PLOS ONE

RESEARCH ARTICLE

Repeatability analysis improves the reliability of behavioral data

Juliane Rudeck[1], **Silvia Vogl**[1], **Stefanie Banneke**[1], **Gilbert Schönfelder**[1,2], **Lars Lewejohann**[1,3]*

1 German Centre for the Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment (BfR), Berlin, Germany, **2** Departement of Toxicology, Institute of Clinical Pharmacology and Toxicology, Charité–Universitätsmedizin Berlin, cooperate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Berlin, Germany, **3** Department of Veterinary Medicine, Institute of Animal Welfare, Animal Behavior and Laboratory Animal Science, Freie Universität Berlin, Berlin, Germany

* Lars.Lewejohann@bfr.bund.de

Citation: Rudeck J, Vogl S, Banneke S, Schönfelder G, Lewejohann L (2020) Repeatability analysis improves the reliability of behavioral data. PLoS ONE 15(4): e0230900. https://doi.org/10.1371/journal.pone.0230900

Editor: Giuseppe Biagini, University of Modena and Reggio Emilia, ITALY

Received: January 7, 2020

Accepted: March 11, 2020

Published: April 2, 2020

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files. We confirm that the video files are NOT part of our minimal data set needed for the analysis. Nevertheless, video files are available on request from the author.

Funding: This work was supported by the Federal Ministry of Education and Research (BMBF) '[grant number 031A262D]' and by the Deutsche Forschungsgemeinschaft (DFG) [FOR 2591; LE 2356/5-1]. Prof. Schoenfelder wrote the BMBF proposal and raised the BMBF funding. Prof.

Abstract

Reliability of data has become a major concern in the course of the reproducibility crisis. Especially when studying animal behavior, confounding factors such as novelty of the test apparatus can lead to a wide variability of data which may mask treatment effects and consequently lead to misinterpretation. Habituation to the test situation is a common practice to circumvent novelty induced increases in variance and to improve the reliability of the respective measurements. However, there is a lack of published empirical knowledge regarding reasonable habituation procedures and a method validation seems to be overdue. This study aimed at setting up a simple strategy to increase reliability of behavioral data measured in a familiar test apparatus. Therefore, exemplary data from mice tested in an Open Field (OF) arena were used to elucidate the potential of habituation and how reliability of measures can be confirmed by means of a repeatability analysis using the software R. On seven consecutive days, male C57BL/6J, BALB/cJ and 129S1/SvImJ mice were tested in an OF arena once daily and individual mouse behavior was recorded. A repeatability analysis was conducted with regard to repeated trials of habituation. Our data analysis revealed that monitoring animal behavior during habituation is important to determine when individual differences of the measurements are stable. Repeatability values from distance travelled and average activity increased over the habituation period, revealing that around 60% of the variance of the data can be explained by individual differences between mice. The first day of habituation was significantly different from the following 6 days. A three-day habituation period appeared to be sufficient in this study. Overall, these results emphasize the importance of habituation and in depth analysis of habituation data to define the correct starting point of the experiment for improving the reliability and reproducibility of experimental data.

Lewejohann wrote the DFG proposal and raised the DFG funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Animal experimentation always requires an ethical evaluation, balancing the scientific benefit and possible constraints to animal welfare. Without a doubt, the generation of high-quality research data is crucial for a better translation of experimental data to humans. Regarding the welfare, stress in laboratory animals should be limited to a minimum. Fortunately, these two goals are often working perfectly together because it is well known that data derived from non-stressed animals is of higher quality expressed, for example, in more robust and/ or more consistent scientific data [1, 2].

One well known possibility to minimize stress for the laboratory animal is the prior habituation to unknown experimental equipment and/or handling procedures [1, 3]. The term habituation originates from behavioral biology and describes the diminution of a response induced by constant or repeated exposure to a novel stimulus and displays a simple form of non-associative learning [4]. A distinction is made between habituation over time (intrasession) and over repeated exposures (intersession) [4]. In this original field of behavioral biology, the habituation process is the primary research objective and serves as a model for non-associative learning and memory mechanisms as well as for pharmacological studies [4–6]. One of the most frequently assessed parameters of habituation is a decrease in exploratory behavior measured in the Open Field (OF) test as a change in distance travelled or activity. Once the animal is familiar with the new environment, the explorative behavior is reduced and it is considered that the laboratory animal has habituated and the learning and memory process is completed [4].

Knowledge about habituation can be used to introduce a wide variety of unknown behavioral tests to the animal and to increase the intended performance in the actual test situation which then is familiar to the animal [1, 3, 7, 8]. For example, in rodents, physical activity can be measured under voluntary or under forced condition. On the one hand, voluntary exercise has the disadvantage of very variable volume and intensity differing from animal to animal. On the other hand, rodents in forced running wheel tests show low level of performance. Habituation and training with the running wheel have been shown to improve locomotor performance in the forced running wheel system in rats and consequently increased the validity of the scientific data [3]. Although, such effects of habituation are frequently applied, the documentation in the literature is heterogeneous. In case of the OF, there are research groups which use this test without documented habituation (e.g., [9–13]). Some of them stated that they want to test for novelty induced activity [13]. Other researchers establish only one day of habituation to either reduce novelty induced activity or to measure the effect of habituation from trial one to trial two [14–16]. Frequently, experimenters place the animals 30 min prior to testing in the corresponding room or arena and define this as habituation [17–19]. Rarely, a previous habituation is mentioned in the literature which is carried out for more than two days [20–22]. If the specific research question requires testing in a known environment, there is usually no well-founded derivation of the habituation period which is stated in the literature. In fact, a strategy how to determine an effective habituation period for a certain experiment would be favorable.

The repeatability analysis is one possible method to assess the accuracy of measurements which can be normally or not normally distributed [23] and it allows identifying an explanation for the occurring variances. Repeatability describes the proportion of the total variation that is reproducible among the repeated measurements of the same group [23–25]. Nakagawa and Schielzeth published a practical guide for biologists to bring forward the method of repeatability analysis in the scientific community [23]. This method of data analysis is commonly employed in the field of ecology and evolutionary biology where the repeatability of

morphological, physiological and behavioral traits is of special interest [26]. Some examples include the wing morphology of drosophila [27], cancer evolution [28], or mate choice for reproduction [29–31]. More recently, the concept of animal personality that focuses on consistent between-individual differences in behavior was investigated by applying repeatability analysis [32–37]. In this context, the repeatability between several trials on one day or successive days as well as between several independent experiments of different research groups is meant. Bell et al. showed in a meta-analysis of repeatability of independent experiments from different research groups for animal behavior that many behavioral types are more consistent within individuals than previously assumed [2]. In addition, Mazzamuto et al. underline the importance of method validation to ensure that the intended parameter is really captured [38]. Given that there is growing concern regarding reproducibility of experimental data in biomedical science [39], we are convinced that a documented validation of the habituation data is a beneficial approach in all fields of animal behavioral science to increase the reliability and reproducibility of experimental data. In the context of this manuscript, we define experimental data as reliable if the method by which it was collected is scientifically valid and robust. This means that the method should measure what was intended minimizing any disturbing influences. A mathematically justified and comprehensible explanation for the choice of a certain habituation period would strengthen the scientific outcome, as it is shown that the intended parameter can actually be captured. In addition, valuable information can be obtained from the habituation data e.g., the source of variance, which might be helpful for the interpretation of the primary experimental results.

Therefore, the aim of our study was to elucidate the putative hidden information of habituation data in combination with the repeatability analysis for research questions which requires a familiar environment as test situation (for example wheel running, nest building, burrowing behavior or baseline locomotion [3, 22, 40]). As an example, we investigated retrospectively the effectiveness of intersession habituation of male mice of C57BL/6J, BALB/cJ and 129S1/SvImJ inbred strains to an OF arena as familiar environment within two different experiments carried out at the same institute (one in winter and one in summer time) [41]. For this purpose, we analyzed video data to assess locomotor parameters in the OF (distance travelled, average activity). We conducted a repeatability analysis to check the reliability of our data from trial to trial and to evaluate if the chosen habituation period was sufficient or could be reduced in future studies. In order to identify the primary source of variance in locomotion data, we examined the factors experiment/ batch, strain, and the individual animal (animal ID). All these factors are known from the literature to affect locomotion [2, 42, 43].

Material and methods

Ethics statement

The data for this study is based on two experiments which were approved by the Berlin state authority, *Landesamt für Gesundheit und Soziales*, under license No. G 0309/15 and G 0194/16 and were conducted in accordance with the German Animal Protection Law (TierSchG, TierSchVersV). The main purpose of these two studies was the evaluation of the analgesic effect of buprenorphine in three different mouse strains [41]. All decisions regarding strain, sex and age of the animals were made with regard to this research question. At the end of this main study, animals were sacrificed by gradual CO_2 filling according to Annex IV of Directive 2010/63/EU. Here, we exclusively present the data of the habituation period of the animals.

Animals and animal care

Male C57BL/6J mice (n = 38) from Charles River (Sulzfeld, Germany, breeding the original Jackson strain from the USA) and male BALB/cJ mice (n = 15) and male 129S1/SvImJ mice

(n = 15) from Jackson Laboratory (Bar Harbor, USA, imported via Charles River) were obtained after weaning (three to four weeks old), to familiarize them with the housing conditions. The mice were free of all viral, bacterial, and parasitic pathogens listed in the Federation of European Laboratory Animal Science Associations (FELASA) guidelines. During familiarization time, mice were housed in groups of 5–8 animals per cage in Eurostandard type III polycarbonate cages with filter tops (Tecniplast, Hohenpeissenberg, Germany), autoclaved bedding and nesting material (LASbedding^{TM}PG2, LASvendi, Soest, Germany), a mouse house consisting of cardboard (LBS Biotechnology, United Kingdom) and gnawing material as environmental enrichment (J. Rettenmaier & Söhne GmbH + Co KG, Rosenberg, Germany). The animals had free access to autoclaved food pellets (LASQCdiet^{TM} Rod16, LASvendi, Soest, Germany) and water acidified with HCL (pH 2.5–3.0 to prevent growth of algae and pathogens). The room temperature was maintained at 21 ± 1˚C, with a relative humidity of 55 ± 10%. The light/ dark cycle in the room consisted of 12/ 12 h artificial light with lights on from 5.00 am to 5.00 pm in winter time and 6.00 am to 6.00 pm in summer time. All animals were handled by cupping the mouse in open hands.

Habituation

After two weeks of familiarization to the housing conditions, all mice were single housed to prevent aggressive behavior. With the start of the habituation period, mice were six to seven weeks old. On seven consecutive days, each mouse was habituated once a day to the following experimental conditions in a randomized order. Each mouse was fixated by grip in the neck to mimic the stress of a subcutaneous (s.c.) injection and was subsequently placed in a round OF arena (ø 30 cm, 40 cm height, 109 lux) for 5 min, monitored by a video camera from above. Afterwards, the mouse was placed on the warm (at 35˚C) Incremental Hot Plate (IHP, IITC Inc. Life Science, Woodland Hills, USA) for 4 min, monitored by two video cameras (frontal and lateral). Habituation and testing took place between 9.00 am to 11.00 am in winter time and 10.00 am to noon in summer time to avoid influence by circadian rhythm. On day one and six, body weight of each mouse was measured within the routine animal check as a general welfare indicator (S1 Fig). Since there are hints in literature that the sex of the experimenter might have an influence [44], the whole habituation and experiments were performed by the same female experimenter. The results from pain assessment on the IHP will not be part of this publication.

Data collection and analysis

Locomotor activity, monitored as distance travelled (cm), and average activity (%, threshold 0.1 cm/s) were assessed for every mouse for 3 min in the round OF arena and analyzed with the Viewer Software (version 4, Biobserve GmbH, Bonn, Germany). We have chosen the mean of the last three minutes of observation for the evaluation to limit the influence of possible initial stress behavior.

Statistical analysis and plotting of the graphs were conducted with the freely available software R, version 3.5.1 and R-studio, version 1.1.383 (www.r-project.org) and with the software Graph Pad Prism (version 6, Graph Pad Software, San Diego, USA). To get an overview of the data set, all data points of distance travelled and average activity were plotted with R and statistically analyzed with Graph Pad Prism using the Friedman test with Dunn's multiple comparison test compared to day one. To test the repeatability of the data and to identify the source of the variance within the data, a repeatability analysis was performed using the "rptR" library to quantify the constancy of phenotypes [23, 45]. First, data sets were tested for normal distribution by Q-Q-norm-plot (S2 Fig). If data did not show substantial deviation from normal

distribution the repeatability value R was calculated using a linear mixed-effect model (LMM) based on Gaussian distribution. Animal ID, strain or experiment/ batch were included as random factors, whereas, distance travelled (cm) or average activity (%) served as fixed factors. The confidence interval (CI) [2.5%, 97.5%] is based on 500 bootstrapping runs and 100 permutations. To test for statistical significance of the repeatability of distance travelled or average activity between the selected numbers of trials, the likelihood ratio test was performed. To identify the most important random factor, R was first calculated over the seven-day habituation period for all fixed factors including animal ID, strain and experiment as multiple random grouping factors. To observe the progress of repeatability over the whole time, R was calculated additionally over three adjacent measurements (days), including the random grouping factors animal ID or strain. In a second step, the identified effect covariate strain that explain only a part of the variances in the data was included to adjust the repeatability for the predominating random factor animal ID. Data and basic steps to perform the analysis are available in the supplement (S1 Text and S1 Table) in order to ensure the comprehensibility of our analysis.

Results

Effects of habituation on OF behavior of single mice

In the initial step, the raw data of each animal were examined.

From the second day of habituation onwards, a considerable reduction of distance travelled and average activity occurred compared to day one (Fig 1). In addition, the widths of confidence intervals of distance travelled and average activity decreased from day one until reaching stable plateaus at day five (Table 1).

Repeatability analysis of explorative behavior

In a second step, a repeatability analysis was carried out in order to obtain reliable evidence for the successful habituation and additional information on the variance of the data.

In order to identify the main factor by which the observed variance in locomotion data can be explained, a repeatability analysis with multiple grouping factors was conducted over the seven-day habituation period (Table 2). The distance travelled as well as the average activity served as fixed factors and animal ID, strain, and experiment/ batch functioned as random multiple grouping factors. The variances observed for distance travelled and average activity is primarily explained by animal ID and strain. The fact that we included data from two different experiments did not contribute significantly to the observed variances of distance travelled or average activity (Table 2).

For this reason, repeatability values for distance travelled and average activity were calculated including both, animal ID and strain as a random factor. To record changes in repeatability over the whole period of habituation, each repeatability value R was calculated over three adjacent days resulting in five groupings (S2 and S3 Tables). Within the first grouping (day 1–3) repeatability values of 0.027 (distance travelled, random factor: ID) and 0.177 (distance travelled, random factor: strain) as well as 0.199 (activity, random factor: ID) and 0.321 (activity, random factor: strain) were calculated for distance travelled and average activity, respectively (Fig 2). This indicates that only 2.7% of the variance of distance travelled data and 19.9% of the variance of the average activity data can be explained by the factor animal ID. 17.7% of the variance of distance travelled data and 32.1% of the average activity data can be explained by the factor strain. The repeatability of the results within this period is considered unlikely. In addition, most of the variance cannot be explained by definable factors and is therefore considered random noise likely induced by other factors, e.g., stress or anxiety-related behavior. From the second (day 2–4) to the fifth grouping (day 5–7) stable repeatability values of 0.556,

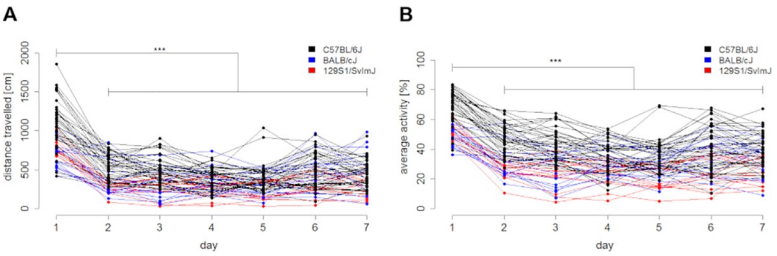

Fig 1. Frequency of observed behavior decreases during habituation. (A) Distance travelled [cm] and (B) average activity [%] are depicted. Values are presented as exact value per each animal (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). The Friedman test with Dunn's multiple comparisons test was used for statistical evaluation (n.s. = not significant (A) p-value = 0.7119, (B) p-value > 0.9999, **** p-value < 0.0001 (A, B).

https://doi.org/10.1371/journal.pone.0230900.g001

0.617, 0.505, and 0.498 for distance travelled (random factor: ID) and 0.634, 0.652, 0.627, and 0.604 for average activity (random factor: ID) were calculated (Fig 2A and 2B). This means that roughly 50 to 60% of the variance can be explained by individual differences between the mice (S2 and S3 Tables). In contrast, only 27.6, 22.9, 18.2, 19.0% and 40.1, 36.3, 29.7, 30.7% of the observed variance in distance travelled and average activity data, respectively, can be traced back to strain (Fig 2C and 2D, S2 and S3 Tables).

From this repeatability analysis it can be concluded that strain is an important factor, but the influence of the individually stable behavior of each mouse predominates. Hence, estimations of repeatability over the whole time were calculated again including only animal ID as random factor with adjustment for strain (Fig 3, S2 Table). The resulting adjusted repeatability values are comparable to animal ID repeatability values for distance travelled and average activity showing only slightly reduced (~15%) repeatability values.

Discussion

Measured data of each individual mouse as well as adjusted repeatability values showed a stable pattern after three days of habituation for distance travelled and average activity. Moreover, additional habituation days did not lead to a further reduction of explorative behavior or increased repeatability of investigated parameters and reached a plateau in the second grouping from day two to four. Hence, a habituation time of three days seemed to be sufficient for this experimental set up. Furthermore, the individual personality of the mouse has a greater influence on the variance of the data than the strain to which it belongs. On the basis of these results a simple strategy can be established to introduce habituation as a preliminary step in behavioral experiments which requires a known environment (Fig 4). First, plan and set up the experiment and habituate the animals to the testing situation. The simultaneous or retrospective checking of the measured data with repeatability analysis allows the determination of the optimal starting point of the experiment.

Table 1. Variances of distance travelled and average activity within the habituation period.

	day 1	day 2	day 3	day 4	day 5	day 6	day 7
distance travelled [cm]	[863.5, 1016]	[402.2, 497.1]	[353.9, 449.3]	[317.1, 387.8]	[297.9, 379.3]	[389.6, 497.9]	[372.7, 468.3]
average activity [%]	[58.09, 64.65]	[35.68, 42.03]	[31.33, 38.04]	[29.15, 34.54]	[28.04, 33.66]	[32.81, 40.12]	[33.08, 39.02]

The [2.5%. 97.5%] confidence interval was calculated for the factor travelled distance and average activity for ever day of the habituation period (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice).

https://doi.org/10.1371/journal.pone.0230900.t001

Table 2. Repeatability values with multiple grouping factors: animal ID, strain and experiment.

	repeatability for animal ID			repeatability for strain			repeatability for experiment		
	R	CI	p-value	R	CI	p-value	R	CI	p-value
distance travelled	0.101	[0.034, 0.184]	1.31E-4	0.1	[0, 0.298]	0.00219	0.0124	[0, 0.094]	0.296
average activity	0.137	[0.065, 0.241]	3.6E-08	0.173	[0, 0.443]	7.59E-05	0.053	[0, 0.232]	0.0606

For every factor, the repeatability R, the [2.5%, 97.5%] confidence intervals (CI) and the p-values, calculated by likelihood ratio test, were displayed over the seven-day habituation period (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution. The CI resulted from 500 bootstrapping runs and 100 permutations.

https://doi.org/10.1371/journal.pone.0230900.t002

In this study, we used locomotion data measured in a familiar OF arena instead of home cage based data. In recent years, home cage RFID-based monitoring has become a promising tool to detect unbiased behavior of rodents in social groups over long periods [46, 47]. In fact, it has been shown that home cage activity is repeatable over long periods of time and indicating the emergence of stable individual differences in activity [32]. After habituation to the test arena activity data in the OF also became repeatable and it would be tempting to speculate that individual differences indeed correlate between home cage and habituated OF arena. Our initial research purpose, which is not the main focus of this manuscript, was to investigate the analgesic effect of buprenorphine in three different mouse strains [41]. In this context, we aimed at the comparison of baseline locomotion with the locomotion after buprenorphine administration. Individual tracking in separate cage systems or arenas is a common practice e.g., to study the influence of different xenobiotics [20–22]. In this study, we analyzed a relative short time frame of three minutes due to the restricted time frame to prevent putative influences of the circadian rhythm. To ensure that we really detected baseline locomotion in the OF arena we introduced the longer habituation period of seven consecutive days and analyzed it

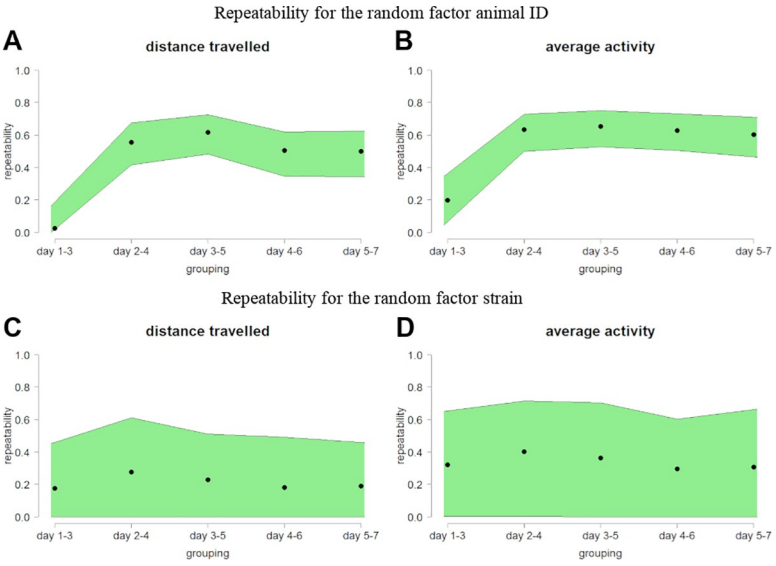

Fig 2. Animal ID and strain repeatability increase during habituation. (A, B) Calculated animal ID repeatability value and (C, D) strain repeatability value for the factors (A, C) distance travelled and (B, D) average activity were presented. Each repeatability value (R, black points) was calculated over three adjacent days resulting in five groupings (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution. The [2.5%, 97.5%] confidence intervals (CI) were displayed in green, resulting from 500 bootstrapping runs and 100 permutations.

https://doi.org/10.1371/journal.pone.0230900.g002

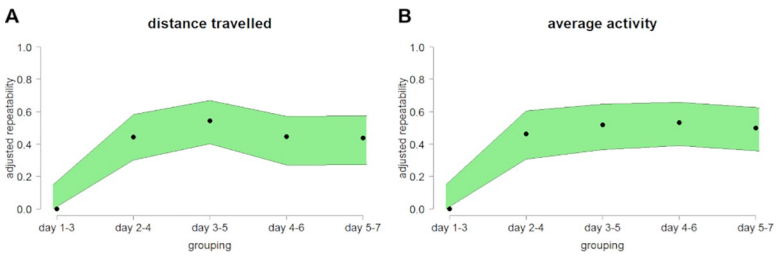

Fig 3. Animal ID repeatability adjusted for strain increase during habituation. Calculated animal ID repeatability values adjusted for the fixed factor strain were presented for (A) distance travelled and (B) average activity. Each repeatability value (R, black points) was calculated over three adjacent days resulting in five groupings (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model for locomotion as well as activity and with a generalized linear mixed-effect model based on Gaussian distribution for rearing and sniffing behavior. The [2.5%, 97.5%] confidence intervals (CI) were displayed in green, resulting from 500 bootstrapping runs and 100 permutations.

https://doi.org/10.1371/journal.pone.0230900.g003

with the repeatability analysis. However, data from home cage based observations would have been equally suitable for this kind of analysis. In addition, longer observation periods would have been proficient for the analysis of different relevant time frames.

In this study, a direct measurement of stress-related behavior or stress hormones was not performed. As reduction in body weight can give a hint of the general well-being of the animal, we measured the body weight on day one and six of habituation as part of the routine check-up. No loss of body weights was detected and, hence, we conclude that multiple habituation days to the least did not influence this general welfare parameter. Comparable to this result, Bodden et al. investigated the impact of repeated versus single open-field testing on welfare in C57BL/6J mice and also found no body weight changes or other signs of compromised welfare [48]. The anxiety of an animal is usually connected to time spent in the center and time spent

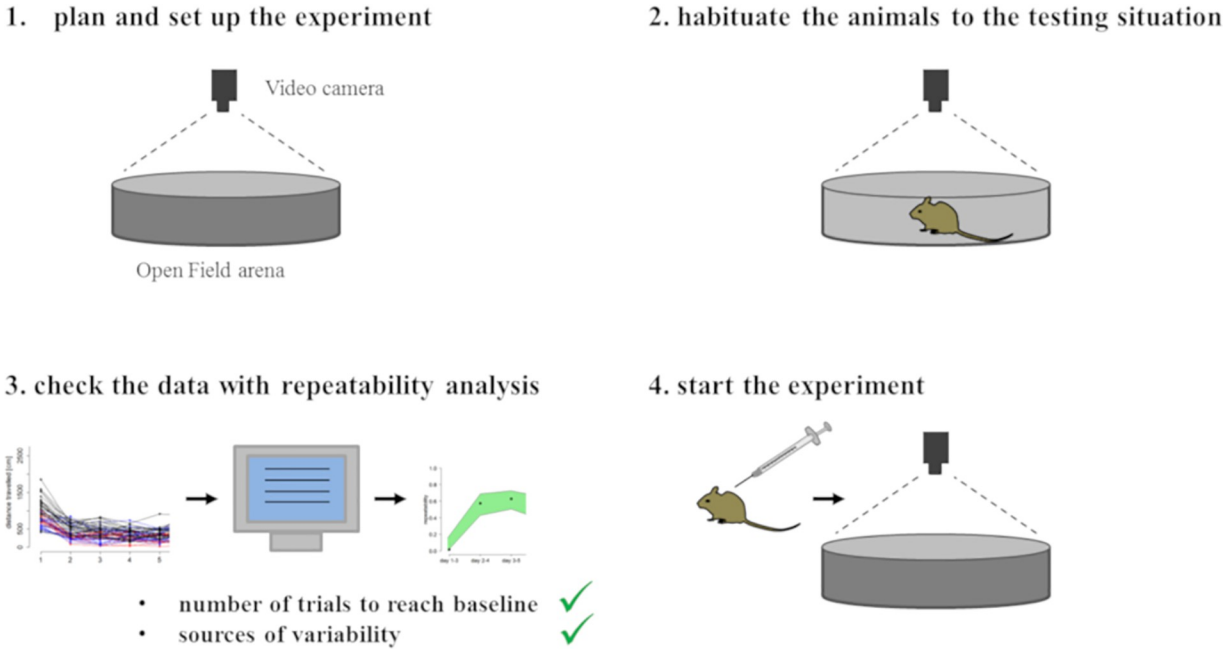

Fig 4. Strategy for the implementation of behavioral experiments. This basic strategy can be applied to all experiments with laboratory animals to determine the optimal starting point of your experiment which needs a familiar environment.

https://doi.org/10.1371/journal.pone.0230900.g004

nearby the wall of an OF [4]. We have not assessed these parameters because of the small size of the OF used and for this reason cannot comment on the anxiety behavior of the animals. In our study, we were able to demonstrate that the habituation approach was successful in decreasing the distance travelled and average activity.

Furthermore, we could show that repeatability analysis is a powerful tool to assess reproducible animal behavior and to identify the source of variance within the measured data [23]. Even if the variance of the data cannot be reduced completely, knowing the sources of variability allows a much better interpretation of behavioral data. In connection with the reproducibility crisis, Baker reported that more than half of the researchers participating in a Nature's survey have failed to reproduce their own experiments [39]. The two experiments used here as examples were carried out with an interval of nine month. Our analysis did not reveal any significant effect of time of experimentation, indicating that our approach yielded reproducible results. Even if the variance of locomotion data can be only reduced to a certain extent, the variance becomes explainable from the second day on. Based on the results of the repeatability analysis, the influence of data from two different experiments could be excluded as the source of variation within this data set. For this reason, we were able to pool the data and concentrated on the other two scientifically relevant factors, namely strain and the individual animal.

Interestingly, within the first grouping (day one to three) the variance of the data could not be traced back to either strain or the individual animal and is therefore more likely induced by unknown other factors, e.g., stress or anxiety. Only after the second grouping (day two to four) reliable parameters have been identified by the repeatability analysis as the source of variance. From this we conclude that, at the minimum, one day, better three days of habituation would be advisable for this or comparable set ups to be able to observe reproducible behavior. Dealing with reproducible behavior, on the other hand, supports a reliable interpretation of the variance of the data with regard to contribution of strain or individual animal personality.

Mouse strain comparisons are commonly implemented in behavioral and genetic studies with widely use of C57BL/6 and BALB/c mice [49]. However, previous reports on explorative behavior of C57BL/6J and BALB/c mice are inconsistent [49–52]. C57BL/6J mice were found to be engaged in more exploratory activity than BALB/cJ mice [50, 51]. In contrast, other studies reported higher anxiety levels associated with lower activity in C57BL/6 than BALB/c mice [53]. An and colleagues reported no differences in locomotor activity between male C57BL/6J and BALB/cJ mice in a novel cage observation test [50]. In our data set only a small percentage of the variance in explorative data could be traced back to strain. From this we concluded that strain differences were not the primary source of variance using these three strains even though they derived from two different breeding facilities which could have added additional variation. The large confidence interval might be explained by the fact that only three different strains have been used.

Another possible explanation is that there is a wide variation in the individual personality of mice, regardless of the strain. In recent years, behavioral studies focused more on consistent between-individual differences giving evidence that the animal personality in addition to strain is a crucial factor to explain variance of measurements [32–36, 54]. This assumption is supported by a meta-analysis for animal behavior of Bell and colleagues. They demonstrated that many behavioral types are more consistent within individuals than previously assumed [2]. Interestingly, such individual differences are also found in genetically homogeneous inbred mice housed under highly standardized conditions. This indicates that individual differences tend to escape standardization approaches and might reflect an evolutionary intrinsic value that perpetuates inter-individual differentiation [54]. In addition, Brust and colleagues showed that the personality of a mouse stabilizes with age and is highly repeatable over their lifespan [32]. In our study, we observed highly consistent between-individual differences for

distance travelled and average activity from the second grouping onwards. Interestingly, the individual differences between the mice, i.e., their "animal personality" [54] was identified as the main contributor to variability in explorative behavior in the repeatedly conducted OF test. However, if only a single OF test was conducted, the variability in the data could not have been fully explained as the results of the OF test were only reasonably repeatable after habituation. Consequently, testing without habituation might have had a considerable effect on the recently found lack of reproducibility, known as the reproducibility crisis. In accordance to these findings, Chappell and colleagues observed variable locomotion behavior but highly repeatable between-individual differences in mice [55]. Integrating the concept of animal personality not only offers the possibility to improve the interpretation and thus the quality of scientific data, but also to enhance animal welfare science by solving welfare problems on an individual level [56].

Beside this investigated influencing factors, i.e., two different experiments, three different strains, and the personality of the individual animal, other factors are also important to consider in such kind of analysis. Since this study was a side project of a main explorative research study dealing with the analgesic effect of buprenorphine in different mouse strains we primarily focused on male mice [41]. However, scientific data from literature show that behavioral patterns can differ tremendously between male and female mice [57]. Hence, investigating the influence of different sexes on specific parameters compared to the individual mouse behavior would be of special interest. In addition to include more influencing factors, the use of this repeatability analysis to investigate other behavioral parameters within different behavioral set ups is promising. To keep this study as comprehensible as possible, we used locomotion data, e.g., distance travelled and average activity, of an OF test as an example. However, we also tested this method on the following parameters and provided them as supporting information: number of ambulations and average velocity in the OF, rearing and stretched sniffing behavior on the IHP (S3 and S4 Figs, S4–S6 Tables).

Conclusion

We demonstrated a simple strategy to improve the data quality of those behavioral experiments requiring a familiar environment, such as wheel running, nest building, burrowing behavior or baseline locomotion. This was realized by implementing the repeatability analysis method to prove successful habituation as well as explaining underlying factors for variability in the data. While the overall variability is not necessarily reduced, detailed analysis of habituation data reveals the sources of variability. Furthermore, analyzing the habituation data by means of a repeatability analysis helps to determine the optimal start point of an experiment. In our data, the variances observed in the behavior patterns were predominantly due to the individual personality of the mice rather than a strain dependent effect. Behavioral data was repeatable after successful habituation and calculated repeatability values reached a plateau after only a few habituation trials. Hence, as a refinement strategy, the habituation period in comparable experimental designs can be reduced to three days in future experiments. Taken together, this exemplary study underlines the importance of publishing and evaluating habituation data for all kind of behavioral experiments. Importantly, declaration of the effectiveness of habituation procedures could furthermore enhance the validity of experimental data and fortify reproducibility.

Supporting information

S1 Fig. Slightly increasing body weight during habituation period. Body weight [g] at day one and six of habituation period per strain were presented as box plot with median and

whiskers [2.5, 97.5%] (n = 38 C57BL/6J: p-value = 0.0018, n = 15 BALB/cJ: p-value = 0.0004 and n = 15 129S1/SvImJ: p-value < 0.0001, Wilcoxon matched-pairs signed rank test, two-tailed).
(TIF)

S2 Fig. Testing for normal distribution of distance travelled and average activity data. Data sets of distance travelled and average activity were checked for normal distribution using Q-Q-norm plot (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice).
(TIF)

S3 Fig. Repeatability of average velocity and ambulations in the open field arena as well as rearing and sniffing behavior on the 35˚C warm incremental hot plate. (A—D) Calculated animal ID repeatability value and (E—H) strain repeatability value for the factors (A, E) average velocity, (B, F) number of ambulations, (C, G) number of rearings and (D, H) time spent with stretched sniffing were presented. Each repeatability value (R, black points) was calculated over three adjacent days resulting in five groupings for velocity and ambulations and in four groupings for rearing and sniffing behavior (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution for velocity and ambulations and with a generalized linear mixed-effect model based on Poisson distribution for rearing and sniffing behavior. The [2.5%, 97.5%] confidence intervals (CI) were displayed in green, resulting from 500 bootstrapping runs and 100 permutations.
(TIF)

S4 Fig. Testing for normal distribution of average velocity and ambulations in the open field arena as well as rearing and sniffing behavior on the 35˚C warm incremental hot plate. Data sets of average velocity, sum of ambulations, number of rearings and time spent with stretched sniffing were checked for normal distribution using Q-Q-norm plot (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice).
(TIF)

S1 Table. Raw data used for the repeatability analysis with the software R.
(XLSX)

S2 Table. Repeatability values for animal ID as random factor with and without adjustment for the factor strain for distance travelled and average activity. Each repeatability value (R) was calculated over three adjacent days resulting in five groupings. For every factor R, the [2.5%, 97.5%] confidence intervals (CI) and p-values calculated by likelihood ratio test were displayed (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution. The CI resulted from 500 bootstrapping runs and 100 permutations.
(PDF)

S3 Table. Repeatability values for strain as random factor for distance travelled and average activity. Each repeatability value (R) was calculated over three adjacent days resulting in five groupings. For every factor R, the [2.5%, 97.5%] confidence intervals (CI) and p-values calculated by likelihood ratio test were displayed (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution. The CI resulted from 500 bootstrapping runs and 100 permutations.
(PDF)

S4 Table. Repeatability values with multiple grouping factors: animal ID, strain and experiment for average velocity, ambulations as well as rearing and sniffing behavior. For every factor, the repeatability R, the [2.5%, 97.5%] confidence intervals (CI) and the p-values, calculated by likelihood ratio test, were displayed over the seven-day habituation period for average velocity and number of ambulations and over a six-day habituation period for rearing and sniffing behavior (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution for average velocity and number of ambulations and with a generalized linear mixed-effect model based on Poisson distribution for rearing and sniffing behavior. The CI resulted from 500 bootstrapping runs and 100 permutations.
(PDF)

S5 Table. Repeatability values for animal ID as random factor for average velocity, ambulations as well as rearing and sniffing behavior. Each repeatability value (R) was calculated over three adjacent days resulting in five groupings for average velocity and number of ambulations and in four groupings for rearing and sniffing behavior. For every factor R, the [2.5%, 97.5%] confidence intervals (CI) and p-values calculated by likelihood ratio test were displayed (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution for average velocity and ambulations and with a generalized linear mixed-effect model based on Poisson distribution for rearing and sniffing behavior. The CI resulted from 500 bootstrapping runs and 100 permutations.
(PDF)

S6 Table. Repeatability values for strain as random factor for average velocity, ambulations as well as rearing and sniffing behavior. Each repeatability value (R) was calculated over three adjacent days resulting in five groupings for average velocity and number of ambulations and in four groupings for rearing and sniffing behavior. For every factor R, the [2.5%, 97.5%] confidence intervals (CI) and p-values calculated by likelihood ratio test were displayed (n = 38 C57BL/6J, n = 15 BALB/cJ and n = 15 129S1/SvImJ male mice). Estimation of repeatability was conducted with a linear mixed-effect model based on Gaussian distribution for average velocity and ambulations and with a generalized linear mixed-effect model based on Poisson distribution for rearing and sniffing behavior. The CI resulted from 500 bootstrapping runs and 100 permutations.
(PDF)

S1 Text. R-script for the repeatability analysis.
(PDF)

Author Contributions

Conceptualization: Juliane Rudeck, Lars Lewejohann.

Data curation: Juliane Rudeck.

Funding acquisition: Gilbert Schönfelder, Lars Lewejohann.

Investigation: Juliane Rudeck.

Visualization: Juliane Rudeck.

Writing – original draft: Juliane Rudeck.

Writing – review & editing: Silvia Vogl, Stefanie Banneke, Gilbert Schönfelder, Lars Lewejohann.

References

1. Gouveia K, Hurst JL. Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. Scientific reports. 2017; 7:44999. Epub 2017/03/23. https://doi.org/10.1038/srep44999 PMID: 28322308; PubMed Central PMCID: PMC5359560.
2. Bell AM, Hankison SJ, Laskowski KL. The repeatability of behaviour: a meta-analysis. Anim Behav. 2009; 77(4):771–83. https://doi.org/10.1016/j.anbehav.2008.12.022 WOS:000264325300003. PMID: 24707058
3. Toval A, Banos R, De la Cruz E, Morales-Delgado N, Pallares JG, Ayad A, et al. Habituation Training Improves Locomotor Performance in a Forced Running Wheel System in Rats. Frontiers in behavioral neuroscience. 2017; 11:42. Epub 2017/03/25. https://doi.org/10.3389/fnbeh.2017.00042 PMID: 28337132; PubMed Central PMCID: PMC5340750.
4. Leussis MP, Bolivar VJ. Habituation in rodents: a review of behavior, neurobiology, and genetics. Neuroscience and biobehavioral reviews. 2006; 30(7):1045–64. Epub 2006/06/16. https://doi.org/10.1016/j.neubiorev.2006.03.006 PMID: 16774787.
5. Pavkovic Z, Milanovic D, Ruzdijic S, Kanazir S, Pesic V. The influence of propofol anesthesia exposure on nonaversive memory retrieval and expression of molecules involved in memory process in the dorsal hippocampus in peripubertal rats. Pediatr Anesth. 2018; 28(6):537–46. https://doi.org/10.1111/pan.13396 WOS:000434418900007. PMID: 29752843
6. Evans DA, Stempel AV, Vale R, Branco T. Cognitive Control of Escape Behaviour. Trends Cogn Sci. 2019; 23(4):334–48. https://doi.org/10.1016/j.tics.2019.01.012 WOS:000461850900011. PMID: 30852123
7. Bowen RS, Cates BE, Combs EB, Dillard BM, Epting JT, Foster BR, et al. Stabilization of the wheel running phenotype in mice. Physiology & Behavior. 2016; 155:149–56. https://doi.org/10.1016/j.physbeh.2015.12.006 WOS:000369455200018. PMID: 26687894
8. Moreton E, Baron P, Tiplady S, McCall S, Clifford B, Langley-Evans SC, et al. Impact of early exposure to a cafeteria diet on prefrontal cortex monoamines and novel object recognition in adolescent rats. Behavioural Brain Research. 2019; 363:191–8. https://doi.org/10.1016/j.bbr.2019.02.003 WOS:000460822600024. PMID: 30735761
9. Sichova K, Pinterova N, Zidkova M, Horsley RR, Lhotkova E, Stefkova K, et al. Mephedrone (4-Methylmethcathinone): acute Behavioral effects, hyperthermic, and Pharmacokinetic Profile in rats. Frontiers in Psychiatry. 2018;8. https://doi.org/10.3389/fpsyt.2018.00008 WOS:000419702000001. PMID: 29449817
10. Teixeira LV, Almeida RF, Rohden F, Martins LAM, Spritzer PM, de Souza DOG. Neuroprotective Effects of Guanosine Administration on In Vivo Cortical Focal Ischemia in Female and Male Wistar Rats. Neurochemical Research. 2018; 43(7):1476–89. https://doi.org/10.1007/s11064-018-2562-3 WOS:000435605500017. PMID: 29855847
11. Wirz A, Mandillo S, D'Amato FR, Giuliani A, Riviello MC. Response, use and habituation to a mouse house in C57BL/6J and BALB/c mice. Experimental Animals. 2015; 64(3):281–93. WOS:000358753700007. https://doi.org/10.1538/expanim.14-0104 PMID: 25854626
12. Singh JCH, Kakalij RM, Kshirsagar RP, Kumar BH, Komakula SSB, Diwan PV. Cognitive effects of vanillic acid against streptozotocin-induced neurodegeneration in mice. Pharmaceutical Biology. 2015; 53(5):630–6. https://doi.org/10.3109/13880209.2014.935866 WOS:000351180700002. PMID: 25472801
13. Breu M, Zhang JY, Porambo M, Pletnikov MV, Goeral K, Kakara M, et al. Diffusion Tensor Imaging Abnormalities in the Cerebral White Matter Correlate with Sex-Dependent Neurobehavioral Deficits in Adult Mice with Neonatal Ischemia. Developmental Neuroscience. 2016; 38(2):83–95. https://doi.org/10.1159/000442943 WOS:000379162600001. PMID: 26977597
14. de Melo M, Pereira DE, Moura RD, da Silva EB, de Melo F, Dias CDQ, et al. Maternal Supplementation With Avocado (Persea americana Mill.) Pulp and Oil Alters Reflex Maturation, Physical Development, and Offspring Memory in Rats. Frontiers in Neuroscience. 2019;13. https://doi.org/10.3389/fnins.2019.00013 WOS:000456566900001. PMID: 30760975
15. de Sousa LP, de Almeida RF, Ribeiro-Gomes FL, Carvalho LJD, Souza TME, de Souza DOG, et al. Long-term effect of uncomplicated Plasmodium berghei ANKA malaria on memory and anxiety-like behaviour in C57BL/6 mice. Parasites & Vectors. 2018;11. https://doi.org/10.1186/s13071-018-2778-8 WOS:000428288700001. PMID: 29554958

16. Simoes LR, Abreu R, Generoso JS, Goularte JA, Collodel A, Giridharan VV, et al. Prevention of Memory Impairment and Neurotrophic Factors Increased by Lithium in Wistar Rats Submitted to Pneumococcal Meningitis Model. Mediators of Inflammation. 2017. https://doi.org/10.1155/2017/6490652 WOS:000413590800001. PMID: 29200666

17. Aso-Someya N, Narikiyo K, Masuda A, Aou S. The functional link between tail-pinch-induced food intake and emotionality and its possible role in stress coping in rats. Journal of Physiological Sciences. 2018; 68(6):799–805. https://doi.org/10.1007/s12576-018-0596-6 WOS:000449788000010. PMID: 29423592

18. Nikolaus S, Beu M, Silva MAD, Huston JP, Hautzel H, Mattern C, et al. Relationship Between L-DOPA-Induced Reduction in Motor and Exploratory Activity and Striatal Dopamine D-2 Receptor Binding in the Rat. Frontiers in Behavioral Neuroscience. 2016;9. https://doi.org/10.3389/fnbeh.2016.00009 WOS:000367574800001. PMID: 26903826

19. Pagani JH, Avram SKW, Cui Z, Song J, Mezey E, Senerth JM, et al. Raphe serotonin neuron-specific oxytocin receptor knockout reduces aggression without affecting anxiety-like behavior in male mice only. Genes Brain and Behavior. 2015; 14(2):167–76. https://doi.org/10.1111/gbb.12202 WOS:000351629000004. PMID: 25677455

20. Chellammal HSJ, Alagarsamy V, Ramachandran D, Gummadi SB, Manan MM, Yellu NR. Neuroprotective effects of 1 'delta-1 '-acetoxyeugenol acetate on A beta((25–35)) induced cognitive dysfunction in mice. Biomedicine & Pharmacotherapy. 2019; 109:1454–61. https://doi.org/10.1016/j.biopha.2018.10.189 PubMed PMID: WOS:000452539100159. PMID: 30551397

21. Nasehi M, Alaghmandan-Motlagh N, Ebrahimi-Ghiri M, Nami M, Zarrindast MR. The interaction between hippocampal GABA-B and cannabinoid receptors upon spatial change and object novelty discrimination memory function. Psychopharmacology. 2017; 234(20):3117–28. https://doi.org/10.1007/s00213-017-4688-4 WOS:000410745900010. PMID: 28779310

22. Milanovic D, Pesic V, Loncarevic-Vasiljkovic N, Pavkovic Z, Popic J, Kanazir S, et al. The Fas Ligand/Fas Death Receptor Pathways Contribute to Propofol-Induced Apoptosis and Neuroinflammation in the Brain of Neonatal Rats. Neurotoxicity Research. 2016; 30(3):434–52. https://doi.org/10.1007/s12640-016-9629-1 WOS:000383727300010. PMID: 27189477

23. Nakagawa S, Schielzeth H. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biological reviews of the Cambridge Philosophical Society. 2010; 85(4):935–56. Epub 2010/06/24. https://doi.org/10.1111/j.1469-185X.2010.00141.x PMID: 20569253.

24. Lessells CM, Boag PT. Unrepeatable Repeatabilities: A Common Mistake. The Auk. 1987; 104(1):116–21. https://doi.org/10.2307/4087240 WOS:A1987F617400014.

25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin. 1979; 86(2):420–8. Epub 1979/03/01. https://doi.org/10.1037//0033-2909.86.2.420 PMID: 18839484.

26. Araya-Ajoy YG, Mathot KJ, Dingemanse NJ. An approach to estimate short-term, long-term and reaction norm repeatability. Methods Ecol Evol. 2015; 6(12):1462–73. https://doi.org/10.1111/2041-210x.12430 WOS:000368517700010.

27. Houle D, Mezey J, Galpern P, Carter A. Automated measurement of drosophila wings. BMC Evol Biol. 2003; 3:12. https://doi.org/10.1186/1471-2148-3-12 WOS:000188122100025. PMID: 12773209

28. Taylor TB, Johnson LJ, Jackson RW, Brockhurst MA, Dash PR. First steps in experimental cancer evolution. Evol Appl. 2013; 6(3):535–48. https://doi.org/10.1111/eva.12041 WOS:000316957800012. PMID: 23745144

29. Burley NT, Hamedani E, Symanski C. Mate choice decision rules: Trait synergisms and preference shifts. Ecol Evol. 2018; 8(5):2380–94. https://doi.org/10.1002/ece3.3831 WOS:000426725900001. PMID: 29531661

30. Haneke-Reinders M, Reinhold K, Schmoll T. Sex-specific repeatabilities and effects of relatedness and mating status on copulation duration in an acridid grasshopper. Ecol Evol. 2017; 7(10):3414–24. https://doi.org/10.1002/ece3.2937 WOS:000402554700011. PMID: 28515877

31. Pogany A, Szurovecz Z, Vincze E, Barta Z, Szekely T. Mate preference does not influence reproductive motivation and parental cooperation in female zebra finches. Behaviour. 2014; 151(12–13):1885–901. https://doi.org/10.1163/1568539x-00003221 WOS:000343841200010.

32. Brust V, Schindler PM, Lewejohann L. Lifetime development of behavioural phenotype in the house mouse (Mus musculus). Frontiersinzoology 2015;12 Suppl 1:S17. Epub 2016/01/28. https://doi.org/10.1186/1742-9994-12-s1-s17 PMID: 26816516; PubMed Central PMCID: PMC4722345.

33. Brent LJN, Semple S, MacLarnon A, Ruiz-Lambides A, Gonzalez-Martinez J, Platt ML. Personality Traits in Rhesus Macaques (Macaca mulatta) Are Heritable but Do Not Predict Reproductive Output. Int J Primatol. 2014; 35(1):188–209. https://doi.org/10.1007/s10764-013-9724-6 WOS:000330954400011. PMID: 24659840

34. Chen BJ, Liu K, Zhou LJ, Gomes-Silva G, Sommer-Trembo C, Plath M. Personality differentially affects individual mate choice decisions in female and male Western mosquitofish (Gambusia affinis). PloS one. 2018; 13(5):23. https://doi.org/10.1371/journal.pone.0197197 WOS:000432118800029. PMID: 29763435

35. Koski SE. Social personality traits in chimpanzees: temporal stability and structure of behaviourally assessed personality traits in three captive populations. Behav Ecol Sociobiol. 2011; 65(11):2161–74. https://doi.org/10.1007/s00265-011-1224-0 WOS:000295939100014.

36. Roche DG, Careau V, Binning SA. Demystifying animal 'personality' (or not): why individual variation matters to experimental biologists. J Exp Biol. 2016; 219(24):3832–43. https://doi.org/10.1242/jeb.146712 WOS:000391279300006. PMID: 27852750

37. Dingemanse NJ, Dochtermann NA. Quantifying individual variation in behaviour: mixed-effect modelling approaches. The Journal of animal ecology. 2013; 82(1):39–54. Epub 2012/11/23. https://doi.org/10.1111/1365-2656.12013 PMID: 23171297.

38. Mazzamuto MV, Cremonesi G, Santicchia F, Preatoni D, Martinoli A, Wauters LA. Rodents in the arena: a critical evaluation of methods measuring personality traits. Ethology Ecology & Evolution. 2019; 31(1):38–58. https://doi.org/10.1080/03949370.2018.1488768 WOS:000452773800004.

39. Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016; 533(7604):452–4. Epub 2016/05/27. https://doi.org/10.1038/533452a PMID: 27225100.

40. Jirkof P, Tourvieille A, Cinelli P, Arras M. Buprenorphine for pain relief in mice: repeated injections vs sustained-release depot formulation. Lab Anim. 2015; 49(3):177–87. Epub 2014/12/10. https://doi.org/10.1177/0023677214562849 PMID: 25488320.

41. Rudeck J, Vogl S, Heinl C, Steinfath M, Fritzwanker S, Kliewer A, et al. Analgesic treatment with buprenorphine should be adapted to the mouse strain. Pharmacology, biochemistry, and behavior. 2020. Epub 2020/02/24. https://doi.org/10.1016/j.pbb.2020.172877 PMID: 32088361.

42. Podhorna J, Brown RE. Strain differences in activity and emotionality do not account for differences in learning and memory performance between C57BL/6 and DBA/2 mice. Genes, brain, and behavior. 2002; 1(2):96–110. Epub 2003/07/30. https://doi.org/10.1034/j.1601-183x.2002.10205.x PMID: 12884980.

43. Leppanen PK, Ravaja N, Ewalds-Kvist SB. Twenty-three generations of mice bidirectionally selected for open-field thigmotaxis: selection response and repeated exposure to the open field. Behav Processes. 2006; 72(1):23–31. Epub 2006/01/03. https://doi.org/10.1016/j.beproc.2005.11.010 PMID: 16386379.

44. Bohlen M, Hayes ER, Bohlen B, Bailoo JD, Crabbe JC, Wahlsten D. Experimenter effects on behavioral test scores of eight inbred mouse strains under the influence of ethanol. Behav Brain Res. 2014; 272:46–54. Epub 2014/06/17. https://doi.org/10.1016/j.bbr.2014.06.017 PMID: 24933191; PubMed Central PMCID: PMC4968576.

45. Stoeffel MA, Nakagawa S, Schielzeth H. An introduction to repeatability estimation with rptR 2018 [updated 2018/01/042018/06/12]. Available from: https://cran.r-project.org/web/packages/rptR/vignettes/rptR.html.

46. Redfern WS, Tse K, Grant C, Keerie A, Simpson DJ, Pedersen JC, et al. Automated recording of home cage activity and temperature of individual rats housed in social groups: The Rodent Big Brother project. Plos One. 2017; 12(9):e0181068. Epub 2017/09/07. https://doi.org/10.1371/journal.pone.0181068 PMID: 28877172; PubMed Central PMCID: PMC5587114.

47. Freund J, Brandmaier AM, Lewejohann L, Kirste I, Kritzler M, Kruger A, et al. Emergence of individuality in genetically identical mice. Science. 2013; 340(6133):756–9. Epub 2013/05/11. https://doi.org/10.1126/science.1235294 PMID: 23661762.

48. Bodden C, Siestrup S, Palme R, Kaiser S, Sachser N, Richter SH. Evidence-based severity assessment: Impact of repeated versus single open-field testing on welfare in C57BL/6J mice. Behav Brain Res. 2018; 336:261–8. Epub 2017/08/27. https://doi.org/10.1016/j.bbr.2017.08.029 PMID: 28842269.

49. Crawley JN, Davis LG. Baseline exploratory activity predicts anxiolytic responsiveness to diazepam in five mouse strains. Brain research bulletin. 1982; 8(6):609–12. Epub 1982/06/01. https://doi.org/10.1016/0361-9230(82)90087-9 PMID: 6890398.

50. Brown SM, Holtzman M, Kim T, Kharasch ED. Buprenorphine metabolites, buprenorphine-3-glucuronide and norbuprenorphine-3-glucuronide, are biologically active. Anesthesiology. 2011; 115(6):1251–60. Epub 2011/11/01. https://doi.org/10.1097/ALN.0b013e318238fea0 PMID: 22037640; PubMed Central PMCID: PMC3560935.

51. Augustsson H, Meyerson BJ. Exploration and risk assessment: a comparative study of male house mice (Mus musculus musculus) and two laboratory strains. Physiology & Behavior. 2004; 81(4):685–98. https://doi.org/10.1016/j.physbeh.2004.03.014 WOS:000222130700018. PMID: 15178164

52. Belzung C, Griebel G. Measuring normal and pathological anxiety-like behaviour in mice: a review. Behav Brain Res. 2001; 125(1–2):141–9. Epub 2001/10/30. https://doi.org/10.1016/s0166-4328(01)00291-1 PMID: 11682105.

53. Avgustinovich DF, Lipina TV, Bondar NP, Alekseyenko OV, Kudryavtseva NN. Features of the genetically defined anxiety in mice. Behavior Genetics. 2000; 30(2):101–9. https://doi.org/10.1023/a:1001999020138 WOS:000088971700003. PMID: 10979600

54. Lewejohann L, Zipser B, Sachser N. "Personality" in laboratory mice used for biomedical research: a way of understanding variability? Developmental psychobiology. 2011; 53(6):624–30. Epub 2011/08/26. https://doi.org/10.1002/dev.20553 PMID: 21866543.

55. Chappell MA, Garland T, Rezende EL, Gomes FR. Voluntary running in deer mice: speed, distance, energy costs and temperature effects. J Exp Biol. 2004; 207(22):3839–54. https://doi.org/10.1242/jeb.01213. WOS:000225669600006. PMID: 15472015

56. Richter SH, Hintze S. From the individual to the population—and back again? Emphasising the role of the individual in animal welfare science. Applied Animal Behaviour Science. 2019; 212:1–8. https://doi.org/10.1016/j.applanim.2018.12.012 WOS:000462804600001.

57. An XL, Zou JX, Wu RY, Yang Y, Tai FD, Zeng SY, et al. Strain and sex differences in anxiety-like and social behaviors in C57BL/6J and BALB/cJ mice. Experimental animals. 2011; 60(2):111–23. Epub 2011/04/23. https://doi.org/10.1538/expanim.60.111 PMID: 21512266.