a continuous evolution of rodent behavioral tests and there are well over 100 tests in contemporary use, exhaustively summarized in Supplementary Table 1 of the present review. In recent years, genetically modified mice have become readily available and the use of mice in behavioral testing recently surpassed that of rats (**Figure 1**). However, the increasing availability of genetically modified rats (Jacob et al., 2010) may shift the tide again.

Unfortunately, rodent behavioral testing in the laboratory setting have proved difficult and test results may vary depending on the person performing the experiment (Chesler et al., 2002), in which laboratory the experiments are performed (Crabbe et al., 1999) and environmental factors including for example animal housing (Richter et al., 2010). To further advance the field of rodent behavioral testing and achieve reliable and reproducible results, these issues must be resolved. Fortunately, recent technological improvements facilitate the design and construction of sophisticated automated test equipment. New techniques

include 3D printing which can construct structures of almost any shape (Jones, 2012), user friendly electronic microcontrollers such as Arduino boards, which can control gates, sensors and reward delivery (D'Ausilio, 2012), as well as Radio Frequency Identification (RFID), which is used to detect the position and identity of individual rats and mice (Lewejohann et al., 2009). Combined with sound ethological principles and a good understanding of rodent biology, these techniques may facilitate the design and construction of novel behavioral tests with improved characteristics.

Unfortunately, there is no such thing as a perfect behavioral test and the most suitable one has to be chosen depending on the goals of the project. There are, however, a large number of factors to take into account when making this decision. To facilitate the process of finding strengths and weaknesses of existing behavioral test, the present review evaluates a multitude of important aspects of behavioral testing which are summarized in **Table 1**.
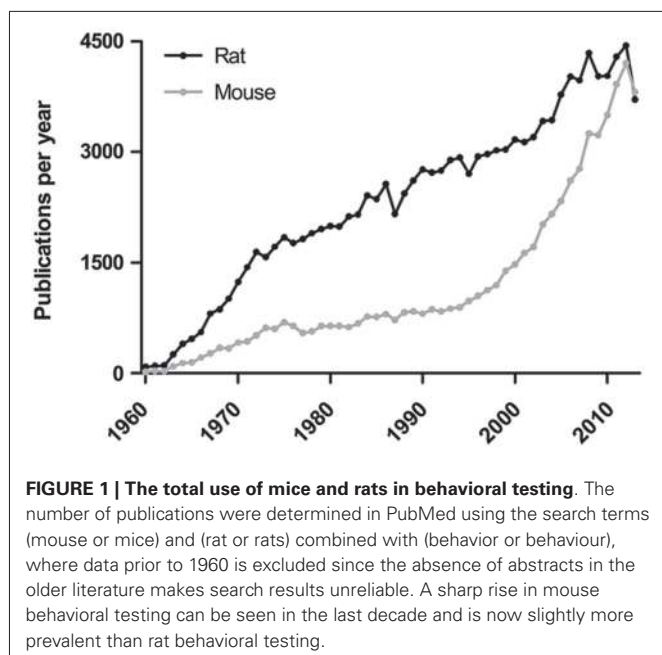
MOTIVATION

Most behavioral tests used to evaluate sensory-motor function as well as learning and memory aim to measure an animal's ability to solve a task. On the other hand, behavioral tests used in rodent drug dependence research such as self administration and conditioned place preference (see Supplementary Table 1), focus on measuring the motivation to perform a selected action. The measured performance in a test will invariably include a combination of both the ability and the motivation of the animal to solve the task. To detect differences in ability, the level of motivation should therefore be equal among individual animals. To solve the problem of variability in motivation, behavioral tests typically attempt to provide sufficient motivation to make each animal perform at the height of its ability. However, a variable level of motivation is a potential problem in many tests, including the rotarod and grip strength test (Balkaya et al., 2013). A commonly chosen source of motivation is fear, such as fear of drowning in the forced swim test (FST; Porsolt et al., 1977), fear of falling in the rotarod (Dunham and Miya, 1957) or fear of receiving electrical shocks during active avoidance learning (Moscarello and LeDoux, 2013). Fear can be studied independently, using for example predator odors, to gain insights into for instance post-traumatic stress disorder (Staples, 2010; Johansen et al., 2011), which differs from its use as a motivator

Table 1 | Points to consider when evaluating a behavioral test.

Motivation	How are the animals motivated to perform the task? Is the level of motivation high and stable? Can the source of the motivation interfere with the disease model?
Animal-experimenter interaction	Are the experimenter and the animal in close contact during testing? Can handling be performed prior to testing to allow the animals to adjust to human contact?
Dynamic range	Can the test make accurately measure the ability of both naïve and severely impaired animals? Is there a risk for flooring or ceiling effects? Can the test difficulty be adjusted within or between trials?
Repeatability	Can the test be repeated to assess changes in ability over time?
Interaction with other tests	Is there a risk that the test experience can affect the behavior in other test?
Data collection	How are the results collected? Is there a risk for subjective effects in the evaluation? Can the results be collected automatically?
Result evaluation	Which statistical tests can be used to evaluate the results? How are the results presented?
Result interpretation	Can the results be attributed to a single domain or can, for example, changes in general activity level interfere with the measurements?
Automation	How much of the test procedure is automated? Does the test have the potential to be fully automated?
Variability	Can the results be related to baseline performance to mitigate the effects of variability? Can the estrous cycle of female animals be measured to restrict testing to a single day in the cycle?
Experimental design	Can the test be performed in a blinded fashion? Can the test be used for both mice and rats of different strains and stocks to obtain more robust results?
Sensory modality	Which sensory modality does the animal use to solve the task? Is it possible to confirm intact sensory functions?
Predictive validity	Do drugs approved for human use improve test results for rodents?
Construct validity	Does the test rely on brain structures used for this psychological construct in humans?
Ethological validity	Does the test resemble natural rodent behavior?
Face validity	Is it immediately apparent what the test is intended to measure?
Intrinsic validity	Does the test give the same result when experiments are repeated?
Extrinsic validity	Does the test give the same result when performed in, for example, different strains, age group or species?
Housing	Does the test cause restrictions on how the animals can be housed? Can the social status be assessed for group housed animals to allow the creation of balanced experimental groups?
Throughput	How long does it take to run the test? Do the animals have to be trained before performing the test? Can several animals be tested in parallel? Is the collection and evaluation of the results time consuming?
Costs	How much does the equipment cost? How much lab space has to be devoted to the test? How much staff time does the test require?
Practical considerations	Can the equipment be stowed away when not in use? Is extensive training of the experimenter required to carry out the test? Does the test have to be performed on several consecutive days which may overlap with weekends, holidays and vacations? Can the test be run by a substitute in case of sick leave? Can the test equipment be easily cleaned and disinfected?

The points are ordered as they appear in the text and the importance of each point has to be assessed based on the goals of individual projects.

FIGURE 1 | The total use of mice and rats in behavioral testing. The number of publications were determined in PubMed using the search terms (mouse or mice) and (rat or rats) combined with (behavior or behaviour), where data prior to 1960 is excluded since the absence of abstracts in the older literature makes search results unreliable. A sharp rise in mouse behavioral testing can be seen in the last decade and is now slightly more prevalent than rat behavioral testing.

in tests which evaluate other traits. Fear must be used cautiously as a source of motivator since it may cause unwanted effects such as freezing or panic-like behavior, and fear-induced stress may negatively influence cognitive performance (Harrison et al., 2009). When a test situation is perceived as dangerous to the animal, motivation can also be provided by introducing an escape route to the home cage when the task is solved (Blizard et al., 2006).
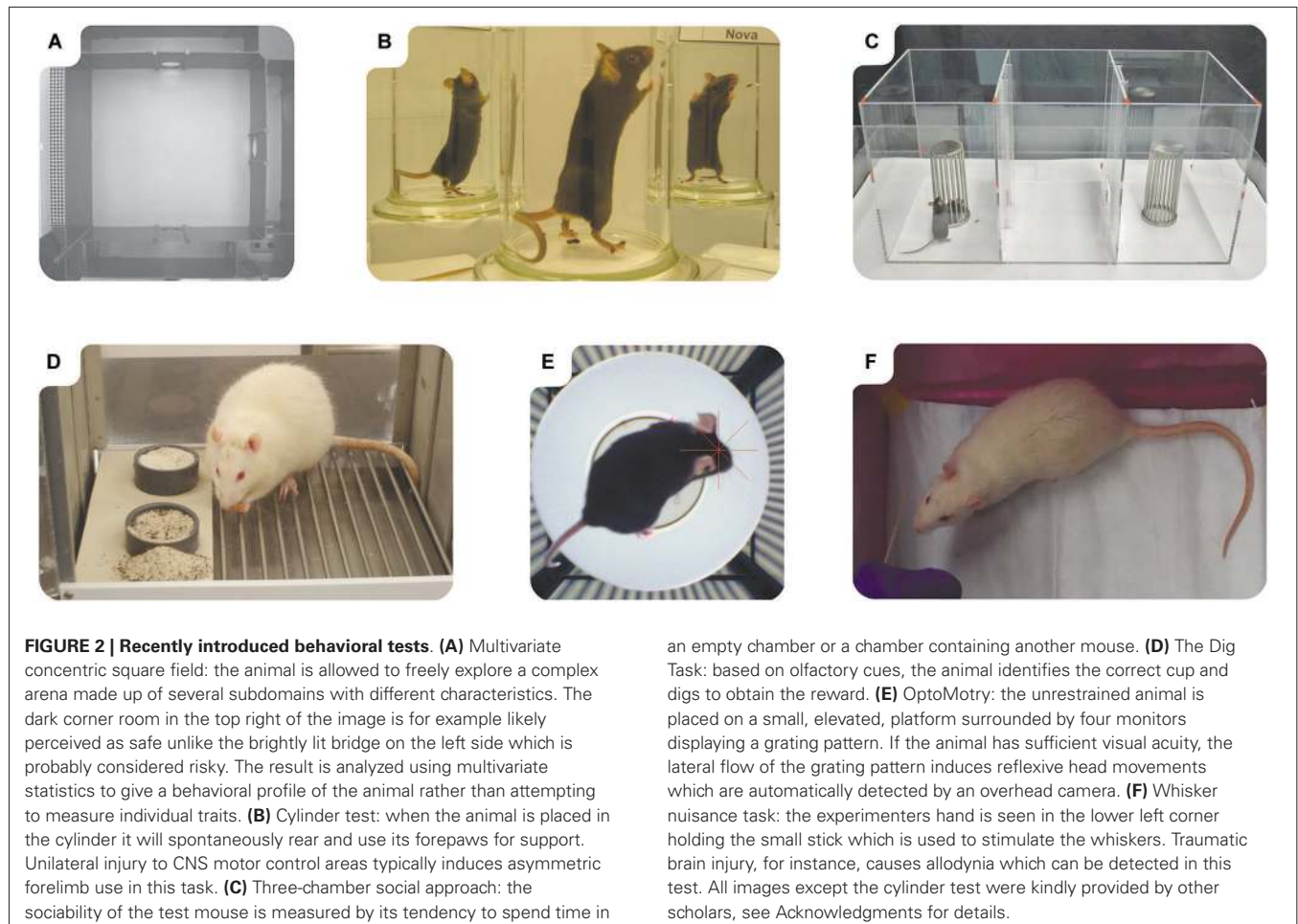
Hunger is another commonly used motivator. Rodents typically feed throughout the day although mainly in the beginning and end of the dark phase (Clifton, 2000), and a sufficient level of hunger for motivation is induced at a level of food deprivation which causes a 10–20% reduction in body weight. It must be noted that fear, stress and anxiety inhibit food consumption (Petrovich et al., 2009) and the animals may have to be habituated to the test arena prior to initiation of the experiment. The need for food deprivation might be alleviated or avoided by using a highly palatable food, by using the natural tendency of rats and mice to forage for food and hoard it in their nests (Whishaw et al., 1995) or by relying on thirst rather than hunger. In addition, rodent diet may also influence test results. Free access to food cause obesity in laboratory rodents (Martin et al., 2010) and dietary restriction was observed to be beneficial in models of stroke, addiction and excitotoxicity (Bruce-Keller et al., 1999; Yu and Mattson, 1999; Guccione et al., 2013).

Several tests rely on spontaneous rodent behavior, such as the exploration of a novel environment in the open field (Hall, 1934), multivariate concentric square field (**Figure 2A**; Meyerson et al., 2006) and the cylinder test (**Figure 2B**; Schallert et al., 2000). Furthermore, social interaction tests such as the three-chamber social approach (**Figure 2C**; Nadler et al., 2004), can be used to evaluate memory function as well as sociability. Other spontaneous behaviors include nest building and burrowing which can

be used to assess cognitive function (Deacon, 2012). Relying on spontaneous behaviors may reduce the need of strong motivators and likely attenuates the stress level of the animal in a test in contrast to the release of stress hormones induced by, for instance, the Morris Water Maze (MWM; Engelmann et al., 2006). Motivation to repeatedly perform simple tasks may be provided by operant conditioning. This can, for example, be used to motivate animals to press a lever to receive a sucrose pellet and is often performed in a Skinner box. Operant conditioning does, however, typically require some degree of food deprivation as well as an intact learning ability. It can be assumed that the use of positive motivators is not only ethical but also less likely to result in stress-induced aberrant behaviors. Insufficient interest to perform tasks without a strong motivator is still a potential caveat in behavioral testing though this problem might be mitigated by the longer test sessions enabled by automated testing (see *Automated testing* below).

ANIMAL-EXPERIMENTER INTERACTION

In virtually all behavioral tests, there is a degree of interaction between experimenter and animal which potentially influences the obtained results. The importance of this interaction was established by the observation that experimenter identity had a greater influence than genotype on hot plate test results (Chesler et al., 2002). Although different interpretations of test instructions might be one explanation, differences between researchers in their amount of animal work experience and anxiety towards rodents is also likely to be important. It has also recently been demonstrated that the presence of a male, but not female, experimenter induce analgesia in rodents (Sorge et al., 2014). In addition, rats are able to distinguish between, and results may be affected by the level of rodent familiarity with, the individual experimenters (McCall et al., 1969; van Driel and Talling, 2005) and any remaining odor traces from predatory pets such as cats will induce stress in rodents (Burn, 2008). Individual human experimenters may also display some day to day variation in for example stress level, mood and/or odor, which potentially increases the variability of the results. Individual rodents also differ in their response to humans and typically initially avoid human contact but gradually accept it following repeated exposure (Schallert et al., 2003; Hurst and West, 2010). Handling of laboratory animals prior to any behavioral testing may therefore reduce the effects of animal-experimenter interaction and potentially reduce variability (Schmitt and Hiemke, 1998; Hurst and West, 2010). Additionally, lack of handling before a series of repeated testing may cause altered results over time as the animals get more and more used to human contact. Note, however, that human presence always influences animal behavior and reduced fear of the experimenter may even decrease the motivation to perform some tests. Animal-experimenter interaction is not limited to fear reactions, and may also be caused by curiosity and the anticipation of reward. The importance of animal-experimenter interaction likely differs between tests, and is likely of little concern in for example operant chambers, while it potentially significantly affects several tests of neurological function where the animal is held by the experimenter throughout the test.

FIGURE 2 | Recently introduced behavioral tests. **(A)** Multivariate concentric square field: the animal is allowed to freely explore a complex arena made up of several subdomains with different characteristics. The dark corner room in the top right of the image is for example likely perceived as safe unlike the brightly lit bridge on the left side which is probably considered risky. The result is analyzed using multivariate statistics to give a behavioral profile of the animal rather than attempting to measure individual traits. **(B)** Cylinder test: when the animal is placed in the cylinder it will spontaneously rear and use its forepaws for support. Unilateral injury to CNS motor control areas typically induces asymmetric forelimb use in this task. **(C)** Three-chamber social approach: the sociability of the test mouse is measured by its tendency to spend time in an empty chamber or a chamber containing another mouse. **(D)** The Dig Task: based on olfactory cues, the animal identifies the correct cup and digs to obtain the reward. **(E)** OptoMotry: the unrestrained animal is placed on a small, elevated, platform surrounded by four monitors displaying a grating pattern. If the animal has sufficient visual acuity, the lateral flow of the grating pattern induces reflexive head movements which are automatically detected by an overhead camera. **(F)** Whisker nuisance task: the experimenters hand is seen in the lower left corner holding the small stick which is used to stimulate the whiskers. Traumatic brain injury, for instance, causes allodynia which can be detected in this test. All images except the cylinder test were kindly provided by other scholars, see Acknowledgments for details.

In the attempts made to evaluate reproducibility between different laboratories the experiments are usually performed by different persons (Crabbe et al., 1999; Mandillo et al., 2008), with experimenter identity reported as an important source of variability (Mandillo et al., 2008). Animal-experimenter interaction may thus be one of the reasons for the difficulties in achieving consistent results in behavioral tests. By relying on fully automated testing (see *Automated testing* below) human contact is avoided and the reproducibility of the test is potentially increased. Since fully automated, high quality, behavioral test are still sparse, the problem of animal-experimenter interaction is instead typically addressed by having the same person perform all testing within a study.
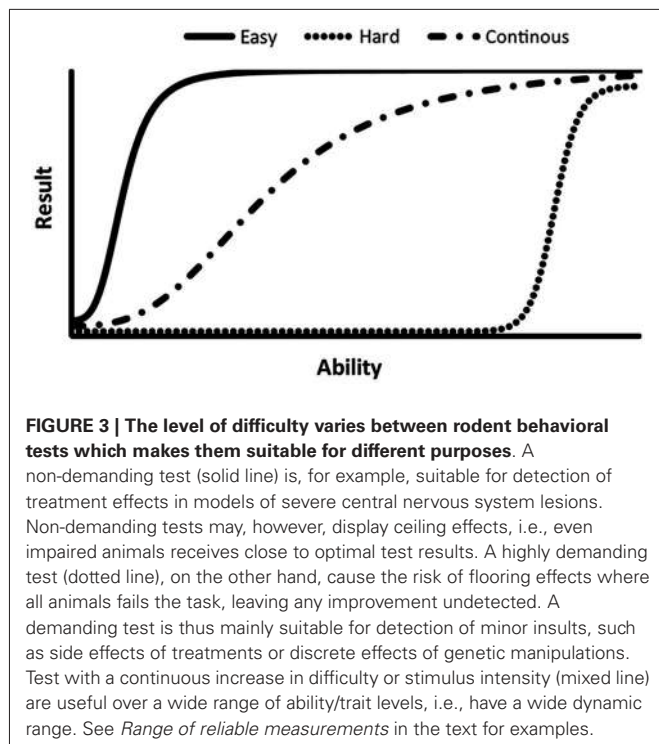
RANGE OF RELIABLE MEASUREMENTS

A behavioral test should ideally enable accurate and precise measurements without flooring and ceiling effects (**Figure 3**) when testing both highly impaired and normally functioning animals. A reliable assessment over a wide range of ability levels can be achieved by continuously increasing the difficulty or stimulus intensity during a test session, such as protocols using accelerating, rather than constant, speed in the rotarod. In this test, the rotating drum starts at a very slow speed, demanding only for severely impaired animals, and then accelerates up to a speed challenging even for naïve mice (Dunham and Miya, 1957; Brooks and Dunnett, 2009). The same principle is applied in the ledged tapered beam, a modification of traditional beam walking tests. Here, the beam is initially wide and easy to traverse but gradually tapers and becomes increasingly narrow and challenging (Schallert et al., 2002). This approach to test design is not limited to motor function tests and is for example also used in the successive alleys test of anxiety-like behavior. This test consists of four alleys which are increasingly narrow and open, and thus more anxiogenic, to enable assessment of anxiety-like behavior over a wide range (Deacon, 2013). The demands of a test can also be controlled using parameters within the test, for example by adjusting the platform size in the MWM (Vorhees and Williams, 2006) or by changing the temperature in the hot plate test (Neelakantan and Walker, 2012). Correct parameter settings are important since differences between two groups cannot be detected if a test is either too difficult or too easy (**Figure 3**).

REPEATABILITY AND INTERACTION BETWEEN TESTS

Repeated testing is desirable in the study of diseases with a dynamic and prolonged course as well as in development and ageing research. The measurement of baseline performance also

FIGURE 3 | The level of difficulty varies between rodent behavioral tests which makes them suitable for different purposes. A non-demanding test (solid line) is, for example, suitable for detection of treatment effects in models of severe central nervous system lesions. Non-demanding tests may, however, display ceiling effects, i.e., even impaired animals receives close to optimal test results. A highly demanding test (dotted line), on the other hand, cause the risk of flooring effects where all animals fails the task, leaving any improvement undetected. A demanding test is thus mainly suitable for detection of minor insults, such as side effects of treatments or discrete effects of genetic manipulations. Test with a continuous increase in difficulty or stimulus intensity (mixed line) are useful over a wide range of ability/trait levels, i.e., have a wide dynamic range. See *Range of reliable measurements* in the text for examples.

allows treatment groups to be equally balanced based on performance prior to, for example, drug administration or injury induction (Lenzlinger et al., 2005). Repeating a test may, however, not always be possible since the experience of one test session may influence subsequent testing. Most behavioral tests are likely affected to some extent by repeated testing, and practice effects have for example been observed in the zero maze (Cook et al., 2002) and the hot plate test (Espejo and Mir, 1994). Repeated testing can also be influenced by the test interval, demonstrated by the functional recovery seen with daily, although not weekly, assessment in the rotarod following traumatic brain injury (O'Connor et al., 2003). These results also suggest that intensive behavioral testing can function as rehabilitation therapy which would cause the testing itself to influence the obtained results. Furthermore, learning effects may influence the results when tests are repeated and impairments in learning and memory may thus influence the evaluation of other brain functions. Test repeatability is also desirable when extensive training prior to the actual testing is required, an approach validated in for example the 5-choice continuous performance test (Young et al., 2013).

Apart from repeating a single test, animals are also commonly subjected to several different behavioral tests. This practice may be problematic since participation in one test potentially influences the results obtained from subsequent testing. Thus, the order in which the tests are carried out is important and performing the tests on separate days potentially reduces the interaction between tests. However, this strategy has to our knowledge not been verified. A possible solution may be to combine several tests into a single test. For instance has the open field, elevated plus maze (EPM) and light-dark box tests been integrated into

a single test arena (Ramos et al., 2008). Another strategy is to use a test that does not measure a single behavioral trait in the animal, but instead determines a behavioral profile. An example of this strategy is the multivariate concentric square field test, which uses a complex arena to enable simultaneous evaluation of several aspects of rodent behavior and subsequent analysis of the results using multivariate statistics (Meyerson et al., 2006; Ekmark-Lewén et al., 2010). Another example is a modified version of the hole board test, which measure several behaviors related to anxiety-like behavior, cognition and social interactions (Ohl and Keck, 2003).

DATA COLLECTION AND RESULT INTERPRETATION

Specific rodent behaviors are typically difficult to describe using a single continuous variable and categorical scales are therefore commonly used. Manual scoring using this type of scales may be subjective and should therefore be performed by a researcher blinded to the treatment, disease and genetic status of the evaluated animal. Preferably, the evaluation should also be preceded by an evaluation of the inter- and intra-rater reliability (Shrout and Fleiss, 1979; Rousson, 2011). The use of automated data collection such as video tracking software for mazes and the use of hind paw attached magnets to measure inactivity in the FST (Shimamura et al., 2007), is not always feasible but may assure objective data sampling and likely reduces the work load (see *Automated testing* below). One caveat caused by automated data collection is the increased risk of false positive findings when evaluating a large number of behavioral variables from one or several tests. Proper use of corrections for multiple testing, multivariate statistics, a clearly defined hypothesis and replication of key findings in independent experiments are potential solutions for this problem.

When behavioral test results have been collected, the interpretation of them is rarely obvious. For instance, mice which floats in a single location in the MWM fails to find the platform although this behavior may not reflect impairments in learning and memory capacity (Wahlsten et al., 2003c). Additionally, if a rodent clings onto the rod in the rotarod test instead of running on top of it, no conclusions about its balance and coordination can be made (Wahlsten et al., 2003b). In the interpretation of behavioral test results, an important issue is to understand the cause of the observed behavior. By studying natural rodent behavior, evaluating the ethological validity of test (see *Validity* below), determining the source of motivation in the test (see *Motivation* above) and using the knowledge of rodents' sensory capacity (see *Sensory modality* below) to view the test from a rodent's perspective, increased understanding of the rodent behavior in a test may be achieved.

Redesigning existing tests can potentially resolve interpretation issues in some cases. One such example is the elevated zero maze, a modification of the EPM, in which the center zone has been removed since the amount of time spent in it is difficult to interpret (Shepherd et al., 1994; Braun et al., 2011). Reduced levels of paw licking and escape behaviors in the hot plate test can be attributed to both anesthetic, anxiolytic and sedative drug effects (Yezierski and Vierck, 2011; Casarrubea et al., 2012). A potential solution to this interpretation issue

is to use the double plate test where the animal is allowed to freely explore two plates, one plate at room temperature and the other maintained at an aversive temperature. The fraction of the time spent on each plate may then be used to quantify anesthetic effects (Walczak and Beaulieu, 2006). High variability is another common problem in behavioral testing, likely mainly caused by factors outside the test situation such as social status and past experiences of the animal. However, some of the variability may be caused by imprecise measurements and more extensive testing may mitigate this problem. For example, the evaluation of more rears in the cylinder test or performing more test runs in the beam walk test may increase the reliability of these tests. Behavioral test sessions often also have a limited duration, typically only a few minutes, which may be insufficient. Accordingly, altered behavioral patterns have been observed when using extended test session durations in the open field test (Fonio et al., 2012). Although beyond the scope of the present review, adequate use of statistics is obviously crucial in behavioral testing and a detailed discussion about core concepts can be found in the *Points of significance* article series (Krzywinski and Altman, 2013). When interpreting beneficial pre-clinical treatment effects, it is, however, also important to not only consider the statistical significance of the effect but also if the magnitude is sufficient to translate into a clinically meaningful improvement. Finally, if behavioral test results are described with a single, easily interpreted variable, it is easier for other scientist to draw correct conclusions from them. However, this strategy has to be balanced against the benefit of a detailed description of the observed behavior.

AUTOMATED TESTING

The continuous advancement in electronics and image analysis makes automated behavioral testing procedures increasingly feasible. The use of automated procedures has several potential advantages such as objective scoring, avoidance of animal-experimenter interaction, extended test durations and reduced work effort. Fully automated systems are, however, still sparingly used although automated data collection is rather common (see *Data collection and result interpretation* above) and operant chambers typically only require the animal to be placed in the testing chamber while the rest of the procedure is automated. Simultaneous tracking of individual animals within a group can be achieved by the application of different fluorescent dyes to the animals' fur which allows automated evaluation of social dynamics over extended time periods (Shemesh et al., 2013). As previously correctly pointed out (Crabbe and Morris, 2004), it is unfeasible to design automated tests which rely on a robotic system to capture and ferry animals from their home cage to the test arena. Fully automated testing may instead be carried out in the animals' home cage (de Visser et al., 2006; Krackow et al., 2010) or in a test module attached to the home cage via an automated sorting system (Schaefers and Winter, 2011; Winter and Schaefers, 2011). Integration of the test equipment and home cage into a single unit is for example used in the IntelliCage (Krackow et al., 2010) and the PhenoTyper (de Visser et al., 2006). This strategy does, however, have certain limitations. The test equipment can, for example, only be used

by one group of animals at a time, and to test a group of animals in several different in-cage test systems, the animals have to be transferred between them. These limitations may be resolved by adding an automated sorting system between the home cage and the test arena, a strategy which has been successfully implemented for both an operant chamber (Winter and Schaefers, 2011) and an automated T-maze (Schaefers and Winter, 2011). In this setup, the animals are tagged using RFID-chips which can be identified by the sorting system which make sure that only one animal at a time enters the test arena. This approach can potentially be extended by connecting several cages to the same test arena or by connecting one cage to several test arenas.

The creation of fully automated tests which precisely mimic currently used tests is, however, an unconceivable task in most cases. Instead, thinking along new lines may be required to allow computers to make measurements and control the test situation. An elegant example is the automated assessment of grooming behavior, where the vibrations caused by the grooming are detected using a highly sensitive scale and translated to meaningful information using advanced algorithms (Chen et al., 2010). Attempting to mimic a human observer by relying on image analysis of video-captured grooming behavior is considerably more difficult, even though machine learning potentially makes this possible (Kabra et al., 2013). Furthermore, rodents are stressed by human contact and avoiding it would therefore be an example of refinement in the 3R (Replacement, Reduction and Refinement) system (Russell and Burch, 1959). Finally, regardless of how tempting it can be to use automated tests to speed up the work process, obtaining high throughput at the expense of high quality may result in unreliable data of limited value. Frequently it may, therefore, be advantageous to use a more labor intensive test to achieve a higher validity and increased chance of successful clinical translation.

VARIABILITY

Rodent behavior usually varies considerably from animal to animal, likely caused by a combination of genetic, environmental and experimental factors. Variability may be present even prior to the initiation of an experiment, for example, when using outbred animals which are meant to be genetically diverse. Outbred rat stocks are frequently used and include, for example, Sprague-Dawley, Long-Evans and Wistar rats. Although the mice used in medical research usually are inbred, the CD1 and Swiss Webster mice are outbred and it is reasonable to assume that the genetic diversity in outbred animals contributes to behavioral variability (Chia et al., 2005). All individuals in an inbred strain are meant to be genetically identical but may still differ in their minisatellite regions, short repetitive DNA sequences with highly polymorphic copy numbers, which potentially affect gene expression and behavior (Lathe, 2004). Since variability in preferred activities potentially alters life experiences, variability may also increase over time (Freund et al., 2013).

Maternal behavior also varies between individuals and high levels of pup licking and grooming leads to altered stress reactivity and improved performance in the MWM when the pups reach adulthood (Liu et al., 2000). Maternal behavior is also inherited in

a non-genomic fashion making it a possible source of variability affecting future generations (Meaney, 2001). Rodent embryos excrete sex hormones in the uterus where they are lined up like beads on a string. This means that the sex hormones an embryo is exposed to depends on the sex of the adjacent embryos, which potentially affects adult behavior (Ryan and Vandenbergh, 2002). Rodents have a well-developed immune system and obvious symptoms of post-surgery infections are rare. However, infection causes an immune response, which may affect experimental outcome and increase variability. For instance, infection following surgery was shown to exacerbate ischemic brain injury (Yousuf et al., 2013). The composition of commensal bacteria in the rodent intestine can also influence behavior adding yet another potential source of variability (Foster and McVey Neufeld, 2013).

The estrous cycle of female animals also potentially affects behavior (Meziane et al., 2007), although it may fortunately be measured fairly easily using vaginal swab cytology (Caligioni, 2009). Female mice previously unexposed to male mice coordinate their estrous cycle when exposed to male pheromones, the Whitten effect (Whitten, 1956). By introducing male urine-soaked bedding into a cage with female mice for a few days, experiments can be performed using individuals in the same estrous stage (Dalal et al., 2001). The commonly held belief that female animals have a higher variability due to their estrous cycle has recently been questioned after a meta-analysis found that males and females had similar levels of variability on a wide range of outcome measures (Prendergast et al., 2014).

Experimental disease models often require surgery or substance administration and the level of sustained injury or disease severity may vary from subject to subject (Kim et al., 2011). The dose of administrated drug and the resulting plasma level may also vary between individuals (Kääriäinen et al., 2011). Variability may also be introduced during the actual testing, for example by the order in which the animals in a cage are tested (Chesler et al., 2002), and test order should therefore be randomized. The effects of variability may be limited by measuring behavior before an intervention to allow animals with aberrant behavior to be excluded as well as enabling presentation of the results normalized to the baseline value. Since variability reduces the ability to detect significant differences, the identification and removal of sources of variability is important in the development and refinement of behavioral tests and testing procedures. Since decreasing variability enables smaller experimental group sizes it is not only practical but also ethical since it reduces the number of animals used (Russell and Burch, 1959).

EXPERIMENTAL DESIGN

The risk of bias is ever present in medical research and failure to take measures to avoid it resulted in overestimated treatment effects in preclinical stroke trials (Sena et al., 2007). Several guidelines outlining important measures to avoid bias have been prepared including the Camrades and ARRIVE guidelines (Kilkenny et al., 2010). Particularly, the assessment of animal behavior risks being subjective and should always be performed by a person blinded to the genotype, surgical intervention or drug treatment. Given the possibility for animal-experimenter interactions, it is

also prudent to perform actual testing in a blinded manner. If the person performing genotyping, surgery or treatment administration is also performing behavioral testing, blinding becomes a practical problem. If the animals are unlabeled prior to the intervention, it may be possible to have them labeled by another researcher afterwards. Replacement of cage cards may also be a possible way to conceal group assignment in some cases. RFID tags used to identify animals contain a number which is paired with the text displayed on the RFID reader, which means that blinding can be achieved by pairing the number with a new text. Group size is another important aspect of the experimental design and it has been suggested that neuroscience studies commonly use an insufficient number of animals. Power calculations are therefore recommended to determine adequate group sizes (Button et al., 2013).

Group assignment would not be an issue if all rodents in an experiment were identical. However, given the arguments in the section on variability (see *Variability* above), individual rodents are likely unique and randomization is required to avoid bias. However, unrestricted randomization may lead to, for instance, a cage containing only control animals, which introduces a risk of systematic errors. This risk can be avoided by dividing the study into blocks consisting of, for example, a single litter or a cage of animals and then randomly allocate the animals in each block to treatment groups (Festing and Altman, 2002). Given the importance of blinding and randomization, it is recommended to always describe it in scientific reports (Macleod et al., 2009).

Most rodent behavioral testing is done to understand human physiology and to find treatments for diseases, and experiments should therefore be designed to maximize the potential for successful translation of the results into patient benefit. It seems reasonable that a drug treatment which is effective only in single strain during a narrow age span is unlikely to translate into an effective clinical treatment. The chance for successful translation is on the other hand likely higher for treatments which are effective over a wide range of parameters like species, strain, sex, age and environmental factors. Evaluating treatments in the traditional way with a treatment group and a control group for each combination of parameters would unfortunately be prohibitively expensive and time consuming. By instead using a pair of animals for each set of parameters, where one animal per pair is given the evaluated treatment and the other serves as the control, robust effects could be detected without using large amounts of animals. Since performance is likely to be influenced by numerous factors, in particular the strain, the results would have to be evaluated using paired non-parametric statistics. This way of using a paired design to test animals of different species, strains, ages and sex within a single study has to our knowledge not previously been evaluated. It may, however, be a way to avoid the risk of obtaining idiosyncratic results caused by using inbred animals of a single age and sex which are housed and raised under identical conditions.

SENSORY MODALITY

To fully understand animal behavior, it is crucial to recognize that their view of the world can differ drastically from ours (Burn, 2008). Extreme examples include the sea turtles reliance on the

earth's magnetic field for navigation and the ability of various snake species to detect their prey using infrared radiation. Rodents perceive their environment primarily by using their excellent olfactory system and vomero-nasal organ and do not, unlike humans, rely heavily on vision (Ache and Young, 2005). Rodent vision is adapted to a nocturnal lifestyle with a dominance of rods in the retina. Furthermore, both rats and mice only have two types of cones limiting them to dichromatic vision, though one type of cones enables detection of ultraviolet light (Huberman and Niell, 2011). Rodents also have highly developed whiskers with large cortical representation, and actively move their whiskers over objects to examine them (Diamond et al., 2008). Another difference between human and rodent sensory systems is the ability of rodents to both detect ultrasound and use it for communication (Wöhr et al., 2008, 2011). It has also been suggested that mice can detect magnetic fields (Muheim et al., 2006) and that rats are able to echolocate (Rosenzweig et al., 1955).

Vision-based rodent behavioral test can closely mimic clinical test procedure which has been suggested to facilitate the translation of pre-clinical test results to the clinic (Horner et al., 2013), while others have cautioned against adopting an anthropocentric view (Wynne, 2004). Olfaction based test procedures are an alternative which may have better ethological validity (see *Validity* below). Such tests are still infrequently used relative to other types of behavioral tests, but there are for example several procedures where the animals dig for rewards in scented media (See Supplementary Table 1). One of them is the Dig Task (**Figure 2D**; Martens et al., 2013), although several other protocols are available for this versatile cognitive evaluation system, where reward location can be indicated not only by the scent and type of the digging medium but also the surface structure and location of the containers (Wood et al., 1999; Birrell and Brown, 2000; Gilmour et al., 2013).

Other factors to consider are that several commonly used mouse strains have restricted hearing abilities and that the barren environment in which laboratory animals are reared can impair sensory development (Cancedda et al., 2004; Turner et al., 2005). Accordingly, it is a desirable quality control for any behavioral test to verify that the animals have sufficient sensory capacity to detect the provided sensory cues. Such sensory evaluation is of particular importance when using genetically modified animals which may have unexpected sensory deficits as well as in disease models which may alter sensory functions. Fortunately, visual function can be assessed automatically using the Optomotry system (**Figure 2E**; Prusky et al., 2004), anosmia can be rapidly detected using the buried food test (Yang and Crawley, 2009) and hearing can be evaluated using pre-pulse inhibition of the acoustic startle response (Clause et al., 2011). The whisker sensory system can also be assessed using the Whisker nuisance task (**Figure 2F**; McNamara et al., 2010). Detailed assessment of sensory functions can also be performed using operant conditioning procedures (See Supplementary Table 1).

TEST VALIDITY

There are several types of validity which have to be considered when evaluating behavioral test results. The predictive validity is arguably one of the most important and is typically defined as

the ability of a rodent behavioral test to predict the effect of a drug in humans. Determining the predictive validity requires an existing clinical treatment which can be back-tested in the rodent behavioral test under evaluation (Schallert, 2006). An example is benzodiazepines, which are widely used to treat anxiety in patients and also reduce the extent of rodents' anxiety-like behavior in both the light-dark box test (Crawley and Goodwin, 1980) and the EPM (Pellow and File, 1986). Predictive validity can, however, frequently not be assessed due to the lack of available efficient treatments and construct validity (Strauss and Smith, 2009) may be used instead. This type of validity depends on whether the test measures the intended psychological construct or not and may for example be assessed by evaluating whether the involved neural systems and neurotransmitters are similar in rodents and humans.

In a test with high face validity it is obvious what the test is intended to measure. The rotarod test is for example easily and immediately recognized by most scientists as a motor function test. The ethological validity on the other hand depends on how well the test resembles natural animal behavior. It is for example easy to see how the use of predator odors to elicit fear responses resembles a situation which may be experienced by wild rodents as well. On the contrary, mice suspended by their tails are not observed in nature, although the tail suspension test is commonly used as an indicator of depression-like behavior (Cryan et al., 2005).

The internal validity, reproducibility, of a test depends on the similarity of results obtained from repeated experiments. The external validity, generalizability, of a test is determined by its ability to predict results in other strains and species, including humans. Standardization of behavioral tests and testing environment may increase internal validity although it may, on the other hand, decrease the external validity (Würbel, 2002; van der Staay and Steckler, 2002). Though standardization of test parameters may increase reproducibility, the choice of parameter setting is not always obvious and optimal settings for one disease model is not necessarily ideal in models of other diseases. Systematic variation of test parameters may on the other hand increase the external validity (Richter et al., 2010) and thus increases the chance of successful translation of the result to the clinic.

Using a test with a high predictive validity is likely the best choice if available. There is, however, a risk that the demonstrated validity only holds for other drugs of the same class. Relying on untried behavioral tests might on the other hand lead to the discovery of drugs acting via novel mechanisms. If fully validated tests are not available, a potential alternative to increase the reliability of the conclusions is to measure a trait using several different tests in the same domain. Tests and testing procedures may also be validated by verifying that known strain differences and/or drug effects can be replicated. In drug discovery research it is also important to separately consider the validity of the disease model to determine the reliability of the obtained test results.

ANIMAL HOUSING

Housing is crucial when testing behavior and may influence the results in several ways. For example, rodents within a cage will form a social hierarchy where a lower rank is associated with

increased levels of stress hormones and anxiety-like behavior, potentially adding to behavioral variability (Barnum et al., 2008; Costa-Pinto et al., 2009; Davis et al., 2009; Prendergast et al., 2014). Several tests are available to measure the social status including barbering, urination patterns, the tube confrontation test, the visible burrow system and the food contest task (Merlot et al., 2004; Wang et al., 2011; see Supplementary Table 1), which allows the formation of balanced experimental groups prior to an experiment. A further complication of group housing is the tendency of male mice to fight and wound each other (Emond et al., 2003), though this is less likely with littermates or among animals who have been sharing a cage since early age (Costa-Pinto et al., 2009). On the other hand, since both mice and rats are social animals, single housing may affect the results in several behavioral tests (Võikar et al., 2005). For the majority of studies, group housing is likely preferable since results obtained using socially deprived single housed animals may have a reduced validity.

Another aspect of rodent housing is the potential inability of a rodent cage to provide sufficient stimuli for normal development. Various forms of enrichment such as plastic tunnels, running wheels, rubber balls and nesting material can be used to mitigate this problem (Nithianantharajah and Hannan, 2006). The effects of enrichment are, however, complex and conditional knock out of a subtype of the NMDA receptor in the hippocampus was found to affect non-spatial memory and learning in mice reared in standard although not in enriched environments (Rampon et al., 2000). Rodent cages must also be cleaned at regular intervals, preferably on days without behavioral testing since the transfer to a new cage transiently increases stress hormones (Castelhano-Carlos and Baumans, 2009). Lightning conditions also needs to be considered since testing at different time points in the light-dark cycle may alter the outcome of behavioral testing (Hopkins and Bucci, 2010). Contrary to humans, rats and mice have their resting period during the day and a reversed light-dark cycle (Bertoglio and Carobrez, 2002) enables testing when the animals are active which is likely to be the best alternative in most behavioral studies.

PRACTICAL CONSIDERATIONS

Rodent behavioral tests vary considerably in the required work effort they require, which needs to be balanced against the reliability and importance of the obtained results. There are also ever more efficient methods available for the development of genetic models and antibodies as well as for the quantitation of mRNA synthesis and protein expression levels. If behavioral testing fails to match these improvements, it risks becoming a bottleneck in the drug development process (Brunner et al., 2002). If fully automated systems cannot be implemented, throughput can be increased by testing several animals in parallel, such as when using rotarod devices constructed with several lanes. Test equipment of a small size and reasonable cost also allows the use of several test chambers in parallel to make the work more efficient, a strategy applied for example in fear conditioning testing (Maren, 2008) and when using operant chambers. Relying on test procedures which can be performed by new personnel without extensive training or technical skills is also of practical value (Brooks and

Dunnett, 2009). The cost of laboratory premises can also be a substantial part of a research budget, which needs to be considered when planning to use physically large equipment like the MWM or the EPM. Furthermore, if tests are not used continuously, the possibility to disassemble and store test equipment can be of great practical value.

Behavioral tests must frequently be performed at a specific time depending on estrous cycle, time after injury or disease onset, or as part of a project schedule including several batches of animals subjected to several different tests. Such rigorous schemes are often needed, although problematic when faced with sick leave, vacations and holidays. One way to overcome such practical problems is to use fully automated tests or tests with low animal-experimenter interaction which may allow for the testing to be performed by a substitute investigator.

Cleaning of behavioral test equipment serves to avoid the spread of contagious diseases and to remove odor traces which otherwise might influence test results. Apart from obvious olfactory cues like feces and urine, rodents secrete fluids from their foot pads which might influence subsequent testing (Quatrale and Laden, 1968). Rats are also viewed as predators by mice (Quatrale and Laden, 1968) and if the two species share equipment, thorough cleaning is required. To facilitate the cleaning process the equipment should be devoid of narrow passages and corners, made of a readily cleanable material and preferably able to withstand sterilization procedures.

CONCLUDING REMARKS

Although genetics, electrophysiology and histology are very important tools for understanding underlying mechanisms of novel drug treatments, behavior represents the final output of the CNS and should be the basis for the final conclusion of preclinical evaluations of novel drugs or genetic modifications. Unfortunately, behavioral testing is very labor intensive as well as sensitive to environmental factors and the translation from preclinical to clinical studies has proven difficult in the fields of stroke, brain trauma, spinal cord injury, pain and Alzheimer's disease (Lo, 2008; Mao, 2009; Loane and Faden, 2010; Filli and Schwab, 2012; Savonenko et al., 2012). Novel behavioral tests may be needed to overcome this crucial problem but it is at least potentially mitigated by careful selection and execution of existing behavioral tests. Initial considerations prior to behavioral testing include selection of animal species, strain, gender and age as well as a determination of sample size, order of testing, type of housing and whether to use a reversed light-dark cycle or not. The best choice of a test depends not only on the scientific goals of the project, the intended measure and possible interpretations of the results, but also on practical and economical constraints, and may therefore differ between projects. In most cases, however, the ideal test is one which not only measures a clinically relevant trait and has a high validity, but also is practically feasible and ethically acceptable. The finding of such a test may be challenging but the process is potentially facilitated by the evaluation structure presented in **Table 1** and the listing of available tests in Supplementary Table 1. Finally, given the incomplete understanding of the brain and the limited treatment options for neurological disorders, rodent behavioral testing will likely continue to evolve

and to be an indispensable part of neuroscience for the foreseeable future.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Journal/10.3389/fnbeh.2014.00252/abstract

REFERENCES

Ache, B. W., and Young, J. M. (2005). Olfaction: diverse species, conserved principles. *Neuron* 48, 417–430. doi: 10.1016/j.neuron.2005.10.022

Balkaya, M., Kröber, J. M., Rex, A., and Endres, M. (2013). Assessing post-stroke behavior in mouse models of focal ischemia. *J. Cereb. Blood Flow Metab.* 33, 330–338. doi: 10.1038/jcbfm.2012.185

Barnum, C. J., Blandino, P., and Deak, T. (2008). Social status modulates basal IL-1 concentrations in the hypothalamus of pair-housed rats and influences certain features of stress reactivity. *Brain Behav. Immun.* 22, 517–527. doi: 10.1016/j.bbi.2007.10.004

Bertoglio, L. J., and Carobrez, A. P. (2002). Behavioral profile of rats submitted to session 1-session 2 in the elevated plus-maze during diurnal/nocturnal phases and under different illumination conditions. *Behav. Brain Res.* 132, 135–143. doi: 10.1016/s0166-4328(01)00396-5

Birrell, J. M., and Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J. Neurosci.* 20, 4320–4324.

Blizard, D. A., Weinheimer, V. K., Klein, L. C., Petrill, S. A., Cohen, R., and McClearn, G. E. (2006). "Return to home cage" as a reward for maze learning in young and old genetically heterogeneous mice. *Comp. Med.* 56, 196–201.

Braun, A. A., Skelton, M. R., Vorhees, C. V., and Williams, M. T. (2011). Comparison of the elevated plus and elevated zero mazes in treated and untreated male sprague-dawley rats: effects of anxiolytic and anxiogenic agents. *Pharmacol. Biochem. Behav.* 97, 406–415. doi: 10.1016/j.pbb.2010.09.013

Brooks, S. P., and Dunnett, S. B. (2009). Tests to assess motor phenotype in mice: a user's guide. *Nat. Rev. Neurosci.* 10, 519–529. doi: 10.1038/nrn2652

Bruce-Keller, A. J., Umberger, G., McFall, R., and Mattson, M. P. (1999). Food restriction reduces brain damage and improves behavioral outcome following excitotoxic and metabolic insults. *Ann. Neurol.* 45, 8–15. doi: 10.1002/1531-8249(199901)45:1<8::aid-art4>3.3.co;2-m

Brunner, D., Nestler, E., and Leahy, E. (2002). In need of high-throughput behavioral systems. *Drug Discov. Today* 7, S107–S112. doi: 10.1016/s1359-6446(02)02423-6

Burn, C. C. (2008). What is it like to be a rat? Rat sensory perception and its implications for experimental design and rat welfare. *Appl. Anim. Behav. Sci.* 112, 1–32. doi: 10.1016/j.applanim.2008.02.007

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Caligioni, C. S. (2009). Assessing reproductive status/stages in mice. *Curr. Protoc. Neurosci.* Appendix 4, Appendix 4I. doi: 10.1002/0471142301.nsa04is48

Cancedda, L., Putignano, E., Sale, A., Viegi, A., Berardi, N., and Maffei, L. (2004). Acceleration of visual system development by environmental enrichment. *J. Neurosci.* 24, 4840–4848. doi: 10.1523/jneurosci.0845-04.2004

Casarrubea, M., Sorbera, F., Santangelo, A., and Crescimanno, G. (2012). The effects of diazepam on the behavioral structure of the rat's response to pain in the hot-plate test: anxiolysis vs. pain modulation. *Neuropharmacology* 63, 310–321. doi: 10.1016/j.neuropharm.2012.03.026

Castelhano-Carlos, M. J., and Baumans, V. (2009). The impact of light, noise, cage cleaning and in-house transport on welfare and stress of laboratory rats. *Lab. Anim.* 43, 311–327. doi: 10.1258/la.2009.0080098

Chen, S.-K., Tvrdik, P., Peden, E., Cho, S., Wu, S., Spangrude, G., et al. (2010). Hematopoietic origin of pathological grooming in Hoxb8 mutant mice. *Cell* 141, 775–785. doi: 10.1016/j.cell.2010.03.055

Chesler, E. J., Wilson, S. G., Lariviere, W. R., Rodriguez-Zas, S. L., and Mogil, J. S. (2002). Influences of laboratory environment on behavior. *Nat. Neurosci.* 5, 1101–1102. doi: 10.1038/nn1102-1101

Chia, R., Achilli, F., Festing, M. F. W., and Fisher, E. M. C. (2005). The origins and uses of mouse outbred stocks. *Nat. Genet.* 37, 1181–1186. doi: 10.1038/ng1665

Clause, A., Nguyen, T., and Kandler, K. (2011). An acoustic startle-based method of assessing frequency discrimination in mice. *J. Neurosci. Methods* 200, 63–67. doi: 10.1016/j.jneumeth.2011.05.027

Clifton, P. G. (2000). Meal patterning in rodents: psychopharmacological and neuroanatomical studies. *Neurosci. Biobehav. Rev.* 24, 213–222. doi: 10.1016/s0149-7634(99)00074-3

Cook, M. N., Crounse, M., and Flaherty, L. (2002). Anxiety in the elevated zero-maze is augmented in mice after repeated daily exposure. *Behav. Genet.* 32, 113–118. doi: 10.1023/A:1015249706579

Costa-Pinto, F. A., Cohn, D. W. H., Sa-Rocha, V. M., Sa-Rocha, L. C., and Palermo-Neto, J. (2009). Behavior: a relevant tool for brain-immune system interaction studies. *Ann. N Y Acad. Sci.* 1153, 107–119. doi: 10.1111/j.1749-6632.2008.03961.x

Crabbe, J. C., and Morris, R. G. M. (2004). Festina lente: late-night thoughts on high-throughput screening of mouse behavior. *Nat. Neurosci.* 7, 1175–1179. doi: 10.1038/nn1343

Crabbe, J. C., Wahlsten, D., and Dudek, B. C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science* 284, 1670–1672. doi: 10.1126/science.284.5420.1670

Crawley, J., and Goodwin, F. K. (1980). Preliminary report of a simple animal behavior model for the anxiolytic effects of benzodiazepines. *Pharmacol. Biochem. Behav.* 13, 167–170. doi: 10.1016/0091-3057(80)90067-2

Cryan, J. F., Mombereau, C., and Vassout, A. (2005). The tail suspension test as a model for assessing antidepressant activity: review of pharmacological and genetic studies in mice. *Neurosci. Biobehav. Rev.* 29, 571–625. doi: 10.1016/j.neubiorev.2005.03.009

D'Ausilio, A. (2012). Arduino: a low-cost multipurpose lab equipment. *Behav. Res. Methods* 44, 305–313. doi: 10.3758/s13428-011-0163-z

Dalal, S. J., Estep, J. S., Valentin-Bon, I. E., and Jerse, A. E. (2001). Standardization of the whitten effect to induce susceptibility to Neisseria gonorrhoeae in female mice. *Contemp. Top. Lab. Anim. Sci.* 40, 13–17.

Davis, J. F., Krause, E. G., Melhorn, S. J., Sakai, R. R., and Benoit, S. C. (2009). Dominant rats are natural risk takers and display increased motivation for food reward. *Neuroscience* 162, 23–30. doi: 10.1016/j.neuroscience.2009.04.039

Deacon, R. (2012). Assessing burrowing, nest construction and hoarding in mice. *J. Vis. Exp.* e2607. doi: 10.3791/2607

Deacon, R. (2013). The successive alleys test of anxiety in mice and rats. *J. Vis. Exp.* e2705. doi: 10.3791/2705

de Visser, L., van den Bos, R., Kuurman, W. W., Kas, M. J. H., and Spruijt, B. M. (2006). Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes Brain Behav.* 5, 458–466. doi: 10.1111/j.1601-183x.2005.00181.x

Dews, P. B., Estes, W. K., Morse, W. H., Quine, W. V., and Herrnstein, R. J. (1994). Memorial minute for B. F. Skinner. *Behav. Anal.* 17, 2–5.

Diamond, M. E., von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., and Ahissar, E. (2008). "Where" and "what" in the whisker sensorimotor system. *Nat. Rev. Neurosci.* 9, 601–612. doi: 10.1038/nrn2411

Dunham, N. W., and Miya, T. S. (1957). A note on a simple apparatus for detecting neurological deficit in rats and mice. *J. Am. Pharm. Assoc. Am. Pharm. Assoc. (Baltim)* 46, 208–249. doi: 10.1002/jps.3030460322

Ekmark-Lewén, S., Lewén, A., Meyerson, B. J., and Hillered, L. (2010). The multivariate concentric square field test reveals behavioral profiles of risk taking, exploration and cognitive impairment in mice subjected to traumatic brain injury. *J. Neurotrauma* 27, 1643–1655. doi: 10.1089/neu.2009.0953

Emond, M., Faubert, S., and Perkins, M. (2003). Social conflict reduction program for male mice. *Contemp. Top. Lab. Anim. Sci.* 42, 24–26.

Engelmann, M., Ebner, K., Landgraf, R., and Wotjak, C. T. (2006). Effects of Morris water maze testing on the neuroendocrine stress response and intrahypothalamic release of vasopressin and oxytocin in the rat. *Horm. Behav.* 50, 496–501. doi: 10.1016/j.yhbeh.2006.04.009

Espejo, E. F., and Mir, D. (1994). Differential effects of weekly and daily exposure to the hot plate on the rat's behavior. *Physiol. Behav.* 55, 1157–1162. doi: 10. 1016/0031-9384(94)90404-9

Festing, M. F. W., and Altman, D. G. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J.* 43, 244–258. doi: 10. 1093/ilar.43.4.244

Filli, L., and Schwab, M. E. (2012). The rocky road to translation in spinal cord repair. *Ann. Neurol.* 72, 491–501. doi: 10.1002/ana.23630

Fonio, E., Golani, I., and Benjamini, Y. (2012). Measuring behavior of animal models: faults and remedies. *Nat. Methods* 9, 1167–1170. doi: 10.1038/nmeth. 2252

Foster, J. A., and McVey Neufeld, K.-A. (2013). Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci.* 36, 305–312. doi: 10.3410/f. 718013229.793477291

Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., et al. (2013). Emergence of individuality in genetically identical mice. *Science* 340, 756–759. doi: 10.1126/science.1235294

Gilmour, G., Arguello, A., Bari, A., Brown, V. J., Carter, C., Floresco, S. B., et al. (2013). Measuring the construct of executive control in schizophrenia: defining and validating translational animal paradigms for discovery research. *Neurosci. Biobehav. Rev.* 37, 2125–2140. doi: 10.1016/j.neubiorev.2012.04.006

Gray, P. H. (1973). Comparative psychology and ethology: a saga of twins reared apart. *Ann. N Y Acad. Sci.* 223, 49–53. doi: 10.1111/j.1749-6632.1973. tb41417.x

Guccione, L., Djouma, E., Penman, J., and Paolini, A. G. (2013). Calorie restriction inhibits relapse behaviour and preference for alcohol within a two-bottle free choice paradigm in the alcohol preferring (iP) rat. *Physiol. Behav.* 110C–111C, 34–41. doi: 10.1016/j.physbeh.2012.11.011

Hall, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J. Comp. Psychol.* 18, 385–403. doi: 10.1037/h0071444

Harrison, F. E., Hosseini, A. H., and McDonald, M. P. (2009). Endogenous anxiety and stress responses in water maze and Barnes maze spatial memory tasks. *Behav. Brain Res.* 198, 247–251. doi: 10.1016/j.bbr.2008.10.015

Hopkins, M. E., and Bucci, D. J. (2010). Interpreting the effects of exercise on fear conditioning: the influence of time of day. *Behav. Neurosci.* 124, 868–872. doi: 10.1037/a0021200

Horner, A. E., Heath, C. J., Hvoslef-Eide, M., Kent, B. A., Kim, C. H., Nilsson, S. R. O., et al. (2013). The touchscreen operant platform for testing learning and memory in rats and mice. *Nat. Protoc.* 8, 1961–1984. doi: 10.1038/nprot. 2013.122

Huberman, A. D., and Niell, C. M. (2011). What can mice tell us about how vision works? *Trends Neurosci.* 34, 464–473. doi: 10.1016/j.tins.2011.07.002

Hurst, J. L., and West, R. S. (2010). Taming anxiety in laboratory mice. *Nat. Methods* 7, 825–826. doi: 10.1038/nmeth.1500

Jacob, H. J., Lazar, J., Dwinell, M. R., Moreno, C., and Geurts, A. M. (2010). Gene targeting in the rat: advances and opportunities. *Trends Genet.* 26, 510–518. doi: 10.1016/j.tig.2010.08.006

Johansen, J. P., Cain, C. K., Ostroff, L. E., and LeDoux, J. E. (2011). Molecular mechanisms of fear learning and memory. *Cell* 147, 509–524. doi: 10.1016/j.cell. 2011.10.009

Jones, N. (2012). Science in three dimensions: the print revolution. *Nature* 487, 22–23. doi: 10.1038/487022a

Kääriäinen, T. M., Käenmäki, M., Forsberg, M. M., Oinas, N., Tammimäki, A., and Männistö, P. T. (2011). Unpredictable rotational responses to L-dopa in the rat model of Parkinson's disease: the role of L-dopa pharmacokinetics and striatal

dopamine depletion. *Basic Clin. Pharmacol. Toxicol.* 110, 162–170. doi: 10. 1111/j.1742-7843.2011.00782.x. [Epub ahead of print].

Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S., and Branson, K. (2013). JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat. Methods* 10, 64–67. doi: 10.1038/nmeth.2281

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., and Altman, D. G. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 8:e1000412. doi: 10.1371/journal.pbio. 1000412

Kim, D.-E., Kim, J.-Y., Nahrendorf, M., Lee, S.-K., Ryu, J. H., Kim, K., et al. (2011). Direct thrombus imaging as a means to control the variability of mouse embolic infarct models: the role of optical molecular imaging. *Stroke* 42, 3566–3573. doi: 10.1161/strokeaha.111.629428

Krackow, S., Vannoni, E., Codita, A., Mohammed, A. H., Cirulli, F., Branchi, I., et al. (2010). Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage. *Genes Brain Behav.* 9, 722–731. doi: 10.1111/j.1601-183x.2010.00606.x

Krzywinski, M., and Altman, N. (2013). Points of significance: importance of being uncertain. *Nat. Methods* 10, 809–810. doi: 10.1038/nmeth.2613

Lathe, R. (2004). The individuality of mice. *Genes Brain Behav.* 3, 317–327. doi: 10. 1111/j.1601-183X.2004.00083.x

Lenzlinger, P. M., Shimizu, S., Marklund, N., Thompson, H. J., Schwab, M. E., Saatman, K. E., et al. (2005). Delayed inhibition of Nogo-A does not alter injury-induced axonal sprouting but enhances recovery of cognitive function following experimental traumatic brain injury in rats. *Neuroscience* 134, 1047–1056. doi: 10.1016/j.neuroscience.2005.04.048

Lewejohann, L., Hoppmann, A. M., Kegel, P., Kritzler, M., Krüger, A., and Sachser, N. (2009). Behavioral phenotyping of a murine model of Alzheimer's disease in a seminaturalistic environment using RFID tracking. *Behav. Res. Methods* 41, 850–856. doi: 10.3758/BRM.41.3.850

Liu, D., Diorio, J., Day, J. C., Francis, D. D., and Meaney, M. J. (2000). Maternal care, hippocampal synaptogenesis and cognitive development in rats. *Nat. Neurosci.* 3, 799–806. doi: 10.1038/77702

Lo, E. H. (2008). Experimental models, neurovascular mechanisms and translational issues in stroke research. *Br. J. Pharmacol.* 153(Suppl. 1), S396–S405. doi: 10.1038/sj.bjp.0707626

Loane, D. J., and Faden, A. I. (2010). Neuroprotection for traumatic brain injury: translational challenges and emerging therapeutic strategies. *Trends Pharmacol. Sci.* 31, 596–604. doi: 10.1016/j.tips.2010.09.005

Macleod, M. R., Fisher, M., O'Collins, V., Sena, E. S., Dirnagl, U., Bath, P. M. W., et al. (2009). Good laboratory practice: preventing introduction of bias at the bench. *Stroke* 40, e50–e52. doi: 10.1161/STROKEAHA.108.525386

Mandillo, S., Tucci, V., Hölter, S. M., Meziane, H., Banchaabouchi, M. A., Kallnik, M., et al. (2008). Reliability, robustness and reproducibility in mouse behavioral phenotyping: a cross-laboratory study. *Physiol. Genomics* 34, 243–255. doi: 10. 1152/physiolgenomics.90207.2008

Mao, J. (2009). Translational pain research: achievements and challenges. *J. Pain* 10, 1001–1011. doi: 10.1016/j.jpain.2009.06.002

Maren, S. (2008). Pavlovian fear conditioning as a behavioral assay for hippocampus and amygdala function: cautions and caveats. *Eur. J. Neurosci.* 28, 1661–1666. doi: 10.1111/j.1460-9568.2008.06485.x

Martens, K. M., Vonder Haar, C., Hutsell, B. A., and Hoane, M. R. (2013). The dig task: a simple scent discrimination reveals deficits following frontal brain damage. *J. Vis. Exp.* e50033. doi: 10.3791/50033

Martin, B., Ji, S., Maudsley, S., and Mattson, M. P. (2010). "Control" laboratory rodents are metabolically morbid: why it matters. *Proc. Natl. Acad. Sci. U S A* 107, 6127–6133. doi: 10.1073/pnas.0912955107

McCall, R. B., Lester, M. L., and Corter, C. M. (1969). Caretaker effect in rats. *Dev. Psychol.* 1, 771. doi: 10.1037/h0028199

McCoy, G. W. (1909). A preliminary report on tumors found in wild rats. *J. Med. Res.* 21, 285–296.

McNamara, K. C. S., Lisembee, A. M., and Lifshitz, J. (2010). The whisker nuisance task identifies a late-onset, persistent sensory sensitivity in diffuse brain-injured rats. *J. Neurotrauma* 27, 695–706. doi: 10.1089/neu.2009.1237

Meaney, M. J. (2001). Maternal care, gene expression and the transmission of individual differences in stress reactivity across generations. *Annu. Rev. Neurosci.* 24, 1161–1192. doi: 10.1146/annurev.neuro.24.1.1161

Merlot, E., Moze, E., Bartolomucci, A., Dantzer, R., and Neveu, P. J. (2004). The rank assessed in a food competition test influences subsequent reactivity

to immune and social challenges in mice. *Brain Behav. Immun.* 18, 468–475. doi: 10.1016/j.bbi.2003.11.007

Meyerson, B. J., Augustsson, H., Berg, M., and Roman, E. (2006). The Concentric square field: a multivariate test arena for analysis of explorative strategies. *Behav. Brain Res.* 168, 100–113. doi: 10.1016/j.bbr.2005.10.020

Meziane, H., Ouagazzal, A. M., Aubert, L., Wietrzych, M., and Krezel, W. (2007). Estrous cycle effects on behavior of C57BL/6J and BALB/cByJ female mice: implications for phenotyping strategies. *Genes Brain Behav.* 6, 192–200. doi: 10.1111/j.1601-183x.2006.00249.x

Moscarello, J. M., and LeDoux, J. E. (2013). Active avoidance learning requires prefrontal suppression of amygdala-mediated defensive reactions. *J. Neurosci.* 33, 3815–3823. doi: 10.1523/JNEUROSCI.2596-12.2013

Muheim, R., Edgar, N. M., Sloan, K. A., and Phillips, J. B. (2006). Magnetic compass orientation in C57BL/6J mice. *Learn. Behav.* 34, 366–373. doi: 10.3758/bf03193201

Nadler, J. J., Moy, S. S., Dold, G., Trang, D., Simmons, N., Perez, A., et al. (2004). Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav.* 3, 303–314. doi: 10.1111/j.1601-183x.2004.00071.x

Neelakantan, H., and Walker, E. A. (2012). Temperature-dependent enhancement of the antinociceptive effects of opioids in combination with gabapentin in mice. *Eur. J. Pharmacol.* 686, 55–59. doi: 10.1016/j.ejphar.2012.04.042

Nithiananantharajah, J., and Hannan, A. J. (2006). Enriched environments, experience-dependent plasticity and disorders of the nervous system. *Nat. Rev. Neurosci.* 7, 697–709. doi: 10.1038/nrn1970

O'Connor, C., Heath, D. L., Cernak, I., Nimmo, A. J., and Vink, R. (2003). Effects of daily versus weekly testing and pre-training on the assessment of neurologic impairment following diffuse traumatic brain injury in rats. *J. Neurotrauma* 20, 985–993. doi: 10.1089/089771503770195830

Ohl, F., and Keck, M. E. (2003). Behavioural screening in mutagenised mice—in search for novel animal models of psychiatric disorders.. *Eur. J. Pharmacol.* 480, 219–228. doi: 10.1016/j.ejphar.2003.08.108

Pellow, S., and File, S. E. (1986). Anxiolytic and anxiogenic drug effects on exploratory activity in an elevated plus-maze: a novel test of anxiety in the rat. *Pharmacol. Biochem. Behav.* 24, 525–529. doi: 10.1016/0091-3057(86)90552-6

Petrovich, G. D., Ross, C. A., Mody, P., Holland, P. C., and Gallagher, M. (2009). Central, but not basolateral, amygdala is critical for control of feeding by aversive learned cues. *J. Neurosci.* 29, 15205–15212. doi: 10.1523/JNEUROSCI.3656-09.2009

Porsolt, R. D., Bertin, A., and Jalfre, M. (1977). Behavioral despair in mice: a primary screening test for antidepressants. *Arch. Int. Pharmacodyn. Ther.* 229, 327–336.

Prendergast, B. J., Onishi, K. G., and Zucker, I. (2014). Female mice liberated for inclusion in neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* 40, 1–5. doi: 10.1016/j.neubiorev.2014.01.001

Prusky, G. T., Alam, N. M., Beekman, S., and Douglas, R. M. (2004). Rapid quantification of adult and developing mouse spatial vision using a virtual opto-motor system. *Invest. Ophthalmol. Vis. Sci.* 45, 4611–4616. doi: 10.1167/iovs.04-0541

Quatrale, R. P., and Laden, K. (1968). Solute and water secretion by the eccrine sweat glands of the rat. *J. Invest. Dermatol.* 51, 502–504. doi: 10.1038/jid.1968.81

Ramos, A., Pereira, E., Martins, G. C., Wehrmeister, T. D., and Izídio, G. S. (2008). Integrating the open field, elevated plus maze and light/dark box to assess different types of emotional behaviors in one single trial. *Behav. Brain Res.* 193, 277–288. doi: 10.1016/j.bbr.2008.06.007

Rampon, C., Tang, Y. P., Goodhouse, J., Shimizu, E., Kyin, M., and Tsien, J. Z. (2000). Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice. *Nat. Neurosci.* 3, 238–244. doi: 10.1038/72945

Richter, S. H., Garner, J. P., Auer, C., Kunert, J., and Würbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nat. Methods* 7, 167–168. doi: 10.1038/nmeth0310-167

Rosenzweig, M. R., Riley, D. A., and Krech, D. (1955). Evidence for echolocation in the rat. *Science* 121:600. doi: 10.1126/science.121.3147.600

Rousson, V. (2011). Assessing inter-rater reliability when the raters are fixed: two concepts and two estimates. *Biom. J.* 53, 477–490. doi: 10.1002/bimj.201000066

Russell, W. M. S., and Burch, R. L. (1959). *The Principles of Humane Experimental Technique.* Wheathampstaed: Universities Federation for Animal Welfare.

Ryan, B. C., and Vandenbergh, J. G. (2002). Intrauterine position effects. *Neurosci. Biobehav. Rev.* 26, 665–678. doi: 10.1016/s0149-7634(02)00038-6

Samoilov, V. O. (2007). Ivan petrovich pavlov(1849–1936). *J. Hist. Neurosci.* 16, 74–89. doi: 10.1080/09647040600793232

Savonenko, A. V., Melnikova, T., Hiatt, A., Li, T., Worley, P. F., Troncoso, J. C., et al. (2012). Alzheimer's therapeutics: translation of preclinical science to clinical drug development. *Neuropsychopharmacology* 37, 261–277. doi: 10.1038/npp.2011.211

Schaefers, A. T. U., and Winter, Y. (2011). Rapid task acquisition of spatial-delayed alternation in an automated T-maze by mice. *Behav. Brain Res.* 225, 56–62. doi: 10.1016/j.bbr.2011.06.032

Schallert, T. (2006). Behavioral tests for preclinical intervention assessment. *NeuroRx* 3, 497–504. doi: 10.1016/j.nurx.2006.08.001

Schallert, T., Fleming, S. M., Leasure, J. L., Tillerson, J. L., and Bland, S. T. (2000). CNS plasticity and assessment of forelimb sensorimotor outcome in unilateral rat models of stroke, cortical ablation, parkinsonism and spinal cord injury. *Neuropharmacology* 39, 777–787. doi: 10.1016/s0028-3908(00)00005-8

Schallert, T., Woodlee, M. T., and Fleming, S. M. (2002). "Disentangling multiple types of recovery from brain injury recovery of function," in *Pharmacology of Cerebral Ischemia*, eds J. Krigelstein and S. Klumpp (Stuttgart: Medpharm Scientific Publishers), 201–216.

Schallert, T., Woodlee, M. T., and Fleming, S. M. (2003). Experimental focal ischemic injury: behavior-brain interactions and issues of animal handling and housing. *ILAR J.* 44, 130–143. doi: 10.1093/ilar.44.2.130

Schmitt, U., and Hiemke, C. (1998). Strain differences in open-field and elevated plus-maze behavior of rats without and with pretest handling. *Pharmacol. Biochem. Behav.* 59, 807–811. doi: 10.1016/s0091-3057(97)00502-9

Sena, E., van der Worp, H. B., Howells, D., and Macleod, M. (2007). How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci.* 30, 433–439. doi: 10.1016/j.tins.2007.06.009

Shemesh, Y., Sztainberg, Y., Forkosh, O., Shlapobersky, T., Chen, A., and Schneidman, E. (2013). High-order social interactions in groups of mice. *Elife* 2:e00759. doi: 10.7554/eLife.00759

Shepherd, J. K., Grewal, S. S., Fletcher, A., Bill, D. J., and Dourish, C. T. (1994). Behavioural and pharmacological characterisation of the elevated "zero-maze" as an animal model of anxiety. *Psychopharmacology (Berl)* 116, 56–64. doi: 10.1007/bf02244871

Shimamura, M., Kuratani, K., and Kinoshita, M. (2007). A new automated and high-throughput system for analysis of the forced swim test in mice based on magnetic field changes. *J. Pharmacol. Toxicol. Methods* 55, 332–336. doi: 10.1016/j.vascn.2006.11.003

Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420

Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., et al. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* 11, 629–632. doi: 10.1038/nmeth.2935

Staples, L. G. (2010). Predator odor avoidance as a rodent model of anxiety: learning-mediated consequences beyond the initial exposure. *Neurobiol. Learn. Mem.* 94, 435–445. doi: 10.1016/j.nlm.2010.09.009

Strauss, M. E., and Smith, G. T. (2009). Construct validity: advances in theory and methodology. *Annu. Rev. Clin. Psychol.* 5, 1–25. doi: 10.1146/annurev.clinpsy.032408.153639

Steensma, D. P., Kyle, R. A., and Shampo, M. A. (2010). Abbie Lathrop, the "mouse woman of Granby": rodent fancier and accidental genetics pioneer. *Mayo Clin. Proc.* 85:e83. doi: 10.4065/mcp.2010.0647

Thierry, B. (2010). Darwin as a student of behavior. *C. R. Biol.* 333, 188–196. doi: 10.1016/j.crvi.2009.12.007

Turner, J. G., Parrish, J. L., Hughes, L. F., Toth, L. A., and Caspary, D. M. (2005). Hearing in laboratory animals: strain differences and nonauditory effects of noise. *Comp. Med.* 55, 12–23.

van der Staay, F. J., and Steckler, T. (2002). The fallacy of behavioral phenotyping without standardisation. *Genes Brain Behav.* 1, 9–13. doi: 10.1046/j.1601-1848.2001.00007.x

van Driel, K. S., and Talling, J. C. (2005). Familiarity increases consistency in animal tests. *Behav. Brain Res.* 159, 243–245. doi: 10.1016/j.bbr.2004.11.005

Võikar, V., Polus, A., Vasar, E., and Rauvala, H. (2005). Long-term individual housing in C57BL/6J and DBA/2 mice: assessment of behavioral consequences. *Genes Brain Behav.* 4, 240–252. doi: 10.1111/j.1601-183x.2004.00106.x

Vorhees, C. V., and Williams, M. T. (2006). Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nat. Protoc.* 1, 848–858. doi: 10.1038/nprot.2006.116

Wahlsten, D., Metten, P., and Crabbe, J. C. (2003a). A rating scale for wildness and ease of handling laboratory mice: results for 21 inbred strains tested in two laboratories. *Genes Brain Behav.* 2, 71–79. doi: 10.1034/j.1601-183x.2003.00012.x

Wahlsten, D., Metten, P., Phillips, T. J., Boehm, S. L., Burkhart-Kasch, S., Dorow, J., et al. (2003b). Different data from different labs: lessons from studies of gene-environment interaction. *J. Neurobiol.* 54, 283–311. doi: 10.1002/neu.10173

Wahlsten, D., Rustay, N. R., Metten, P., and Crabbe, J. C. (2003c). In search of a better mouse test. *Trends Neurosci.* 26, 132–136. doi: 10.1016/s0166-2236(03)00033-x

Walczak, J.-S., and Beaulieu, P. (2006). Comparison of three models of neuropathic pain in mice using a new method to assess cold allodynia: the double plate technique. *Neurosci. Lett.* 399, 240–244. doi: 10.1016/j.neulet.2006.01.058

Wang, F., Zhu, J., Zhu, H., Zhang, Q., Lin, Z., and Hu, H. (2011). Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science* 334, 693–697. doi: 10.1126/science.1209951

Whishaw, I. Q., Coles, B. L., and Bellerive, C. H. (1995). Food carrying: a new method for naturalistic studies of spontaneous and forced alternation. *J. Neurosci. Methods* 61, 139–143. doi: 10.1016/0165-0270(95)00035-s

Whitten, W. K. (1956). Modification of the oestrous cycle of the mouse by external stimuli associated with the male. *J. Endocrinol.* 13, 399–404. doi: 10.1677/joe.0.0130399

Winter, Y., and Schaefers, A. T. U. (2011). A sorting system with automated gates permits individual operant experiments with mice from a social home cage. *J. Neurosci. Methods* 196, 276–280. doi: 10.1016/j.jneumeth.2011.01.017

Wöhr, M., Houx, B., Schwarting, R. K. W., and Spruijt, B. (2008). Effects of experience and context on 50-kHz vocalizations in rats. *Physiol. Behav.* 93, 766–776. doi: 10.1016/j.physbeh.2007.11.031

Wöhr, M., Roullet, F. I., and Crawley, J. N. (2011). Reduced scent marking and ultrasonic vocalizations in the BTBR T+tf/J mouse model of autism. *Genes. Brain. Behav.* 10, 35–43. doi: 10.1111/j.1601-183x.2010.00582.x

Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature* 397, 613–616. doi: 10.1038/17605

Würbel, H. (2002). Behavioral phenotyping enhanced—beyond (environmental) standardization. *Genes. Brain. Behav.* 1, 3–8. doi: 10.1046/j.1601-1848.2001.00006.x

Wynne, C. D. L. (2004). The perils of anthropomorphism. *Nature* 428:606. doi: 10.1038/428606a

Yang, M., and Crawley, J. N. (2009). Simple behavioral assessment of mouse olfaction. *Curr. Protoc. Neurosci.* Chapter 8:Unit 8.24. doi: 10.1002/0471142301.ns0824s48

Yezierski, R. P., and Vierck, C. J. (2011). Should the hot-plate test be reincarnated? *J. Pain* 12, 936–937; author reply938–939. doi: 10.1016/j.jpain.2011.05.003

Young, J. W., Meves, J. M., and Geyer, M. A. (2013). Nicotinic agonist-induced improvement of vigilance in mice in the 5-choice continuous performance test. *Behav. Brain Res.* 240, 119–133. doi: 10.1016/j.bbr.2012.11.028

Yousuf, S., Atif, F., Sayeed, I., Wang, J., and Stein, D. G. (2013). Post-stroke infections exacerbate ischemic brain injury in middle-aged rats: immunomodulation and neuroprotection by progesterone. *Neuroscience* 239, 92–102. doi: 10.1016/j.neuroscience.2012.10.017

Yu, Z. F., and Mattson, M. P. (1999). Dietary restriction and 2-deoxyglucose administration reduce focal ischemic brain damage and improve behavioral outcome: evidence for a preconditioning mechanism. *J. Neurosci. Res.* 57, 830–839. doi: 10.1002/(sici)1097-4547(19990915)57:6<830::aid-jnr8>3.3.co;2-u