Behavioural Brain Research 336 (2018) 261–268

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Evidence-based severity assessment: Impact of repeated versus single open-field testing on welfare in C57BL/6J mice

Carina Bodden[a,b], Sophie Siestrup[a], Rupert Palme[c], Sylvia Kaiser[a,b], Norbert Sachser[a,b], S. Helene Richter[a,b,*]

[a] Department of Behavioural Biology, University of Münster, Badestr. 13, 48149 Münster, Germany
[b] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Fliednerstr. 21, 48149 Münster, Germany
[c] Unit of Physiology, Pathophysiology and Experimental Endocrinology, Department of Biomedical Sciences, University of Veterinary Medicine, Veterinärplatz 1, 1210 Vienna, Austria

ARTICLE INFO

Keywords:
Animal welfare
Severity assessment
Open-field test
Repeated testing
Refinement
Behavioral tests

ABSTRACT

According to current guidelines on animal experiments, a prospective assessment of the severity of each procedure is mandatory. However, so far, the classification of procedures into different severity categories mainly relies on theoretic considerations, since it is not entirely clear which of the various procedures compromise the welfare of animals, or, to what extent. Against this background, a systematic empirical investigation of the impact of each procedure, including behavioral testing, seems essential. Therefore, the present study was designed to elucidate the effects of repeated versus single testing on mouse welfare, using one of the most commonly used paradigms for behavioral phenotyping in behavioral neuroscience, the open-field test. In an independent groups design, laboratory mice (*Mus musculus* f. domestica) experienced either repeated, single, or no open-field testing – procedures that are assigned to different severity categories. Interestingly, testing experiences did not affect fecal corticosterone metabolites, body weights, elevated plus-maze or home cage behavior differentially. Thus, with respect to the assessed endocrinological, physical, and behavioral outcome measures, no signs of compromised welfare could be detected in mice that were tested in the open-field repeatedly, once, or, not at all. These findings challenge current classification guidelines and may, furthermore, stimulate systematic research on the severity of single procedures involving living animals.

1. Introduction

For many years, ethical concerns have been raised regarding the use of animals for research purposes [1]. In particular, the conflict between the need for animal welfare, on the one hand, and the desire for new scientific insights, on the other, have led the debate. In this context, estimating the severity of a procedure involving animals is one of the most important, but also one of the most difficult tasks that experimenters have to face when balancing harm and benefits of an experiment [2]. The classification of procedures according to their severity is therefore an important tool to ensure effective prediction and minimization of animal suffering. Meanwhile, such classification guidelines have become an integral part of legislation on animal research in many countries, although it remains difficult to measure and objectively quantify severity in living animals. The current Directive regulating animal use within the European Union (2010), for example, requires all procedures to be assigned to a severity category, ranging from “mild” through “moderate” and “severe” to “non recovery” [3]. However, the classification into different categories so far mainly relies on theoretic considerations, highlighting the increasing demand for a systematic and evidence-based severity assessment (e.g., see [4]). Only in this way will it be possible to identify and define clear criteria that reflect the actual severity of a specific procedure [1].

Over the past years, a tremendous number of mutant lines have been generated, with particular emphasis on transgenic and knockout laboratory mice (*Mus musculus* f. domestica). As robust mouse phenotypes hold great promise as translational tools for discovering effective treatments for a variety of human diseases, a systematic characterization of behavioral traits becomes increasingly important. Especially in neurological and psychiatric research, there is a growing demand for high-throughput techniques to comprehensively characterize mouse behavior. Given that most studies in the field of behavioral phenotyping therefore involve several behavioral paradigms, the question arises how severe these standard procedures are and to what extent they compromise animal welfare.

According to the European Commission Working Group report on

* Corresponding author at: Department of Behavioural Biology, University of Münster, Badestr. 13, 48149 Münster, Germany.
E-mail address: richterh@uni-muenster.de (S.H. Richter).

http://dx.doi.org/10.1016/j.bbr.2017.08.029
Received 27 June 2017; Received in revised form 11 August 2017; Accepted 17 August 2017
Available online 23 August 2017
0166-4328/

severity classification [5], the application of a single behavioral test is considered below the threshold for regulation, whereas a combination or accumulation of more than one behavioral test is classified as a procedure of "mild" severity [3]. As the combination of three or more complementary tests is strongly recommended for a systematic identification of behavioral phenotypes as well as for the sake of reproducibility [6–10], most behavioral phenotyping studies currently fall at least into the "mild" severity category.

Officially, the statistics on the severity of procedures reveal that most procedures used in animal experiments have been categorized as "mild", while smaller proportions of experiments have been classified as "moderate", "severe", or "non recovery" (e.g., statistics on the severity of procedures in Germany, 2015, "mild": 60%, "moderate": 24%, "severe": 5%, "non recovery": 11%, [11]). However, as inferred from Annex VIII of the European guidelines, this "mild" category subsumes a striking variety of different procedures of heterogeneous application areas, including, for example, the induction of tumors with no detectable clinical adverse effects as well as short-term deprivation of social partners. With respect to behavioral procedures, it is explicitly stated that the "combination or accumulation of open-field testing", also falls into this severity category [3].

The open-field (OF) test is one of the most common tests used in behavioral phenotyping studies [12,13]. It was originally described by Hall [14] for the study of emotionality in laboratory rats (*Rattus norvegicus* f. domestica) and subjects an animal for a certain period of time to an unknown, illuminated arena with surrounding walls [15]. Different versions exist, differing, for example, in shape of the arena (circular, square or rectangular), size, color as well as illumination level (for reviews see [13,15,16]). The test has been shown to be sensitive to the anxiolytic-like effects of classical benzodiazepines and 5-HT1A receptor agonists and is therefore a validated procedure to assess anxiety-like behavior in rodents (for a review see [13]). Apart from this, many researchers use repeated OF testing to study the habituation processes to a novel environment [17]. Although physiological changes during single OF testing have been assessed via infrared thermography [18], as far as we know, no experiments have been conducted that systematically investigated acute and long-term effects of the exposure to repeated versus single OF testing on the welfare of laboratory animals.

Therefore, the present study aimed at studying the impact of repeated versus single OF testing on welfare-related parameters in mice. In particular, we studied corticosterone metabolite concentrations in the feces, body weights, anxiety-like as well as home cage behavior to gain as comprehensive a picture as possible of the individuals' welfare. Both the activity of the hypothalamic-pituitary-adrenal (HPA) axis (for a review see [19]) as well as changes in body weight [20–22] have previously been utilized as highly sensitive indicators for assessing the degree of stress. Also behavioral measures, such as spontaneous behavior in the home cage (e.g. activity, play, stereotypic behavior) as well as the performance in specific behavioral paradigms (e.g. elevated plus-maze), have frequently been assessed to study welfare-related questions (e.g. [23]). In an independent groups design, C57BL/6J mice of two experimental groups performed the OF test either repeatedly or once, while mice of two control groups were either exposed to a novel environment or received no specific treatment. By systematically investigating the aforementioned welfare-related measures prior to and after the treatments we sought to test the hypothesis that mice exposed to single or repeated OF testing differ in endocrinological, physical, as well as behavioral measures.

2. Animals and methods

2.1. Animals and housing conditions

In the present study, 48 male mice of the C57BL/6J strain were used, which were provided by Charles River Laboratories (Research Models and Services, Germany GmbH, Sulzfeld). This inbred strain was selected because of its widespread use in neurobehavioral studies. Since the experiment was conducted in two independent batches at an interval of two weeks, mice were aged either 3 or 5 weeks at delivery. To avoid any effects of age-dependent previous experiences, mice of each batch were counterbalanced with respect to the four treatment groups (see Section 2.3). Upon arrival at our institute, mice were housed in an open cage system in groups of four individuals until postnatal day (PND) 55. From then on, mice were kept individually to exclude aggressive interactions. All cages (transparent Makrolon type III; dimensions: 38 cm × 23 cm × 15 cm) contained wood shavings (Allspan, Höveler GmbH & Co.KG, Langenfeld, Germany) as bedding material and a paper towel as nesting material. Furthermore, a transparent red plastic mouse house (Mouse House™, Tecniplast Deutschland GmbH, Hohenpeißenberg, Germany) and a wooden stick (approximately 1.5 cm × 1.5 cm × 10 cm) were provided for each cage. Food pellets (Altromin 1324, Altromin Spezialfutter GmbH & Co. KG, Lage, Germany) and tap water were provided *ad libitum*. Housing rooms were maintained at a 12/12 h light/dark cycle with lights off at 9:00 a.m., a temperature of about 22 °C, and a relative humidity of about 50%.

2.2. Ethics statement

All procedures complied with the regulations covering animal experimentation within the EU (European Communities Council DIRECTIVE 2010/63/EU) and were approved by the national and local authorities (Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen "LANUV NRW", reference number: 84-02.04.2015.A245). Our study involved behavioral testing only and did not cause any distress or pain to the animals. After the study, the animals remained in the animal facility of the institute for further behavioral studies.

2.3. Experimental design

The whole experiment was divided into three experimental phases: a pre-treatment phase, a treatment phase, and a post-treatment phase (Fig. 1). The treatment phase comprised three treatment days, during which mice experienced a specific treatment depending on their group. Allocation to one of the following four groups (n = 12/group) was randomized: repeated open-field testing (RT), single open-field testing (ST), control group 1 (C1), and control group 2 (C2; Fig. 2). The pre- and post-treatment phase served to assess a series of different welfare-related measures to investigate the effects of repeated versus single behavioral testing on the welfare of laboratory mice.

All treatments, behavioral tests, observations, and collections of feces were performed during the dark phase. The sequence of animals during these procedures was pseudo-randomized. The experimenter was blind to the treatments at any time during the pre- and post-treatment phase.

2.4. Procedures during the treatment phase

During the treatment phase, mice of the two experimental groups were exposed to the OF either three times (RT: PND 76, 78, and 80) or once (ST: PND 80; see Fig. 2). According to the European Commission Working Group report on severity classification [5], RT would reflect an accumulation of behavioral tests, and thus, a procedure of mild severity, whereas ST would fall below the threshold for regulation. Testing the animals in the OF [16] was thus part of the experimental treatment. The OF consisted of a white square arena (80 cm × 80 cm × 42 cm) and was illuminated with an intensity of 40 lx. The test was performed in a testing room a few meters away from the housing room. During the transport, the home cage was protected from light. Before the test, each mouse was placed individually inside a cylinder (11 cm diameter, 20 cm high) in one corner of the OF apparatus. After 1 min the cylinder was lifted and the mouse was allowed to

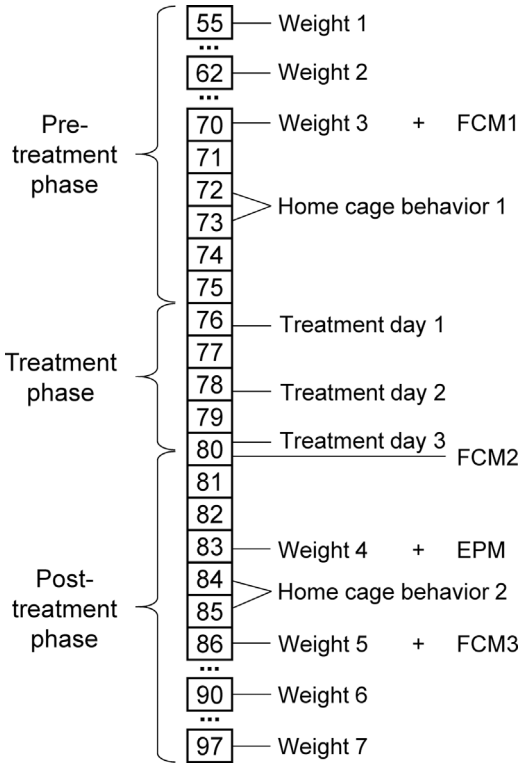

Fig. 1. Time-line diagram.
The experiment was divided into three phases: The pre-treatment phase (postnatal day (PND) 55–75), the treatment phase (PND 76–80), and the post-treatment phase (PND 80–97). FCM = Fecal corticosterone metabolite measurements; FCM1 = Baseline level, FCM2 = Short-term response to treatment, FCM3 = Long-term response to treatment, EPM = elevated plus-maze test.

freely explore the arena for 15 min. The test equipment was thoroughly cleaned with 70% ethanol and dried between subjects.

Mice of control group C1 were transferred to a new cage on the three treatment days, while mice of group C2 received no specific treatment on these days (PND 76, 78, and 80; see Fig. 2). The transfer to the new cage not only reflected a routine cage cleaning process, but also confronted the animals with a new environment. During this procedure, mice were placed in a new Makrolon type III cage with fresh wood shavings and a fresh paper towel. The cage enrichment (mouse house and the wooden stick) was transferred from the previous to the new cage and thus remained the same during the complete treatment phase. Animals stayed in the new cage until the next treatment day or until the next cage cleaning. Mice of group C2 experienced no specific treatment, but were handled on the three treatment days to rule out any effects possibly induced by the handling itself. The handling here simulated a standard weighing procedure, during which the animal was picked up, placed in an empty plastic box positioned on a digital scale, and put back in the home cage immediately afterwards.

2.5. Assessment of welfare indicators

As welfare indicators, corticosterone metabolite concentrations in the feces, body weights, anxiety-like as well as home cage behavior were assessed.

2.5.1. Corticosterone metabolite concentrations

The stress hormone level of mice was monitored non-invasively by measuring fecal corticosterone metabolites (FCM) at three time points [24–26]. Previous studies have shown that variations in FCM reliably reflect effects of external stimuli, such as housing conditions, social interactions, and even unfamiliar bedding [27–30]. Both the short-term (PND 80: FCM2) and long-term response (PND 86: FCM3) to the treatment phase were analyzed during the post-treatment phase (see Fig. 1). Additionally, a reference baseline level was assessed during the pre-treatment phase (PND 70: FCM1; see Fig. 1). Since Touma and colleagues [25] demonstrated that, during the dark phase, a peak of FCM can be found in the feces 4–6 h after the exposure to a stressor, treatment-response fecal samples were collected 3.5–6.5 h after the beginning of the treatment. Thereby, it was ensured that feces comprised only FCM as a response to the treatment but not to the sampling procedure. Feces for the baseline measurement and the long-term response were collected at the same time without prior treatment. For the sample collection, mice of all groups were placed individually in Makrolon cages type III equipped with a thin layer of wood shavings, a paper towel, mouse house, wooden stick, and food and water *ad libitum*. After the expiration of the 3 h sampling period, each mouse was transferred back to its individual home cage. Subsequently, all feces defecated during the 3 h were collected and frozen at −20 °C. Samples were dried and homogenized, and aliquots of 0.05 g were extracted with 1 ml of 80% methanol. For the analysis of the samples, a 5α-pregnane-3β,11β,21-triol-20-one enzyme immunoassay was used, which was established and successfully validated to measure FCM in mice (for details see [25,26]). Intra- and inter-assay coefficients of variation were below 10% and 12%, respectively.

2.5.2. Body weights

Measuring the development of body weights allows for the detection of welfare-related effects [20]. Therefore, each mouse was weighed repeatedly over the course of the pre-treatment (PND 55, 62, 70; Weight 1–3) and the post-treatment phase (PND 83, 86, 90, 97; Weight 4–7; see Fig. 1) using a digital scale (accuracy: 0.1 g; CM 150-1N, Kern, Ballingen, Germany).

2.5.3. Anxiety-like and exploratory behavior on the elevated plus-maze

To assess anxiety-like and exploratory behavior, the elevated plus-maze (EPM) [31] test was performed on PND 83 during the post-

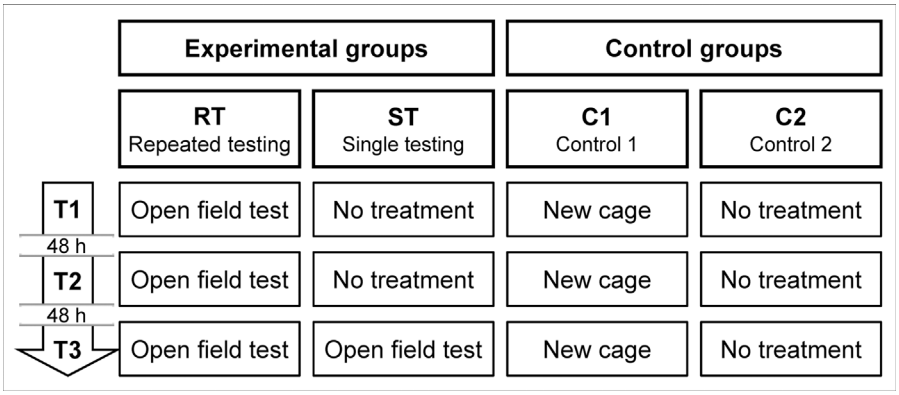

Fig. 2. Schematic overview of the two experimental and the two control groups and their treatments on three treatment days.
Between the three treatment days, there were time intervals of 48 h. Groups RT and ST served as experimental groups. While mice of group RT performed the OF three times, mice of group ST went through the OF only once. Groups C1 and C2 served as controls. Mice of these groups either experienced repeated exposure to a novel environment (C1) or no specific treatment (C2). For details of C1 and C2 see Section 2.4. T1–T3 = Treatment days.

treatment phase (see Fig. 1). The plus-shaped apparatus was elevated 50 cm above the ground. It consisted of two opposing open arms (30 cm × 5 cm) and two opposing closed arms (30 cm × 5 cm) with 20 cm high walls that extended from a central square (5 cm × 5 cm). The two open arms were surrounded by a small lip (4 mm) preventing the mice from falling off. Illumination intensity was 25 lx. The test was performed at the beginning of the dark phase in a testing room a few meters away from the housing room. During the transport, the home cage was protected from light. After spending 1 min in an empty cage, each mouse was individually placed on the central platform facing a closed arm and allowed to freely explore the apparatus for 5 min. The parameters measured were the percentage of time spent on the open arms, the percentage of entries into the open arms, and the percentage of distance traveled on the open arms to assess anxiety-like behavior. The sum of entries into the open and closed arms as well as the total distance were assessed as indicators of exploratory locomotion. The apparatus was thoroughly cleaned with 70% ethanol and dried between subjects. The animal's movements were recorded by a webcam (Webcam Pro 9000, Logitech, Europe S.A., Lausanne, Switzerland) and analyzed by the video tracking system ANY-maze (Version 4.99, Stoelting Co., Wood Dale, USA).

2.5.4. Home cage behavior

Observations of home cage behaviors were performed in the housing room at two time points, during the pre-treatment phase (PND 72 and 73: Home cage behavior 1) and during the post-treatment phase (PND 84 and 85: Home cage behavior 2; see Fig. 1). The observations were conducted over the course of the whole dark phase under red light conditions by an experienced observer (S.S.). Altogether, the behavior of each mouse was observed 40 times during the pre-treatment phase and 40 times during the post-treatment phase. One-zero sampling was performed to record home cage behaviors (Table 1). Observation intervals for each mouse (focal animal sampling) lasted 20 s [32,33]. At the end of the 20 s intervals, the general activity of the mouse at this particular time point was recorded using instantaneous sampling [33] (Table 1). The order in which the mice were observed was pseudo-randomized. For data analysis, the percentage of scans or intervals, respectively, in which each behavior occurred were calculated. For definitions of observed behaviors see Table 1.

2.6. Statistics

The group size was set to n = 12, based on previous studies and recommendations for behavioral phenotyping experiments [9].

Table 1
Definitions of home cage behaviors.

General activity (Instantaneous sampling)	
Active	The mouse is *active* when it is not *inactive*.
Inactive	The mouse is lying or sitting motionlessly, except for tiny whisker, ear or tail movements.
Home cage behaviors (One-zero sampling)	
Stereotypic behavior[a]	
Circling	Climbing in tight circles at the cage lid.
Exploratory behaviors	
Climbing on lid	The mouse does not touch the ground with any paws and holds to the cage lid. The tail can still touch the ground.
Rearing	A mouse raises itself on its hindpaws and stretches its snout into the air.
Maintenance behaviors	
Drinking	A mouse nibbles at a water bottle.
Feeding	A mouse ingests food.
Self-grooming	A mouse scratches, grooms or licks its own body.

Definitions of behaviors are based on previous publications [32,34].
[a] In mice, different forms of stereotypic behaviors exist, but in our study only *circling* was observed in considerable quantities.

Analyses of variance (ANOVAs) were used to analyze corticosterone metabolite concentrations, body weights, and anxiety-like as well as exploratory behaviors on the EPM. To meet the assumptions of parametric analysis, residuals were graphically examined for homoscedasticity and outliers and the Lilliefors corrected Kolmogorov-Smirnov Test was applied. In particular, univariate ANOVA was used to analyze several dependent variables (FCM, anxiety-like and exploratory behaviors on the EPM) with fixed between-subject factor 'group'. ANOVA with repeated measures (RM ANOVA) was performed for the analysis of body weight with within-subjects factor 'time' (PND), fixed between-subject factor 'group', and the interaction of 'group' and 'time'. In order to account for possible violations to sphericity, the Greenhouse-Geisser or Huynh-Feldt correction was applied. Furthermore, to present the magnitude of the reported effects in a standardized metric, effect sizes were calculated as partial eta squared (η^2_p) [35], and all raw data were summarized as means with standard deviations in Supplementary Table S1. Since not all home cage behavior data were normally distributed, non-parametric statistics was applied (Kruskal-Wallis Test). All main effects and interaction terms were tested on local significance level alpha = 0.05, respectively. Data are presented either as bars with means and standard error (SEM) or box plots with medians, 10th, 25th, 75th and 90th percentiles.

All statistical analyses were conducted using the statistical software IBM SPSS Statistics (IBM Version 23, Release 2015). Graphs were created using the software SigmaPlot 12.5 for Windows (Build 12.5.0.38, Systat Software, Inc. 2011).

3. Results

3.1. Fecal corticosterone metabolites

The statistical analysis of fecal corticosterone metabolite (FCM) concentrations did not reveal a significant main effect of group within any sampling point (FCM1: $F_{(3,44)} = 1.203$, $p = 0.320$, $\eta^2_p = 0.076$; FCM2: $F_{(3,44)} = 1.021$, $p = 0.392$, $\eta^2_p = 0.065$; FCM3: $F_{(3,44)} = 0.935$, $p = 0.432$, $\eta^2_p = 0.060$; Fig. 3). Thus, the four groups did neither differ significantly in FCM concentrations before the treatment-phase, nor in

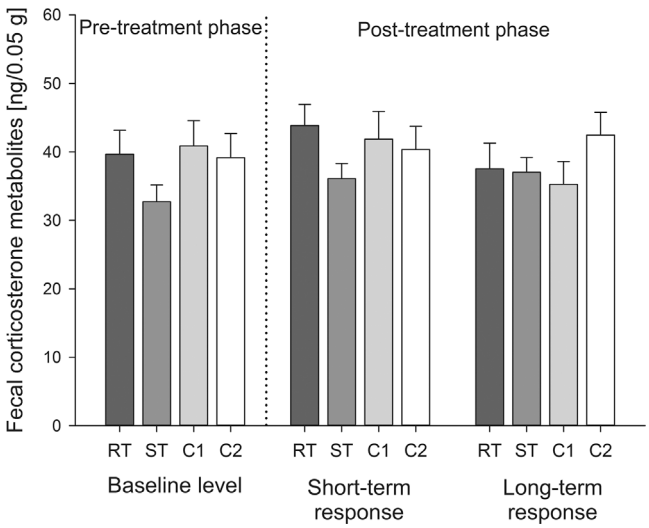

Fig. 3. Fecal corticosterone metabolite concentrations over the course of the experiment. Baseline FCM levels were assessed during the pre-treatment phase (PND 70), while short-term (PND 80) and long-term (PND 86) responses to the treatment phase were assessed subsequently to the third treatment or one week later, respectively. During the treatment phase, mice performed the OF either three times (RT), once (ST), or not at all, but were transferred to a new cage (C1) or received no specific treatment (C2). Data are presented as bars with mean and SEM. Statistics: ANOVA; for details see Section 2.6. Sample size: n = 12/group. There were no significant main effects of group within baseline, short-term and long-term response.

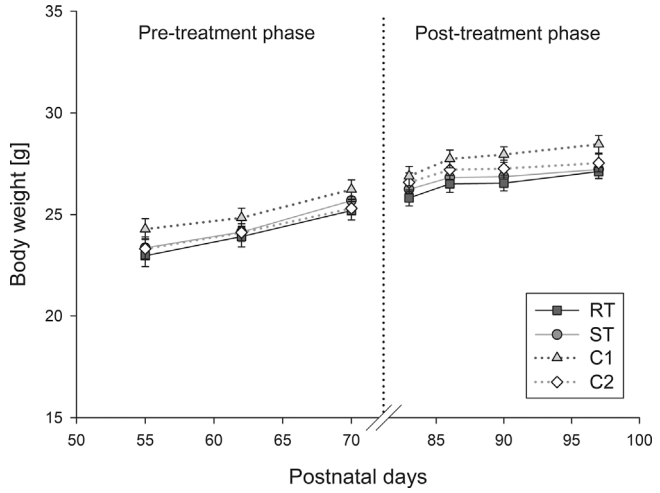

Fig. 4. Body weights during the pre- and post-treatment phase.
Body weights were measured repeatedly over the course of the pre-treatment (PND 55, 62, 70; Weight 1–3) and the post-treatment phase (PND 83, 86, 90, 97; Weight 4–7). Over the course of the treatment phase, mice of group RT performed the OF three times, while mice of group ST performed the OF only once. Mice of control group C1 were exposed to a new cage, whereas mice of control group C2 received no specific treatment. Data are shown as mean + SEM. Statistics: RM ANOVA; for details see Section 2.6. Sample size: n = 12/group. There were neither significant main effects of group within the pre-treatment or post-treatment phase nor group-by-time interactions.

their short-term and long-term responses after the treatment.

3.2. Body weights

Repeated measures ANOVA was performed for both the pre- and the post-treatment phase. Both analyses detected a significant main effect of time, with weights increasing over time (pre-treatment phase: $F_{(1.4,60.3)} = 101.033$, $p < 0.001$; post-treatment phase: $F_{(2.6,112.8)} = 97.607$, $p < 0.001$). Neither during the pre- ($F_{(3,44)} = 1.057$, $p = 0.377$, $\eta^2_p = 0.067$) nor during the post-treatment phase ($F_{(3,44)} = 1.714$, $p = 0.178$, $\eta^2_p = 0.105$), was a difference in body weight between the groups found (Fig. 4). There were also no significant group-by-time interactions during the pre- ($F_{(4.1,60.3)} = 0.294$, $p = 0.885$, $\eta^2_p = 0.020$) and post-treatment phase ($F_{(7.7,112.8)} = 1.624$, $p = 0.129$, $\eta^2_p = 0.100$).

3.3. Anxiety-like and exploratory behavior

The groups did not differ significantly concerning both anxiety-like behavior and exploratory locomotion on the EPM (anxiety-like behavior: percentage of time on open arms: $F_{(3,44)} = 1.134$, $p = 0.346$, $\eta^2_p = 0.072$; Fig. 5A; percentage of entries into open arms: $F_{(3,44)} = 2.463$, $p = 0.075$, $\eta^2_p = 0.144$; Fig. 5B; percentage of distance traveled on open arms: $F_{(3,44)} = 0.997$, $p = 0.403$, $\eta^2_p = 0.064$; exploratory locomotion: sum of entries: $F_{(3,44)} = 0.807$, $p = 0.497$, $\eta^2_p = 0.052$; Fig. 5C; total distance traveled: $F_{(3,44)} = 2.273$, $p = 0.093$, $\eta^2_p = 0.134$; Fig. 5D).

3.4. Home cage behavior

Both during the pre- and post-treatment phase, the four groups did not differ significantly in their general activity (active: pre-treatment: $\chi^2_{(3)} = 6.711$, $p = 0.082$; post-treatment: $\chi^2_{(3)} = 5.251$, $p = 0.154$; Fig. 6), stereotypic behavior (circling: pre-treatment: $\chi^2_{(3)} = 2.047$, $p = 0.563$; post-treatment: $\chi^2_{(3)} = 4.278$, $p = 0.233$; Fig. S1A), exploratory behaviors (climbing on lid: pre-treatment: $\chi^2_{(3)} = 1.864$, $p = 0.601$; post-treatment: $\chi^2_{(3)} = 0.402$, $p = 0.940$; Fig. S1B; rearing: pre-treatment: $\chi^2_{(3)} = 1.083$, $p = 0.781$, post-treatment: $\chi^2_{(3)} = 1.615$, $p = 0.656$; Fig. S1C), and maintenance behaviors (drinking: pre-treatment: $\chi^2_{(3)} = 0.994$, $p = 0.803$, post-treatment: $\chi^2_{(3)} = 4.564$, $p = 0.207$; feeding: pre-treatment: $\chi^2_{(3)} = 6.492$, $p = 0.090$, post-treatment: $\chi^2_{(3)} = 3.461$, $p = 0.326$; self-grooming: pre-treatment: $\chi^2_{(3)} = 2.354$, $p = 0.502$, post-treatment: $\chi^2_{(3)} = 5.493$, $p = 0.139$; Fig. S1D).

4. Discussion

For the purpose of refining animal experiments, a prospective assessment of the severity of each procedure is mandatory. Since there are growing concerns on the severity of less invasive studies, a systematic investigation of the impact of repeated behavioral testing is necessary and, so far, widely missing. Here, we concentrated on one of the most frequently used tests in behavioral phenotyping, the open-field (OF). The aim of the present study was to elucidate whether repeated OF testing has any effects on the welfare of laboratory mice compared to single OF testing. For a comprehensive picture of the animals' welfare, we measured several established welfare indicators (for a review see [36]), including physiological, physical, and behavioral measures in C57BL/6J mice, the most widely used laboratory strain. Overall, we did not detect significant differences between mice that performed the OF three times, once or not at all in any of the 13 welfare-related measures. With respect to our hypothesis, there is thus no evidence that group differences were not due to chance.

A widely used and highly sensitive indicator for the degree of stress is the activity of the hypothalamic-pituitary-adrenal (HPA) axis (for a review see [19]). In the present study, we performed a non-invasive method using fecal samples, from which corticosterone metabolites were extracted [24–26]. By applying this technique, previous studies have shown that slight environmental manipulations cause significant variations in fecal corticosterone metabolites (FCM) in mice. More specifically, exposure of pregnant females to unfamiliar males' soiled bedding induced a significantly higher increase in FCM compared to fresh bedding [29]. Furthermore, increased housing density as well as social defeat profoundly elevated FCM levels [28,30]. The impact on FCM was also demonstrated for more invasive procedures, such as intra-bone marrow transplantation, which caused significantly increased FCM values [37]. Notably, such variations in HPA axis activity are often reflected by changes in body weight. A stress-induced increase in corticosterone levels thereby frequently correlates with a decrease in body weight [20–22]. In contrast to these examples, no differences between the four groups were detected with respect to FCM and body weights in the present study. Thus, neither single nor repeated OF testing was found to cause an acute or long-term activation of the adrenocortical system. Given that FCM measurement has been proven to be a sensitive method to detect even minor treatment effects, these findings underline the assumption that the treatments did not affect the animals' physiological state, although we cannot prove a null effect on the basis of this study. With respect to the EU guidelines, this may question the classification of repeated versus single testing in two different severity categories (i.e., mild versus below threshold).

Elevated levels of corticosterone are commonly regarded as evidence for compromised welfare. There are, however, limitations to this interpretation as levels do not only rise in response to adverse experiences but also to positive experiences, such as sexual activity or anticipation of a reward [23,36,38,39]. Therefore, a comprehensive welfare assessment should also consider behavioral measures, including the observation of spontaneous behavior in the home cage as well as the performance in welfare-related behavioral paradigms [23]. One frequently used behavioral test for assessing state anxiety in mice is the elevated plus-maze (EPM) test [31]. The EPM is not only pharmacologically validated and sensitive in the acquisition of anxiety-like behavior, but also known for the detection of subtle welfare-related changes. For example, it is well documented that environmental enrichment influences the animals' behavior in the EPM in an anxiolytic way [40–43]. Also social experiences are well-known for their impact on

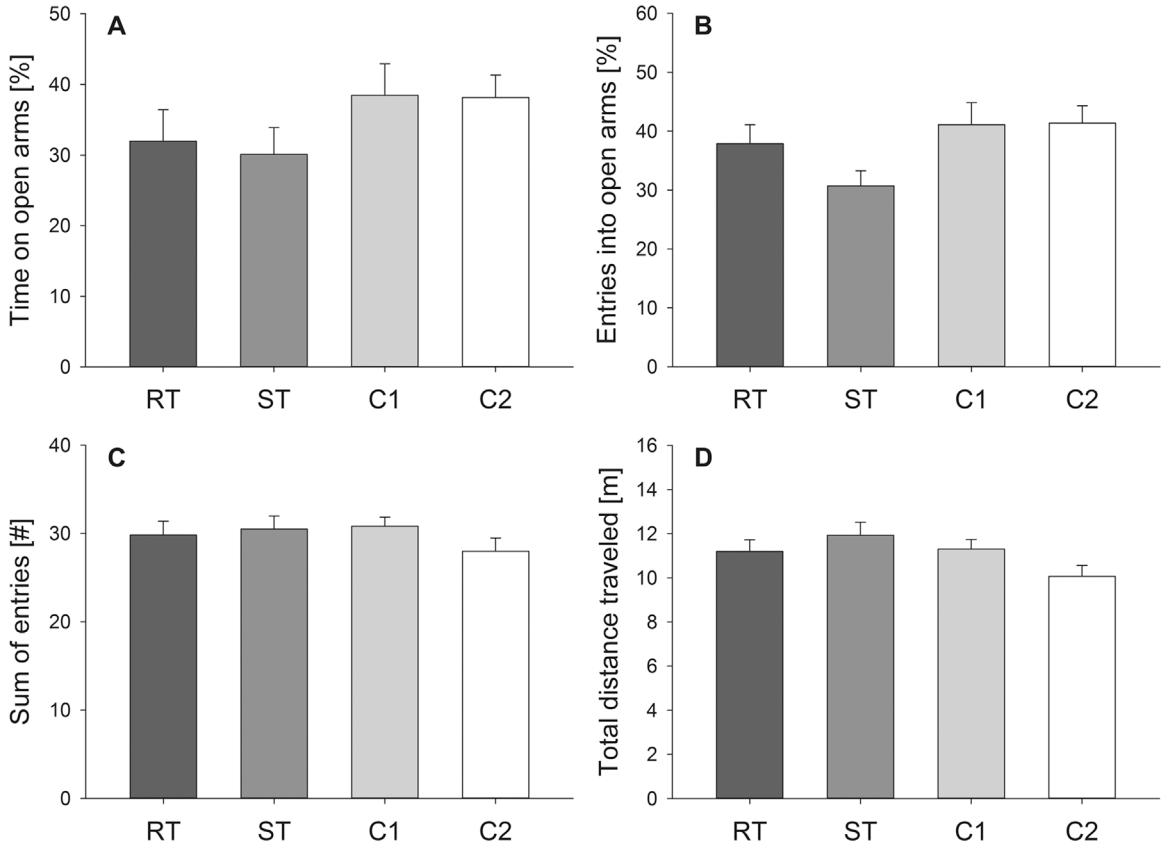

Fig. 5. Anxiety-like behavior and exploratory locomotion measured in the elevated plus-maze.
(A) Percentage of time on open arms, (B) Percentage of entries into open arms, (C) Sum of entries into open and closed arms, and (D) Total distance traveled. Mice of the RT group performed the OF three times, while mice of the ST group performed the OF only once. Mice of control group C1 were transferred to a new cage on each treatment day, while mice of control group C2 received no specific treatment. Data are presented as bars with mean and SEM. Statistics: ANOVA; for details see Section 2.6. Sample size: n = 12/group. There were no significant main effects of group.

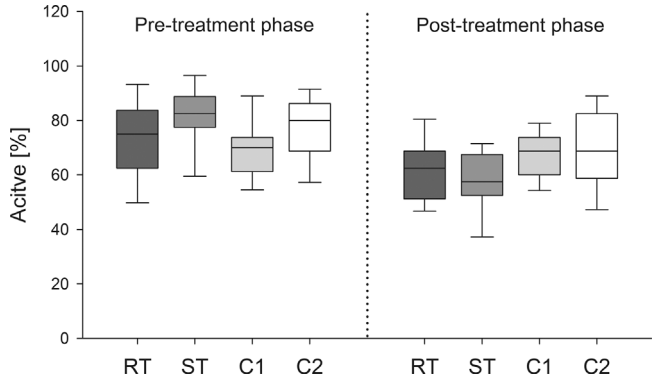

Fig. 6. Home cage activity during the pre- and post-treatment phase.
Percentage of scans in which mice were *active*. Over the course of the treatment phase, mice performed the OF either three times (RT), once (ST) or not at all, but were transferred to a new cage (C1) or received no specific treatment (C2). Data are shown as box plots with median, 10th, 25th, 75th and 90th percentiles. Statistics: Kruskal-Wallis Test; for details see Section 2.6. Sample size: n = 12/group. There were no significant main effects of group within the pre- and post-treatment phase.

EPM performance, with, for example, social defeat causing increased levels of anxiety-like behavior [28]. Furthermore, even the way of handling the animals has an effect on the behavior in the EPM [44,45]. In the present study, however, no significant differences in anxiety-like behavior on the EPM were detected between the four groups, arguing against any treatment-specific changes in emotionality.

The observation of spontaneous behavior in the familiar home cage yields further insights into animal welfare. In this context, stereotypies are widely discussed. Defined as repetitive, invariant behavior patterns without any obvious goal or function, they are indicative of impaired welfare and should therefore be taken seriously as a warning signal of potential suffering [46,47]. With estimated prevalence rates of 50%, mice develop several different types of stereotypy under laboratory conditions, including repetitive bar-mouthing or repetitive jumping [47]. Similarly, compromised welfare can be indicated by fluctuations in the activity pattern, since low frequencies and durations of sleep behavior were found to correlate with indicators of elevated physiological and physical stress (e.g., rats [48]). Also, a high amount of stationary behavior (awake but inactive) was shown to be indicative of poor welfare [42]. In the present study, the overall level of stereotypies was remarkably low. Furthermore, none of the home cage behaviors, including stereotypic behavior and activity level, were significantly influenced by repeated or single OF testing or control procedures. While these findings do not prove equality of treatment groups, they blend in perfectly with the overall picture.

Taken together, testing male C57BL/6J mice in an OF either once or repeatedly may be regarded as a humane behavioral procedure. Before drawing general conclusions, however, results should be confirmed in females as well as in other strains and species (e.g., see [49,50]). From a broader perspective, these findings are not only important for the classification of procedures in behavioral phenotyping studies, but also underline the fundamental need for a systematic and evidence-based severity assessment of any procedures involving living animals. Many of the opinions concerning the classification of standard techniques are not based upon an objective and reproducible assessment of welfare, but upon a subjective evaluation. With respect to the 3R-concept (see [51]), such studies also provide the basis of refinement approaches.

Only when it is clear, how severe a procedure is, can refinements be addressed that reduce the degree of pain, suffering, distress or lasting harm in an experiment. Furthermore, since data quality can be significantly impaired by poor welfare, refining animal experiments would also improve the scientific validity [52].

Acknowledgements

This work was supported by a grant from the German Research Foundation (DFG) to S.H.R. (RI 2488/3-1: http://gepris.dfg.de/gepris/projekt/283089959) and to N.S. (SFB/TRR58, Project A01: http://www.dfg.de/foerderung/programme/listen/projektdetails/index.jsp?id=44541416). The authors thank Edith Ossendorf and Edith Klobetz-Rassam for excellent technical assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbr.2017.08.029.

References

[1] C.M. Sherwin, S.B. Christiansen, I.J. Duncan, H.W. Erhard, D.C. Lay, J.A. Mench, C.E. O'Connor, J.C. Petherick, Guidelines for the ethical use of animals in applied ethology studies, Appl. Anim. Behav. Sci. 81 (2003) 291–305, http://dx.doi.org/10.1016/S0168-1591(02)00288-5.

[2] J. Jones, An applied approach to the assessment of severity, in: C.F.M. Hendrikson, D.B. Morton (Eds.), Hum. Endpoints Anim. Exp. Biomed. Res. Royal Society for Medicine Press, London, 1999, pp. 40–47.

[3] Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes, Off. J. Eur. Union (OJ) (2010).

[4] K. Hohlbaum, B. Bert, S. Dietze, R. Palme, H. Fink, C. Thöne-Reineke, Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice? Assessing the degree of distress, PLoS One 12 (2017) e0179588, http://dx.doi.org/10.1371/journal.pone.0179588.

[5] Expert Working Group on Severity Classification Criteria, Expert working group on severity classification of scientific procedures performed on animals: final report, Bruss. Eur. Com. (2009). http://ec.europa.eu/environment/chemicals/lab_animals/pdf/report_ewg.pdf (accessed June 27, 2017).

[6] D.C. Rogers, E.M.C. Fisher, S.D.M. Brown, J. Peters, A.J. Hunter, J.E. Martin, Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment, Mamm. Genome 8 (1997) 711–713, http://dx.doi.org/10.1007/s003359900551.

[7] J.N. Crawley, R. Paylor, A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice, Horm. Behav. 31 (1997) 197–211, http://dx.doi.org/10.1006/hbeh.1997.1382.

[8] J.N. Crawley, Behavioral phenotyping of rodents, Comp. Med. 53 (2003) 140–146.

[9] J.N. Crawley, Behavioral phenotyping strategies for mutant mice, Neuron 57 (2008) 809–818, http://dx.doi.org/10.1016/j.neuron.2008.03.001.

[10] F.J. Van Der Staay, T. Steckler, The fallacy of behavioral phenotyping without standardisation, Genes Brain Behav. 1 (2002) 9–13, http://dx.doi.org/10.1046/j.1601-1848.2001.00007.x.

[11] Federal Ministry of Food and Agriculture, Number of experimentalanimals 2015. Severity of experiments (2016). http://www.bmel.de/DE/Tier/Tierschutz/_texte/TierschutzTierforschung.html?docId=8596776 (accessed June 27, 2017).

[12] C. Belzung, Chapter 4.11: Measuring rodent exploratory behavior, in: R.T. Crusio, W.E. Gerlai (Eds.), Tech. Behav. Neural Sci., 1999, pp. 738–749, doi:http://dx.doi.org/10.1016/S0921-0709(99)80057-1.

[13] L. Prut, C. Belzung, The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review, Eur. J. Pharmacol. 463 (2003) 3–33, http://dx.doi.org/10.1016/S0014-2999(03)01272-X.

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

[15] R.N. Walsh, R.A. Cummins, The open-field test: a critical review, Psychol. Bull. 83 (1976) 482–504, http://dx.doi.org/10.1037/0033-2909.83.3.482.

[16] J. Archer, Tests for emotionality in rats and mice: a review, Anim. Behav. 21 (1973) 205–235, http://dx.doi.org/10.1016/S0003-3472(73)80065-X.

[17] V.J. Bolivar, B.J. Caldarone, A.A. Reilly, L. Flaherty, Habituation of activity in an open field: a survey of inbred strains and F1 hybrids, Behav. Genet. 30 (2000) 285–293.

[18] B. Lecorps, H.G. Rödel, C. Féron, Assessment of anxiety in open field and elevated plus maze using infrared thermography, Physiol. Behav. 157 (2016) 209–216, http://dx.doi.org/10.1016/j.physbeh.2016.02.014.

[19] V. Riley, Psychoneuroendocrine influences on immunocompetence and neoplasia, Science 212 (1981) 1100–1109, http://dx.doi.org/10.1126/science.7233204.

[20] N. Sachser, C. Lick, Social experience, behavior, and stress in guinea pigs, Physiol. Behav. 50 (1991) 83–90, http://dx.doi.org/10.1016/0031-9384(91)90502-f.

[21] K.L.K. Tamashiro, M.M.N. Nguyen, M.M. Ostrander, S.R. Gardner, L.Y. Ma, S.C. Woods, R.R. Sakai, Social stress and recovery: implications for body weight and body composition, Am. J. Physiol. Regul. Integr. Comp. Physiol. 293 (2007) R1864–R1874, http://dx.doi.org/10.1152/ajpregu.00371.2007.

[22] J.Y. Jeong, D.H. Lee, S.S. Kang, Effects of chronic restraint stress on body weight, food intake, and hypothalamic gene expressions in mice, Endocrinol. Metab. Seoul 28 (2013) 288–296, http://dx.doi.org/10.3803/EnM.2013.28.4.288.

[23] M.S. Dawkins, Using behaviour to assess animal welfare, Anim. Welf. 13 (2004) S3–S7.

[24] M. Lepschy, C. Touma, R. Palme, Faecal glucocorticoid metabolites: how to express yourself – comparison of absolute amounts versus concentrations in samples from a study in laboratory rats, Lab. Anim. 44 (2010) 192–198, http://dx.doi.org/10.1258/la.2009.009082.

[25] C. Touma, N. Sachser, E. Möstl, R. Palme, Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice, Gen. Comp. Endocrinol. 130 (2003) 267–278, http://dx.doi.org/10.1016/S0016-6480(02)00620-2.

[26] C. Touma, R. Palme, N. Sachser, Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones, Horm. Behav. 45 (2004) 10–22, http://dx.doi.org/10.1016/j.yhbeh.2003.07.002.

[27] V. Marashi, A. Barnekow, E. Ossendorf, N. Sachser, Effects of different forms of environmental enrichment on behavioral, endocrinological, and immunological parameters in male mice, Horm. Behav. 43 (2003) 281–292, http://dx.doi.org/10.1016/S0018-506X(03)00002-3.

[28] F. Jansen, R.S. Heiming, L. Lewejohann, C. Touma, R. Palme, A. Schmitt, K.P. Lesch, N. Sachser, Modulation of behavioural profile and stress response by 5-HTT genotype and social experience in adulthood, Behav. Brain Res. 207 (2010) 21–29, http://dx.doi.org/10.1016/j.bbr.2009.09.033.

[29] R.S. Heiming, C. Bodden, F. Jansen, L. Lewejohann, S. Kaiser, K.-P. Lesch, R. Palme, N. Sachser, Living in a dangerous world decreases maternal care: a study in serotonin transporter knockout mice, Horm. Behav. 60 (2011) 397–407, http://dx.doi.org/10.1016/j.yhbeh.2011.07.006.

[30] A. Nicholson, R.D. Malcolm, P.L. Russ, K. Cough, C. Touma, R. Palme, M.V. Wiles, The response of C57BL/6J and BALB/cJ mice to increased housing density, J. Am. Assoc. Lab. Anim. Sci. 48 (2009) 740–753.

[31] R.G. Lister, The use of a plus-maze to measure anxiety in the mouse, Psychopharmacology (Berl.) 92 (1987) 180–185.

[32] A.N. Gross, S.H. Richter, A.K.J. Engel, H. Würbel, Cage-induced stereotypies, perseveration and the effects of environmental enrichment in laboratory mice, Behav. Brain Res. 234 (2012) 61–68, http://dx.doi.org/10.1016/j.bbr.2012.06.007.

[33] P. Martin, M. Bateson, Measuring Behaviour: An Introductory Guide, Cambridge University Press, Cambridge, 2007.

[34] L. Lewejohann, V. Kloke, R.S. Heiming, F. Jansen, S. Kaiser, A. Schmitt, K.P. Lesch, N. Sachser, Social status and day-to-day behaviour of male serotonin transporter knockout mice, Behav. Brain Res. 211 (2010) 220–228, http://dx.doi.org/10.1016/j.bbr.2010.03.035.

[35] D. Lakens, Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs, Front. Psychol. 4 (2013), http://dx.doi.org/10.3389/fpsyg.2013.00863.

[36] E.S. Paul, E.J. Harding, M. Mendl, Measuring emotional processes in animals: the utility of a cognitive approach, Neurosci. Biobehav. Rev. 29 (2005) 469–491, http://dx.doi.org/10.1016/j.neubiorev.2005.01.002.

[37] U. Pfeiffenberger, T. Yau, D. Fink, A. Tichy, R. Palme, M. Egerbacher, T. Rülicke, Assessment and refinement of intra-bone marrow transplantation in mice, Lab. Anim. 49 (2015) 121–131, http://dx.doi.org/10.1177/0023677214559627.

[38] B. Buwalda, J. Scholte, S.F. de Boer, C.M. Coppens, J.M. Koolhaas, The acute glucocorticoid stress response does not differentiate between rewarding and aversive social stimuli in rats, Horm. Behav. 61 (2012) 218–226, http://dx.doi.org/10.1016/j.yhbeh.2011.12.012.

[39] A.S. Mallien, R. Palme, J. Richetto, C. Muzzillo, S.H. Richter, M.A. Vogt, D. Inta, M.A. Riva, B. Vollmayr, P. Gass, Daily exposure to a touchscreen-paradigm and associated food restriction evokes an increase in adrenocortical and neural activity in mice, Horm. Behav. 81 (2016) 97–105, http://dx.doi.org/10.1016/j.yhbeh.2016.03.009.

[40] H. Prior, N. Sachser, Effects of enriched housing environment on the behaviour of young male and female mice in four exploratory tasks, J. Exp. Anim. Sci. 37 (1995) 57–68.

[41] P. Chapillon, C. Manneche, C. Belzung, J. Caston, Rearing environmental enrichment in two inbred strains of mice: 1. Effects on emotional reactivity, Behav. Genet. 29 (1999) 41–46.

[42] U.A. Abou-Ismail, H.D. Mahboub, The effects of enriching laboratory cages using various physical structures on multiple measures of welfare in singly-housed rats, Lab. Anim. 45 (2011) 145–153, http://dx.doi.org/10.1258/la.2011.010149.

[43] V. Roy, C. Belzung, C. Delarue, P. Chapillon, Environmental enrichment in BALB/c mice: effects in classical tests of anxiety and exposure to a predatory odor, Physiol. Behav. 74 (2001) 313–320, http://dx.doi.org/10.1016/S0031-9384(01)00561-3.

[44] J.L. Hurst, R.S. West, Taming anxiety in laboratory mice, Nat. Methods 7 (2010) 825–826, http://dx.doi.org/10.1038/nmeth.1500.

[45] K. Gouveia, J.L. Hurst, Reducing mouse anxiety during handling: effect of experience with handling tunnels, PLoS One 8 (2013) e66401, http://dx.doi.org/10.1371/journal.pone.0066401.

[46] G.J. Mason, Stereotypies: a critical review, Anim. Behav. 41 (1991) 1015–1037, http://dx.doi.org/10.1016/S0003-3472(05)80640-2.

[47] G.J. Mason, N.R. Latham, Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? Anim. Welf. 13 (2004) S57–S69.

[48] U.A. Abou-Ismail, O.H.P. Burman, C.J. Nicol, M. Mendl, Can sleep behaviour be used as an indicator of stress in group-housed rats (*Rattus norvegicus*)? Anim. Welf. 16 (2007) 185–188.
[49] J.C. DeFries, J.P. Hegmann, M.W. Weir, Open-field behavior in mice: evidence for a major gene effect mediated by the visual system, Science 154 (1966) 1577–1579, http://dx.doi.org/10.1126/science.154.3756.1577.
[50] D.A. Blizard, Situational determinants of open-field behaviour in Mus musculus, Br. J. Psychol. 62 (1971) 245–252, http://dx.doi.org/10.1111/j.2044-8295.1971.tb02034.x.
[51] W.M. Russell, R.L. Burch, The Principles of Humane Experimental Technique, Methuen, London, 1959.
[52] J.P. Garner, Stereotypies and other abnormal repetitive behaviors: potential impact on validity, reliability, and replicability of scientific outcomes, ILAR J. 46 (2005) 106–117, http://dx.doi.org/10.1093/ilar.46.2.106.