Supplementary material

- Systematic analysis of severity in a widely-used cognitive depression model for mice

Mallien et al.

Table 1 Overview of groups in each experiment.

Experiment1: FCM

Group	Description
home cage control	remain in the colony room during LH procedure
handling	enter the setups, but remain unshocked
non-trained	enter the setups, but only receive escapable shocks in the shuttle box (day 3)
trained	enter the setups, and receive inescapacle shocks (day 1&2) and escapable shocks in the shuttle box (day 3)

Experiment 2: typical indicator of wellbeing

Group	Description
home cage control	remain in the colony room during LH procedure
handling	enter the setups, but remain unshocked
non-trained	enter the setups, but only receive escapable shocks in the shuttle box (day 3)
trained	enter the setups, and receive inescapacle shocks (day 1&2) and escapable shocks in the shuttle box (day 3)

Experiment 3: VWR

Group	Description
home cage control	remain in the colony room during LH procedure
handling	enter the setups, but remain unshocked
trained	enter the setups, and receive inescapacle shocks (day 1&2) and escapable shocks in the shuttle box (day 3)
DSS	Colitis model, received 1% DSS for 5 consecutive days, remain in the colony room during the LH procedure

Animals and housing

Eight weeks old male C57BL/6N mice were provided by Charles River Laboratories (Sulzfeld, Germany), and allowed to adapt to an 12 h/12 h inverted dark-light cycle of the colony room for at least 2 weeks. The animals were pseudorandomly assigned to the respective experimental group. Locomotion and pain thresholds were used to stratify mice into groups

and avoid confounding effects. The room temperature was 23 ± 2°C, the relative humidity 50 ± 5%. Unless further specified, mice were single-housed in conventional macrolon cages (Type II) with bedding (Aspen granule 2-3 cm x 3mm, Abedd, Vienna, Austria), nesting material (unbleached cellulose tissue), tap water and food (LASQC diet Rod16 HiHyg, Altromin, Lage) *ad libitum*. Mice were tail handled. For each experiment, we used used the power analysis program G*Power 3.1 to calculate a priori the required sample size.

Open field test

Mice were placed into an unfamiliar white 50 x 50 cm^2 arena illuminated with 25 lux. Activity was recorded for 10 min and analyzed with image processing system EthoVision XT 8.0 (Noldus Information Technology, Wageningen, the Netherlands). Total distance moved, velocity and time spent in center zone (area 10 cm distant from walls) was evaluated as described earlier [1].

Hot plate test

The pain threshold was assessed in the hot plate test (ATLab, Vendargues, France) by placing the mice (for max. 45 seconds) on a 53 ± 0.3 °C plate until it showed a first coping reaction, e.g. licking the hind paws or jumping.[2]

Cognitive depression

The Learned Helplessness paradigm was conducted as previously described[2]. Briefly, mice were transferred to the experimental room where they were introduced into plexiglas chambers (18 x 18 x 30 cm^3) with stainless steel grid floors (Coulborn Instruments, Düsseldorf, Germany). Trained subjects received 360 unpredictable and inescapable foot shocks (0.150 mA) on two consecutive days, respectively. On day 3, trained and non-trained control subjects were analyzed for helpless behavior in shuttle boxes consisting of two compartments

connected via an opening. Each chamber contained a signaling light, which announced foot shocks (0.150 mA) for 5 s. The escape performance was analyzed during 30 trials. The inter-trial interval was set to 30 s and total testing time was max. 22 min.

General information on the assessed welfare parameters

The behavioral readouts for the well-being were chosen, because they are common and well-established. All of them are performed within the home cage. On one hand, this implies the possibility to integrate rating into the daily routine, since it avoids complex assessments including transfers to other facilities and is therefore feasible. On the other hand, the tests are particularly non-invasive and consequently are beneficial to the animal welfare and consti-tute only a low risk for confounding the scientific outcome. Similarly, the use fecal corticosterone metabolite circumvents the stressful collection of blood samples and therefore the preferred method to assess the hormonal response.

The body weight is even a commonly used parameter in score sheets to determine the humane endpoint. Falling below a specific body weight (e.g. 20% of initially detected body weight) is used as an indicator for severe strains and stresses. Hence, we included this measure in our analysis to compare the depression model with somatic models.

Experiment 1: Fecal corticosterone metabolites (FCMs)

48 mice were assigned to the respective groups: home cage n=10, handling n=10, non-trained n=12 and trained n=16. Data of handling and non-trained were merged for analysis until the shuttle box test, since both groups were treated identically until then.

We sampled feces to determine FCM concentrations before the onset (*pre*), during each training session (*training days 1* & *2*), during the *Shuttle Box Test* (day 3) and one week after the LH (*post LH*). For each day two samples were collected, representing i) acute

corticosterone levels during the foot shock exposition or sham treatment and ii) a delayed idle period, to detect persistent effects. The gut passage delay time has been shown to be 4-6 hours during the dark phase and 8-12 hours in the light phase.[3] Hence, we collected acute samples 3-7 hours and delayed 15-19 hours after the respective treatment. For feces collection, mice were placed into individual secondary home cages in the colony room for the respective periods.

Fecal samples were processed as described in Mallien et al. [4]. Briefly, 50 mg of a dried and homogenized fecal sample per mouse were mixed with 1 ml 80% methanol, shaken for 30 min and subsequently centrifuged at 2500 g for 10 min. An aliquot of the supernatant was used to analyze FCM using a well-established and validated 5α-pregnane-3β,11β,21-triol-20-one enzyme immunoassay (EIA).[3, 5]

Experiment 2: typical indicators of wellbeing

28 mice were assigned to the groups: home cage n=6, handling n=6, non-trained n=8 and trained n=8. The nest test was performed before and 3 weeks after the LH procedure and the TINT 1 and 20 days afterwards. Burrowing was analyzed 2, 8 and 21 days after the LH test.

The body weight is even a commonly used parameter in score sheets to determine the humane endpoint. Falling below a specific body weight (e.g. 20% of initially detected body weight) is used as an indicator for severe strains and stresses. Hence, we included this measure in out analysis to compare the depression model with somatic models.

Nesting and TINT

A cotton nestlet (Zoonlab, Castrop-Rauxel, Germany), was introduced into the home cages one hour before the dark phase. Nest building was evaluated after 13 hours and 24 hours according to a rating scale based on cohesion and shape. The scale was as follows:

The nest score was modified from Deacon [6].

Table 2 Nest score

Score	Description
0	Nestlet untouched.
1	Nestlet nearly untouched (>90% intact).
2	Nestlet partly torn (50-90% intact).
3	Nestlet mostly torn (50-90%). ○ <50% of the nestlet is still intact. ○ <90% of the nestlet within one quarter of the cage. ○ The cotton is not build into a nest, but spread in the cage.
4	Identifiable, flat nest (>90% torn) ○ Material in nest shape within one quarter of the cage. ○ Less than 50% of the nests' margin is higher than the body height of the coiled up mouse.
5	Nearly perfect nest (>90% torn) ○ Bowl-shaped nest: more than 50% of the nests' margin is higher than the body height of the *convolute* mouse.
6	Perfect nest.

Additionally, we frequently assessed the nest quality daily at 10 a.m., to track this parameter throughout the experiment. Hence, this cohort received cotton nestlets instead of cellulose as nesting material. The time to integrate novel material into the nest (TINT) was assessed in the home cage within the first three hours of the light phase. For this, sizzle material (Zoonlab, Castrop-Rauxel, Germany) was introduced in the diagonally opposing corner of the nest site and latency to integrate was measured for max. 10 min.[7]

Burrowing

We placed food pellet-filled bottles (14 cm long x 5.5 cm Ø) in the rear end of the home cage one hour before the dark phase and observed the amount that was burrowed out of the bottle (% of total weight) after 6h and 24h. All mice were accustomed to the procedure one week before the LH procedure on four consecutive days.[7]

Experiment 3: Voluntary wheel running

48 mice were assigned to the groups: home cage n=9, handling n=11, trained n=10 and DSS-colitis n=10 (due to technical problems only 40 could be included in the evaluation).

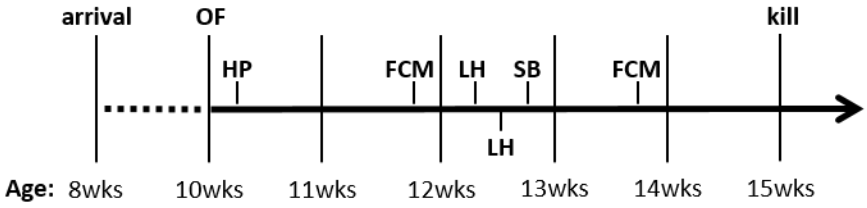
The colitis group received 1% dextran sulfate sodium (DSS) (mol wt 36,000-50,000; MP Biomedicals, Eschwege, Germany) for 5 consecutive days (day 1 to 5) and remained in the colony room. In experiment 3 only, each home cage (Type III, 42.5 x 26.6 x 18.5 cm) was enriched by a running wheel with a sensor to detect the running distance and duration via a bicycle computer (NX-8441-675, Pearl, Buggingen, Germany). Both measures were recorded daily at the beginning of the dark phase throughout the experiment. To determine the steady state running performance, an adaptation phase of 16 days was chosen before the experimental onset (see supplementary material – Figure 2). LH procedure was performed on day 2 to 4. Body weight was consistently assessed at 12 pm. Due to malfunctions of running wheel systems and the consequential imprecision we decided to exclude unreliable results. Two cages of each group were affected.

Statistical analyses

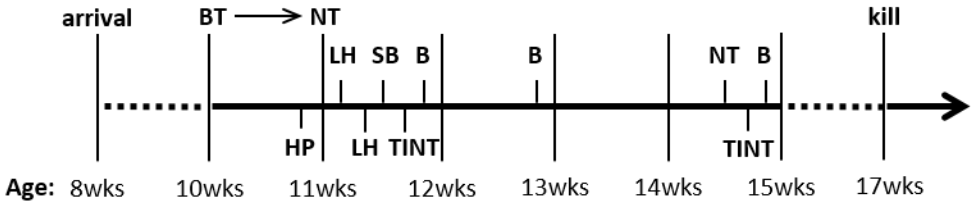
Statistical analyses were carried out using IBM SPSS Statistics 24 . The experimental unit was the single animal. Differences were considered to be significant at $p \leq 0.05$. Data from time to integrate into nest, burrowing, body weight and VWR were analyzed with repeated measures

ANOVAs. Single time points as well as open field and hot plate test were analyzed by univariate ANOVAs. If necessary, pairwise comparisons were corrected using the Holm-Bonferroni-value adjustment. FCMs and Nest Test analyses were performed using non-parametrical analysis techniques. Effects of dependent samples were calculated either by the Friedman test (if more than two sample dates were considered) or by the Wilcoxon signed-rank test for direct comparisons of two consecutive sample dates. Differences between groups were calculated using the Kruskal-Wallis H test. Pairwise comparisons were Holm-Bonferroni corrected. We used a new approach to assess individual severity levels as previously published in the collaborative work 'Running in the wheel: Defining individual severity levels in mice' (Häger et al. 2018[8]). Using the combined VWR and body weight change data of a DSS colitis model (Biernot 2017[9]) an unsupervised k-means algorithm was applied to estimate unbiased severity classes in the data. In Häger et al. 2018[8], data were further obtained from single-housed C57BL/6J.129P2-Il10tm1Cgn+/+ (B6 IL10+/+) mice. The optimum number of three clusters was determined by Scree-analysis (the trade-off between the within-groups sum of squares of the algorithm and the number of clusters) and led to the marking of classes (i.e. in VWR direction) which were hypothesized as potential severity levels. The statistical rationale behind this infers that new data points from different experiments can be allocated to these clusters (levels), thereby indicating putative severity. This also allows severity comparisons between individual animals or experimental groups – given that the same variables were measured. Level borders were determined at VWR = 87.37% and 50.16% (Fig. 1) with 95% confidence. Also, the resulting levels were proposed to classify the severity of each individual animal either as no (level 0), mild (level 1) or moderate (2).

Experiment 1: FCM



Experiment 2: typical indicators of wellbeing



Experiment 3: VWR

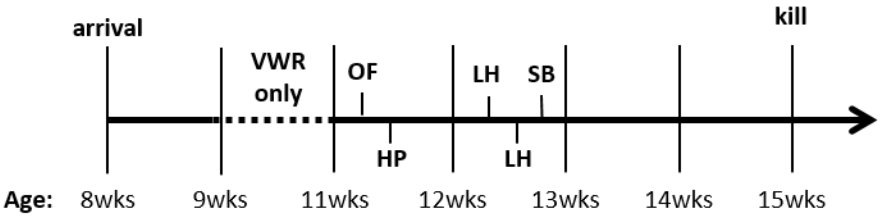


Figure 1 Time-line diagrams for each experiment. OF: Open Field Test, HP: Hot Plate Test, FCM: Fecal corticosterone metabolite sampling, LH: Learned helplessness (day 1 &2); SB: Shuttle Box Test, BT: Burrowing Training, TINT: Time to integrate into nest test, B: Burrowing assessment, VWR: voluntary wheel running.

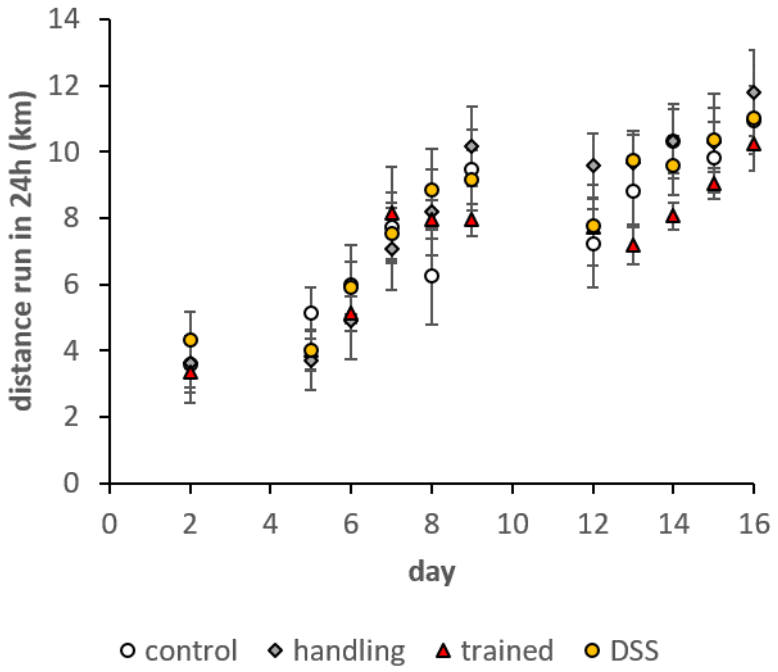


Figure 2 Steady state running performance during the adaptation phase (16 days) prior to experimental onset. Mice received running wheel equipped home cages upon arrival.

References

1. Zueger M, Urani A, Chourbaji S, et al. Olfactory bulbectomy in mice induces alterations in exploratory behavior. *Neurosci Lett* 2005; 374: 142-146. 2005/01/13. DOI: S0304-3940(04)01316-3 [pii]

10.1016/j.neulet.2004.10.040.

2. Chourbaji S, Zacher C, Sanchis-Segura C, et al. Social and structural housing conditions influence the development of a depressive-like phenotype in the learned helplessness paradigm in male mice. *Behav Brain Res* 2005; 164: 100-106. DOI: 10.1016/j.bbr.2005.06.003.

3. Touma C, Sachser N, Mostl E, et al. Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice. *Gen Comp Endocrinol* 2003; 130: 267-278. 2003/02/28. DOI: S0016648002006202 [pii].

4. Mallien AS, Palme R, Richetto J, et al. Daily exposure to a touchscreen-paradigm and associated food restriction evokes an increase in adrenocortical and neural activity in mice. *Horm Behav* 2016; 81: 97-105. DOI: 10.1016/j.yhbeh.2016.03.009.

5. Touma C, Palme R and Sachser N. Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones. *Horm Behav* 2004; 45: 10-22. 2004/01/22. DOI: S0018506X03001727 [pii].

6. Deacon RM. Assessing nest building in mice. *Nature protocols* 2006; 1: 1117-1119. 2007/04/05. DOI: nprot.2006.170 [pii]

10.1038/nprot.2006.170.

7. Jirkof P. Burrowing and nest building behavior as indicators of well-being in mice. *J Neurosci Methods* 2014; 234: 139-146. DOI: 10.1016/j.jneumeth.2014.02.001.

8. Häger C, Keubler LM, Talbot SR, et al. Running in the wheel: Defining individual severity levels in mice. *PLoS Biol* 2018; 16: e2006159. 2018/10/20. DOI: 10.1371/journal.pbio.2006159.

9. Biernot S. *Beurteilung von Belastungsschweregraden mittels Phänotypisierung und klinischer Überwachung bei Versuchstieren*. Tierärztliche Hochschule, https://nbn-resolving.org/urn:nbn:de:gbv:95-110357, 2017.