Title: Testing cognitive functions in rodent disease models: present pitfalls and future perspectives

Author: Heikki Tanila

PII: S0166-4328(17)30634-4
DOI: http://dx.doi.org/doi:10.1016/j.bbr.2017.05.040
Reference: BBR 10889

To appear in: *Behavioural Brain Research*

Received date: 12-4-2017
Revised date: 14-5-2017
Accepted date: 15-5-2017

Please cite this article as: Tanila Heikki.Testing cognitive functions in rodent disease models: present pitfalls and future perspectives.*Behavioural Brain Research* http://dx.doi.org/10.1016/j.bbr.2017.05.040

Testing cognitive functions in rodent disease models: present pitfalls and future perspectives

Heikki Tanila, A. I. Virtanen Institute, University of Eastern Finland, Kuopio, Finland

Heikki Tanila, MD, PhD
Professor of Translational Neuroscience
A. I. Virtanen Institute
University of Eastern Finland
Neulaniementie 2, 70210 Kuopio
FINLAND

email: Heikki.Tanila@uef.fi
tel: +358-40-3552084

Abstract

Testing of cognitive functions in rodent disease models constitutes a substantial sector of behavioral neuroscience. It is most often needed in phenotyping genetically modified new rodent (usually mouse) lines or in preclinical testing of cognitive effects of new CNS drugs. This review concerns present pitfalls and future perspectives in this large field, with an emphasis on memory testing in CNS disease models and their preclinical drug testing. It is important to realize that no behavioral test is specific for a single cognitive domain. There are numerous noncognitive factors that may lead to impaired performance in most widely applied memory tasks. It is important to rule these out by applying a battery of test that should include at least tests for motor functions, spontaneous activity and anxiety besides cognitive aspects. In addition, considering and reporting all task-relavant details will help to resolve the common problem that certain behavioral findings cannot be reproduced by other laboratories. More collaboration between molecular and behavioral neuroscience laboratories and systematic training of young neuroscientist on behavioral techniques will help ensure quality of behavioral studies in the future.

Key words: Genetically modified mice, neuropharmacology, learning, memory, cognition, behavioral methods

1. Introduction

Testing of cognitive functions in rodent disease models constitutes a substantial sector of behavioral neuroscience. Yet, only a small fraction of cognitive testing of rodents aims at detailed understanding of underlying cognitive processes. It is much more often needed in the analysis and phenotyping of genetically modified new rodent (usually mouse) lines, and especially in the validation of a subset of those lines as CNS disease models. In addition, the preclinical testing of cognitive effects of new CNS drugs with established rodent models involves thousands of scientists and students worldwide. This review concerns present pitfalls and future perspectives in this large field, with an emphasis on memory testing CNS phenotyping of disease models and their preclinical drug testing.

2. Behavioral test results are often difficult to reproduce

There have been repeated concerns for poor reproducibility of preclinical behavioral studies and their translation to clinics [1, 2, 3] as well as of behavioral phenotyping studies of mutant mouse lines [4]. This is indeed a big problem, and especially true for testing of cognitive functions. One obvious solution to this problem has been attemps to standardize the test environment and procedures. Several EU funded projects have been run to improve standardization of behavioral test procedures, and detailed description of the most common behavioral test protocols can be easily found in the literature. On the other hand, the reason for poor test reproducibility is not only in details of the test procedure itself. When it comes to work in mice, one well established source of variation are big differences between widely used inbred mouse strains in their cognitive performance [5]. However, a pioneering study elegantly demonstrated that even in a highly standardized test, such as the elevated plusmaze, differences between laboratories can be even bigger than those between different mouse strains [6]. This raises the concern that issues like conditions of husbandry, the handling history of animals, the broader testing environment outside the apparatus, and the human-animal interactions, which are usually not standardized or not even reported, make a significant contribution to the test outcome [4].

One confounding factor, indeed, is that the conditions where cognitive testing are done in rodents may vary a great deal [2, 3]. In one extreme are laboratories that have a long history of behavioral testing and staff devoted to this field of science. They have solid experience in the behavioral

repertoire of mice or rats and a large reference material to immediately see if some results deviate too much from the expected. These laboratories are often asked to assist research groups with the focus in molecular or cellular aspects rather than behavioral neuroscience. In the other extreme are molecular biology laboratories that need yet some *in vivo* evidence for their almost ready-to-go submission for a high-impact journal and assign one laboratory member, usually a student, to run a standard cognitive test as described in the literature, assuming it runs like an ELISA kit according to the manufacturer's instructions. The problem arises when both laboratories report the study methods only briefly often referring to a previous publication, which usually also lacks many critical details about the test. The test outcome is highly dependent on small but important details that are seldom reported, such as handling history of the animals, illumination, environmental landmarks, noise level in the test room, number and gender of testers, cleaning of the test arena between trials and/or animals, and location of the tester and waiting animals in relation to the animal to be tested. Thus two ostensibly similar test descriptions may conceal numerous factors that differ between laboratories. In the case of a devoted behavioral test laboratory all these details may be have been well taken into account, and there may be detailed description of the test conduction internally in the laboratory, but in the worst case, these details have been ignored; yet this will not show up in the brief report.

3. Behavioral tests are seldom specific to a certain cognitive function

A common fallacy, especially among scientists who are not experienced in behavioral testing, is that a particular common behavioral test specifically measures a specific cognitive function. However, this is seldom the case, if ever. Let's take some most widely used cognitive tests as examples.

3.1. Passive avoidance. This is a classic test for long-term memory. Since it is based on single-trial learning and requires no pretraining, it is a high-throughput behavioral task, which explains its continued popularity, especially in behavioral pharmacology. A curious feature of this task is that animal tested immediately after the experience show little memory, and that memory retention increases with time [7]. Therefore, the established practise is to take the animal back to its home cage after the exposure and return it for testing the next day. In the test, the animal is placed in a well-lit side of a two-compartment box and offered access through a narrow hole to a closed dark compartment. When the animal enters the dark side following its natural instinct, the opening is closed and a mild foot shock is given. Then the animal is taken to its home cage and returned to the arena after a delay, usually 24 h. The latency to enter is taken as a measure of the memory strength. This may be true, but differences between animals may also arise from a number of noncognitive reasons.

For instance, genetic or drug induced hyperactivity or reduced anxiety will lead to shortened latency to enter, while hypoactivity and increased anxiety will ostensibly appear as superior memory. In addition, the groups may also differ in their sensitivity to the shock itself. The task is very powerful when used to test drug effects on memory consolidation. In this case, the drug administration follows the training session, and thus no significant concentration of the drug is present either during the training or the test phase. It is also an outstanding test to define whether a drug influences memory encoding, consolidation or retrieval based on the time of the drug administration. However, when used for comparing effects of chronic drug treatments or genetically modified mouse lines, the results need to be interpreted cautiously.

3.2. Novel object recognition. This is a relatively new test that has gained popularity, because it is based on animals' natural preferences instead of a learned response. Thus, it does not require food or water deprivation for motivation or a lengthy training. Nevertheless, it does require thorough habituation to the test environment. The test is based on rodents' natural curiosity to explore their immediate environment. The tests consists of a sample phase and a test phase. First, the animal is given 10-15 min to freely explore two identical objects (here object A) in the test arena, after which it will be returned to the home cage. After a variable delay (from 5 min to 48 h) the animal is returned to the test arena where one of the sample object has been replaced by a new one (here object B). During the test phase that usually lasts 5-10 min, the time in contact with the objects is measured and a preference index is calculated. The assumption is that the animal would spend more time exploring the novel object than the familial one, while equal time on exploration of both objects suggests that the animal does not remember seeing the sample object. One key question in interpreting the test result is whether the animals prefer one object because it is novel or because they find it for some other reason more appealing. A devoted behavioral test laboratory might have spent months in pilot experiments in order to find objects toward which the control mice or rats show the same amount of general interest. In addition, to ensure that the critical feature is novelty, one should use object A as the sample for half of the animals and object B for the other half. If actually done, this is seldom reported. In contrast, a laboratory that has little experience in behavioral testing, may take two objects that appear funtional in the test and never even considers to test the true novelty effect.

The novel object recognition test is relatively insensitive to motor impairment. However, some unspecific factors may still influence the test outcome. For instance, 129/Sv mice, a popular strain for gene-targeting of embryonal stem cells, are in general hypoactive and therefore more impaired that other common mouse straing in this task [8]. In essence, the novel object recogntion test is based on the balance between curiosity and neophobia. The handling history, familiarity with the test areana,

and its size and illumination determine which factor is the main determinant of the animals' choice. Especially genetically modified mouse strains are often reported to show significantly lower than change preference to the novel object. Such a negative preference may simply reflect increased anxiety leading to avoidance of the novel object.

Since the test was adopted to the rodent behavioral test battery from studies in human infants and monkeys, the general assumption is that also the rodents mainly discriminate the objects based on their vision. However, as described by Leinonen & Tanila in this issue, the rodent vision is not particularly well suited for observing object details, and therefore rodents more likely explore the objects with their whiskers or use their very sensitive olfaction. Thus objects with similar surface texture but different color are not particularly appropriate test objects. Researchers usually want to keep the object recognition task separate from odor recognition and try to minimize the use of odor cues in this task. Therefore, good laboratory practise is to present an identical sample copy rather than one of the original samples in the test phase to avoid the possibility that the animal has left scent marks on the objects. This conduct is often but not always reported. Finally, to form any kind of memory trace of the sample object, the animal must have a decent time of exploration. The duration of the period of sample object exploration is quite uniform between reported studies, but some laboratories only consider the actual exploration time while others simply count the time the animal spends in the same arena where the objects are located. If the real exploration time during a 10-min session is limited to 1-2 s, one would not expect a strong memory for the sample object! Therefore, an important but often ignored parameter to report is the time in contact with the objects during the sampling phase.

One advantage of novel object recognition is the one-trial learning. The performance of the animals deteriorates as the delay between the sample and the test trial increases. By comparing the performance at a short and a long retention interval, for instance, between 5 min and 24 h, one can in many instances distinguish a true memory impairment from nonmnemonic cause of impaired performance [8].

3.3. Spontaneous alteration in a Y-maze. This task is often described as the Y-maze task, but this is an ambiguous name, since many different exploration or avoidance tasks employing a similar maze structure have been described. This task is also easy to conduct as it requires no special training of the animal, and can be fully automated. Again, the test is based on the animals natural curiosity for exploration. The basic assumption is that this task draws on working memory, the ability to keep a memory item active in mind until a response is made, after which another item will replace it from

the active memory. In practise, the animal is given free access to all three maze arms and its spontanous choices are followed until a predefined number of choices (often 10) have been reached or a cut-off time (usually 3-5 min) has been reached . If the animal chooses a different arm than the one it arrived from, this choice is called an alteration. This is considered the correct response, whereas returning to the previous arm is considered an error. The test results is usually reported as % alternations. One inherent problem with the task is that it is self-paced. If the animal is scared, tired or otherwise unwilling to explore, it can dwell on in each arm for minutes, making the memory of the previous location weaker and also less dependent on true working memory. On the other hand, if the animal is hyperactive, it can start to make rotations. In fact, a stereotypic rotation around the maze will yield 100% alteration score in the task without posing any demand on working memory! Therefore, an essential output parameter in this task in addition to % alternations is the direction bias (e.g. number of turns to the dominat direction as % of all turns,[9]). Unfortunately, this is seldom reported.

Interestingly, spontaneous alternation in a T-maze, but not in a Y-maze, has been proven to be sensitive to hippocampal lesion [9]. The difference cannot be ascribed to the difference in the angle between the arms (90 vs. 120 deg), but most likely depends on the fact that the typical alternation task is continuous in the Y-maze while discrete trials, paced by the experimenter, are most often used in the spontaneous alternation in the T-maze.

3.4. Morris swim navigation task. This task has become a gold standard task for spatial learning and memory. It is conventionally called Morris watermaze [10], although the test arena is in fact a wading pool, not a maze (labyrinth). Since a later modification of the task (so-called radial water maze [11]) employs a true maze structure, the term Morris swim navigation task or simply Morris swim task would be recommended. This task is based on the rodent's aversion for water environment. Therefore, it is not difficult to motivate the animals to swim around the pool to look for a submerged platform that provides a temporary escape from the water. The beauty of the task is that the water prevents the animals from using their proprioception (number of steps) to assess the traversed distance and forces them to rely on visual landmarks of the environment. The crux of the task is that the animals is released from four different sites, so that an egocentric response ('straight ahead and then a gentle left turn') will not help to find the platform in more than one of the starting locations. Typically the task acquisition takes 3-5 days with 3-5 daily trials. The advantage of 5 daily trials is that the 1st and the last trial of the day starts from the same location, which allows separate estimation for the within-day or between-days learning/memory [12]. After the escape latency to the hidden platform reaches

a plateau, the spatial memory is separately assessed in a probe trial without the platform. This is usually done 24 h after the last acquisition trial. Here, the swim pattern is monitored by a video-tracking system and the time in the target quadrant or the former platform area is calculated.

Among critical variables in the task are the water temperature and the recovery time from the water exposure. Whereas rats with a thick fur and a subcutanous fat layer seldom show substantial hypothermia in the task, this is a serious confound in mice. When the task was novel, a preferred test protocol in pharmacological studies was to give all daily trials in a row to speed up training and keep drug concentration stable during the exposure. However, such a procedure in mice resulted in grave hypothermia, up to 9 ^{o}C drop in core body temperature [13]. Since our report, a paradigm shift has taken place, and nowadays all swim navigation testing in mice includes a 5-10 min warm-up between individual trials. Hypothermia still remains a possible confound if the genetic mutation or drug administration (such as NMDA-antagonists) by itself lowers the body temperature. On the other hand, raising the water temperature substantially over the ambient room temperature makes the test less aversive for the animals and may lead to floating.

Still nowadays, many laboratories run the Morris swim task without video tracking, by just manually taking the time to reach the escape platform, and during the probe test, by counting the number of platform crossings. This approach may produce misleading results when the genetic or drug intervention primarily affects the swimming speed. Also, when the swimming speed is measured with video-tracking and a slowing is observed, escape latency becomes an unreliable measure of learning. In such case, the swim path length would work better as an outcome measure, except for the very first trials when the animals usually do not find the platform without guidance. In that case, a shorter distance simply reflects slower swim speed ! On the probe trial, the number of platform crossings give a realistic idea how likely the animal would have found the platform. But it does not tell whether the animal missed the platform location by 1 cm or 30 cm.

The Morris swim task is considered a specific task for the hippocampal function. This is only partially true. It is well documented that the search bias during the probe trial depends on the intact circuitry in the septal third of the hippocampus [14]. In contrast, having the animal navigate from a single starting location [15] or from multiple starting location to a visually marked platform [16] does not require an intact hippocampus at all. Further, the task acquisition phase involves learning of many aspects of the task simultaneously, most importantly to avoid the instinctive wall-hugging strategy (thigmotaxis) and to take the courage to search for the platform in the pool center. Brain lesions that lead to decreased behavioral flexibility or increased anxiety may prolong this phase. It is also possible for an animal to demonstrate a perfect learning curve in terms of escape latency by adopting a strategy

to swim at a fixed distance from the pool wall. Thus, impairment in the acquisition phase may result from damage to several different brain structures [17], and to make the claim that one tests hippocampal function requires demonstration of impaired search bias in the probe test. One practical problem is that mice do not tend to show a clear search bias as the rats do. One reason for this is the pool size. First studies in mice tended to downscale the pool with the animal body size (rat pool 180 – 200 cm in diameter, mouse pool 60 – 90 cm). This led to increased tendency of mice to swim around the pool in circles. Once it became obvious that mice are equally good swimmers as rats, larger pools (100 – 150 cm in diameter) were adopted for mice as well. However, there is also a species difference in the swimming pattern. Rather than showing a focal search around the platform area, mice tend to swim in large elipsoid rounds that yet concentrate on the former platform location. The best outcome measure to reveal such a diffuse search bias has proved to be the mean distance to the platform center during the probe trial [18].

Since the Morris swim task is designed to be dependent on visual landmark cues, a possible visual impairment becomes a major confound in the task. A traditional approach to rule out a visual impairment has been to run a separate test with a visible platform while the room cues have been blocked by curtains around the pool. However, the ability of the cued platform test to rule out visual impairment has been questioned [19]. More importantly, the cued platform test may seriously interfere with spatial learning. If run before the hidden platform test with curtains around the pool to eliminate all distal cues, it encourages the animal to ignore the distal landmarks, which in the next phase become fundamentally important. Therefore, visible platform task should be run after the hidden platform version. Maybe an even better strategy is to altogether avoid the use of Morris swim task on animals that are known to have visual impairment. Albino rats and especially mice (see Leinonen & Tanila in this issue) loose most of their vision when kept unprotected under normal fluorescent light, and several inbred mouse strains carry a mutatation leading to early retinal degeneration. This information is easily available from the breeders.

The most important landmark that the rats and mice base their navigation on is the experimenter him/herself. And yet, hardly any published study where Morris swim task has been applied reports what the experimenter was actually doing during the task. In the worst case, the experimenter releases the animal in one of the four sites and then quickly moves to sit or stand next to the hidden platform to be able to catch the animal as soon as it reaches the platform. In such a case, the experimenter acts as a beacon for the animal and mitigates the basic principle of triangulation as a requirement for the successful navigation. Another common pattern for the experimenter is to move to a sector next to the platform quadrant. In this case, especially mice tend to swim toward the experimenter once they

have failed to find the platform, with the hope of being picked up faster. This results in obscure looking search bias toward the neighboring quadrant to the target one. In the best case, the experimenter stays hidden behind a screen during the trial and follows the animals swim pattern indirectly through the video monitor.

4. *Cognition is more than just memory*

In the present rodent neuroscience literature, especially when the focus is at the molecular or cellular level, cognition and memory are used almost as synonyms. The reason is obvious. We have a large battery of well designed and validated test to assess memory in rodents, while much fewer tests are available for other cognitive functions. However, to be able to discuss cognitive impairment, one should assess at least the following functions besides memory: (1) attention, (2) behavioral flexibility, (3) problem solving, and (4) action planning.

Attention includes two mutually exclusive cognitive operations: focused and divided attention. While focused attention requires all stimuli but the focus of action to be ignored, divided attention implies surveillance of the environment for potential targets. However, both types of attention share a common element by requiring behavioral vigilance. Focused attention can be studied to some extent with simple operant reaction time tasks, which require constant monitoring of a cue light that indicates the possibility of obtaining a food or liquid reward upon a prompt lever press in response [20]. However, such a task cannot make the distinction between impaired attention and slowed motor reaction. Having two or more choice options, like in the classic 5-choice serial reaction time task [21], helps to dissect out motor and perceptual components by assessing not only reaction time but also the choice accuracy. Especially by manipulating the duration or brightness of the visual cue while maintaining the motor response the same, one can demonstrate purely attentional defects. The downside of these tasks is the operant nature and long training required (usually 1 months of daily training). Thus even when fully automated, these kind of task will never be applicable as screening tests.

Behavioral flexibility is usually tested in various reversal tasks. Perhaps the simplist and fastest one for rodents to learn is spatial reversal in a T- or Y-maze. First, one needs to determine the spontaneous turning bias of the animal by giving it a free choice to either arm from the stem of the maze. Then the animal will be consistently rewarded only for making a turn choice to the opposite arm until the preset criterion (usually 90% correct) is achieved. Then on the following session (usually on a separate day), the opposite response will be consistently rewarded [12]. The same idea can be applied to choices

between two objects hiding a food reward or between two odors. In operant tasks, an often applied version of the reversal principle is the Go-Nogo task, where only one stimulus is presented at a time. A positive stimulus (light, tone, odor, object) indicates the possibility of obtaining a reward when approached, while a negative stimulus implies that the response will not be rewarded or even be punished. Go-Nogo tasks are seldom used in pure behavioral research but are common when combined with, for instance, electrophysiology. Perhaps the most complex reversal task successfully applied to rodents is attention set shifting [22]. The earlier versions of the task applied rats' natural tendency to dig for food in a cup filled with bedding. The presense of the food reward in this case could be implicated by the cup odor or the type of the medium. First, the animal is trained to make the distinction between the cup odors, one being positive and the other negative. Once learned, the critical stimulus dimension all the sudden becomes the cup medium instead of the odor. More recently, the touchscreen has become the most popular way of testing attentional set shifting in rodents [23]. However, despite the advantage of fully automated training procedure, the time required for the task acquisition (~1 month) is far too long for enabling this kind of task to be used as a screening test or part of a larger behavioral test battery. All these reversal type tasks typically draw on the function of the medial frontal cortex.

There is no separate task that can be claimed to be specific for problem solving. On the other hand, the learning phase of any task from the simplest operant lever pressing to escape from an aversive situation involves problem solving. In contrast, planning has be specifically tested in humans and monkeys as stepwise series of actions to reach the desired goal. Unfortunately, no such task to my knowledge is available for rodents.

5. Future perspectives

An increasing trend in behavioral neuroscience is the development of new automated tasks, including tests for cognition. This development arises from two main motives, which in turn have an independent history. The first motive, entertained by cognitive psychologists, has been to increase the translation between experimental animal and human studies by making the tasks as identical as possible. This development started with automated cognitive testing in operant chambers and has now largely shifted to tasks utilizing touchscreens for both animal and human subjects [24]. The second motive, entertained mainly by ethologists, has been to get rid of the confounding presence of the human experimenter and to make the testing devices as part of the rodents home environment. This development has culminated in the development of large automated cages ('social homecage') where

animals can be tested for days and in groups, thanks to identification of individuals with implanted transponders [25]. These two kinds of automated testing devices are almost orthogonal in terms of their pros and cons, as well as areas of application. The touchscreen tests are purely cognitive tests and have a high translational validity. On the other hand, their ecological validity is poor, especially when used for rodents, since they force the animals to work in a human environment. As a consequence, these tasks require massive daily training (around a month) before a rat or a mouse has learned the basic task principle to allow drug testing or manipulation of test parameters. Therefore, these kind of tests are mainly suited to within-subject testing of various drugs or the same drug in different doses. To obtain enough statistical power for group comparisons, tests should be ideally run with at least a dozen of test devices in parallel, which is a major initial investment on equipment. On the other hand, the training and testing can be done with little manpower. In contrast, social homecage type of automated testing has high ecological validity by making the test a part of an enriched environment rather than a separate test occasion, and by allowing rodents to stay group-housed. On the other hand, so far the recorded behavioral parameters have had little correspondence to human cognitive testing. Also the social homecage approach works best when a larger number of animals can be tested in parallel. This requires substantial investments on equipment, but again little manpower in the data acquisition phase. However, due to the massive amount of data recorded, the analysis phase may be time consuming. These kind of tests are therefore better suited to basic phenotyping of new genetically manipulated mouse lines. It looks likely that both types of automated testing approached will increase in popularity and eventually find they niche in the field of cognitive behavioral neuroscience. However, at present it still looks unlikely that automated tests will replace classic and well established cognitive tests in the near future.

Behavioral neuroscientists are always eager to develop new and better tasks. However, with the exceptions of a quickly learned and conducted attention task, and a task assessing planning in rodents, the current big challenges in behavioral testing of cognitive functions, when applied to behavioral pharmacology and phenotyping of genetically modified mice, are not primarily due to lack of suitable behavioral tasks. In my opinion, thanks to decades of intensive research on cognitive psychology in rodents and fruitful interaction between cognitive psychologist and ethologists, we have at present an outstanding toolkit of various validated behavioral tasks at our disposal. Instead, the biggest challenge at present and for years to come is to use this toolkit the corrent way. Areas that deserve attention include: (1) sufficient training of researchers running behavioral tasks, (2) optimal test selection, (3) proper conduct of the tests, and (4) reporting of the procedures in sufficient detail.

One future vision is that at least the majority of behavioral testing on cognition in rodent models would be carried out by laboratories devoted to behavioral studies and possessing sufficient reference material for the tests to be run. This can only be achieved by collaboration. Laboratories with the focus in molecular biology would certainly benefit from consultation of behavioral neuroscientist or their participation in projects. But also the other way round, behavioral neuroscientist will definitely benefit from better understanding of molecular and cellular mechanisms underlying certain kinds of behavior. In addition, constant efforts are needed in including training of behavioral techniques in the curricula of young neuroscientists independent of their specialization. The present system, where the behavioral study provides the final proof-of-principle evidence in elegant molecular studies in top-tier scientific journals, but the manuscript is submitted and reviewed by molecular biologists with little experience in behavioral neuroscience, will not enforce quality on the behavioral section.

In the future, we should get rid of published studies that aim at assessing cognitive impact of a genetic or pharmacological modification but present only a single behavioral assay of memory, usually even such a weak one as passive avoidance, spontaneous alternation in a Y-maze or novel object recognition. The cognitive assessment of a rodent model should be like the neuropsychological examination of the a human patient. The evaluation consists of a battery of tests surveying different domains of cognition. If the hypothesis to be tested is as wide as cognitive improvement, the battery should consists tests of short-term and long-term memory, and a separate test for behavioral flexibility. In addition, the most common confounds of memory tests, alterations in motor activity and anxiety, should be ruled out as an explantion to the observed changes. Ideally, the test protocol should have space for an ad-hoc control test based on the results.

During the recent years, researches performing behavioral studies on genetically modified mice have been able to come up with guiding principles for laboratory animal science (http://www.lasa.co.uk/PDF/LASA_BAP_BNA_ESSWAP_GP_Behavioural_LAS_Nov13.pdf) and even detailed recommendations on how to conduct the screening tests, such as the IMPReSS (International Mouse Phenotyping Resource of Standardised Screens) guidelines (http://www.mousephenotype.org/impress). In addition, ARRIVE guidelines (http://www.mousephenotype.org/about-impc/arrive-guidelines) on the principles of reporting research on experimental animals in general have been well received. Further, researchers working on experimental studies on stroke rehabilitation have developed consensus statements for selection and conduction of tests for motor recovery in rodents [26]. Nevertheless, there is an obvious need for similar guidelines for behavioral cognitive studies in rodent models. It will be unrealistic to find a consensus on which individual behavioral task should be applied, for instance, to preclinical trials

aiming at reversing memory impairment in Alzheimer model mice. However, the behavioral neuroscience community should be relatively unanimous as to the proper conduct of common cognitive tasks and their reporting in such a way that most important known caveats in the tests will be covered. Establishments of such guidelines and their distribution to the neuroscience community and publishers would be the most effective way to improve the quality of the bulk of cognitive behavioral testing in the decades to come.

References:

[1] Steward O, Balice-Gordon R. Rigor or mortis: best practices for preclinical research in neuroscience. Neuron. 2014;84(3):572-81.

[2] Sukoff Rizzo SJ, Silverman JL. Methodological Considerations for Optimizing and Validating Behavioral Assays. Curr Protoc Mouse Biol. 2016;6(4):364-379.

[3] Mogil JS. Laboratory environmental factors and pain behavior: the relevance of unknown unknowns to reproducibility and translation. Lab Anim (NY). 2017;46(4):136-141.

[4] Van der Staay FJ, Steckler T. The fallacy of behavioral phenotyping without standardisation. Genes Brain Behav. 2002;1(1):9-13.

[5] Crusio WE, Schwegler H, Lipp HP. Radial-maze performance and structural variation of the hippocampus in mice: a correlation with mossy fibre distribution. Brain Res. 1987;425(1):182-5.

[6] Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. Science. 1999;284(5420):1670-2.

[7] McGaugh JL. Time-dependent processes in memory storage. Science. 1966; 153(3742):1351-8.

[8] Dere E, Huston JP, De Souza Silva MA. The pharmacology, neuroanatomy and neurogenetics of one-trial object recognition in rodents. Neurosci Biobehav Rev. 2007;31(5):673-704.

[9] Deacon RM, Bannerman DM, Kirby BP, Croucher A, Rawlins JN. Effects of cytotoxic hippocampal lesions in mice on a cognitive test battery. Behav Brain Res. 2002;133(1):57-68.

[10] Morris RG, Garrud P, Rawlins JN, O'Keefe J. Place navigation impaired in rats with hippocampal lesions. Nature. 1982;297(5868):681-3.

[11] Buresová O, Bures J, Oitzl MS, Zahálka A. Radial maze in the water tank: an aversively motivated spatial working memory task. Physiol. Behav. 1985; 34:1003-5.

[12] Liu L, Ikonen S, Heikkinen T, Heikkilä M, Puoliväli J, van Groen T, Tanila H. Effects of fimbria-fornix lesion and amyloid pathology on spatial learning and memory in transgenic APP+PS1 mice. Behav Brain Res. 2002;134(1-2):433-45.

[13] Iivonen H, Nurminen L, Harri M, Tanila H, Puoliväli J. Hypothermia in mice tested in Morris water maze. Behav Brain Res. 2003;141(2):207-13.

[14] Moser MB, Moser EI, Forrest E, Andersen P, Morris RG. Spatial learning with a minislab in the dorsal hippocampus. Proc Natl Acad Sci U S A. 1995;92(21):9697-701.

[15] Eichenbaum H, Stewart C, Morris RG. Hippocampal representation in place learning. J Neurosci. 1990;10(11):3531-42.

[16] McDonald RJ, White NM. Parallel information processing in the water maze: evidence for independent memory systems involving dorsal striatum and hippocampus. Behav Neural Biol. 1994;61(3):260-70.

[17] D'Hooge R, De Deyn PP. Applications of the Morris water maze in the study of learning and memory. Brain Res Brain Res Rev. 2001;36(1):60-90.

[18] Maei HR, Zaslavsky K, Teixeira CM, Frankland PW. What is the Most Sensitive Measure of Water Maze Probe Test Performance? Front. Integr. Neurosci. 2009:3,4.

[19] Lindner MD, Plone MA, Schallert T, Emerich DF. Blind rats are not profoundly impaired in the reference memory Morris water maze and cannot be clearly discriminated from rats with cognitive deficits in the cued platform task. *Brain Res. Cogn. Brain Res*. 1997; 5, 329-33.

[20] Moore H, Dudchenko P, Bruno JP, Sarter M. Toward modeling age-related changes of attentional abilities in rats: simple and choice reaction time tasks and vigilance. Neurobiol Aging. 1992;13(6):759-72.

[21] Carli M, Robbins TW, Evenden JL, Everitt BJ. Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats;implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. Behav Brain Res. 1983;9(3):361-80.

[22] Birrell JM, Brown VJ. Medial frontal cortex mediates perceptual attentional set shifting in the rat. J Neurosci. 2000;20(11):4320-4.

[23] Brigman JL, Bussey TJ, Saksida LM, Rothblat LA. Discrimination of multidimensional visual stimuli by mice: intra- and extradimensional shifts. Behav Neurosci. 2005;119(3):839-42.

[24] Bussey TJ, Padain TL, Skillings EA, Winters BD, Morton AJ, Saksida LM. The touchscreen cognitive testing method for rodents: how to get the best out of your rat. Learn Mem. 2008;15(7):516-23.

[25] Vannoni E, Voikar V, Colacicco G, Sánchez MA, Lipp HP, Wolfer DP. Spontaneous behavior in the social homecage discriminates strains, lesions and mutations in mice. J Neurosci Methods. 2014;234:26-37.

[26] Bernhardt J, Borschmann K, Boyd L, Thomas Carmichael S, Corbett D, Cramer SC, Hoffmann T, Kwakkel G, Savitz SI, Saposnik G, Walker M, Ward N. Moving rehabilitation research forward: Developing consensus statements for rehabilitation and recovery research. Int J Stroke. 2016;11(4):454-8.