Journal of Neuroscience Methods 293 (2018) 375–388

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Research paper

An improved procedure for integrated behavioral z-scoring illustrated with modified Hole Board behavior of male inbred laboratory mice

M. (Maaike) Labots[a,b,*,1], M.C. (Marijke) Laarakker[a,b,2], D. (Dustin) Schetters[a,3], S.S. (Saskia) Arndt[a,b], H.A. (Hein) van Lith[a,b]

[a] Department of Animals in Science and Society, Division of Animal Welfare & Laboratory Animal Science, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands
[b] Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

HIGHLIGHTS

- Using pooled data as reference group in calculating integrated behavioral z-scores.
- Applying new method of calculating integrated z-scores in mouse behavior analyses.
- Improving generalizability in measuring mouse behavior.
- Integrating effect sizes for improving the practical importance of the results.

ARTICLE INFO

Article history:
Received 9 May 2017
Received in revised form
10 September 2017
Accepted 12 September 2017
Available online 20 September 2017

Keywords:
Activity
Anxiety
Improved procedure
Inbred strain differences
Integrated behavioral z-scoring
Mice
Modified Hole Board

ABSTRACT

Background: Guilloux et al. introduced: integrated behavioral z-scoring, a method for behavioral phenotyping of mice. Using this method multiple ethological variables can be combined to show an overall description of a certain behavioral dimension or motivational system. However, a problem may occur when the control group used for the calculation has a standard deviation of zero or when no control group is present to act as a reference group.
New method: In order to solve these problems, an improved procedure is suggested: taking the pooled data as reference. For this purpose a behavioral study with male mice from three inbred strains was carried out. The integrated behavioral z-scoring methodology was applied, thereby taking five different reference group options. The outcome regarding statistical significance and practical importance was compared.
Results: Significant effects and effect sizes were influenced by the choice of the reference group. In some cases it was impossible to use a certain population and condition, because one or more behavioral variables in question had a standard deviation of zero. Based on the improved method, male mice from the three inbred strains differed regarding activity and anxiety.
Comparison with existing method: Taking the method described by Guilloux et al. as basis, the present procedure improved the generalizability to all types of experimental designs in animal behavioral research.
Conclusions: To solve the aforementioned problems and to avoid getting the diagnosis of data manipulation, the pooled data (combining the data from all experimental groups in a study) as reference option is recommended.

Abbreviations: ANOVA, analysis of variance; EDTA, Ethylenediaminetetraacetic acid; EPM, elevated plus maze; *i.p.*, intra-peritoneal; LD, light-dark box; mHB, modified Hole Board; MPD, Mouse Phenome Database; OF, open field; RIA, radio-immunoassay; SD, standard deviation; SEM, standard error of the mean; SPF, specified pathogen free.

* Corresponding author at: Department of Animals in Science and Society, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 2, P.O. Box 80166, 3508 TD Utrecht, The Netherlands.
E-mail addresses: M.Labots@uu.nl (M. Labots), Marijke.Laarakker@gmail.com (M.C. Laarakker), D.Schetters@vumc.nl (D. Schetters), S.S.Arndt@uu.nl (S.S. Arndt), H.A.vanLith@uu.nl (H.A. van Lith).
[1] Current address: Department of Animals in Science and Society, Division of Behavioural Neuroscience, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands.
[2] Current address: Boston Scientific Nederland B.V., Nieuwegein, The Netherlands.
[3] Current address: Department of Anatomy and Neurosciences, VU University Medical Center, Amsterdam, The Netherlands.

http://dx.doi.org/10.1016/j.jneumeth.2017.09.003
0165-0270/

1. Introduction

Unconditioned ethological tests, that are used to analyze anxiety and activity profiles in mice, often generate many behavioral variables (Ennaceur and Chazot 2016). For example, in the multidimensional modified Hole Board (mHB) 35 behavioral variables can be measured or calculated (Laarakker et al., 2008). To organize multiple related behavioral variables into more manageable and meaningful information, composite variables can be created. Furthermore, using composite variables is also a common practice for controlling Type I error rate (Song et al., 2013). A composite variable is a variable made up of two or more variables that are highly related to one another conceptually or statistically. Principal component and factor analysis are one method for creating composite variables. Z transformation and subsequently summing the obtained z-scores is another method to create a composite variable. For both methods the original variables are assumed to be scaled at the numeric level (*i.e.* continuous scale of measurement or count based measures). However, the use of z-scores does not require the strict criteria one would need for a principal component/factor analysis. For example, the subject to variable ratio should be at least 3:1 and the principal component/factor analysis usually require a sample size >100 (Budaev 2010 and references cited therein), implying that for many rodent behavioral studies a principal component/factor analysis cannot be justified.

To circumvent the difficulties associated with principal component/factor analyses, Guilloux et al. (2011) introduced a method regarding behavioral phenotyping in mice based on z-scores, termed: integrated behavioral z-scoring. Using this method, multiple ethological variables measuring different aspects of behavior can be combined to show an overall description of that behavioral dimension or motivational system. The design of the behavioral study in mice by Guilloux et al. (2011) was such that in their animal experiments there was always a control group, which served as reference group. However, for some animal studies it is not possible to identify a clear-cut control group. For example, in a study where inbred strains of mice are behaviorally phenotyped and compared (*e.g.* Trullas and Skolnick 1993: single factor design; between-group factor, *inbred strain*), it is difficult to discern which experimental group will serve as the reference group. Another example may be a behavioral mouse study with both male and female inbred animals (*e.g.* Laarakker et al., 2011: $2 \times 2 \times 2$ mixed factorial design with repeated measures on the third factor; between-group factors, *sex* and *inbred strain*; within-group factor, *treatment*), or a behavioral study with genetically modified mouse models with different genetic background (*e.g.* Holmes et al., 2003: $2 \times 2 \times 3$ factorial design; between-group factors, *sex*, *inbred strain* and *genotype of serotonin transporter gene*). An additional problem arises when in the control group, the standard deviation of a certain behavioral variable is zero. In this case, integrated behavioral z-score calculation is not possible. This prompted us to perform a behavioral study (2×3 mixed factorial design with repeated measures on the first factor; within-group factor, *treatment*; between-group factor, *inbred strain*) with mice from three inbred strains using different reference groups in order to address these issues.

In the present study, the A/J, BALB/cJ and C57BL/6J mouse inbred strains were behaviorally characterized using the mHB test (Labots et al., 2015). These strains show distinct behaviors in tests of anxiety, and are therefore often used in preclinical anxiety studies. The behavior of male animals before and after an acute aversive stimulus (intra-peritoneal (*i.p.*) injection with saline) was compared to test for the robustness of the behavioral phenotypes. Subsequently, the integrated behavioral z-scoring methodology was applied for the obtained ethological results and five different reference group options were selected. The outcomes regarding statistical significance and practical importance were compared. To overcome the aforementioned problems we suggest to take the pooled data as reference. Since the circulating corticosteroid level has often been found to correlate with the intensity of the anxiogenic situation (Korte 2001), corticosterone levels were measured in blood plasma collected before and after the behavioral test.

To illustrate that choosing a reference population may also be of relevance for preclinical drug studies that include several vehicle controls, data from Laarakker et al. (2010: 3×3 factorial design; between-group factors, *inbred strain* and *treatment*) were re-analyzed using the integrated behavioral z-scoring methodology. Furthermore, the present inbred strain differences regarding anxiety-related behavior and activity in the mHB were compared with published results for these behaviors in the elevated plus maze (EPM), light-dark box (LD) and open field (OF), thereby using our improved procedure. Finally, the improved integrated behavioral z-scoring method gave us also the opportunity to study substrain differences. For this purpose we compared the activity and anxiety-related behavioral phenotypes in the mHB of A/J, BALB/cJ and C57BL/6J male mice (present study) with those of A/JCrl, BALB/cAnNCrl and C57BL/6NCrl male animals (our previously published data; Arndt et al., 2009).

2. Materials and methods

2.1. Ethical note

The protocol of the experiment was approved by the Ethics Committee for Animal Experiments of Utrecht University & University Medical Center Utrecht, Utrecht-The Netherlands (approval number: 0408.1201/vervolg 1). Further, all animal experiments followed the Dutch '*Code of Practice for monitoring the welfare of laboratory animals*' and refer to the '*Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research*' (National Research Council of the National Academies, 2003). To the best of our abilities the present animal study is reported in accordance with the so-called ARRIVE guidelines (http://www.nc3rs.org.uk/arrive-guidelines; Kilkenny et al., 2010).

2.2. Animals and housing

The study was performed with naïve male mice from the following inbred strains: A/J ($n = 15$, albino), BALB/cJ ($n = 15$, albino) and C57BL/6J ($n = 15$, pigmented/black). The number of animals used in this study was calculated using the Russ Lenth's Power and Sample Size software (http://www.stat.uiowa.edu/~rlenth/Power/). Before behavioral testing $n = 3$ BALB/cJ mice were taken out of the experiment due to health related problems not caused by the experimental set-up. To avoid possible effects of estrous cycle on behavioral variables (Meziane et al., 2007), only male mice were used in this study. The animals were bred by and purchased from The Jackson Laboratory (Bar Harbor, ME, USA; A/J stock# 000646, BALB/cJ stock# 000651, C57BL/6J stock# 000664). Charles River Nederland B.V. (Maastricht, The Netherlands) coordinated the shipping of the animals from The Jackson Laboratory to the Utrecht University. All mice studied were purchased as specified pathogen free (SPF) animals. The mice were 4–5 weeks old upon arrival (body weight (g), mean ± standard error of the mean [SEM] and range: A/J, 17.3 ± 0.5 and 12.8–19.4; BALB/cJ, 18.0 ± 0.5 and 15.2–20.9; C57BL/6J, 13.2 ± 0.3 and 11.4–16.4). All animals were housed at the Central Laboratory Animal Research Facility of Utrecht University (conventional, non-SPF) for three weeks (pre-experimental period) to habituate prior to behavioral testing. Testing took place in the same room in which the habituation took place. Testing equipment had been installed in this room prior to arrival of the animals. The animal room was sound-attenuated. Relative humidity

was kept at a constant level of approximately 50 ± 5%, the ambient temperature was maintained at 21 ± 2 °C and the ventilation rate was 15–20 air changes per hour. Background noise was provided by a constantly playing radio. All mice were housed in social groups (three animals of the same inbred strain per cage) directly after arrival in enriched, wire topped Macrolon® Type II L cages (size: 365 × 207 × 140 mm, floor area 530 cm^2; Techniplast S.p.A., Buguggiate, Italy) (randomized cage position). Within a cage, all animals were earmarked (ear notches) for later recognition and numbering. Enrichment, besides standard bedding material (Aspen chips: Abedd-Dominik Mayr KEG, Köflach, Austria), included a cardboard shelter (Tecnilab-BMI B.V., Someren, The Netherlands), a tissue and a small amount (less than a hand full) of paper shreds (EnviroDri®: Tecnilab-BMI B.V.). The mice had *ad libitum* access to tap water and standard mouse chow [Rat and Mouse Breeder and Grower Expanded--RM(E)), Special Diet Services, Essex, UK]). The light:dark cycle was reversed (white light: 19:00–07:00 h, maximal 150 lx; red light: 07:00–19:00 h, maximal 5 lx). During the habituation period, all mice were handled at least four times a week for a few minutes by an experienced person (D. Schetters) who also performed the behavioral tests. Handling included picking up the animal at the tail base, placing it on the hand or arm and restraining it by hand for a few seconds at random times of the day.

2.3. Modified Hole Board (mHB) test

The behavioral tests were all performed in the mHB (Labots et al., 2015) in a randomized manner. This set-up combines the features of an OF, a hole board test and a LD (Laarakker et al., 2011; Labots et al., 2015). Behavioral testing was performed between 10:00 and 14:00 h (*i.e.* during the active phase of the animals) under red-light conditions. A detailed description of the set-up can be found elsewhere (Laarakker et al., 2011). As in previous studies (*e.g.* Laarakker et al., 2008, 2011), several variables indicative for avoidance behavior, risk assessment, arousal, locomotion, exploration, immobility and social affinity were scored. However, only the results for the variables associated with anxiety-related behavior and activity have been used here. An overview of these variables can be found in supplementary Table S1. Avoidance behavior is assessed using the variables *'total number of board entries'*, *'latency until the first board entry'*, and *'percentage of time on the board'*. The *'total number of stretched attends'* and the *'latency until the first stretched attend'* are measures of risk assessment. Behaviors like *'self-grooming'* (total number, latency and percentage) and *'defecation'* (total number and latency) indicate the arousal state of an animal. *'Line crossings'* (total number and latency) were counted to assess horizontal locomotion, whereas *'rearing'* (total number and latency, in the box or on the board) is a measure for vertical locomotion. *'Motionless'* (total number, latency and percentage) was also scored.

For obvious reasons, it was not possible to perform the behavioral tests blinded with respect to the *strain* (due to the coat color of the animals, see section *Materials and methods – Animals and housing*: 2.2) and the *treatment* (*i.p.* injection or control, see section *Materials and methods – Experimental protocol and blood sampling*: 2.4). However, due to the large number of behaviors observed and the absence of specific predictions regarding strain and/or treatment effects, we feel that – although the trained observer (D. Schetters) was not blind to the experimental conditions – it seems unlikely that any consequential bias was introduced.

2.4. Experimental protocol and blood sampling

At the start of behavioral testing, all mice of one cage were placed in the group-compartment of the mHB to allow habituation (10 min). The mice were subsequently placed in the experimental compartment and allowed free exploration for 5 min while a trained observer (D. Schetters) scored the behavior (*'basal measures'*) by hand, using the computer software Observer 4.1 (Noldus, Wageningen, The Netherlands; Noldus 1991). Between behavioral tests, feces was removed from the set-up, urine was wiped up, and the experimental compartment was cleaned with a moistened towel with only water. Body weight (g) of the mice (mean ± SEM and range) were: A/J, 23.9 ± 0.5 and 19.5–26.8; BALB/cJ, 23.7 ± 0.3 and 21.7–25.6; C57BL/6J, 24.2 ± 0.4 and 22.5–27.2).

Thirty min after behavioral testing, about 50–100 μl blood was collected dropwise by tail incision (Dürschlag et al., 1996) for determination of plasma corticosterone levels (see *section Materials and methods – Corticosterone response*: 2.5). Blood samples were collected in pre-chilled Ethylenediaminetetraacetic acid (EDTA) coated microvettes (Sarstedt, Nümbrecht, Germany), subsequently stored on ice and centrifuged at 4000 rpm (diameter of the rotor: 17 cm) for 15 min in a refrigerated centrifuge (IEC Microlite/Microlite RF®: Thermo Electron Cooperation; West Sussex, UK) set at 4 °C. Samples were centrifuged within 3 h of collection and no hemolysis was observed. After centrifugation plasma was stored at −20 °C until analysis. Seven days later, the test procedure was repeated, but this time all animals also received an *i.p.* injection with 0.2 ml saline (0.9% w/v of NaCl; B. Braun, Melsungen, Germany) 30 min before being placed in the box compartment of the mHB (*'after i.p. measures'*). Body weight (g) of the mice (mean ± SEM and range) was: A/J, 24.7 ± 0.6 and 19.5–28.3; BALB/cJ, 24.6 ± 0.4 and 22.0–26.7; C57BL/6J, 25.5 ± 0.4 and 23.3–28.2).

2.5. Corticosterone response

Blood plasma corticosterone levels were determined by radioimmunoassay (RIA) according to the protocol of the Coat-A-Count® Rat Corticosterone kit (DPC®, Los Angeles, CA, USA). RIA was performed according to the manufacturers' protocol, using EDTA plasma from blood collected 30 min after each behavioral test. Due to technical problems during blood sampling or during the laboratory assay there were missing data (in total $n = 7$) of the plasma corticosterone levels after *i.p.* injection. Blood plasma samples were analyzed in randomized sequence, coded by number.

2.6. Behavioral z-score calculation

To obtain comprehensive and integrated behavioral measures, anxiety-related and activity-related mHB data were transformed using a z-score methodology. Details and rationale have been described by Guilloux et al. (2011). Briefly, for each behavioral mHB measure from Table S1, z-scores for individual animals were calculated using the following equation, which indicates how many standard deviation (SD, σ) an observation (X) is above or below the mean (μ) of a reference population + reference condition.

$$z = \frac{X - \mu}{\sigma}$$

Although it is not common for discrete numerical data, the means and SD for 'total number-variables' were also calculated and z-scores were subsequently computed. These so-called count based measures were treated as continuous data as suggested by Fagerland et al. (2011). The following reference groups were chosen: *i)* basal values from the A/J inbred strain; *ii)* basal values from the BALB/cJ inbred strain; *iii)* basal values from the C57BL/6J inbred strain; *iv)* basal values from the three inbred strains; or *v)* basal and after *i.p.* values from the three inbred strains. The z-scores for behavioral mHB measures were first averaged within a behavioral dimension, and then across behavioral dimensions to ensure equal weighting of the behavioral dimensions comprising the final combined z-score for a specific motivational system. The direction-

ality of scores was adjusted so that increased score values reflected increased values for that behavioral dimension or motivational system.

2.7. Statistical analyses

2.7.1. General

All statistical analyses were carried out using an IBM® SPSS® Statistics for Windows (version 22.0) computer program (IBM Corp., Armonk, NY, USA) and according to Field (2009). Two-sided, exact (*i.e.* for the non-parametric tests) probabilities were estimated throughout. The integrated behavioral z-score data and values of the corticosterone determinations were summarized as means with SEM.

2.7.2. Integrated behavioral z-score data from this study

Integrated behavioral z-score data were tested for significant differences by a multivariate repeated measures analysis of variance (ANOVA) with between-subject factor *strain* and within-subject factor *treatment* (general linear model module). Using the present experimental design (2×3 mixed factorial design with repeated measures on the first factor; http://www.3rs-eduction.co.uk/html/9_experimental_designs.html) the possibility cannot be excluded that one is dealing with an effect of repeated testing rather than the effect of the *i.p.* injection. However, there was a one-week interval between the two behavior tests; it has been shown that as long as sufficient time is allowed before retesting in the same paradigm, the same animal can be tested repeatedly in tests for unconditioned behavior (Bouwknecht and Paylor 2008; Paylor et al., 2006). Therefore, we conclude that the treatment effects on anxiety-related behavior and activity variables are the result of the injection. The choice of a multivariate instead of an univariate statistic in the repeated measures ANOVA is based on the criteria given by Algina and Keselman (1997). Tests of significance were derived using the Wilk's lambda criterion. In order to make the statistical results comparable for the five different reference options, the integrated z-scores per behavioral dimension or motivational system were first rank-transformed before an ANOVA was run (Conover and Iman 1981). An (one-way) ANOVA performed on ranked data is also known as the Scheirer-Ray-Hare extension of the Kruskal-Wallis test (Scheirer et al., 1976). To achieve statistical comparability for the five reference options, all between-subject *post hoc* comparisons (*i.e.* comparing two inbred strains) were performed with the Wilcoxon-Mann-Whitney test and the non-transformed data were used. To aid this statistical comparability for the within-subject *post hoc* comparisons (*i.e.* per inbred strain comparing *'basal behavior'* with *'after i.p. behavior'*) the Wilcoxon matched-pairs signed ranks test (on the non-transformed data) was used.

2.7.3. Blood plasma corticosterone levels

The Kolmogorov-Smirnov one-sample test was used to check Gaussianity of the blood plasma corticosterone data. We found that this variable was normally distributed in each group. The corticosterone values were tested for significant differences by multivariate (based on Wilk's lambda statistic) repeated measures ANOVA with between-subject factor *strain* and within-subject factor *treatment*. The residuals were normally distributed per group (Kolmogorov-Smirnov one-sample test). Homoscedasticity was tested by the Levene's test, which is a powerful and robust test based on the *F* statistic. The variances were not homogeneous and they were therefore equalized by a logarithmic transformation of the data ($y = {}^{10}Log[x - 87]$). After transformation, the within-group data as well as the residuals were still normally distributed (Kolmogorov-Smirnov one-sample test).

Since the ANOVA detected significant effects, group means were further compared. Between-subject *post hoc* comparisons (strain differences) were performed with unpaired Student's *t* tests, using pooled (for equal variances) or separate (for unequal variances) variance estimates. The equality of the variances was tested with the Levene's test. For the unpaired Student's *t* test with separate variance estimates, IBM® SPSS® Statistics for Windows uses the Welch-Satterthwaite correction. Within-subject *post hoc* comparisons were made using a paired Student's *t* test because the difference between basal and after *i.p.* values within groups was normally distributed (Kolmogorov-Smirnov one-sample test).

2.7.4. Correlations

The Spearman coefficient of rank correlation (R_S) was calculated between the blood plasma corticosterone levels, z-scores for anxiety-related behavior and z-scores for activity; significance was assessed by a two-tailed test based on the *t* statistic. The correlations were calculated for the basal and after *i.p.* values.

2.7.5. Preclinical drug study

To illustrate that choosing a reference population may also be of relevance for preclinical drug studies that include several vehicle controls, data from Laarakker et al. (2010) were taken (3×3 factorial design; between-group factors, *inbred strain* and *treatment*). The integrated behavioral z-scoring methodology was applied for the obtained ethological results and four different reference group options (A/J, *i.p.* saline; C57BL/6J, *i.p.* saline; C57BL/6J-Chr $19^{A/J}$/NaJ, *i.p.* saline; pooled data of all the experimental groups) were selected. The outcomes regarding statistical significance and practical importance were compared. Integrated z-scores for anxiety-related behavior and activity were tested for significant inbred strain and treatment differences using two-way ANOVA. The integrated z-scores per motivational system were first rank-transformed before the ANOVA was performed. All *post hoc* comparisons were performed with the Wilcoxon-Mann-Whitney test and the non-transformed data were used.

2.7.6. Comparison of mHB with three traditional behavioral anxiety paradigms

Traditional behavioral anxiety paradigms include EPM, LD and OF (Ennaceur and Chazot 2016). The behavioral data obtained with these unconditioned tests for male mice from the A/J, BALB/cJ and C57BL/6J inbred strains are available in the Mouse Phenome Database (MPD, http://www.phenome.jax.org). For these paradigms, integrated z-scores for anxiety-related behavior and activity were calculated (Table S2) and compared these results with the mHB (basal values). The z-scores were tested per behavioral test (EPM, LD, OF, or mHB) for significant inbred strain differences by a one-way ANOVA. In order to make the statistical results comparable between the four different behavioral tests, the integrated z-scores per motivational system were first rank-transformed before ANOVA was performed. To achieve statistical comparability for the four behavioral paradigms, all *post hoc* comparisons (*i.e.* comparing two inbred strains) were performed with the Wilcoxon-Mann-Whitney test and the non-transformed data were used.

2.7.7. Substrain comparison

Which substrain is used can have influence on the behavioral data acquired (Kang et al., 2015; Labots et al., 2016). Therefore the present mHB data for A/J, BALB/cJ and C57BL/6J mice (all purchased from The Jackson Laboratory (see section *Materials and methods – Animals and housing*: 2.2) were compared with those for A/JCrl, BALB/cAnNCrl and C57BL/6NCrl male mice (all purchased from Charles River Nederland B.V.). The results for the latter mice

were based on Arndt et al. (2009). Integrated z-scores for anxiety-related behavior and activity were calculated and these z-scores were tested for significant (standard) inbred strain and vendor differences using two-way ANOVA. The integrated z-scores per motivational system were first rank-transformed before ANOVA was performed. All *post hoc* comparisons (*i.e.* comparing two (sub)strains) were performed with the Wilcoxon-Mann-Whitney test and the non-transformed data were used.

2.7.8. Thresholds for significance

To take the greater probability of a Type I error as a result of multiple comparisons into account, a more stringent criterion should be used for statistical significance (*i.e.* for the Wilcoxon-Mann-Whitney tests, Wilcoxon matched-pairs signed ranks tests, unpaired and paired Student's *t* tests). We addressed this issue by calculating a Dunn-Šidák correction ($\alpha = 1 - [1 - 0.05]^{1/\gamma}$; γ = number of times a group is used for a meaningful comparison; section *Materials and methods – Statistical analyses: preclinical drug study (2.7.5)*: $\gamma = 4 \rightarrow \alpha = 0.012741$; section *Materials and methods – Statistical analyses: z-scores (2.7.2), blood plasma corticosterone (2.7.3) and substrain comparison (2.7.7)*: $\gamma = 3 \rightarrow \alpha = 0.016952$; section *Materials and methods – Statistical analyses: comparison behavioral anxiety paradigms (2.7.6)*: $\gamma = 2 \rightarrow \alpha = 0.025321$).

Calculating numerous correlations also increases the risk of a Type I error. To avoid this, the level of statistical significance of Spearman correlation coefficients were also adjusted using the Dunn-Šidák method ($\alpha = 1 - [1 - 0.05]^{1/\kappa}$; κ = number of times a variable is used for a correlation; section *Materials and methods – Statistical analyses – Correlations (2.7.4)*: $\kappa = 2 \rightarrow \alpha = 0.025321$). In all other cases, the probability of a Type I error < 0.05 was taken as the criterion of significance. Significant *P* values are reported with six decimal places, whereas non-significant *P* values are reported with two digits after the decimal point.

2.7.9. Effect sizes

Statistical significance represented by *P* values may not necessarily predicate practical importance (Festing 2014; Lovell 2013). Rather, our opinion is that scientists should also be more interested in the size of the observed effects than whether the result is statistically significant. Effect sizes reported comprise the partial eta squared values (η_p^2) within the (repeated measures) ANOVAs. For this purpose, non-transformed data were used. The following cutoffs for the η_p^2 effect size coefficients were used: small effect, $\eta_p^2 \leq 0.03$; medium effect, $0.03 < \eta_p^2 < 0.10$; large effect, $0.10 \leq \eta_p^2 < 0.20$; very large effect, $\eta_p^2 \geq 0.20$. To estimate the relative magnitude of the differences in the *post hoc* comparisons, Cohen's *d* effect size coefficients can be used. The Cohen's *d* score is here defined as the difference between the two means (non-transformed data) divided by the pooled SD. On the basis of a comprehensive review of a wide variety of phenotypes by Wahlsten (2011), guidelines have been provided for absolute values of Cohen's *d* ($|d|$) that correspond to what are generally regarded as small, medium, large and very large effects in mouse neurobehavioral studies: small effect, $|d| \leq 0.5$; medium effect, $0.5 < |d| < 1.0$; large effect, $1.0 \leq |d| < 1.5$; very large effect, $|d| \geq 1.5$.

3. Results

3.1. Statistical significance

3.1.1. General

Supplementary Tables S3 and S4 contain a summary of the results regarding z-scores for anxiety-related behavior and activity in the mHB, respectively. The significant results from the statistical analyses (repeated measures ANOVA and *post hoc* comparisons) for the five reference options are summarized in Tables 1 and 2. Significant effects and/or differences are influenced by the choice of the reference. In some cases, it was not possible to use a certain population and condition as a reference, because one or more behavioral variables in question for that population and condition had a SD of zero.

3.1.2. Anxiety-related behavior (Table 1)

The ANOVA revealed a significant interaction effect for avoidance behavior, irrespective of the reference option. In the within-subject *post hoc* analyses, the treatment effect was only significant in mice from the C57BL/6J strain.

Analysis of risk assessment with the ANOVA showed for all the reference options a significant treatment effect. For BALB/cJ and C57BL/6J mice, this treatment effect was also significant in the within-subject *post hoc* analyses, independent of the reference option. In contrast, for A/J mice the difference between basal and after *i.p.* risk assessment was only significant when *'A/J-basal'*, *'C57BL/6J-basal'* or the *'three inbred strains-basal plus after i.p.'* were taken as a reference.

For arousal, there were significant strain differences in the ANOVA and this strain effect was independent of the reference option. Between-subject *post hoc* analyses revealed that under basal conditions male A/J mice were significantly different from C57BL/6J males, regardless of the reference option. However, after *i.p.* injection, this inbred strain difference was only significant when the *'three inbred strains-basal'* was taken as reference. BALB/cJ and C57BL/6J mice only significantly differed under basal conditions and when the *'A/J-basal'* was used as a reference. The significance of the treatment effect in the ANOVA was influenced by the choice of the reference option. With *'three inbred strains-basal plus after i.p.'* values as reference option there was no significant treatment effect in the ANOVA, whereas with the other four options there was. In the within-subject *post hoc* analyses for all five reference conditions the only treatment effect that reached statistical significance was within the C57BL/6J inbred strain.

Analysis of anxiety-related behavior with the ANOVA showed a significant strain effect for all the reference options. Between-subject *post hoc* analyses demonstrated that only A/J and C57BL/6J males were significantly different from each other, both under basal and after *i.p.* injection conditions, independent from the reference choice. The treatment effect in the ANOVA was influenced by the reference choice. With *'C57BL/6J-basal'* as a reference, there was no significant treatment effect, whereas with the other four options this effect was present. This pattern could also be observed in the within-subject *post hoc* analyses of male C57BL/6J mice.

3.1.3. Activity (Table 2)

Based on the ANOVA results, male mice from the three inbred strains were found to differ regarding locomotion. This was found for four out of five reference options: with *'C57BL/6J-basal'* as reference it was not possible to calculate z-scores, because for this option some of the behavioral measures had a SD of zero. Between-subject *post hoc* analyses revealed a significant difference in locomotion between A/J and C57BL/6J mice, both under basal conditions and after *i.p.* injection. This difference was found with four reference options. Again, with *'C57BL/6J-basal'* as a reference, it was not possible to calculate z-scores. BALB/cJ and C57BL/6J males also differed in locomotion, but this effect was only significant under basal conditions. Within-subject *post hoc* analyses also showed a treatment effect in the A/J males, but only when *'A/J-basal'* was taken as reference.

Z-scores for immobility could not be calculated when *'BALB/cJ-basal'* were used as a reference. For the other four reference options, there was a strain, but no treatment or interaction effect in the ANOVA. With *'C57BL/6J-basal'* as a reference, there was a significant

Table 1
Overview of significant effects or differences regarding anxiety-related behavior in the mHB of male mice from three inbred strains under basal conditions and after an intra-peritoneal (*i.p.*) injection using different reference populations and conditions.

Motivational system / Behavioral dimension	*Reference population and condition*	*ANOVA*[1]			*Basal vs after i.p.*[2]			*A vs C*[3]		*A vs B6*[3]		*C vs B6*[3]	
		S	*T*	*SxT*	*A*[6]	*C*[6]	*B6*[6]	*Basal*	*After i.p.*	*Basal*	*After i.p.*	*Basal*	*After i.p.*
Anxiety-related behavior													
- Avoidance													
	A/J (A)[6] *basal*	-[4]	-	X[4]	-	-	X	-	-	-	-	-	-
	BALB/cJ (C)[6] *basal*	-	-	X	-	-	X	-	-	-	-	-	-
	C57BL/6J (B6)[6] *basal*	-	-	X	-	-	X	-	-	-	-	-	-
	A + C + B6 *basal*	-	-	X	-	-	X	-	-	-	-	-	-
	A + C + B6 *basal* and *after i.p.*	-	-	X	-	-	X	-	-	-	-	-	-
- Risk assessment													
	A/J (A) *basal*	-	X	-	X	X	X	-	-	-	-	-	-
	BALB/cJ (C) *basal*	-	X	-	-[5]	X	X	-	-	-	-	-	-
	C57BL/6J (B6) *basal*	-	X	-	X	X	X	-	-	-	-	-	-
	A + C + B6 *basal*	-	X	-	-	X	X	-	-	-	-	-	-
	A + C + B6 *basal* and *after i.p.*	-	X	-	X	X	X	-	-	-	-	-	-
- Arousal													
	A/J (A) *basal*	X	X	-	-	-	X	-	-	X	-	X	-
	BALB/cJ (C) *basal*	X	X	-	-	-	X	-	-	X	-	-	-
	C57BL/6J (B6) *basal*	X	X	-	-	-	X	-	-	X	-	-	-
	A + C + B6 *basal*	X	X	-	-	-	X	-	-	X	X	-	-
	A + C + B6 *basal* and *after i.p.*	X	-	-	-	-	X	-	-	X	-	-	-
Anxiety-related behavior													
	A/J (A) *basal*	X	X	-	-	-	X	-	-	X	X	-	-
	BALB/cJ (C) *basal*	X	X	-	-	-	X	-	-	X	X	-	-
	C57BL/6J (B6) *basal*	X	-	-	-	-	-	-	-	X	X	-	-
	A + C + B6 *basal*	X	X	-	-	-	X	-	-	X	X	-	-
	A + C + B6 *basal* and *after i.p.*	X	X	-	-	-	X	-	-	X	X	-	-

[1]Significance ($P < 0.05$) based on a repeated measures ANOVA with main between-subject factors *strain* and within-subject factor *treatment*. S indicates effect of strain; T, effect of injection (treatment); SxT, interaction. The ANOVA's were performed after ranking of the data.
[2]Contrast significance (Wilcoxon matched-pairs signed ranks test; $P < 0.016952$).
[3]Contrast significance (Wilcoxon-Mann-Whitney test; $P < 0.016952$).
[4]X = significant effect or difference, – = no significant effect or difference.
[5]Deviant results from the other reference options are indicated in red.
[6]Official abbreviation of the inbred strain in parentheses (see http://www.informatics.jax.org/mgihome/nomen/strains.shtml).

Table 2
Overview of significant effects or differences regarding activity behavior in the mHB of male mice from three inbred strains under basal conditions and after an intra-peritoneal (*i.p.*) injection using different reference populations and conditions.

Motivational system / Behavioral dimension	*Reference population and condition*	*ANOVA*[1]			*Basal vs after i.p.*[2]			*A vs C*[3]		*A vs B6*[3]		*C vs B6*[3]	
		S	*T*	*SxT*	*A*[6]	*C*[6]	*B6*[6]	*Basal*	*After i.p.*	*Basal*	*After i.p.*	*Basal*	*After i.p.*
Activity													
- Locomotion													
	A/J (A)[6] *basal*	X[4]	-[4]	-	X[5]	-	-	-	-	X	X	X	-
	BALB/cJ (C)[6] *basal*	X	-	-	-	-	-	-	-	X	X	X	-
	C57BL/6J (B6)[6] *basal*	*Not possible*[7]											
	A + C + B6 *basal*	X	-	-	-	-	-	-	-	X	X	X	-
	A + C + B6 *basal* and *after i.p.*	X	-	-	-	-	-	-	-	X	X	X	-
- Immobility													
	A/J (A) *basal*	X	-	-	-	-	-	-	-	-	-	-	-
	BALB/cJ (C) *basal*	*Not possible*											
	C57BL/6J (B6) *basal*	X	-	-	-	-	-	-	-	X	-	-	-
	A + C + B6 *basal*	X	-	-	-	-	-	-	-	-	-	-	-
	A + C + B6 *basal* and *after i.p.*	X	-	-	-	-	-	-	-	-	-	-	-
Activity													
	A/J (A) *basal*	X	-	-	X	-	-	-	-	X	X	X	-
	BALB/cJ (C) *basal*	*Not possible*											
	C57BL/6J (B6) *basal*	*Not possible*											
	A + C + B6 *basal*	X	-	-	-	-	-	-	-	X	X	X	-
	A + C + B6 *basal* and *after i.p.*	X	-	-	-	-	-	-	-	X	X	X	-

[1]Significance ($P < 0.05$) based on a repeated measures ANOVA with main between-subject factors *strain* and within-subject factor *treatment*. S indicates effect of strain; T, effect of injection (treatment); SxT, interaction. The ANOVA's were performed after ranking of the data.
[2]Contrast significance (Wilcoxon matched-pairs signed ranks test; $P < 0.016952$).
[3]Contrast significance (Wilcoxon-Mann-Whitney test; $P < 0.016952$).
[4]X = significant effect or difference, – = no significant effect or difference.
[5]Deviant results from the other reference options are indicated in red.
[6]Official abbreviation of the inbred strain in parentheses (see http://www.informatics.jax.org/mgihome/nomen/strains.shtml).
[7]For some activity-related behavioral variables the BALB/cJ or C57BL/6J inbred strain had a SD of zero and thus it was not possible to calculate z-scores.

Table 3
Effect sizes regarding anxiety-related behavior in the mHB of male mice from three inbred strains under basal conditions and after an intra-peritoneal (*i.p.*) injection using different reference populations and conditions.[1]

Motivational system / Behavioral dimension	*Reference population and condition*	*Partial eta squared (η_p^2)*						*Cohen's d*					
		ANOVA[1]			*Basal vs after i.p.*[2]			*A vs C*[3]		*A vs B6*[3]		*C vs B6*[3]	
		S	*T*	*SxT*	*A*[2]	*C*[2]	*B6*[2]	*Basal*	*After i.p.*	*Basal*	*After i.p.*	*Basal*	*After i.p.*
Anxiety-related behavior													
- Avoidance													
	A/J (A)[2] *basal*	**0.10**	*0.06*	**0.20**	0.24	-0.29	*0.74*	*0.91*	0.34	0.35	*0.93*	*-0.52*	*0.51*
	BALB/cJ (C)[2] *basal*	*0.09*	*0.05*	***0.23***	0.21	-0.39	*0.80*	*0.92*	0.29	0.23	*0.88*	*-0.65*	*0.55*
	C57BL/6J (B6)[2] *basal*	*0.10*	*0.06*	***0.22***	0.22	-0.35	*0.77*	*0.92*	0.31	0.28	*0.91*	*-0.58*	*0.54*
	A + C + B6 *basal*	*0.10*	*0.05*	***0.22***	0.21	-0.35	*0.77*	*0.91*	0.31	0.29	*0.91*	*-0.57*	*0.55*
	A + C + B6 *basal* and *after i.p.*	*0.10*	*0.05*	***0.22***	0.21	-0.35	*0.77*	0.91	0.31	0.28	*0.91*	*-0.57*	*0.55*
- Risk assessment													
	A/J (A) *basal*	**0.14**	***0.59***	**0.14**	*0.73*	**1.48**	**1.19**	-0.50	0.45	0.44	**1.03**	*0.95*	0.44
	BALB/cJ (C) *basal*	**0.14**	***0.46***	**0.12**	*0.50*	***1.76***	**1.19**	*-0.52*	*0.56*	0.31	**1.04**	*0.84*	0.40
	C57BL/6J (B6) *basal*	**0.13**	***0.62***	**0.14**	*0.81*	**1.38**	**1.18**	-0.47	0.40	0.48	**1.01**	*0.93*	0.46
	A + C + B6 *basal*	**0.14**	***0.57***	**0.14**	*0.69*	***1.53***	**1.20**	-0.51	0.48	0.42	**1.03**	*0.94*	0.43
	A + C + B6 *basal* and *after i.p.*	**0.13**	***0.61***	**0.14**	*0.78*	**1.42**	**1.19**	-0.48	0.42	0.46	**1.02**	*0.94*	0.45
- Arousal													
	A/J (A) *basal*	***0.27***	**0.12**	**0.10**	0.12	-0.44	**-1.08**	*0.64*	-0.08	***2.08***	*0.73*	**1.25**	*0.53*
	BALB/cJ (C) *basal*	***0.25***	**0.16**	**0.11**	0.06	*-0.62*	**-1.14**	*0.92*	-0.10	***2.16***	*0.70*	**1.20**	0.50
	C57BL/6J (B6) *basal*	***0.27***	**0.15**	**0.11**	0.08	*-0.59*	**-1.13**	*0.82*	-0.13	***2.30***	*0.76*	**1.25**	*0.58*
	A + C + B6 *basal*	***0.28***	**0.11**	**0.11**	0.16	-0.43	**-1.07**	*0.67*	-0.07	***2.12***	*0.77*	**1.26**	*0.54*
	A + C + B6 *basal* and *after i.p.*	***0.30***	0.05	**0.11**	0.29	-0.24	*-0.97*	*0.54*	0.00	***1.96***	*0.79*	**1.25**	*0.53*
Anxiety-related behavior													
	A/J (A) *basal*	**0.18**	***0.24***	0.02	*0.57*	0.27	*0.68*	0.46	0.32	**1.03**	**1.17**	0.44	*0.58*
	BALB/cJ (C) *basal*	***0.21***	***0.24***	0.03	0.48	*0.50*	*0.80*	0.38	0.38	*0.96*	**1.17**	*0.66*	*0.57*
	C57BL/6J (B6) *basal*	***0.25***	0.02	*0.03*	*0.55*	-0.06	0.18	*0.61*	0.10	***1.83***	**1.14**	*0.76*	*0.64*
	A + C + B6 *basal*	***0.20***	***0.22***	0.02	*0.56*	0.28	*0.67*	0.46	0.31	**1.10**	**1.18**	*0.53*	*0.61*
	A + C + B6 *basal* and *after i.p.*	**0.19**	***0.30***	0.01	*0.62*	0.36	*0.71*	0.40	0.33	**1.04**	**1.16**	*0.52*	*0.61*

[1] For calculation of the effect sizes non-ranked data were used. Small effects are indicated in normal fonts, medium effects in *italics* and *green*, large effects in **bold** and **yellow**, and very large effects in ***bolditalics*** and ***red***.
[2] Official abbreviation of the inbred strain in parentheses (see http://www.informatics.jax.org/mgihome/nomen/strains.shtml).

Table 4
Effect sizes regarding activity in the mHB of male mice from three inbred strains under basal conditions and after an intra-peritoneal (*i.p.*) injection using different reference populations and conditions.[1]

Motivational system / Behavioral dimension	*Reference population and condition*	*Partial eta squared (η_p^2)*						*Cohen's d*					
		ANOVA			*Basal vs after i.p.*			*A vs C*		*A vs B6*		*C vs B6*[3]	
		S	*T*	*SxT*	*A*[2]	*C*[2]	*B6*[2]	*Basal*	*After i.p.*	*Basal*	*After i.p.*	*Basal*	*After i.p.*
Activity													
- Locomotion													
	A/J (A)[2] *basal*	***0.37***	*0.07*	***0.28***	*-0.80*	-0.20	*0.78*	*-0.71*	-0.34	***-3.27***	**-1.43**	**-1.32**	*-0.57*
	BALB/cJ (C)[2] *basal*	***0.22***	*0.04*	*0.07*	-0.47	-0.24	*0.64*	-0.44	-0.31	*-0.95*	*-0.96*	**-1.10**	*-0.70*
	C57BL/6J (B6)[2] *basal*	*Not possible*[3].....			*Not possible*.....			*Not possible*		*Not possible*		*Not possible*	
	A + C + B6 *basal*	***0.24***	0.02	*0.07*	-0.41	-0.13	*0.67*	-0.23	-0.04	***-1.51***	*-0.98*	**-1.05**	*-0.79*
	A + C + B6 *basal* and *after i.p.*	***0.21***	0.01	*0.06*	-0.37	-0.10	*0.64*	-0.23	-0.03	**-1.30**	*-0.92*	*-0.91*	*-0.80*
- Immobility													
	A/J (A) *basal*	**0.20**	0.02	*0.09*	0.35	*-0.85*	*0.52*	*0.86*	0.41	*0.78*	*0.78*	-0.35	*0.64*
	BALB/cJ (C) *basal*	*Not possible*.....			*Not possible*.....			*Not possible*		*Not possible*		*Not possible*	
	C57BL/6J (B6) *basal*	***0.20***	*0.05*	**0.11**	*0.51*	*-0.85*	*0.52*	*0.76*	*0.63*	*0.79*	*0.76*	-0.35	*0.64*
	A + C + B6 *basal*	***0.20***	0.02	*0.09*	0.36	*-0.85*	*0.52*	*0.85*	0.43	*0.78*	*0.78*	-0.35	*0.64*
	A + C + B6 *basal* and *after i.p.*	***0.20***	0.03	*0.09*	0.38	*-0.85*	*0.52*	*0.83*	0.46	*0.79*	*0.78*	-0.35	*0.64*
Activity													
	A/J (A) *basal*	***0.37***	*0.06*	**0.19**	*-0.59*	-0.07	0.45	*-0.97*	-0.43	***-1.98***	**-1.24**	**-1.13**	*-0.69*
	BALB/cJ (C) *basal*	*Not possible*.....			*Not possible*.....			*Not possible*		*Not possible*		*Not possible*	
	C57BL/6J (B6) *basal*	*Not possible*.....			*Not possible*.....			*Not possible*		*Not possible*		*Not possible*	
	A + C + B6 *basal*	***0.25***	0.03	*0.09*	-0.41	0.14	0.01	*-0.74*	-0.32	**-1.10**	*-0.92*	*-0.70*	*-0.98*
	A + C + B6 *basal* and *after i.p.*	***0.24***	0.03	*0.09*	-0.42	0.16	0.00	*-0.73*	-0.33	**-1.04**	*-0.90*	*-0.63*	*-0.99*

[1] For calculation of the effect sizes non-ranked data were used. Small effects are indicated in normal fonts, medium effects in *italics* and *green*, large effects in **bold** and **yellow**, and very large effects in ***bolditalics*** and ***red***.
[2] Official abbreviation of the inbred strain in parentheses (see http://www.informatics.jax.org/mgihome/nomen/strains.shtml).
[3] For some activity-related behavioral variables the BALB/cJ or C57BL/6J inbred strain had a SD of zero and thus it was not possible to calculate z-scores.

difference in the between-subject *post hoc* analyses between A/J and C57BL/6J mice under basal conditions.

Z-scores for activity could only be calculated with *'A/J-basal'*, *'three inbred strains-basal'* or *'three inbred strains basal plus after i.p.'* as reference. Irrespective of the reference option, there was a significant strain effect with the ANOVA and in the between-subject *post hoc* analyses C57BL/6J male mice were significantly different from both A/J and BALB/cJ males under basal conditions. After an *i.p.* injection A/J males were significantly different from the C57BL/6J counterparts. With *'A/J-basal'* as reference option there was also a significant interaction effect with the ANOVA.

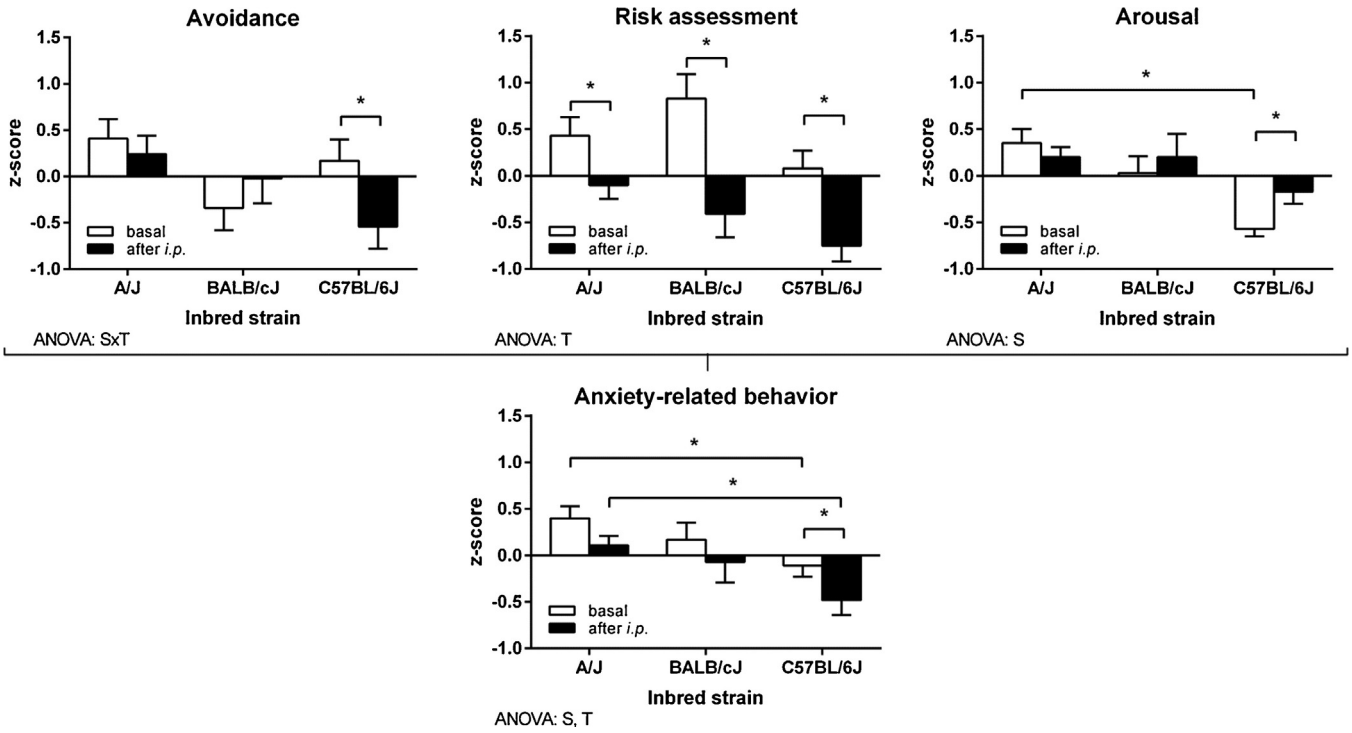

Fig. 1. Z-scores for anxiety-related behavior in the mHB of male mice from three inbred strains.
The mice were tested under basal conditions and after an intra-peritoneal (*i.p.*) injection. The '*three inbred strains-basal plus after i.p.*' pooled values were used as reference option. Results are presented as means plus or minus SEM; A/J $n = 15$, BALB/cJ $n = 12$, C57BL/cJ $n = 15$. ANOVA = a multivariate repeated measures analysis of variance with main between-subject factors *strain* and within-subject factor *treatment*. Effects were significant in the ANOVA when $P < 0.05$. S indicates significant effect of strain; T, significant effect of injection (treatment); SxT, significant interaction. The ANOVA's were performed after ranking of the data. *Post hoc* testing was done by Wilcoxon matched-pairs signed ranks test (within-inbred strain comparison) or Wilcoxon-Mann-Whitney test (between-inbred strain comparison). * = Significant difference ($P < 0.016952$) in the *post hoc* comparison.

3.2. Effect sizes

3.2.1. General

The effect sizes regarding anxiety-related behavior and activity in the mHB of male mice from the three inbred strains for the five reference options are summarized in Tables 3 and 4. Effect sizes were influenced by the choice of the reference. In some cases it was not possible to use a certain population and condition as a reference, because one or more behavioral variables for that population and/or condition had a SD of zero, and thus calculation of effect sizes was not possible.

3.2.2. Anxiety-related behavior (Table 3)

The partial eta squared value (η_p^2) was influenced by the choice of the reference population and condition. For instance, the magnitude of the treatment effect (η_p^2) regarding anxiety-related behavior ranged from small (0.02, *'C57BL/6J-basal'* as reference) to very large (0.30, *'three inbred strains-basal and after i.p.'* as reference). The Cohen's *d* scores were also affected by the reference choice. With *'C57BL/6J-basal'* as reference and anxiety-related behavior as motivational system, the Cohen's *d* for the basal difference between A/J and C57BL/6J males was +1.83 (very large effect), whereas with *'BALB/cJ-basal'* as reference this difference was about 1.9 times lower (+0.96, medium effect). For anxiety-related behavior, the values for the Cohen's *d* scores were in some cases even opposite, depending on the reference choice. The difference between basal and after *i.p.* in BALB/cJ males expressed as Cohen's *d* for this motivational system was +0.50 (medium effect, *'BALB/cJ-basal'* as reference) or −0.06 (small effect, *'C57BL/6J-basal'* as reference).

3.2.3. Activity (Table 4)

The η_p^2 and Cohen's *d* regarding activity were also influenced by the choice of the reference population and condition. For example, the magnitude of the interaction effect (η_p^2) regarding locomotion ranged from medium (*e.g.* 0.06, *'three inbred strains basal plus after i.p.'* as reference) to very large (0.28, *'A/J-basal'* as reference). In the *post hoc* comparisons a similar pattern was found: with *'A/J-basal'* as reference and locomotion as behavioral dimension, the Cohen's *d* for the after *i.p.* difference between A/J and C57BL/6J males was −1.43 (large effect), whereas with *'three inbred strains basal plus after i.p.'* as a reference this difference was about 36% lower (-0.92, medium effect). With locomotion as behavioral dimension and A/J versus C57BL/6J mice as a comparison, the basal effect size ranged from −0.95 (medium effect, *'BALB/cJ-basal'* as reference) through −1.30 (large effect, *'three inbred strains basal plus after i.p.'* as reference) to −3.27 (very large effect, *'A/J-basal'* as reference).

3.3. Integrated behavioral z-scores based on the improved method

3.3.1. General

Taking the '**statistical significance'**- (see section *Results – Statistical significance*: 3.1) and **'effect sizes**-results (see section *Results – Effect sizes:* 3.2) together, it is reasonable to take '*three inbred strains basal plus after i.p.*' as a reference in order to compute integrated behavioral z-scores. Since it is illogical to incorporate a behavioral measure that does not vary between groups and/or over time, with this reference none of the behavioral variables tested will have a SD of zero. Figs. 1 and 2 graphically summarize the results (z-scores) for the two motivational systems (anxiety-related behavior and activity) and the underlying behavioral dimensions (avoidance, risk assessment, arousal, locomotion, immobility).

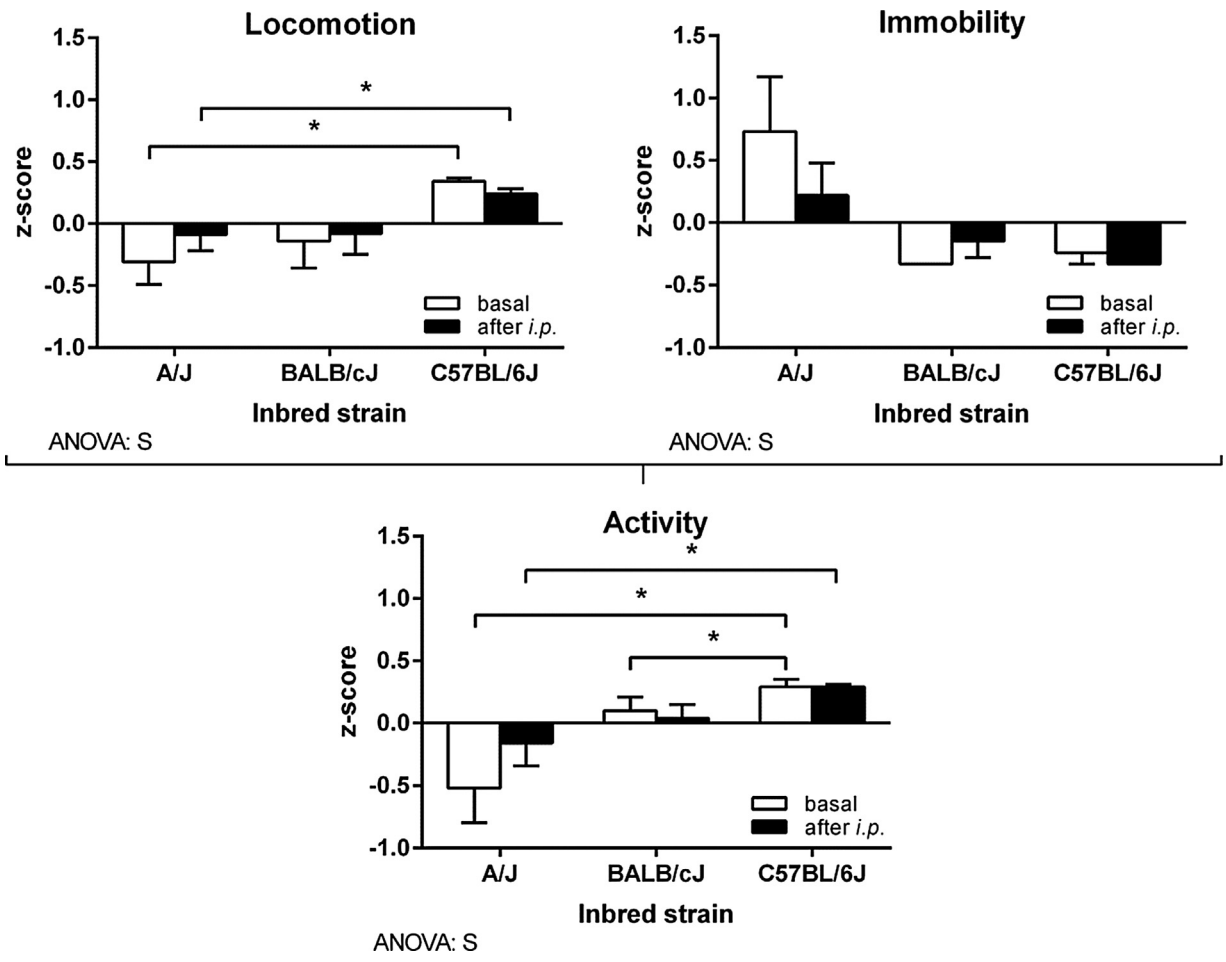

Fig. 2. Z-scores for activity in the mHB of male mice from three inbred strains.
The mice were tested under basal conditions and after an intra-peritoneal (*i.p.*) injection. The '*three inbred strains-basal plus after i.p.*' pooled values were used as reference option. Results are presented as means plus or minus SEM; A/J $n = 15$, BALB/cJ $n = 12$, C57BL/cJ $n = 15$. ANOVA = a multivariate repeated measures analysis of variance with main between-subject factors *strain* and within-subject factor *treatment*. Effects were significant in the ANOVA when $P < 0.05$. S indicates significant effect of strain; T, significant effect of injection (treatment); SxT, significant interaction. The ANOVA's were performed after ranking of the data. *Post hoc* testing was done by Wilcoxon matched-pairs signed ranks test (within-inbred strain comparison) or Wilcoxon-Mann-Whitney test (between-inbred strain comparison). * = Significant difference ($P < 0.016952$) in the *post hoc* comparison.

3.3.2. Anxiety-related behavior (Fig. 1)

Based on the ANOVA results, no significant differences for avoidance behavior between inbred strains were found (ranked data: $F_{(strain)2,39} = 1.852$, $P = 0.17$). The ANOVA revealed a significant interaction effect for avoidance behavior (ranked data: $F_{(strainxtreatment)2,39} = 6.630$, $P = 0.003322$). In the within-subject *post hoc* analyses the treatment effect was only significantly present in the C57BL/6J inbred strain. That is, after *i.p.* injection C57BL/6J mice showed less avoidance behavior in the mHB compared to the baseline condition (Wilcoxon matched-pairs signed ranks test, non-transformed data: $Z = -3.040$, $P = 0.000732$).

There was no significant difference in risk assessment behavior between the three inbred strains (ANOVA, ranked data: $F_{(strain)2,39} = 2.681$, $P = 0.08$). The ANOVA showed a significant treatment effect (ranked data: $F_{(treatment)1,39} = 51.426$, $P < 0.0000005$). In each inbred strain, an *i.p.* injection resulted in less risk assessment behavior. This effect was also significant in the within-subject *post hoc* analyses (Wilcoxon matched-pairs signed ranks test, non-transformed data: A/J, $Z = -2.417$, $P = 0.013428$; BALB/cJ, $Z = -3.059$, $P = 0.000488$; C57BL/6J, $Z = -2.824$, $P = 0.002441$).

For arousal, there were significant strain differences in the ANOVA (ranked data: $F_{(strain)2,39} = 8.448$, $P = 0.000895$). Between-subject *post hoc* analyses revealed that under basal conditions male A/J mice were significantly more aroused than C57BL/6J males (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 23.000$, $W = 143.000$, $Z = -3.763$, $P = 0.000062$). Based on the within-subject *post hoc* analyses, an *i.p.* injection only had a significant effect in the C57BL/6J inbred strain: after *i.p.* injection C57BL/6J mice were more aroused (Wilcoxon matched-pairs signed ranks test, non-transformed data: $Z = -2.667$, $P = 0.004883$).

Analysis of anxiety-related behavior with the ANOVA showed a significant strain effect (ranked data: $F_{(strain)2,39} = 4.485$, $P = 0.017655$). Between-subject *post hoc* analyses demonstrated that only A/J and C57BL/6J males were significantly different from each other, both under basal (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 46.000$, $W = 166.000$, $Z = -2.758$, $P = 0.004940$) and after *i.p.* injection conditions (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 48.000$, $W = 168.000$, $Z = -2.675$, $P = 0.006568$): A/J mice were more anxious than C57BL/6J animals. There also was a treatment effect in the ANOVA (ranked data: $F_{(treatment)1,39} = 13.730$, $P = 0.000653$), but in the within-subject *post hoc* comparisons this was only significant for the C57BL/6J inbred strain (Wilcoxon matched-pairs signed ranks test, non-transformed data: $Z = -2.897$, $P = 0.002014$): after *i.p.* male C57BL/6J mice were less anxious.

3.3.3. Activity (Fig. 2)

Based on the ANOVA results, male mice from the three inbred strains differed regarding locomotion (ranked data: $F_{(strain)2,39} = 8.138$, $P = 0.001112$), immobility (ranked data: $F_{(strain)2,39} = 7.930$, $P = 0.001289$) and activity (ranked data: $F_{(strain)2,39} = 8.273$, $P = 0.001011$). Between-subject *post hoc* analyses revealed that A/J compared to C57BL/6J male mice showed significantly less locomotion, both under basal conditions (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 39.000$, $W = 159.000$, $Z = -3.049$, $P = 0.001665$) and after *i.p.* injection (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 34.000$, $W = 154.000$, $Z = -3.256$, $P = 0.000696$). Under basal conditions, BALB/cJ mice also showed significantly less locomotion compared to C57BL/6J mice (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 30.000$, $W = 108.000$, $Z = -3.256$, $P = 0.002566$). In the between-subject *post hoc* analyses, C57BL/6J mice were significantly more active than A/J (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 44.000$, $W = 164.000$, $Z = -2.841$, $P = 0.003674$) or BALB/cJ males (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 37.000$, $W = 115.000$, $Z = -2.586$, $P = 0.008710$) under basal conditions. After an *i.p.* injection, A/J mice were significantly less active compared to the C57BL/6J counterparts (Wilcoxon-Mann-Whitney test, non-transformed data: $U = 34.000$, $W = 154.000$, $Z = -3.256$, $P = 0.000696$).

3.4. Corticosterone response (Fig. 3)

Analysis of blood plasma corticosterone levels with the ANOVA showed a significant strain and treatment effect (logarithmically transformed data: $F_{(strain)2,32} = 16.575$, $P = 0.000011$; $F_{(treatment)1,32} = 5.134$, $P = 0.030359$). Between-subject *post hoc* analyses showed that BALB/cJ male mice had significantly lower circulating corticosterone levels compared to the A/J or C57BL/6J male mice, both under basal conditions and after an *i.p.* injection (non-transformed data, unpaired Student's *t* test: basal – A/J *vs.* BALB/cJ, $t_{25} = 4.94$, $P = 0.000300$; after *i.p.* – A/J *vs.* BALB/cJ, $t_{11.849} = 3.895$, $P = 0.002180$; basal – BALB/cJ *vs.* C57BL/6J, $t_{25} = -4.811$, $P = 0.000061$; after *i.p.* – BALB/cJ vs. C57BL/6J, $t_{23} = -3.563$, $P = 0.001655$). The treatment effect was not significant in the within-subject *post hoc* analyses (non-transformed data, paired Student's *t* test: A/J, $t_9 = -2.459$, $P = 0.04$; BALB/cJ, $t_{10} = -1.541$, $P = 0.15$; C57BL/6J, $t_{13} = -0.047$, $P = 0.96$).

3.5. Correlations

For both basal and after *i.p.* injection values there were no significant correlations between blood plasma corticosterone level and z-score for anxiety-related behavior (basal: $R_S = -0.095$, $n = 42$, $P = 0.55$; after *i.p.*: $R_S = -0.248$, $n = 35$, $P = 0.15$) or z-score for activity (basal: $R_S = 0.235$, $n = 42$, $P = 0.13$; after *i.p.*: $R_S = 0.090$, $n = 35$, $P = 0.61$). Significant negative correlations were found between the z-score for anxiety-related behavior and that for activity (basal: $R_S = -0.691$, $n = 42$, $P < 0.0000005$; after *i.p.*: $R_S = -0.727$, $n = 42$, $P < 0.0000005$).

4. Discussion

4.1. Pooled values as reference option

Guilloux et al. (2011) previously introduced integrated behavioral z-scoring method for mice. In this paper, the design of the animal experiments was such that there was always a clear control group which served as a reference group. However, for some animal studies it is not possible to readily assign a control group, for example, when inbred mouse strains are compared. The present findings obtained from behavioral testing of three inbred mouse strains show that the outcome of the statistical analyses (statistical significance and effect size) may be influenced by the choice of the reference strain (Tables 1–4). A potential pitfall is that researchers may focus too much on the reference group that give results that fits them best, which can be seen as a form of data manipulation and points to scientific misconduct and thus unethical behavior (Van der Wall 2011). To avoid this pitfall, we suggest to take the pooled data as reference. Another problem that can be circumvented by using pooled values as a reference option is a control group with a SD value of zero. According to the equation used for behavioral z-score calculation (see section *Methods – Behavioral z-score calculation*: 2.6), a SD of zero is problematic because dividing by zero is undefined (Tables 2 and 4). By taking the pooled data as reference, the calculated SD will not be zero, otherwise the behavioral measure in question will not vary. It is therefore not meaningful to incorporate such a measure in the integrated behavioral z-score calculation procedure. In case of repeated measures and crossover designs (http://www.3rs-reduction.co.uk/html/9__experimental_designs.html), we also advise to pool all the data in order to reduce the risk of getting a reference population with a SD of zero. In fact, this is what was done in the present study when the '*three inbred strains-basal plus after i.p.*' pooled values were used as reference option.

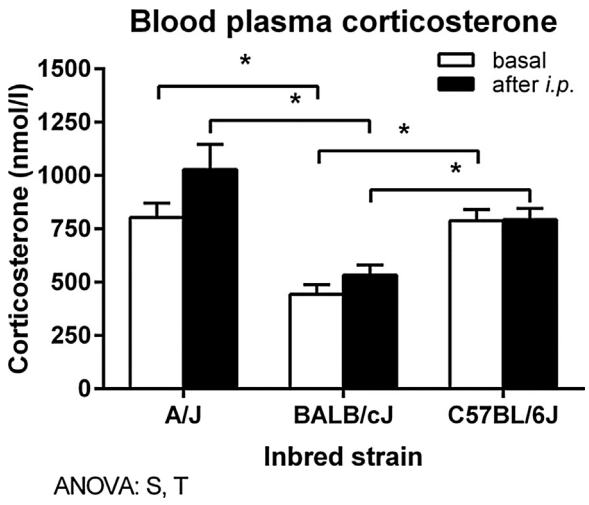

Fig. 3. Blood plasma corticosterone level of male mice from three inbred strains. The mice were tested in the mHB under basal conditions and after an intra-peritoneal (*i.p.*) injection. 30 min after behavioral testing blood was collected. Results are presented as means plus SEM. Number of animals per group: baseline A/J $n = 15$, BALB/cJ $n = 12$, C57BL/6J $n = 15$; after *i.p.* A/J $n = 10$, BALB/cJ $n = 11$, C57BL/6J $n = 14$). ANOVA = a multivariate repeated measures analysis of variance with main between-subject factors *strain* and within-subject factor *treatment*. Effects were significant in the ANOVA when $P < 0.05$. S indicates significant effect of strain; T, significant effect of injection (treatment); SxT, significant interaction. The ANOVA was performed after logarithmic transformation of the data ($y = {}^{10}Log[x - 87]$). *Post hoc* testing was done by paired Student's *t* test (within-inbred strain comparison) or unpaired Student's *t* test (between-inbred strain comparison; Gaussian distributed data + homoscedasticity: unpaired Student's *t* test; Gaussian distributed data + heteroscedasticity: unpaired Student's *t* test with Welch-Satterthwaite correction). * = Significant difference ($P < 0.016952$) in the *post hoc* comparison.

Choosing a reference population may also be of relevance for preclinical drug studies that include several vehicle controls. This can be illustrated, for example, with a drug study that includes two or more inbred strains of mice and for each strain a vehicle control. In the past we investigated the effects of an agonist (dexmedetomidine) and an antagonist (atipamezole) of the alpha 2A-adrenoceptor in male mice from a consomic strain (also called chromosome substitution strain; C57BL/6J-Chr $19^{A/J}$/NaJ = anxious), and the corresponding donor (A/J = anxious) and host (C57BL/6J = non-anxious) strains on mHB

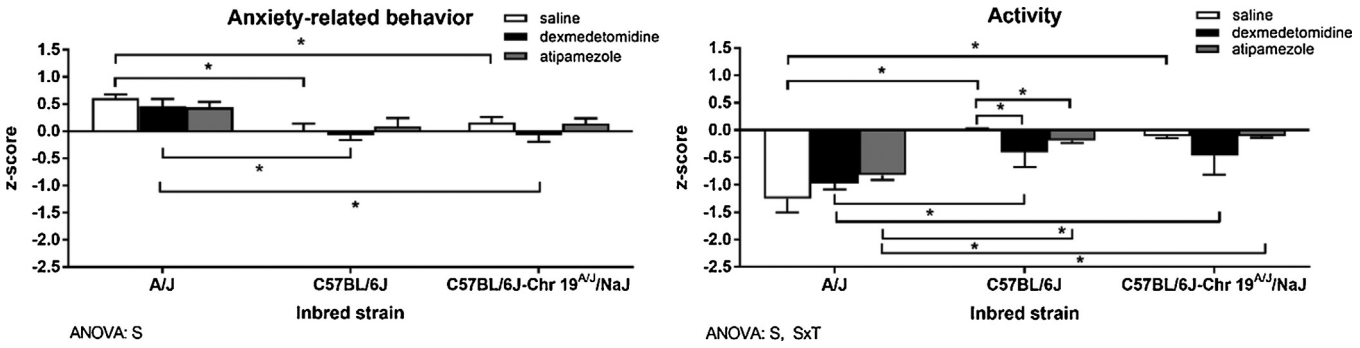

Fig. 4. Z-scores for anxiety-related behavior and activity in the mHB of male mice from three inbred strains with saline, dexmedetomidine or atipamezole *i.p.* injection. The pooled values of the nine groups were used as reference option, followed by subtracting the mean z-score of the C557BL/6J, *i.p.* saline group. Left panel: anxiety-related behavior; right panel: activity. Results are presented as means plus or minus SEM. Inbred strains (from The Jackson Laboratory): A/J, C57BL/6J and C57BL/6J-Chr 19$^{A/J}$/NaJ. Results are based on 15 male mice per group and were taken from Laarakker et al. (2010). Details for this preclinical drug study can be found in this reference. ANOVA = a two-way analysis of variance with main between-subject factors *strain* and *treatment*. Effects were significant in the ANOVA when $P < 0.05$. S indicates effect of strain; T, effect of treatment; SxT, interaction effect. The ANOVA's were performed after ranking of the data. *Post hoc* testing was done by Wilcoxon-Mann-Whitney test. * = Significant difference ($P < 0.012741$) in the *post hoc* comparison.

behavior (Laarakker et al., 2010). From the literature it is thought that the adrenergic alpha 2A receptor is involved in anxiety-related behavior (references cited in Laarakker et al., 2010). The integrated behavioral z-scoring methodology was applied for the obtained ethological results and four different reference group options (A/J, *i.p.* saline; C57BL/6J, *i.p.* saline; C57BL/6J-Chr 19$^{A/J}$/NaJ, *i.p.* saline; pooled data of nine groups) were selected. As would be expected, the outcomes regarding statistical significance and practical importance are influenced by the reference option (supplementary Tables S5 and S6) and clearly illustrates the aforementioned pitfalls. In consomic strain surveys the host strain normally serves as reference strain (Laarakker et al., 2008). Applying this reasoning to the preclinical drug study of Laarakker et al. (2010) one may select the C57BL/6J, *i.p.* saline group as reference. However, for the motivational system 'activity' this was not possible (supplementary Tables S5 and S6) and thus the pooled data of nine groups were taken as reference. If one would like to summarize the results with C57BL/6J, *i.p.* saline as pre-selected control group, then an extra step on the data is necessary: subtracting the mean z-score of the C57BL/6J, *i.p.* saline group. The consequentially obtained integrated behavioral z-score findings for anxiety-related behavior and activity are summarized in Fig. 4.

4.2. Comparison of mHB with EPM, LD and OF data

Behavioral data obtained with EPM, LD and OF for male mice from the inbred A/J, BALB/cJ and C57BL/6J strains are available in the MPD. For these tests, integrated z-scores for anxiety-related behavior and activity were calculated and these results were compared with basal values in the mHB. Based on the mean values the inbred strain ranking regarding anxiety-related behavior for OF and mHB is (from *'low anxious'* to *'high anxious'*): C57BL/6J – A/J – BALB/cJ, and for LD it is: C57BL/6J – BALB/cJ – A/J. For this motivational system the EPM clearly deviates, since the strain ranking is: BALB/cJ – C57BL/6J – A/J. For activity there is no difference in strain ranking between the four tests (from '*low active*' to '*high active*'): A/J – BALB/cJ – C57BL/6J (Fig. 5).

Altogether, we can conclude that the behavioral profile of the different strains obtained in the mHB (*i.e.* regarding the motivational systems anxiety-related behavior and activity) may be a combination of the profiles from the LD and OF. It is evident that the anxiety results obtained in the mHB deviate from those of the EPM. This is not surprising, as the mHB is a combination of the hole board and OF. Moreover, the board is lit with an additional lamp, which is reminiscent of one of the characteristics (a light area) of the LD (Laarakker et al., 2011; Labots et al., 2015; see section *Materials and methods – modified Hole Board (mHB) test*: 2.3). In contrast, the EPM may tax a different component of anxiety-related behavior than the mHB, LD and OF; *e.g.* acrophobia (fear due to height; Kalueff et al., 2008).

4.3. Substrain differences

Closely related substrains of inbred mice often show behavioral phenotype differences that are presumed to be caused by random mutations, genetic drift and/or differences between vendors. In the present study mHB behavior of the A/J, BALB/cJ and C57BL/6J mouse strains was analyzed, whereas in the past this behavior has been studied by our laboratory under the same conditions in A/JCrl, BALB/cAnNCrl and C57BL/6NCrl mice (Arndt et al., 2009). This prompted us to combine the present basal mHB results with those from Arndt et al. (2009) and use our improved procedure for integrated behavioral z-scoring, in order to investigate if there are significant substrain differences. These findings, *i.e.* anxiety-related behavior and activity, are summarized in Fig. 6.

Based on the ANOVA, significant differences for anxiety-related behavior between standard inbred strains were found. Further, the ANOVA revealed a significant vendor effect, as well as an interaction effect for anxiety-related behavior. Between-subject *post hoc* analyses revealed that the vendor difference was only significant for C57BL/6 mice: C57BL/6NCrl animals showed more anxiety-related behavior than their C57BL/6J counterparts. Analysis of activity with the ANOVA showed significant strain, vendor and interaction effects. Between-subject *post hoc* analyses revealed that the vendor difference was significant for A and C57BL/6, but not for BALB/c male mice: Charles River compared to The Jackson Laboratory A and C57BL/6 animals were less active.

C57BL/6 is a commonly used inbred mouse strain in biomedical research, but widespread demand has led to generation of several substrains with different phenotypes, including behavior (*e.g.* Kang et al., 2015; Labots et al., 2016). In a recent study, comparing phenotypic and genomic analysis of C57BL/6J and C57BL/6N substrains, 15 structural variants differentiating C57BL/6J and C57BL/6N were identified encompassing genic regions, as well as 34 SNPs and 2 indels that distinguish these two twin strains (Simon et al., 2013). One of the structural variants that distinguish C57BL/6J and C57BL/6N is the intronic deletion in the *Csmd1* (CUB and Sushi multiple domains 1; located on mouse chromosome 8) gene (Simon et al., 2013). Interestingly, Steen et al. (2013) generated a *Csmd1* knockout by deleting a 1 kb sequence from exon/intron1; these mice developed an increased anxiety-like phenotype in the OF and EPM. Another interesting structural variant that differs between the

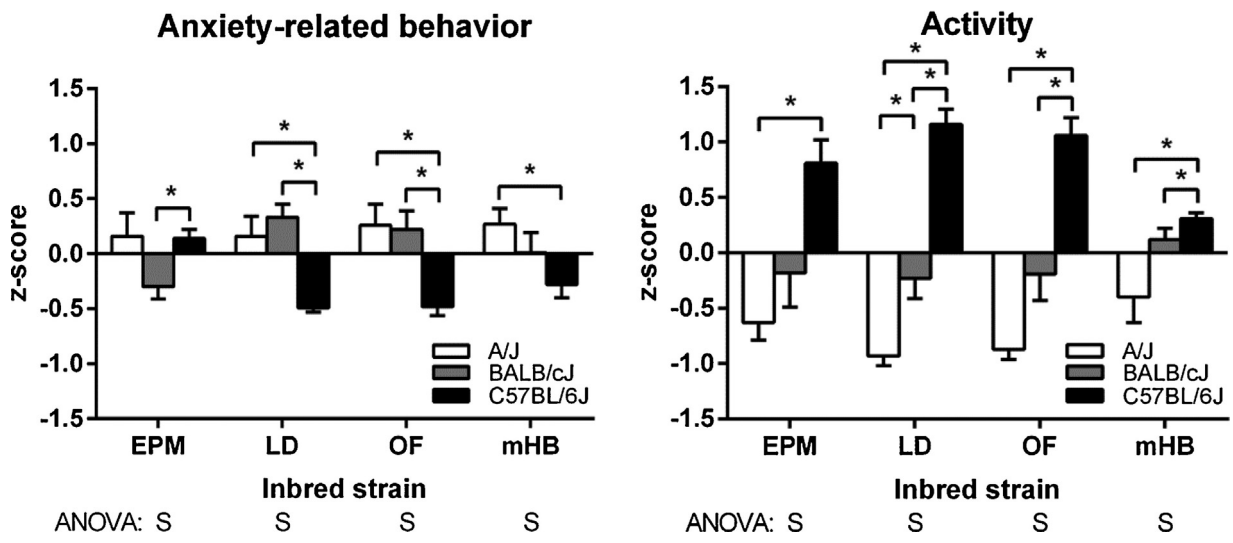

Fig. 5. Z-scores for anxiety-related behavior and activity in the EPM, LD, OF and mHB of male mice from three inbred strains.
The mice were tested under basal conditions. The 'three inbred strains-basal' pooled values were used as reference option. Abbreviations: EPM = elevated plus maze, LD = light-dark box, OF = open field, mHB = modified Hole Board. Results are presented as means plus or minus SEM. Results for EPM, LD and OF are based on 12 mice per inbred strain and were taken from O'Leary et al. (2013) via the MPD (http://www.phenome.jax.org; Brown et al., 2004). Results for mHB are based on 15 (A/J and C57BL/6J inbred strain) or 12 mice (BALB/cJ inbred strain). ANOVA = a one-way analysis of variance with main between-subject factor *strain*. Effects were significant in the ANOVA when $P < 0.05$. S indicates significant effect of strain. The ANOVA's were performed after ranking of the data and per behavioral test. *Post hoc*testing was done by Wilcoxon-Mann-Whitney test (between-inbred strain comparison). * = Significant difference ($P < 0.025321$) in the *post hoc* comparison.

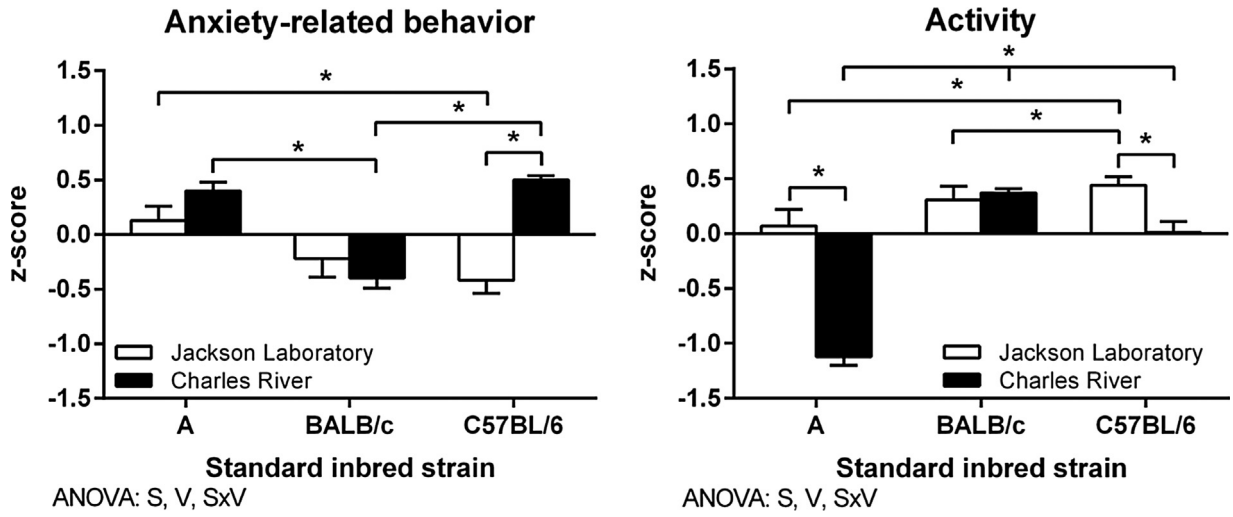

Fig. 6. Z-scores for anxiety-related behavior and activity in the mHB of male mice from three inbred strains purchased from two different vendors.
The mice were tested under basal conditions. The pooled values of the six strains were used as reference option. Results are presented as means plus or minus SEM. Inbred strains from The Jackson Laboratory: A/J, BALB/cJ and C57BL/6J. Inbred strains from Charles River: A/JCrl, BALB/cAnNCrl and C57BL/6NCrl. Results for Charles River are based on 17 (A/JCrl and C57BL/6NCrl) or 18 (BALB/cAnNCrl) mice and were taken from Arndt et al. (2009). Results for The Jackson Laboratory are based on 15 (A/J and C57BL/6J) or 12 mice (BALB/cJ). ANOVA = a two-way analysis of variance with main between-subject factors *strain* and *vendor*. Effects were significant in the ANOVA when $P < 0.05$. S indicates effect of strain; V, effect of vendor; SxV, interaction effect. The ANOVA's were performed after ranking of the data. *Post hoc* testing was done by Wilcoxon-Mann-Whitney test (between-vendor substrain comparison and within-vendor inbred strain comparison). * = Significant difference ($P < 0.016952$) in the *post hoc* comparison.

two twin strains is a deletion in the *Pde4b* (phosphodiesterase 4B, cAMP specific; located on mouse chromosome 4) gene (Simon et al., 2013). *Pde4b* knockout mice show increased anxiety-like behavior in the hole board, LD and OF tests (Zhang et al., 2008). It may therefore be that the aforementioned differences in the *Csmd1* and *Pde4b* genes are responsible for the C57BL/6 substrain difference in anxiety-related behavior in the mHB (Fig. 6). On the other hand, differences between vendors in feed components, bedding type or microbiome could also have contributed. For example, there is preclinical evidence to suggest that gut microbiota have an important role in bidirectional interactions between the gut and the nervous system, which may in turn influence behaviors including anxiety (Foster and McVey Neufeld 2013). In addition, mouse behavior can be influenced by the type of handling at the supplier (Miller and Leach 2016) and the mode of transport to the research facility (Swallow et al., 2005). Since C57BL/6J and C57BL/6N are widely used in mouse knockout studies, these behavioral phenotypic substrain differences – regardless of the basis – have the potential to impact upon penetrance and expressivity of mutational effects on behavior in these twin strains. Further, we think that investigators using C57BL/6 mice as animal models for anxiety should be cautious in comparing data from mice obtained from different vendors.

5. Conclusions

In summary, in the present study the contrasting A/J, BALB/cJ and C57BL/6J inbred mouse strains, which are commonly used in preclinical anxiety research, were behaviorally characterized by assessing their performance in the mHB test. The behavior of male animals before and after an acute aversive stimulus (*i.p.* injection with saline) was compared to test for the robustness of the behavioral phenotypes. Subsequently, the so-called integrated behavioral z-scoring methodology was applied on the obtained ethological results; five different reference options were taken and the outcome regarding statistical significance and effect sizes were compared. In order to avoid the pitfall of data manipulation, it is suggested to take the pooled data (*i.e. 'three inbred strains-basal plus after i.p.'*) as reference option. This will also reduce the risk of getting a reference population with a SD of zero, which is problematic for the calculation of integrated behavioral z-scores. The present procedure will therefore also improve the generalizability to all types of experimental designs – with or without (a) control group(s) – in animal behavioral research. Using this refined method, male mice from the three inbred strains differed regarding activity and anxiety; there also was a treatment effect for anxiety. Based on this improved integrated behavioral z-scoring method the mHB was compared with three other unconditioned behavior tests for anxiety and studied (and detected) also substrain differences. The results suggest that mHB, LD and OF assess a similar component of anxiety-related behavior, but deviate from the EPM.

Acknowledgements

The authors would like to dedicate this article to the memory of prof. dr. Frauke Ohl (chair of the department Animals in Science and Society and professor in Animal Welfare and Laboratory Animal Science) who passed away on January 28th 2016 at a too young age. We are grateful for her contribution to the work leading up to this manuscript. We remember Frauke as an inspiring and beloved colleague and supervisor. We thank prof. dr. Louk Vanderschuren for comments on an earlier version of this manuscript.

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.jneumeth.2017.09.003.

References

Algina, J., Keselman, H.J., 1997. Detecting repeated measures effects with univariate and multivariate statistics. Psychol. Methods 2 (2), 208–218, http://dx.doi.org/10.1037/1082-989X.2.2.208.

Arndt, S.S., Laarakker, M.C., van Lith, H.A., van der Staay, F.J., Gieling, E., Salomons, A.R., van 't Klooster, J., Ohl, F., 2009. Individual housing of mice – impact on behaviour and stress responses. Physiol. Behav. 97 (3–4), 385–393, http://dx.doi.org/10.1016/j.physbeh.2009.03.008.

Bouwknecht, J.A., Paylor, R., 2008. Pitfalls in the interpretation of genetic and pharmacological effects on anxiety-like behaviour in rodents. Behav. Pharmacol. 19 (5–6), 385–402, http://dx.doi.org/10.1097/FBP.0b013e32830c3658.

Brown, R.E., Gunn, R.K., Schellinck, H.M., Wong, A.A., O'Leary, T.P., 2004. Anxiety, exploratory behavior, and motor activity in 14 inbred strains of mice. MPD:Brown1. In: Mouse Phenome Database Web Site. The Jackson Laboratory, Bar Harbor, Maine USA, Accessed 01 August 2017 http://phenome.jax.org.

Budaev, S.V., 2010. Using principal components and factor analysis in animal behaviour research: caveats and guidelines. Ethology 116 (5), 472–480, http://dx.doi.org/10.1111/j.1439-0310.2010.01758.x.

Conover, W., Iman, R., 1981. Rank transformations as a bridge between parametric and nonparametric statistics. Am. Stat. 35 (3), 124–129, http://dx.doi.org/10.2307/2683975.

Dürschlag, M., Würbel, H., Stauffacher, M., Von Holst, D., 1996. Repeated blood collection in the laboratory mouse by tail incision—modification of an old technique. Physiol. Behav. 60 (6), 1565–1568, http://dx.doi.org/10.1016/S0031-9384(96)00307-1.

Ennaceur, A., Chazot, P.L., 2016. Preclinical animal anxiety research-flaws and prejudices. Pharmacol. Res. Perspect. 4 (2), e00223, http://dx.doi.org/10.1002/prp2.223.

Fagerland, M.W., Sandvik, L., Mowinckel, P., 2011. Parametric methods outperformed non-parametric methods in comparisons of discrete numerical variables. BMC Med. Res. Methodol. 11 (44), 1–8, http://dx.doi.org/10.1186/1471-2288-11-44.

Festing, M.F.W., 2014. Extending the statistical analysis and graphical presentation of toxicity test results using standardized effect sizes. Toxicol. Pathol. 42 (8), 1238–1249, http://dx.doi.org/10.1177/0192623313517771.

Field, A., 2009. Discovering Statistics Using SPSS, third edition. SAGE Publications Ltd, London, U.K.

Foster, J.A., McVey Neufeld, K., 2013. Gut-brain axis: how the microbiome influences anxiety and depression. Trends Neurosci. 36 (5), 305–312, http://dx.doi.org/10.1016/j.tins.2013.01.005.

Guilloux, J., Seney, M., Edgar, N., Sibille, E., 2011. Integrated behavioral z-scoring increases the sensitivity and reliability of behavioral phenotyping in mice: relevance to emotionality and sex. J. Neurosci. Methods 197 (1), 21–31, http://dx.doi.org/10.1016/j.jneumeth.2011.01.019.

Holmes, A., Li, Q., Murphy, D.L., Gold, E., Crawley, J.N., 2003. Abnormal anxiety-related behavior in serotonin transporter null mutant mice: the influence of genetic background. Genes Brain Behav. 2, 365–380, http://dx.doi.org/10.1046/j.1601-183X.2003.00050.x.

Kalueff, A.V., Ishikawa, K., Griffith, A.J., 2008. Anxiety and otovestibular disorders: linking behavioral phenotypes in men and mice. Behav. Brain Res. 186 (1), 1–11, http://dx.doi.org/10.1016/j.bbr.2007.07.032.

Kang, M., Ryu, H., Lee, Y., 2015. Comparisons of behavior and synaptic plasticity among three C57BL/6 substrains. Anim. Cells Syst. 19 (3), 181–187, http://dx.doi.org/10.1080/19768354.2015.1023830.

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G., 2010. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol. 8 (6), e1000412, http://dx.doi.org/10.1371/journal.pbio.1000412.

Korte, S.M., 2001. Corticosteroids in relation to fear, anxiety and psychopathology. Neurosci. Biobehav. Rev. 25 (2), 117–142, http://dx.doi.org/10.1016/S0149-7634(01)00002-1.

Laarakker, M.C., Ohl, F., van Lith, H.A., 2008. Chromosomal assignment of quantitative trait loci influencing modified hole board behavior in laboratory mice using consomic strains, with special reference to anxiety-related behavior and mouse chromosome 19. Behav. Genet. 38 (2), 159–184, http://dx.doi.org/10.1007/s10519-007-9188-6.

Laarakker, M.C., van Raai, J.R., van Lith, H.A., Ohl, F., 2010. The role of the alpha 2A-adrenoceptor in mouse stress-coping behaviour. Psychoneuroendocrinology 35, 490–502, http://dx.doi.org/10.1016/j.psyneuen.2009.08.014.

Laarakker, M.C., van Lith, H.A., Ohl, F., 2011. Behavioral characterization of A/J and C57BL/6J mice using a multidimensional test: association between blood plasma and brain magnesium-ion concentration with anxiety. Physiol. Behav. 102 (2), 205–219, http://dx.doi.org/10.1016/j.physbeh.2010.10.019.

Labots, M., van Lith, H.A., Ohl, F., Arndt, S.S., 2015. The modified hole board – measuring behavior, cognition and social interaction in mice and rats. J. Vis. Exp. 98, e52529, http://dx.doi.org/10.3791/52529.

Labots, M., Zheng, X., Moattari, G., Ohl, F., van Lith, H.A., 2016. Effects of light regime and substrain on behavioral profiles of male C57BL/6 mice in three tests of unconditioned anxiety. J. Neurogenet. 30 (3–4), 306–315, http://dx.doi.org/10.1080/01677063.2016.1249868.

Lovell, D.P., 2013. Biological importance and statistical significance. J. Agric. Food Chem. 61 (35), 8340–8348, http://dx.doi.org/10.1021/jf401124y.

Meziane, H., Ouagazzal, A., Aubert, L., Wietrzych, M., Krezel, W., 2007. Estrous cycle effects on behavior of C57BL/6J and BALB/cByJ female mice: implications for phenotyping strategies. Genes Brain Behav. 6 (2), 192–200, http://dx.doi.org/10.1111/j.1601-183X.2006.00249.x.

Miller, A.L., Leach, M.C., 2016. The effect of handling method on the mouse grimace scale in two strains of laboratory mice. Lab. Anim. 50 (4), 305–307, http://dx.doi.org/10.1177/0023677215622144.

National Research Council of the National Academies, 2003. Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research. National Academies Press, Washington D.C., USA.

Noldus, L., 1991. The Observer: a software system for collection and analysis of observational data. Behav. Res. Methods Instrum. Comput. 23, 415–429.

O'Leary, T.P., Gunn, R.K., Brown, R.E., 2013. What are we measuring when we test strain differences in anxiety in mice? Behav. Genet. 43 (1), 34–50, http://dx.doi.org/10.1007/s10519-012-9572-8.

Paylor, R., Spencer, C.M., Yuva-Paylor, L.A., Pieke-Dahl, S., 2006. The use of behavioral test batteries, II: effect of test interval. Physiol. Behav. 87 (1), 95–102, http://dx.doi.org/10.1016/j.physbeh.2005.09.002.

Scheirer, C.J., Ray, W.S., Hare, N., 1976. The analysis of ranked data derived from completely randomized factorial designs. Biometrics 32 (2), 429–434, http://dx.doi.org/10.2307/2529511.

Simon, M.M., Greenaway, S., White, J.K., Fuchs, H., Gailus-Durner, V., Wells, S., Sorg, T., Wong, K., Bedu, E., Cartwright, E.J., 2013. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. Genome Biol. 14 (R82), 1–22, http://dx.doi.org/10.1186/gb-2013-14-7-r82.

Song, M.K., Lin, F.C., Ward, S.E., Fine, J.P., 2013. Composite variables: when and how. Nurs. Res. 62 (1), 45–49, http://dx.doi.org/10.1097/NNR.0b013e3182741948.
Steen, V.M., Nepal, C., Ersland, K.M., Holdhus, R., Nævdal, M., Ratvik, S.M., Skrede, S., Håvik, B., 2013. Neuropsychological deficits in mice depleted of the schizophrenia susceptibility gene CSMD1. PLoS One 8 (11), e79501.
Swallow, J., Anderson, D., Buckwell, A.C., Harris, T., Hawkins, P., Kirkwood, J., Lomas, M., Meacham, S., Peters, A., et al., 2005. Guidance on the transport of laboratory animals. Lab. Anim. 39, 1–39, http://dx.doi.org/10.1258/0023677052886493.
Trullas, R., Skolnick, P., 1993. Differences in fear motivated behaviors among inbred mouse strains. Psychopharmacology (Berl) 111 (3), 323–331, http://dx.doi.org/10.1007/BF02244948.
Van der Wall, E., 2011. Summa cum fraude: how to prevent scientific misconduct. Neth. Heart J. 19 (2), 57–58, http://dx.doi.org/10.1007/s12471-010-0062-4.
Wahlsten, D., 2011. Chapter 5: sample size. In: Mouse Behavioral Testing; How to Use Mice in Behavioral Neuroscience, 1st edn. Academic Press, Elsevier Inc., London, U.K, pp. pp 75–105.
Zhang, H., Huang, Y., Masood, A., Stolinski, L.R., Li, Y., Zhang, L., Dlaboga, D., Jin, S.C., Conti, M., O'Donnell, J.M., 2008. Anxiogenic-like behavioral phenotype of mice deficient in phosphodiesterase 4 B (PDE4B). Neuropsychopharmacology 33 (7), 1611–1623, http://dx.doi.org/10.1038/sj.npp.1301537.

Table S1 Behavioral variables measured in the mHB and used for this study.

Motivational system / Behavioral dimension	*Behavioral variable*	*Description of the mouse behavior*
Anxiety-related behavior		
- *Avoidance [z]*[1]	Total number of board entries *[-z]*[1]	Mouse is on the board (= 'unprotected area') with at least two front paws and its head
	Latency until the first board entry *[z]*	
	Percentage of time on the board *[-z]*	
- *Risk assessment [z]*	Total number of stretched attends *[z]*	Completely stretched body posture, including hind paws; also referred to as 'stretched attends'
	Latency until the first stretched attend *[-z]*	
- *Arousal [z]*	Total number of self-groomings *[z]*	Self-grooming (cleaning fur)
	Latency until the first self-grooming *[-z]*	
	Percentage of time being busy with self-grooming *[z]*	
	Total number of boli produced *[z]*	Defecation
	Latency until the first bolus is produced *[-z]*	
Activity		
- *Locomotion [z]*		
→ Horizontal	Total number of line crossings *[z]*	Line crossing with all its paws in the box (= 'protected area')
	Latency until the first line crossing *[-z]*	
→ Vertical	Total number of rearings in the box *[z]*	Rearing on hind paws in the box; front paws of the mouse are off the floor, either leaning against the side wall or free(the mouse is then only standing on its hind paws), sniffing
	Latency until the first rearing in the box *[-z]*	
	Total number of rearings on the board *[z]*	Rearing on hind-paws on the board (also see description 'rearing in the box')
	Latency until the first rearing on the board *[-z]*	
- *Immobility [-z]*	Total number of being motionless *[z]*	*Being completely motionless*
	Latency until the first time the mouse is motionless *[-z]*	
	Percentage of time being motionless *[z]*	

[1] The directionality of the z-scores was adjusted so that increased score values reflected increased values for that behavioral dimension or motivational system: *[z]* = regular z-score, *[-z]* = adjusted z-score.

Table S2 Selected behavioral variables from three traditional tests measuring anxiety-like and locomotor behavior and used for this paper[1].

Behavioral test / *Motivational system / Behavioral dimension*	*Behavioral variable*
Elevated plus maze (EPM)	
Anxiety-related behavior	
- *Avoidance [z]*[2]	Index of open arm avoidance *[z]*[2] Duration in closed arms *[z]* Total number of closed arm entries *[z]*
- *Risk assessment [z]*	Total number of stretch-attend postures *[z]*
- *Arousal [z]*	Grooming duration *[z]* Total number of defecations *[z]* Total number of urine puddles *[z]*
Activity	
- *Locomotion [z]*	Total number of line crossings plus rearings *[z]*
Light-dark box (LD)	
Anxiety-related behavior	
- *Avoidance [z]*	Duration in dark zone *[z]*
- *Risk assessment [z]*	Total number of stretch-attend postures *[z]*
- *Arousal [z]*	Grooming duration *[z]* Total number of defecations *[z]* Total number of urine puddles *[z]*
Activity	
- *Locomotion [z]*	Total number of line crossings plus rearings *[z]*
Open field (OF)	
Anxiety-related behavior	
- *Risk assessment [z]*	Total number of stretch-attend postures *[z]*
- *Arousal [z]*	Grooming duration *[z]* Total number of defecations *[z]* Total number of urine puddles *[z]*
Activity	
- *Locomotion [z]*	Total number of line crossings plus rearings *[z]*

[1] Variables were taken from O'Leary et al. (2013) via the MPD (http://www.phenome.jax.org; Brown *et al.* 2004).

[2] The directionality of the z-scores was adjusted so that increased score values reflected increased values for that behavioral dimension or motivational system: *[z]* = regular z-score, *[-z]* = adjusted z-score. Here we only have regular z-scores. For the mHB (Table S1) there were also adjusted z-scores.

Table S3 Z-scores for anxiety-related behavior in the mHB of male mice from three inbred strains under basal conditions and after an intra-peritoneal (*i.p.*) injection using different reference populations and conditions.[1]

Motivational system / Behavioral dimension	*Reference population and condition*	*Mouse inbred strain* A/J (A[2]; n = 15)		BALB/cJ (C[2]; n = 12)		C57BL/6J (B6[2]; n = 15)		*ANOVA*[3] *significance*
		Basal	*After i.p.*	*Basal*	*After i.p.*	*Basal*	*After i.p.*	
Anxiety-related behavior								
- Avoidance	A/J (A)[2] *basal*	0.00±0.25	-0.23±0.24	-0.95±0.33	-0.60±0.36	-0.36±0.29[a4]	-1.22±0.30[a]	SxT
	BALB/cJ (C)[2] *basal*	0.80±0.23	0.62±0.22	0.00±0.24	0.36±0.28	0.58±0.25[a]	-0.18±0.25[a]	SxT
	C57BL/6J (B6)[2] *basal*	0.26±0.22	0.08±0.22	-0.55±0.26	-0.21±0.30	0.00±0.25[a]	-0.76±0.26[a]	SxT
	A + C + B6 *basal*	0.31±0.22	0.14±0.21	-0.46±0.25	-0.14±0.28	0.06±0.24[a]	-0.67±0.25[a]	SxT
	A + C + B6 *basal* and *after i.p.*	0.41±0.21	0.24±0.20	-0.34±0.24	-0.02±0.27	0.17±0.23[a]	-0.54±0.24[a]	SxT
- Risk assessment	A/J (A) *basal*	0.00±0.23[a]	-0.58±0.18[a]	0.44±0.26[b]	-0.96±0.29[b]	-0.38±0.22[c]	-1.36±0.21[c]	T
	BALB/cJ (C) *basal*	-0.73±0.45	-1.62±0.47	0.00±0.24[a]	-2.78±0.67[a]	-1.29±0.50[b]	-3.66±0.53[b]	T
	C57BL/6J (B6) *basal*	0.45±0.25[a]	-0.23±0.18[a]	0.96±0.34[b]	-0.58±0.30[b]	0.00±0.23[c]	-0.99±0.20[c]	T
	A + C + B6 *basal*	0.01±0.22	-0.53±0.19	0.42±0.22[a]	-0.93±0.29[a]	-0.34±0.21[b]	-1.32±0.21[b]	T
	A + C + B6 *basal* and *after i.p.*	0.43±0.20[a]	-0.10±0.15[a]	0.83±0.26[b]	-0.41±0.25[b]	0.08±0.19[c]	-0.75±0.17[c]	T
- Arousal	A/J (A) *basal*	0.00±0.15[A]	-0.06±0.12	-0.38±0.18[B]	0.01±0.33	-0.98±0.09[ABa]	-0.46±0.16[a]	S, T
	BALB/cJ (C) *basal*	0.76±0.23[A]	0.71±0.18	0.00±0.22	0.86±0.57	-0.72±0.11[Aa]	0.10±0.26[a]	S, T
	C57BL/6J (B6) *basal*	2.58±0.38[A]	2.47±0.32	1.37±0.43	2.78±0.95	0.00±0.16[Aa]	1.32±0.44[a]	S, T
	A + C + B6 *basal*	0.50±0.16[A]	0.41±0.13[B]	0.07±0.19	0.47±0.34	-0.56±0.09[Aa]	-0.03±0.17[Ba]	S, T
	A + C + B6 *basal* and *after i.p.*	0.35±0.15[A]	0.20±0.11	0.03±0.18	0.20±0.25	-0.57±0.08[Aa]	-0.17±0.13[a]	S
Anxiety-related behavior								
	A/J (A) *basal*	0.00±0.15[A]	-0.29±0.12[B]	-0.29±0.21	-0.52±0.27	-0.57±0.14[Aa]	-1.01±0.19[Ba]	S, T
	BALB/cJ (C) *basal*	0.28±0.21[A]	-0.10±0.20[B]	0.00±0.18	-0.52±0.41	-0.48±0.20[Aa]	-1.25±0.30[Ba]	S, T
	C57BL/6J (B6) *basal*	1.10±0.17[A]	0.77±0.14[B]	0.59±0.29	0.67±0.44	0.00±0.14[A]	-0.14±0.26[B]	S
	A + C + B6 *basal*	0.27±0.14[A]	0.01±0.11[B]	0.01±0.18	-0.20±0.25	-0.28±0.12[Aa]	-0.67±0.18[Ba]	S, T
	A + C + B6 *basal* and *after i.p.*	*0*.40±0.13[A]	0.11±0.10[B]	0.17±0.18	-0.07±0.22	-0.11±0.12[Aa]	-0.48±0.16[Ba]	S, T

[1] Values are means ± SEM; *n* is the number of animals per inbred strain.

[2] Official abbreviation in parentheses (see http://www.informatics.jax.org/mgihome/nomen/strains.shtml).

[3] Significance ($P < 0.05$) based on a multivariate repeated measures ANOVA with main between-subject factors *strain* and within-subject factor *treatment*. S indicates effect of strain; T, effect of injection (treatment); SxT, interaction. The ANOVA's were performed after ranking of the data.

[4] Contrast significance ($P < 0.016952$). Within the same inbred strain and row values bearing the same superscript lowercase letter are significantly different (Wilcoxon matched-pairs signed ranks test). Within the same condition (basal or after *i.p.*) and row values bearing the same superscript uppercase letter are significantly different (Wilcoxon-Mann-Whitney test).

Table S4 Z-scores for activity in the mHB of male mice from three inbred strains under basal conditions and after an intra-peritoneal (*i.p.*) injection using different reference populations and conditions.[1]

Motivational system / Behavioral dimension	*Reference population and condition*	*Mouse inbred strain* *A/J (A[2]; n = 15)* *Basal*	*After i.p.*	*BALB/cJ (C[2]; n = 12)* *Basal*	*After i.p.*	*C57BL/6J (B6[2]; n = 15)* *Basal*	*After i.p.*	*ANOVA[3] significance*
Activity								
- Locomotion	A/J (A)[2] *basal*	0.00±0.15^{Aa4}	0.47±0.16^{aB}	0.55±0.28^{C}	0.76±0.31	1.46±0.07^{AC}	1.20±0.10^{B}	S, SxT
	BALB/cJ (C)[2] *basal*	-0.61±0.44^{A}	-0.05±0.18^{B}	0.00±0.21^{C}	0.15±0.17	0.55±0.04^{AC}	0.44±0.05^{B}	S
	C57BL/6J (B6)[2] *basal*	……*Not possible*[5]……						
	A + C + B6 *basal*	-0.26±0.15^{A}	-0.06±0.11^{B}	-0.12±0.19^{C}	-0.04±0.16	0.35±0.03^{AC}	0.26±0.04^{B}	S
	A + C + B6 *basal* and *after i.p.*	-0.31±0.18^{A}	-0.09±0.13^{B}	-0.14±0.22^{C}	-0.08±0.17	0.34±0.03^{AC}	0.24±0.04^{B}	S
- Immobility	A/J (A) *basal*	0.00±0.25	-0.28±0.16	-0.62±0.00	-0.49±0.09	-0.55±0.07	-0.62±0.00	S
	BALB/cJ (C) *basal*	……*Not possible*……						
	C57BL/6J (B6) *basal*	13.86±6.40^{A}	4.98±2.51	-0.26±0.00	0.37±0.43	0.00±0.26^{A}	-0.26±0.00	S
	A + C + B6 *basal*	0.54±0.36	0.13±0.22	-0.35±0.00	-0.18±0.12	-0.26±0.09	-0.35±0.00	S
	A + C + B6 *basal* and *after i.p.*	0.73±0.44	0.22±0.26	-0.33±0.00	-0.15±0.13	-0.24±0.09	-0.33±0.00	S
Activity								
	A/J (A) *basal*	0.00±0.18^{A}	0.37±0.15^{B}	0.59±0.14^{C}	0.62±0.17	1.00±0.06^{AC}	0.91±0.05^{B}	S, SxT
	BALB/cJ (C) *basal*	……*Not possible*……						
	C57BL/6J (B6) *basal*	……*Not possible*……						
	A + C + B6 *basal*	-0.40±0.23^{A}	-0.09±0.16^{B}	0.12±0.10^{C}	0.07±0.10	0.31±0.05^{AC}	0.31±0.02^{B}	S
	A + C + B6 *basal* and *after i.p.*	-0.52±0.28^{A}	-0.16±0.18^{B}	0.10±0.11^{C}	0.04±0.11	0.29±0.06^{AC}	0.29±0.02^{B}	S

[1] Values are means ± SEM; *n* is the number of animals per inbred strain.

[2] Official abbreviation (see http://www.informatics.jax.org/mgihome/nomen/strains.shtml).

[3] Significance ($P < 0.05$) based on a multivariate repeated measures ANOVA with main between-subject factors *strain* and within-subject factor *treatment*. S indicates effect of strain; T, effect of injection (treatment); SxT, interaction. The ANOVA's were performed after ranking of the data.

[4] Contrast significance ($P < 0.016952$). Within the same inbred strain and row values bearing the same superscript lowercase letter are significantly different (Wilcoxon matched-pairs signed ranks test). Within the same condition (basal or after *i.p.*) and row values bearing the same superscript uppercase letter are significantly different (Wilcoxon-Mann-Whitney test).

[5] For some activity-related behavioral variables the BALB/cJ or C57BL/6J inbred strain had a SD of zero and thus it was not possible to calculate z-scores.

Table S5 Overview of significant effects and effect sizes (ANOVA) regarding anxiety-related behavior and activity in the mHB of male A/J, C57BL/6J and C57BL/6J-Chr $19^{A/J}$/NaJ (consomic) mice after an intra-peritoneal injection of either saline, dexmedetomidine or atipamezole using different reference populations.[1]

		ANOVA[2] *significance*			*Partial eta squared* (η_p^2)[5] *ANOVA*		
Motivational system	*Reference population*	*S*	*T*	*SxT*	*S*	*T*	*SxT*
Anxiety-related behavior							
	A/J, *saline*	X[3]	X[4]	-	***0.276***	*0.054*	0.021
	C57BL/6J, *saline*	X	-	-	***0.393***	*0.031*	0.027
	C57BL/6J-Chr $19^{A/J}$*/NaJ, saline*	X	X	-	***0.413***	*0.046*	0.027
	All nine groups pooled .	X	-	-	***0.211***	0.023	0.016
Activity							
	A/J, *saline*	X	-	X	***0.454***	0.027	*0.038*
	C57BL/6J, *saline*	……..*Not possible*[6]……					
	C57BL/6J-Chr $19^{A/J}$*/NaJ, saline*	……..*Not possible*……					
	All nine groups pooled .	X	-	X	***0.251***	0.023	*0.043*

1 Results are based on 15 male mice per group and were taken from Laarakker *et al.* (2010).

2 Significance ($P < 0.05$) based on a two-way ANOVA with main between-subject factors *strain* and *treatment*. S indicates effect of strain; T, effect of treatment (saline, dexmedetomidine or atipamezole intra-peritoneal injection); SxT, interaction. The ANOVA's were performed after ranking of the data.

3 X = significant effect, - = no significant effect.

4 Deviant results from the other reference options are indicated in purple.

5 For calculation of the effect sizes non-ranked data were used. Small effects are indicated in normal fonts, medium effects in *italics* and *green*, large effects in **bold** and **yellow**, and very large effects in ***bolditalics*** and ***red***.

6 For some activity-related behavioral variables the C57BL/6J and *C57BL/6J-Chr* $19^{A/J}$*/NaJ* inbred strain had a SD of zero and thus it was not possible to calculate z-scores.

Table S6 Overview of significant differences and effect sizes (*post hoc* comparisons) regarding anxiety-related behavior and activity in the mHB of male A/J, C57BL/6J and C57BL/6J-Chr 19$^{A/J}$/NaJ (consomic) mice after an intra-peritoneal injection of either saline, dexmedetomidine or atipamezole using different reference populations.[1]

Motivational system / Reference population	*Post hoc* comparison[2]																	
	1-2	1 -3	1-4	1-7	2-3	2-5	2-8	3-6	3-9	4-5	4-6	4-7	5-6	5-8	6-9	7-8	7-9	8-9
								SIGNIFICANCE[3]										
Anxiety-related behavior																		
A-s[6]	-	-	X[4]	X	-	X	X	-	X	-	-	-	-	-	-	-	-	-
B6-s	-	-	X	X	-	X	X	X[5]	X	-	-	-	-	-	-	-	-	-
CSS19-s	-	-	X	X	-	X	X	X	X	-	-	-	-	-	-	-	-	-
Pooled	-	-	X	X	-	X	X	-	-	-	-	-	-	-	-	-	-	-
Activity																		
A-s	-	-	X	X	-	X	X	X	X	X	X	-	-	-	-	-	-	-
B6-s	……*Not possible*[7]……																	
CSS19-s	……*Not possible*……																	
Pooled	-	-	X	X	-	X	X	X	X	X	X	-	-	-	-	-	-	-
								EFFECT SIZE[8]										
Anxiety-related behavior																		
A-s	*-0.59*	*-0.51*	***-1.78***	***-2.08***	0.11	***-1.70***	***-1.51***	*-0.83*	**-1.32**	-0.34	0.18	0.43	*0.53*	-0.02	-0.07	*-0.93*	-0.34	*0.60*
B6-s	-0.49	*-0.78*	***-2.23***	***-2.13***	-0.13	**-1.50**	**-1.42**	**-1.19**	**-1.32**	-0.19	0.09	0.33	0.29	-0.03	0.13	*-0.59*	-0.13	*0.51*
CSS19-s	-0.41	*-0.79*	***-2.09***	***-2.26***	-0.31	***-2.03***	***-1.86***	**-1.23**	***-1.68***	-0.20	0.00	*0.60*	0.20	-0.07	-0.07	*-0.97*	*-0.89*	0.22
Pooled	-0.39	*-0.53*	***-1.55***	**-1.41**	-0.05	**-1.21**	**-1.12**	*-0.71*	*-0.77*	-0.15	0.16	0.35	0.33	-0.02	0.11	*-0.57*	-0.05	*0.53*
Activity																		
A-s	0.27	*0.53*	***3.34***	***2.83***	0.41	**1.49**	**1.16**	***3.74***	***4.28***	*-0.81*	**-1.32**	*-0.83*	0.31	-0.03	0.42	-0.41	-0.04	0.42
B6-s	……*Not possible*……																	
CSS19-s	……*Not possible*……																	
Pooled	0.40	*0.65*	***2.27***	***2.07***	0.41	*0.79*	*0.59*	***2.47***	***3.13***	*-0.68*	**-1.24**	*-0.76*	0.34	-0.06	*0.58*	-0.49	-0.02	*0.50*

[Table S6, continued]

[1] Results are based on 15 male mice per group and were taken from Laarakker *et al.* (2010).

[2] *Post hoc* comparisons, codes: 1 = A/J, saline; 2 = A/J, dexmedetomidine; 3 = A/J, atipamezole; 4 = C57BL/6J, saline; 5 = C57BL/6J, dexmedetomidine; 6 = C57BL/6J, atipamezole; 7 = C57BL/6J-Chr $19^{A/J}$/NaJ, saline; 8 = C57BL/6J-Chr $19^{A/J}$/NaJ, dexmedetomidine; 9 = C57BL/6J-Chr $19^{A/J}$/NaJ, atipamezole. Example: 1-2 = A/J, saline versus A/J dexmedetomidine.

[3] Contrast significance (Wilcoxon-Mann-Whitney test; $P < 0.012741$).

[4] X = significant difference, - = no significant difference.

[5] Deviant results from the other reference options are indicated in purple.

[6] Reference populations: A-s = A/J, *saline*; B6-s = C57BL/6J, *saline*; CSS19-s = C57BL/6J-Chr $19^{A/J}$/NaJ, *saline*; Pooled: all the nine groups pooled.

[7] For some activity-related behavioral variables the C57BL/6J or C57BL/6J-Chr $19^{A/J}$/NaJ inbred strain had a SD of zero and thus it was not possible to calculate z-scores.

[8] For calculation of the effect sizes non-ranked data were used. Small effects are indicated in normal fonts, medium effects in *italics* and *green*, large effects in **bold** and **yellow**, and very large effects in ***bolditalics*** and ***red***.