Check for updates

RESEARCH ARTICLE

REVISED Lockbox enrichment facilitates manipulative and cognitive activities for mice

[version 2; peer review: 2 approved]

Katharina Hohlbaum[1,2], Niek Andresen[2-4], Paul Mieske[1-3], Pia Kahnau[1], Benjamin Lang[2,3], Kai Diederich[1], Rupert Palme[5], Lars Mundhenk[6], Henning Sprekeler[2,7], Olaf Hellwich[2,4], Christa Thöne-Reineke[2,3*], Lars Lewejohann[1-3*]

[1]German Centre for the Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment (BfR), Berlin, 12277, Germany
[2]Science of Intelligence, Research Cluster of Excellence, Berlin, 10587, Germany
[3]Institute of Animal Welfare, Animal Behavior and Laboratory Animal Science, School of Veterinary Medicine, Freie Universitat Berlin, Berlin, 14163, Germany
[4]Computer Vision and Remote Sensing, Technische Universitat Berlin, Berlin, 10587, Germany
[5]Experimental Endocrinology, Department of Biological Sciences and Pathobiology, University of Veterinary Medicine, Vienna, 1210, Austria
[6]Institute of Veterinary Pathology, School of Veterinary Medicine, Freie Universitat Berlin, Berlin, 14163, Germany
[7]Modeling of Cognitive Processes, Technische Universitat Berlin, Berlin, 10587, Germany

* Equal contributors

v2 **First published:** 04 Jun 2024, **4**:108
https://doi.org/10.12688/openreseurope.17624.1
Latest published: 28 Aug 2024, **4**:108
https://doi.org/10.12688/openreseurope.17624.2

Abstract

Background

Due to the lack of complexity and variety of stimuli, conventional housing conditions of laboratory mice do not allow these animals to fully express their behavioral repertoire, including manipulative and cognitive activities. Therefore, we designed mechanical puzzles, so-called lockboxes, for mice that can be provided in their home cages. We investigated the impact of the lockbox enrichment on their phenotype and affective state when compared to conventional housing (CH) and super-environmental enrichment (SEE).

Methods

Young adult female C57BL/6JCrl mice were examined before and after 2-month exposure to the different types of enrichment in a phenotyping test battery, including tests for trait and state anxiety-

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 28 Aug 2024		✓ view
version 1 04 Jun 2024	✓ view	? view

1. **Rasneer Sonia Bains**, Mary Lyon Centre at MRC Harwell, Oxfordshire OX11 0RD, UK

2. **Jesse D Cushman**, National Institutes of Health, Durham, USA

Any reports and responses or comments on the article can be found at the end of the article.

related behavior, calorimetric measurements, body weight measurements, the analysis of stress hormone metabolite concentrations, and sequential problem-solving abilities with a novel lockbox. At the end of the study, adrenal gland weights were determined and pathohistological evaluation was performed. For all continuous variables, the relative variability was calculated.

Results

While the different types of enrichment affected trait anxiety-related behavior, neither state anxiety-related behavior nor physiological variables (i.e., bodyweight, resting metabolic rate, stress hormone metabolite concentrations, adrenal gland weights) were influenced. LE improved sequential problem-solving (i.e., solving novel lockboxes) when compared to SEE. Regardless of the housing condition, the relative variability increased in most variables over time, although the coefficient of variation decreased for some variables, especially in animals with access to LE. There was no evidence of toxicopathological effects associated with the material from which the lockboxes were made.

Conclusions

All lockboxes are available as open-source tool. LE revealed beneficial effects on the affective state of laboratory mice and their performance in solving novel lockboxes. Neither relevant phenotype of the mice nor reproducibility of the data were compromised by LE, similar to SEE. The lockboxes may also be used as novel approach for assessing cognition in mice.

Keywords
Mice, refinement, animal welfare, enrichment, home cage, cognition, sequential problem-solving, phenotyping

This article is included in the COST Actions gateway.

Corresponding author: Katharina Hohlbaum (katharina.hohlbaum@bfr.bund.de)

Author roles: Hohlbaum K: Conceptualization, Data Curation, Formal Analysis, Investigation, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Andresen N**: Conceptualization, Software, Visualization, Writing – Review & Editing; **Mieske P**: Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Kahnau P**: Formal Analysis, Investigation, Writing – Review & Editing; **Lang B**: Investigation, Writing – Review & Editing; **Diederich K**: Conceptualization, Investigation, Writing – Review & Editing; **Palme R**: Data Curation, Investigation, Resources, Writing – Review & Editing; **Mundhenk L**: Conceptualization, Data Curation, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Sprekeler H** : Conceptualization, Funding Acquisition, Resources, Writing – Review & Editing; **Hellwich O**: Conceptualization, Funding Acquisition, Resources, Writing – Review & Editing; **Thöne-Reineke C**: Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing; **Lewejohann L**: Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Framework Programme for Research & Innovation as part of the COST Action CA20135 - Improving biomedical research by automated behaviour monitoring in the animal home-cage (TEATIME), as supported by the COST Association (European Cooperation in Science and Technology). This work was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project number 390523135.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Hohlbaum K *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hohlbaum K, Andresen N, Mieske P *et al.* **Lockbox enrichment facilitates manipulative and cognitive activities for mice [version 2; peer review: 2 approved]** Open Research Europe 2024, **4**:108
https://doi.org/10.12688/openreseurope.17624.2

First published: 04 Jun 2024, **4**:108 https://doi.org/10.12688/openreseurope.17624.1

REVISED **Amendments from Version 1**

In response to the reviewers' comments, the main conclusions were stated more clearly in the manuscript. Additionally, as suggested by a reviewer, we emphasized the fact that we provide an open-source tool that can be used both as enrichment and as a potential novel approach in behavioral neuroscience.

The section on statistical analysis was restructured and shortened, with the most relevant information moved to Table 1, which also includes the number of animals. Figure 6 and corresponding result descriptions were moved to the supplementary material (https://doi.org/10.5281/zenodo.13269372). The order of the other tables and figures was adjusted accordingly. For figures with multiple plots, we improved clarity by including only one legend. For refining the code used to create these figures, we acknowledged the use of ChatGPT. Definitions of the lines and asterisks indicating significance have been added to the legends of Figure 5 and Figure 8. In Figure 5b, the line indicating a significant difference within the LE group before and after the enrichment phase was corrected.

The discussion on implementing lockbox enrichment in laboratory animal facilities was expanded. We emphasized the limitations, particularly those arising from repeated testing in the behavioral paradigms, in the discussion section. Although the limitation due to differences between the study groups in the latency to explore in the Grid Exploratory Paradigm before the enrichment phase was already highlighted in Version 1 of the manuscript, we have underscored it even further. The conclusion that lockbox enrichment reduces trait anxiety-related behavior is based on the within-group comparison in the Grid Exploratory Paradigm and the results from the Free Exploratory Paradigm.

Finally, typos were corrected, sentences were rephrased, and wording (e.g., "sequential problem-solving with lockboxes") was revised to improve readability and avoid confusion or misunderstanding. Moreover, the abstract was shortened to the maximum of 300 words.

Any further responses from the reviewers can be found at the end of the article

Introduction

In the wild, mice live in complex environments that offer a variety of stimuli. In comparison to natural habitats, the laboratory environment is restricted in space, complexity, and stimuli, preventing the mice from fully expressing their behavioral repertoire[1]. This can diminish the health and well-being of animals, thereby potentially affecting the quality of research data[2]. Therefore, the European legislation stipulates that "all animals shall be provided with space of sufficient complexity to allow expression of a wide range of normal behaviour. […] Establishments shall have appropriate enrichment techniques in place, to extend the range of activities available to the animals and increase their coping activities including physical exercise, foraging, manipulative and cognitive activities, as appropriate to the species"[3].

Within the past decades, a range of different physical/inanimate environmental enrichment items have been introduced to enhance complexity in the laboratory environment[1,4–6]. Additionally, it was proposed to house mice within a semi-natural environment[6,7] and in larger social groups[7]. In contrast, enrichment strategies for laboratory mice that specifically allow manipulative and cognitive activities have been rarely reported. The following examples can be found when searching for "cognitive enrichment" (search term in PubMed: "cognitive enrichment"[All Fields] AND ("mice"[MeSH Terms] OR "mice"[All Fields])): toys of different shapes, sizes, colors, and textures that do not increase physical activity (no further information given in the references)[8], multi-colored blocks, balls, and tunnels[9], positive reinforcement training (i.e., clicker training)[10], touchscreen training[11–13], other operant learning cognitive tasks in operant chambers[14], and spatial memory tasks[15]. Depending on the type of cognitive enrichment, beneficial effects on the animals' well-being[10], postoperative cognitive dysfunction[8], motor function and neuropathology in a Huntington's Disease mouse model[14], and cognitive performance in an Alzheimer's Disease mouse model[15] have already been found. These pieces of evidence suggest that enrichment facilitating manipulative and/or cognitive activities can foster the welfare of laboratory mice. However, some of the above-mentioned examples primarily facilitate manipulative or cognitive activities, while others involve aversive experiences, such as transferring the mice from their home environment to a testing arena. The latter could be seen as contradiction to the legal requirements for enrichment since laboratory animals "shall be given a degree of control and choice over their environment to reduce stress-induced behaviour"[3].

Therefore, we aimed to develop enrichment strategies for both manipulative and cognitive activities that can be presented to the mice in their home cages and used on a voluntary basis. For this, we were inspired by a modified concept of Thorndike's puzzle box for cats[16]. In Thorndike's original experiments, cats were confined in a box and a palatable food reward was placed in front of the box. The cat had to engage with various mechanisms to open the box and reach the food[16]. The original setup of Thorndike's puzzle box had the potential to cause distress in the animals by being locked in the box. Modifications of this concept, where the food reward and not the animal is hidden in the box, and the animal must solve a series of mechanisms to access the reward, appear to better align with the criteria for an effective enrichment strategy. This ensures that the animal is not affected by being confined while still engaging in problem-solving tasks to obtain the reward. Such so-called lockboxes are mechanical puzzles comprising a range of mechanisms (e.g., screws, levers, bolts, doors) which obstruct one another and must be opened sequentially in the correct order to reveal a food reward[17,18]. Lockboxes can be constructed with varying degrees of complexity, comprising of one step to multiple steps. They were used to investigate sequential problem-solving in various species, e.g., cockatoos[19], pigeons[20], keas[21–23], corvids[24–28], and great apes[29–31]. Simple 1-step and 2-step tasks requiring to move/remove objects (e.g., a lid, a petri dish, paper, a window) were previously developed for wild house mice *(Mus musculus)* and striped field mice (*Apodemus agrarius*)[32,33]. However, more sophisticated, and complex lockboxes suitable for mice have not been developed yet.

Against this background, our objective was to design lockboxes for mice and explore various aspects in three substudies. In the substudy presented in this article, we examined the impact of the lockbox enrichment (LE) on the animals' phenotype and affective state, in comparison to mice kept under conventional housing (CH) or super-environmentally enriched housing conditions (SEE). Additionally, we assessed whether the LE affected data variability and reproducibility. In the remaining sub-studies, not covered in this article, we delved into the problem-solving strategies employed by mice when tackling the lockboxes[34] and their facial expressions. Moreover, we aimed to analyze the impact of the different types of enrichment on adult neurogenesis. The findings of the present study will contribute to a better understanding of whether LE can serve as an effective refinement strategy and positively influence the animals' affective state. Additionally, these results will give researchers insights into whether LE may affect behavioral or physiological variables of interest in their studies, which need to be considered before introducing a novel type of enrichment.

Methods

Ethics statement

Research involving animals was performed according to the guidelines of the German Animal Welfare Act and the Directive 2010/63/EU for the protection of animals used for scientific purposes. Approval for animal experimentation and maintenance of the animals was granted by the Berlin State Authority (referred to as "Landesamt für Gesundheit und Soziales", permit number: G0249/19). The pre-registration of the study can be found in the animal study registry (animalstudyregistry.org, doi: 10.17590/asr.0000237).

All efforts were made to ameliorate suffering of the mice: the mice were provided a large floor space, i.e., two Makrolon type III cages for 4 mice. A simulated sunrise smoothly increased the light intensity in the room before the lights turned on. Tunnels were used to handle the mice, which reduces stress and anxiety. The lock boxes were introduced as enrichment items.

Animals, enrichment types, and handling method

A total of 36 female C57BL/6J mice, that were free of pathogens listed in the FELASA recommendations[35], were sourced from Charles River Laboratories (Sulzfeld, Germany). The animals were divided into three batches (i.e., 12 animals per batch) that were delivered at different time points. Each batch included four animals of each study group (three groups: LE, CH and SEE). The individual animals were considered as experimental units. The study planning included both sexes. Due to the time-consuming nature of the study, thus far, we have only been able to examine female animals, which must be considered as limitation of the present article. The estrus cycle phases were not assessed.

Upon reaching approximately four weeks of age, the mice arrived at the animal facility and housed in filter top cages maintained under specifically pathogen free barrier conditions. The environmental conditions were set at a room temperature of 22 ± 1 °C, a relative humidity of 55 ± 10 %, and a 12/12-hour light/dark cycle. The light cycle began at 7 AM (summertime: 8 AM) and a wake-up light (Philips HF 3510, 100–240 vac, 50–60 Hz, Philips GmbH Market DACH, Hamburg, Germany) simulated the sunrise by gradually increasing light intensity starting at 6:30 AM (summertime: 7:30 AM). The mice were accommodated in groups of four (i.e., animals of the same study group were housed together) in two Makrolon type III cages (dimensions: 39 × 23 × 15 cm; designated as cage A and cage B), connected by a gate (Figure 1). The gate comprised three transparent tubes (11.5 × 4 cm, custom-built from GEHR PMMA XT® ACRYL, Mannheim, Germany) with barriers (as described by Habedank *et al.*[36,37]), linked by two doors systems (custom-built, 3D printed; STL files can be found on Zenodo[38]), that could be opened and locked. In both cages, ad libitum access to food pellets (LASvendi, LAS QCDiet, Rod 16, autoclavable) and tap water was provided. Fine wooden bedding material (JRS Lignocel FS14, spruce/ fir, 2,5-4 mm) was used for both cages.

The animals were assigned to three study groups by simple randomization: CH (n = 12), SEE (n = 12), and LE (n = 12). Figure 1 shows the home cage apparatuses of the three study groups. At the beginning of the study, cages of all study groups contained the same resources. Cage A was equipped with a red triangular plastic house with two exits (1 long side: 14.5, 2 short sides: 11.5, height: 5 cm, TheMouseHouse, Tecniplast), a transparent handling tunnel (length: 11.5, outside diameter: 4 cm, custom-built from GEHR PMMA XT® ACRYL,Mannheim, Germany), and three thin (cellulose, unbleached, layers, 20x20cm, Lohmann & Rauscher, Neuwied, Germany) as well as two thick (23 × 24,8cm folded, Essity ZZ Towel) paper tissues for nesting. In addition to this, both the SEE and LE groups were provided with further enrichment items after approximately five weeks (Figure 2). In the SEE group, five cotton rolls (Gr.3, UNIGLOVES, Troisdorf, Germany) and 3–4 g white paper strips (LILLICO, Biotechnology Paper Wool) for nesting, a second transparent tunnel attached to the cage lid with a tunnel clip (polycarbonate, suitable for 38 mm (Dia) tunnels, Datesand, Bredbury, United Kingdom), a red mouse igloo with a yellow running plate (round house: 105 mm in diameter, 55 mm in height; round plate: 150 mm in diameter, ZOONLAB GmbH, Castrop-Rauxel, Germany), a second plane with two holes (dimensions: 23 × 12.5 × 1 cm; diameter of holes: 3 cm; 2 hole lying boards for cage type III, ABEDD, Kalnciems, Latvia), as well as four wooden gnawing blocks were placed into cage A. The LE group had temporary access to different types of lockboxes in cage B, which later became permanent, as described in detail in the section on lockbox training. In this group, the lid grid of cage B was replaced with transparent acrylic glass in an area of 22 × 15 cm. For the other two groups, cage B remained empty and the cage lid was not modified.

The animals in all study groups had unrestricted access to both cages, except for the LE group, whose access to cage B was unrestricted except for the short training periods conducted as part of the temporary LE, as described below in the section on lockbox training.

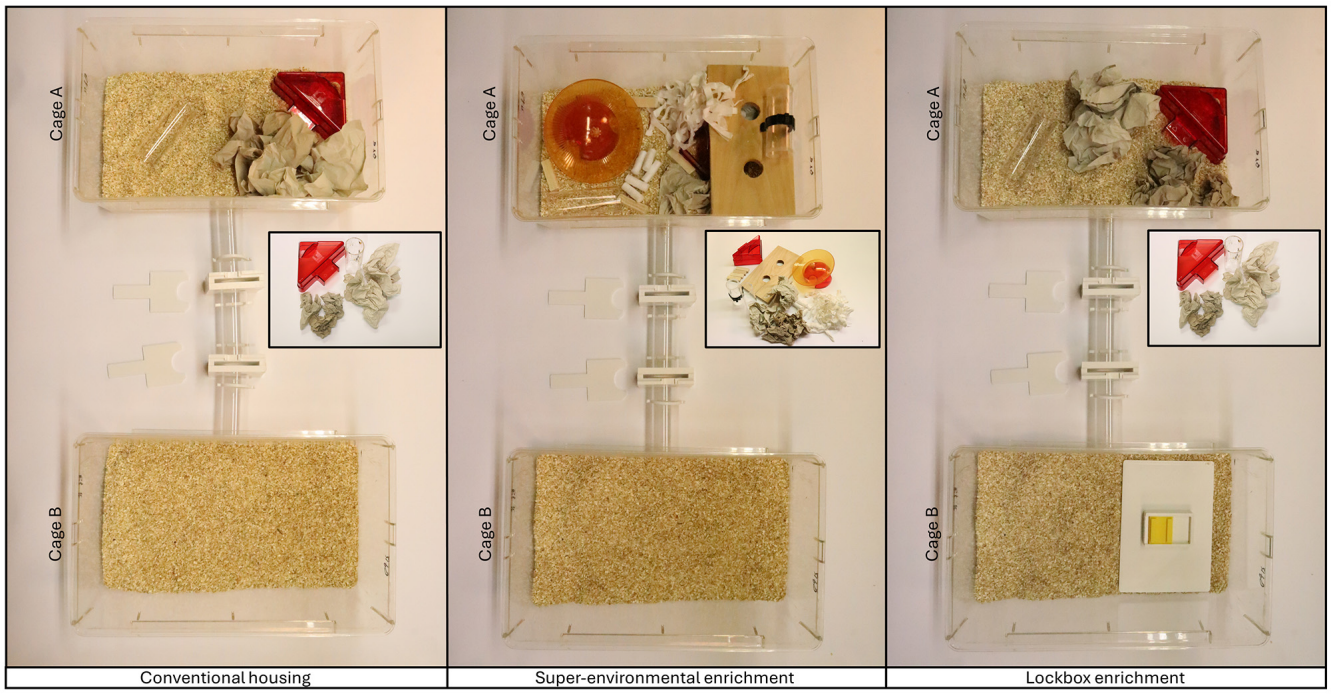


Figure 1. Home cage apparatuses of the study groups. The cage grid with food pellets and water bottle, as well as the filter tops were removed for the photographs. In the super-environmentally enriched cages, the tunnel attached to the cage grid with a clip was positioned on top of the wood board for photography purposes. The inserted images of the enrichment items illustrate the contents of cage A. The cages were connected with a gate comprising three transparent tubes with barriers and two door systems. The doors were placed next to the gate for the photographs and can be used to open or lock the gate. The only difference between the conventional housing and the lockbox enrichment group was the (temporary and permanent) access to the lockboxes in cage B.

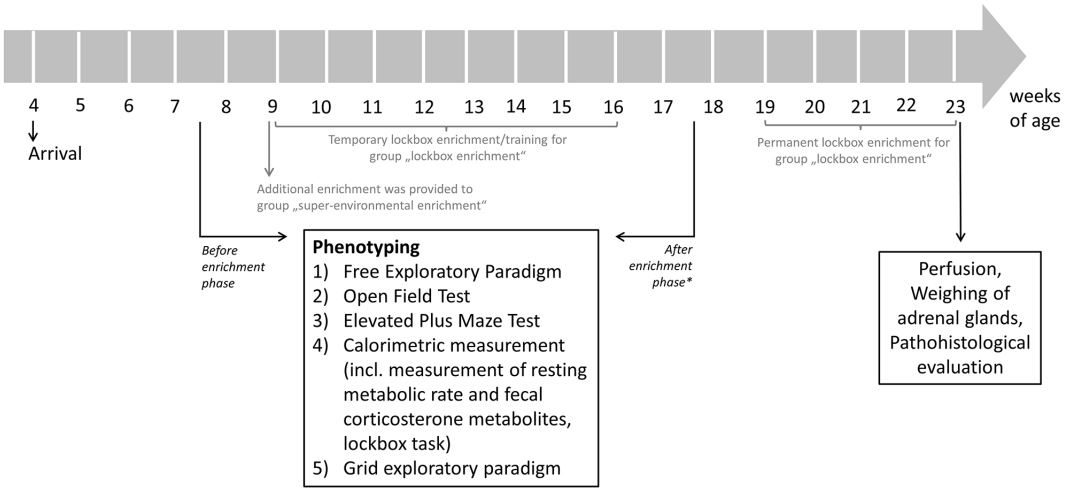


Figure 2. Flow chart of testing schedule. The phenotyping test battery was conducted before and after the 2-month enrichment phase. * While the name of the phase may suggest otherwise, it's important to note that the enrichment was not removed from either the lockbox enrichment group or the super-environmental enrichment group after the mice were tested for the second time.

All animals were handled using a tunnel, i.e., the mice voluntarily entered the tunnel. In the first two weeks, they were habituated to the voluntary tunnel handling by the animal technicians and experimenters. For this, the protocol for voluntary tunnel handling, accessible at the Refinement Wiki of Norecopa, was shortened from 3 to 2 weeks[39].

Animal identification

After acclimating to the animal facility for a week, radio frequency identification (RFID) transponders were subcutaneously implanted into the dorsocervical region using an application cannula (ISO-Compliant Transponder, Peddymark Limited, Redhill, Surrey, United Kingdom; cannula: 2.6 × 0.15 × 40 mm) during a brief isoflurane anesthesia (induction: 4 % isoflurane, maintenance: 1–2 % isoflurane; carrier gas: 100 % oxygen, flow: 1 liter/min; Isofluran CP 1ml/ml, CP-Pharma Handelsgesellschaft mbH, Burgdorf, Germany). In the evening before transponder implantation, 1 mg/kg meloxicam (Loxicom 0.5 mg/ml orale Suspension, Elanco Animal Health, Bad Homburg, Germany or Metacam 0.5 mg/ml Suspension zum Eingeben, Boehringer Ingelheim, Ingelheim am Rhein, Germany) was orally given. In addition to RFID transponders, color marking of the tail (edding 750 paint marker, edding International GmbH, Ahrensburg, Germany) was employed for visual identification of the individual mice.

Testing schedule

Figure 2 demonstrates the testing schedule of the present study. Briefly, the mice arrived at the age of approximately four weeks at the animal facility and habituated to the new environment before the experiments started. Before the additional enrichment was provided the SEE and LE groups, a phenotyping test battery was conducted at the age of 7–8 weeks to measure baseline values of each animal. Until this point in time animals of all groups were housed under the same conditions (i.e., these conditions were equal to the CH). At the age of approximately 9 weeks, the additional environmental enrichment was provided to the SEE group and LE group was allowed temporary access to the lockboxes, which is further described in the section on lockbox training. After the 2-month enrichment phase, the phenotyping test battery was conducted for a second time to evaluate effects of the different types of enrichment. Subsequently, the animals of the LE group had permanent access to lockboxes (i.e., 24 hours per day). A lockbox baited with four oat flakes (kernel oat flakes, Wurzener Nahrungsmittel GmbH, Wurzen, Germany) was placed into cage B and was replaced once a day with another lockbox. The animals of the other study groups were kept under the same housing conditions as during the 2-month enrichment phase, i.e., the enrichment was not removed. Additionally, the same amount of oat flakes, as was filled in the lockboxes, was provided daily to cage B of the other study groups. An acoustic signal (i.e., careful, brief rattling on the gate system) was emitted to signal the animals that the oat flakes had been added to the cage. At the age of approximately 23 weeks, transcardiac perfusion was carried out, adrenal glands were weighted, and animal bodies were prepared for further pathohistological evaluation.

For another substudy, which is not further described in the present article, the brains were removed to examine the impact of the different types of enrichment on adult neurogenesis. As pretreatment for this examination, the animals were intraperitoneally injected with the proliferation marker chlorodeoxyuridine (CldU; 50 mg/kg; 10 µl/g bodyweight) before and the marker iododeoxyuridine (IdU; 50 mg/kg; 10 µl/g bodyweight) after the enrichment phase on three consecutive days. At both time points, the injections were carried out after the phenotyping test battery.

Originally, it was planned to measure home cage activity of the individual mice using the Mouse Position Surveillance System (MoPSS), an RFID based tracking system. However, due to technical malfunctions of the MoPSS devices in two batches, data could not be analyzed. These technical difficulties were solved in the updated version of the MoPSS published by Habedank *et al.*[36].

Lockboxes

3D models of the lockboxes were designed using the computer-aided design software Tinkercad (www.tinkercad.com). Models were transferred to 3D printable files using the slicing application Cura SteamEngine 4.4.0. The lockboxes were 3D printed using an Ultimaker 3 Extended and an Ultimaker S3 (0.4 mm nozzles). Polylactic acid (PLA) of different colors was used as 3D printing material (Filamentworld, Neu-Ulm, Germany: PLA filament white, yellow, red, grey; Formfutura, Nijmegen, The Netherlands: StoneFil PLA Filament Granite). The lockboxes consisted of a white platform and different mechanisms. Details on 3D printing and the STL files of the lockboxes can be found on Zenodo[38]. Briefly, two sets of lockboxes were designed, with each set comprising four single-mechanism lockboxes (1 step) and one combined-mechanism lockbox (4 steps) (Figure 3a). In the combined-mechanism lockbox, the four single mechanisms obstruct each other, requiring removal in the correct sequence to unlock the box.

Both lockbox sets comprised of similar mechanisms of slightly varying forms, i.e., a lever, a stick/drawer combination, a cube or ball, and a rotating disc or sliding door. To open the single-mechanism lockboxes, the mice had to move the lever to the other side, pull out the stick/drawer combination, remove the cube or ball, and finally move the disc or door. For solving the combined-mechanisms lockboxes, the mice had to open the mechanisms in the following order: lever, stick, cube, disc (disc-lockbox) or lever, stick, ball, door (door-lockbox). Short video sequences displaying the mice solving the lockboxes can be found on Zenodo[38].

To encourage the animals to engage with the lockboxes, they were baited with oat flakes as food reward (i.e., single-mechanism lockboxes were filled with half an oat flake, combined-mechanisms lockboxes with one oat flake). If an animal was not able to open the lockbox, it received the oat flakes after the lockbox training was completed for all mice.

Lockbox training

On three consecutive days before the temporary lockbox training for the LE group began, eight oat flakes were added to cage B of all study groups to habituate them to the food reward. An acoustic signal (i.e., careful, brief rattling on the gate

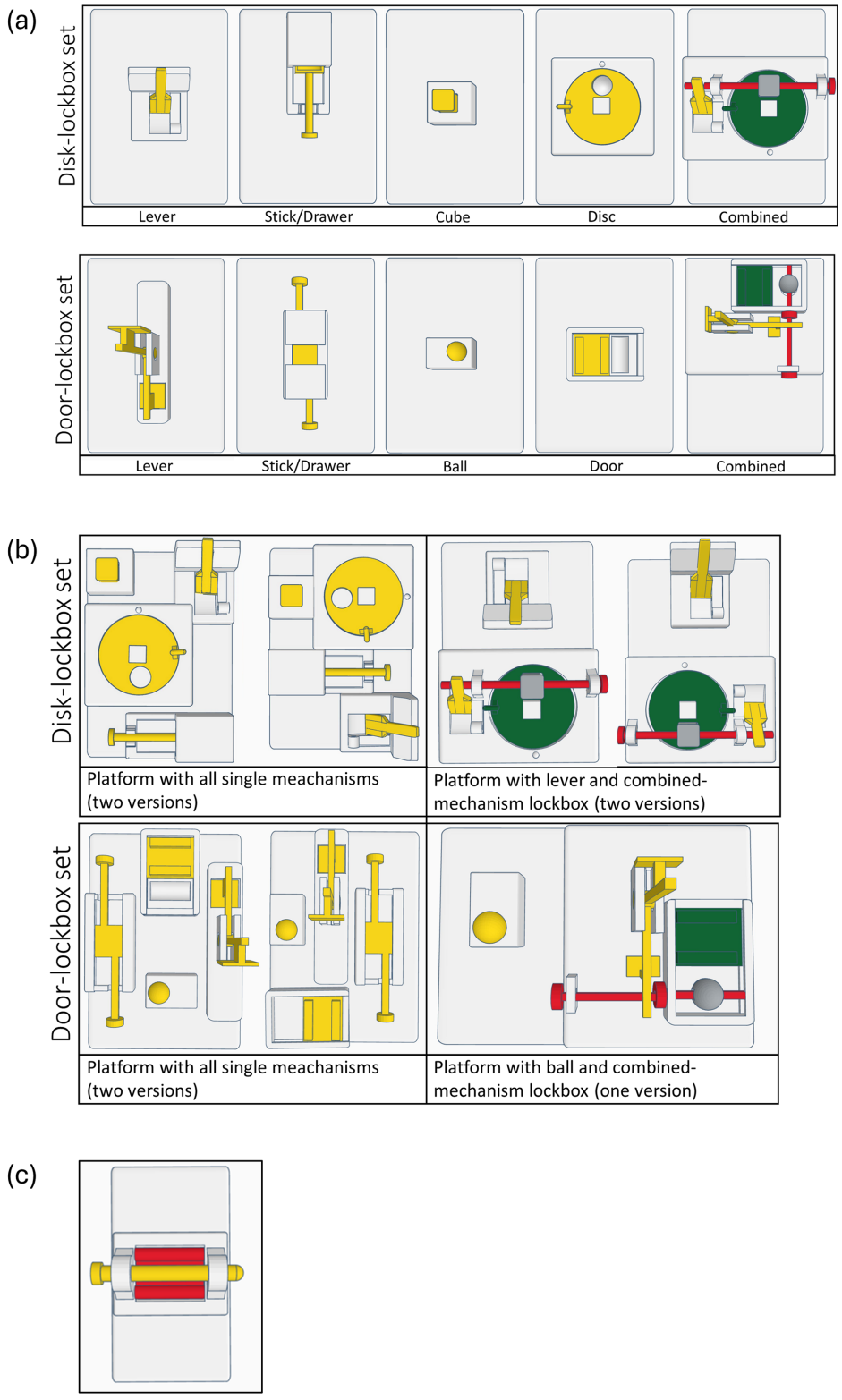


Figure 3. (**a**) Lockbox sets comprising of four single-mechanism lockboxes and a combined-mechanism lockbox. (**b**) Multiple-lockbox platforms. For each lockbox set, two versions of a platform with all single mechanisms were designed, allowing to present all lockboxes simultaneously to the mice. For the door-lockbox set one version and for the disc lockbox set two versions of a platform with a single-mechanism and the combined-mechanism lockbox were created. The different versions allowed the mechanisms to be positioned at a different spot within the cage in each trial. (**c**) 2-step lockbox presented to all mice at the end of the calorimetric measurement.

system) was emitted to signal the animals that the oat flakes had been added to the cage.

The lockbox training started on the following day (Figure 4). The mice were trained in the morning after the lights turned on. Each day, the order of the mice was randomized by drawing cards with the mouse ID on them. To begin the doors of the gate system were closed when all mice were in cage A and cage B was empty. A white screen was placed between cages A and B to prevent the mice from observing each other while solving the lockboxes. After a lockbox was placed into cage B, the first mouse was allowed to enter cage B and interact with the lockbox. For this the door closer to cage A was opened while the door closer to cage B remained locked. All mice usually attempted to access cage B and formed a queue inside the tube. When the mouse to be tested was the first animal in the queue, the door closer to cage A was closed and the door closer to cage B was opened. Once the mice entered cage B, this door was also locked. A trial, defined as the period in which the mouse was allowed to interact with a lockbox, ended when the mouse opened the lockbox or when the maximum trial time expired. The maximum trial time varied in the training phases and was between 15 and 30 minutes (Figure 4). At the end of a trial, the experimenter waited until the mouse stopped interacting with the lockbox, before removing the lockbox from the cage. If the mouse continued interacting with it for more than one minute, the experimenter gently guided it away and removed the lockbox. After a mouse completed all planned trials of a day, it left cage B through the gate system. An acoustic signal (i.e., careful, brief rattling on the gate system) was emitted to signal the mice that the doors were opened, and the next mouse could enter cage B. There were four sets of each lockbox, i.e., a set for each mouse to prevent any odor cues. After the training, the lockboxes were cleaned with 70 % ethanol and left to air dry overnight.

Figure 4 demonstrates the sequence in which the lockboxes were presented to the mice. This sequence was first conducted for the disc-lockbox set starting on day 1, followed by the door-lockbox set beginning on day 29. Briefly, on day 1 (disc-lockbox set) or day 29 (door-lockbox set), the combined lockbox, in a locked state, was presented to the mice (1 trial/day). On day 2–12 (disc-lockbox set) or day 30–40 (door-lockbox set), the animals had access to the single mechanisms of the lockbox set. All four single mechanism lockboxes were presented once a day in a randomized order (1 trial/lockbox/day, i.e., 4 trials/day). Before the first trial of each single-mechanism lockbox on day 2 (disc-lockbox set) or day 30 (door-lockbox set), the animals were allowed to explore the baited lockboxes in an open state for a minute. Then it was rebaited and locked. On days 15 and 16 (disc-lockbox set) or days 43 and 44 (door-lockbox set), all four single-mechanism lockboxes were simultaneously presented to the mice twice (2 trials/day). There were versions of two platforms, each comprising all four single mechanisms, though the mechanisms were arranged differently on each platform (Figure 3b). This allowed the four mechanisms to be located in a different spot within the cage for each trial. After that on days 22–26 (disc-lockbox set) or days 50–54 (door-lockbox set), the mice were trained on the combined lockbox. On day 22 (disc-lockbox set) or day 50 (door-lockbox set), the mice could explore the baited lockbox in an open state for a minute before it was rebaited and locked, and the trial began. While on day 22–25 (disc-lockbox set) or day 50–53 (door-lockbox set) 1 trial/day was conducted, 3 trials/day were performed on day 26 (disc-lockbox set) or day 54 (door-lockbox set). In the first trial, all mechanisms of the lockbox, in the second trial only the disc/door mechanism, and in the third trial only the cube/ball mechanism and disc/door mechanism were locked. On day 27 (disc-lockbox set) or day 55 (door-lockbox set), a platform comprising the combined-mechanism lockbox and a single-mechanism lockbox was presented to the mice twice (2 trials/day). In the second trial, the two lockboxes were rearranged by either rotating the platform by 180 degrees (door-lockbox) or by using the second version of this platform (disc-lockbox).

During the lockbox training, the mice were video-recorded using three Basler Cameras acA1920-40um (Lens LM25HC7, f = 25mm, k = 1.4, Kowa, Nagoya, Japan3) and two Intel

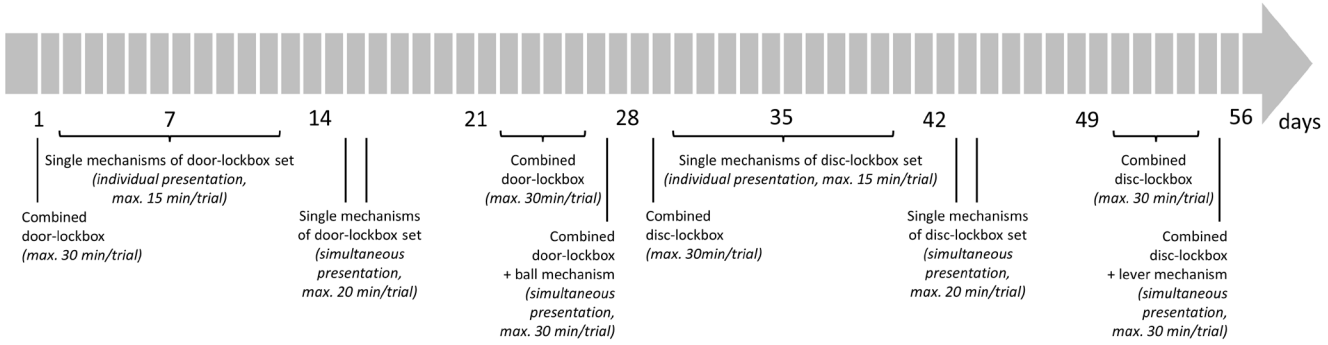


Figure 4. Lockbox training schedule. A trial was defined as the period in which the mouse was allowed the interact with a lockbox. It ended when the mouse opened the lockbox or when the maximum trial time expired.

Real Sense cameras. The video material was used in other substudies to analyze the problem-solving strategies of the mice and their facial expression when solving the lock boxes.

Free Exploratory Paradigm

The Free Exploratory Paradigm was conducted to investigate trait anxiety-related behavior. To enable a repeatability analysis, the Free Exploratory Paradigm included a habituation phase and was conducted over three consecutive days. There were four testing apparatuses for the Free Exploratory Paradigm allowing to simultaneously test the four animals of a cage group, i.e., a test apparatus for each mouse. A testing apparatus consisted of a starting cage (Makrolon type III cage). A handful of used bedding was scattered on top of the clean bedding. Moreover, the cages were equipped with a red triangular plastic house, two thin paper tissues, and a thick paper tissue. Food pellets and tap water were provided in the gridded cage lid. The cage was connected via two tunnels (length: 11.5, outside diameter 4 cm, custom-built from GEHR PMMA XT® ACRYL,Mannheim, Germany) to an arena (50 × 50 cm, height of walls: 50 cm). The two tunnels were linked by a door system (custom-built, 3D printed, STL files can be found on Zenodo[38]) denying access to the arena. The illumination level was set to 90 ± 9 lx in the center and the corners (mean ± standard deviation). After all animals were individually transferred with the tunnel from their group cage to the respective starting cage, the doors remained closed for a minute and were then opened, allowing the mice to voluntarily enter the arena. Every day the Free Exploratory Paradigm was conducted, the mice were allocated to a different testing apparatus from the four available. The 25-minute testing period was recorded from the top with a Basler acA1920-40um camera (Lens LM8HC7, f = 8mm, k = 1.4, Kowa, Nagoya, Japan; 41 monochrome frames per second) using the Basler Video Recording Software (www.baslerweb.com/en/software/video-recording). After 25 minutes had elapsed, the mice were transferred with the tunnel to the group cage. The testing arena was clean with 70 % ethanol at least 12 hours before the test was performed. The experimenter manually counted the number of fecal boli and urinary spots. Additionally, variables of interest were the number of mice entering the arena, the latency to enter the arena, the distance travelled in the arena, and the percentage of time moving in the center or periphery relative to the time moving in both zones.

The number of mice entering the arena and the latency to enter the arena were manually assessed from the videos. A mouse entered the arena when all four paws left the tunnel and were placed on the floor of the arena. The distance travelled in the arena, and the time moving in the center or periphery were automatically analyzed from the videos using DeepLabCut (version 2.3.4, https://github.com/DeepLabCut/DeepLabCut)[40,41] and the DLCAnalyzer[42] (https://github.com/ETHZ-INS/DLCAnalyzer), as described below.

Open Field Test

The Open Field Test was performed to examine state anxiety-related behavior. To enable a repeatability analysis, the Open Field Test included a habituation phase and was conducted over three consecutive days. Every day, the order in which the mice were tested was altered. The animals were individually transferred with the tunnel from their group cage to the testing arena (80 × 80 cm, height of walls: approximately 60 cm, material: polyvinyl chloride). The illumination level was set to 104 ± 3 lux in the center and 59 lux ± 3 in the corners (mean ± standard deviation). The tunnel was attached to the hole (diameter: 4 cm) in the wall of the testing arena and the mouse climbed from the tunnel into the starting zone, which was located in a corner and separated from the rest of the arena by a starting cylinder (15 × 15 cm). Then a plug was used to close the hole in the wall. After a 1-minute period, the starting cylinder was removed, allowing the mouse to freely explore the arena for 10 minutes. The 10-minute testing period was recorded from the top with a Basler acA1920-40um camera (Lens LM8HC7, f = 8mm, k = 1.4, Kowa, Nagoya, Japan; 41 monochrome frames per second) using the Basler Video Recording Software (www.baslerweb.com/en/software/video-recording). After 10 minutes had elapsed, the mouse was transferred with the tunnel to the group cage and the testing arena was cleaned with 70 % ethanol. The experimenter manually counted the number of fecal boli and urinary spots. Additionally, variables of interest were the time spent in the central zone, time spent in the peripheral zone, and distance moved, which were automatically analyzed from the videos using DeepLabCut (version 2.3.4, https://github.com/DeepLabCut/DeepLabCut)[40,41] and the DLCAnalyzer[42] (https://github.com/ETHZ-INS/DLCAnalyzer), as described below.

Elevated Plus Maze Test

The Elevated Plus Maze Test was carried out to measure state anxiety-related behavior. To enable a repeatability analysis, the Elevated Plus Maze Test included a habituation phase and was conducted over three consecutive days. Every day, the order in which the mice were tested was altered. The testing apparatus consisted of two open and two closed arms (dimensions of the arms: 30 × 5 cm, center: 5 × 5 cm, material: polyvinyl chloride). The open arms without side walls were perpendicular to the closed arms with 20-cm-high side walls. The testing apparatus was elevated to a height of 50 cm above the floor. The open arms were surrounded by a raised ledge (6 mm thick, 6 mm high). The illumination level was set to 92 ± 3 lx in the center, 147 ± 3 lux on the open arms, and 31 ± 6 lux on the closed arms (mean ± standard deviation). The animals were individually transferred with the tunnel from their group cage to the center of the testing apparatus, facing a closed arm and away from the experimenter. The test started when the mouse left the tunnel with all four paws. The 5-minute testing period was recorded from the top with a Basler acA1920-40um camera (Lens LM8HC7, f = 8mm, k = 1.4, Kowa, Nagoya, Japan; 41 monochrome frames per second) using the Basler Video Recording Software (www.baslerweb.com/en/software/video-recording). After 5 minutes had elapsed, the mouse was transferred with the tunnel to the cage and the testing arena was cleaned with 70 % ethanol. The experimenter manually assessed defecation the number of fecal boli and urinary spots. Moreover, variables

of interest were the distance travelled, the time spent on the open or closed arms, the number of transitions to the open or closed arms, the ratio of the time spent on the opens arms to the time spent on the closed arms, and the number of protected stretch attend. It must be noted that the center of the Elevated Plus Maze was neither part of the open nor the closed arms, i.e., the time spent in the center of the Elevated Plus Maze Test was not analyzed. DeepLabCut (version 2.3.4, https://github.com/DeepLabCut/DeepLabCut)[40,41] and the DLCAnalyzer[42] (https://github.com/ETHZ-INS/DLCAnalyzer) were used to automatically analyze the distance travelled, the time spent on the open or closed arms, and the number of transitions to the open or closed arms from the videos, as described below. The number of protected stretch attend was manually assessed using BORIS[43].

Videos analysis of Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test

DeepLabCut (version 2.3.4, https://github.com/DeepLabCut/DeepLabCut) was used for body part tracking[40,41]. For the Free Exploratory Paradigm and Open Field Test, the four corners of the testing arena and the body center were labeled. For the Elevated Plus Maze Test, the four corners of the center, the two corners of each arm, and the body center were labeled. The body center was defined as the half-length between the nose and the tail base. For training 30 frames were taken from 10 videos (5 % was used for validation). We used a ResNet50-based neural network[44,45] and trained for 500,000 training iterations (Free Exploratory Paradigm), for 900,000 training iterations (Open Field Test), or 100,000 training iterations (Elevated Plus Maze Test) with otherwise default parameters and validated with one shuffle, which resulted for the Free Exploratory Paradigm in a test error of 2.85 pixels and a train error 3.51 pixels (image size was 1920 by 1080), for the Open Field Test in a test error of 4.49 pixels and a train error 4.28 pixels (image size was 1936 by 1216), and for the Elevated Plus Maze Test in a test error of 2.78 pixels and a train error 3.06 pixels (image size was 1936 by 1216). A p-cutoff of 0.6 was chosen. These networks were then used to analyze videos from all mice tested in the respective tests. The csv files generated in DeepLabCut were further processed with the DLCAnalyzer (i.e., R code for OFT analysis of multiple files for the Free Exploratory Paradigm and Open Field Test, R code for EPM analysis of multiple files for the Elevated Plus Maze Test, https://github.com/ETHZ-INS/DLCAnalyzer)[42] using R (version 4.0.3; www.r-project.org).

Measurement of resting metabolic rate in a calorimetry system

As described previously, the indirect calorimetry was employed to analyze the resting metabolic rate of the mice[7]. Briefly, air was perfused through the four measurement cages of the calorimetry system (TSE PhenoMaster, TSE Systems GmbH, Bad Homburg, Germany) and an empty reference cage. During the measuring period, oxygen levels in the measurement cage were reduced, while carbon dioxide levels increased due to the animals' respiration. After the air flowed through both cages, the air composition in the measurement cages and reference cage was compared. The metabolic rate of the animals was determined by calculating the difference in air compositions. The measurement for each cage lasted 2 min, meaning a measuring cycle of five cages lasted 10 minutes and 6 data points were collected for each of the four animals within an hour.

Since the calorimetry system was located in another room (same light cycle as in the housing room), the mice were transported in group cages (Makrolon type III cages; food pellets and water, bedding, nesting material, and environmental enrichment as described in the section on "animals, enrichment types, and handling methods") the day before the measurement was started to habituate the animals to the room. In the morning of the next day, the mice were weighted and then individually transferred using the tunnel from their group cage to the measurement cages of the calorimetry system (dimensions at floor level: approximately 15 × 30 cm, dimensions at top level: 17.5 × 35.5 cm, height: 13 cm; floor covered with fine wooden bedding material, JRS Lignocel FS14, spruce/ fir, 2,5-4 mm). After all mice had been transferred to the calorimetry system, the measurement was started. In the morning of the next day, after the lights turned on, the 2-step lockbox (Figure 3c) was presented to the mice, as described in the section on "lockbox task in the calorimetric apparatus". Subsequently, the calorimetric measurement was stopped, and the Grid Exploratory Paradigm was conducted. Before the mice were transferred with the tunnel to their group cages and transported to their housing room, the mice were weighted.

Variable of interest was the resting metabolic rate (RMR), which was quantified as the oxygen consumption rate (VO2) during the animals' resting phases. For the RMR analysis, the data were standardized by excluding the active phases of the animals, ensuring the same number of data points for each individual was analyzed. As only four animals could be tested simultaneously in the calorimetry system, the starting time slightly varied between groups. The starting time of the group with the latest measurement start was used as a reference point (11:13 AM). The first hour of recording (i.e., six data points) was omitted due to increased activity during habituation. Additionally, data points after 7:00 AM on the following day were excluded, as the onset of lights increased activity once again and the 2-step lockbox was presented to the mice. This resulted in 113 data points per individual for RMR analysis. To distinguish resting phases from active phases, a plot of cumulative frequency percentage against the measured VO2 was generated. Through segmented linear regression, the threshold between the VO2 levels of the resting and active phases was calculated. Data points falling below this threshold were utilized for determining the RMR (R package 'segmented'; R version 4.2.2; www.r-project.org)[46].

Lockbox task in the calorimetric apparatus

After the animals spent a night in the measurement cages of the calorimetric system, a 2-step lockbox was presented to them (Figure 3c). The 2-step lockbox comprised a stick

that needed to be pulled out first before a block could be removed, unveiling the food reward (i.e., the oat flake). In the morning, after the lights turned on at approximately 7 A.M., the experimenter carefully opened the cages one after the other and, with the gloved hand, shifted bedding material from the front part of each cage to the rear. This was conducted within a 10-min measuring cycle and was then repeated in a second measuring cycle. Subsequently, the open 2-step lockbox was filled with an oat flake and was placed in the front of the cage for a 10-min measuring cycle. Then the lockbox was rebaited with an oat flake and locked. The animals were allowed to explore and interact with the lockbox on a voluntary basis for approximately 30 minutes. This process was considered as habituation. After 30 minutes, the lockbox was rebaited with another oat flake if it was open; otherwise, the experimenter's hand only touched the lockbox to ensure a consistent procedure for all animals, regardless of whether they opened it or not. After the enrichment phase, the 30-minute phase, during which the baited lockbox was presented in a locked state, was eliminated from the habituation period because the animals were already familiar with the lockbox at this timepoint. The animals were recorded with a Basler acA1920-40um camera (Lens LM25HC7, f = 25mm, k = 1.4, Kowa, Nagoya, Japan; 30 monochrome frames per second) and behavioral analysis of the (last) lockbox trial before and after the enrichment phase were manually performed using BORIS[43].

Variables of interest were the interaction time with the lockbox and the time point when it was opened. Moreover, resting phases were observed. These variables were defined, as follows: interactions with both non-removeable and removeable lockbox elements included sniffing, touching with paws or nose, climbing, gnawing, digging, eating or carrying around lock box elements. The lockbox was deemed open when the mouse had removed both stick and the block, allowing access to the oat flake. Resting behavior was evident when a mouse was lying curled up on its side or sitting in a curled-up position and the face tucked into the body[47]. During these resting phases, a mouse remained still, but occasional brief single twitches could be observed[47]. Another variable of interest was the consumption rate (VO2) during lockbox trials. The three values of the oxygen consumption rate measured during the last lockbox trial were averaged for the time point before and after the enrichment phase, respectively. Only mice that interacted with the lockboxes and did not rest were included in the analysis.

Measurement of fecal corticosterone metabolites

All fecal pellets that were excreted when mice were kept for approximately 24 hours in the measuring cages of the calorimetric system were collected using forceps. The samples were frozen at −80°C. Extraction and analysis of fecal corticosterone metabolites (FCMs, i.e., 5α–3β, 11β–corticosterone metabolites in ng/0.05 g feces) were conducted later according to previously established methods[48–50].

Grid Exploratory Paradigm

Grid Exploratory Paradigm was performed to examine trait anxiety-related behavior. In the literature, this test is also called the "Free Exploratory Paradigm"[51] but to avoid confusion with the "Free exploration Paradigm" below, we named it more precisely here. The Grid Exploratory Paradigm was conducted after the calorimetric measurement was completed. The cage lid was removed, and the cage was moved to a rack on the opposite side of the room. Then a ladder (30.5 × 11 cm), constructed from metal grid and a frame (custom-built, 3D printed), was placed into the cage (i.e., start of the test) and the mice were allowed to explore the ladder for 10 minutes. The test was repeated three times with a break of 5 minutes between the trials. The ladders were cleaned with 70 % ethanol after each trial.

The animals were recorded with a Basler acA1920-40um camera (Lens LM25HC7, f = 25mm, k = 1.4, Kowa, Nagoya, Japan; 30 monochrome frames per second). Variables of interest were the latency to explore, the total time of exploration, and the number of explorations. Exploration was defined as all four paws having contact with the ladder or the cage edge, except from rearing events or when the mouse was lifting a leg while walking. Behavioral analysis were manually performed using BORIS[43].

Bodyweight

Mice were weighted weekly during the general health inspection and cage change. For this, they voluntarily entered the tunnel and then climbed from the tunnel onto the scale. After weighing, they voluntarily climbed from the scale into the tunnel and were transferred to their home cage.

Perfusion

At the end of the study, transcardiac perfusion was conducted using the Perfusion ONE System (Leica Biosystems, Nussloch, Germany). To induce anesthesia, a mixture of ketamine and xylazine was administered intraperitoneally at a volume of 10 µL/g bodyweight was administered intraperitoneally using 30 Gauge needles. The initial planned dosage of 80 mg/kg ketamine (Ketabel 100 mg/ml Injektionslösung, bela-pharm, Vechta, Germany) and 16 mg/kg xylazine (Xylazin, 20 mg/ml Injektionslösung, WDT, Garbsen, Germany) had to be adjusted to 180 mg/kg ketamine and 20 mg/kg xylazine since the mice did not reach surgical tolerance. Following the intraperitoneal injection, the loss of the righting reflex, lid reflex of both eyes, and pedal withdrawal reflex of all limbs were assessed. If the lid reflex on one side or pedal withdrawal of at least one limb persisted 15 min after anesthesia induction, an additional dose of ketamine and xylazine was administered intraperitoneally. Once all reflexes were lost, the thorax was opened below the sternum and the heart was dissected free to insert the cannula into the left ventricle. The perfusion pressure was set at 150 mmHg and the right atrium of the heart was opened to drain the perfusion fluid. Animals were perfused with 0.9% physiological saline solution (NaCl) for two minutes to wash out all blood from the vasculature, followed by perfusion with 4% paraformaldehyde. After perfusion, the adrenal glands were carefully removed, freed from adipose fatty tissue under an incident light microscope, and weighted with a precision scale. For further pathological examinations, the animal bodies were stored in a 4 % paraformaldehyde solution.

Pathohistological evaluation
Samples of formalin fixed pars non-glandularis and glandularis of the stomach, small intestine, caecum, colon, mesenteric lymph node, liver, and kidney were paraffin embedded and cut into 3-µm sections followed by routine haematoxylin and eosin staining for pathohistological evaluation.

Method of blinding
Computer-generated random numbers were used to assign animal IDs to a study group (i.e., the IDs did not contain any information on the study group). The persons analyzing the videos and conducting pathological examinations were blinded. When statistical analysis was carried out, the groups were named after numbers and numbers were replaced with the type of enrichment after the analysis was completed. Due to the obvious differences in the enrichment between the study groups, the animal care takers and experimenters could not be blinded when they handled the animals.

Method of randomization
At arrival animals are assigned to a study group by simple randomization (i.e., the first mouse caught was assigned to the CH group, the second mouse to the SEE group, the third mouse to the LE group, etc.). For the phenotyping test battery, the order of the mice was randomized. Moreover, in the LE group, the order of mice was randomized every day. When the mice were trained on the single-mechanism lockboxes, the order of the lockboxes was randomized every day. The lock boxes consisting of several single mechanisms or both a single mechanism and a combined lockbox were presented in way that the location of the different lockboxes varied in each trial.

Sample size calculation
The present study was exploratory. After three batches (i.e., three cages with four animals) per study group (i.e., CH, SEE, LE) were examined, the data were analyzed and effect sizes were determined. It was defined during the study planning that the fourth batch should not be examined if the data show an effect size = 0 or a relevant, statistically significant effect ($p < 0.05$) and a large effect size (e.g. for the difference between more than two means Eta-squared > 0.14, for the difference between two mean Cohens d > 0.8). There was a large effect in the total time of exploration (Eta-squared ≈ 0.18, $p < 0.05$) and the latency to explore (Eta-squared ≈ 0.15, $p < 0.05$) in the grid exploratory paradigm as well as in the number of mice solving a 2-step lockbox in the calorimetric apparatus (Cramer-V ≈ 0.5). Therefore, the fourth batch was not investigated in order to reduce the number of animals used for the experiment.

Statistical analysis
Statistical analysis was carried out using IBM SPSS Version 26 (IBM Corporation, Armonk, NY, USA, https://www.ibm.com/products/spss-statistics) and the open-source statistical software R (version 4.2.2, www.r-project.org)[52]. For data analysis using R, the R packages lme4[53] and lmerTest[54] were utilized to fit linear mixed-effects models (LMMs) of varying complexity. Animal (i.e., the individual mice, n = 36; referred to as MouseID in the data) and cage (i.e., groups of four animals, 9 levels) were treated as random factors in all models (exceptions are described below). Fixed factors included group (i.e., the three different types of enrichment, 3 levels) and time (i.e., before and after the 2-months enrichment phase, 2 levels) in the simpler model, or group, time, and interaction between group and time in the more complex model. Diagnostic plots (Q-Q plots, residuals versus fitted values, and histogram of residuals) were generated using the R packages ggplot2[55], ggsci[56], and ggpubr[57] to assess the assumptions of the models and check for any deviations from normality or homoscedasticity. Post-hoc pairwise comparisons were performed using the R package emmeans[58] to evaluate differences between levels of the fixed effects. The false discovery rate (FDR) method was applied to correct p values. Model fits were compared using ANOVA to determine whether the additional complexity in the models resulted in a significant improvement in fit compared to the simpler models. The model selection was based on p, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values. When comparing model fits, if a significant difference in overall fit statistics was observed between the simpler and more complex models, AIC and BIC values were evaluated. Lower AIC and BIC values indicated a better fit to the data, thus, the model with the lower AIC and BIC values was chosen. If no significant difference was detected between the simpler and more complex models, the simpler model was used.

If normality or homoscedasticity was violated, data were transformed. If data transformation failed, the data distribution was identified and, if possible, generalized linear mixed-effects models (GLMMs) were employed. Model fit comparisons were conducted as described above. For GLMMs, post-hoc pairwise comparisons were performed using the R package car[59]. Alternatively, other non-parametric tests were applied.

Details on the statistical analysis are listed in Table 1 for each variable. In cases where other factors than described above influenced the choice of model (e.g., warnings from R, comprehensibility, and interpretability), the rationale is provided in Table 1. Significant differences revealed by post-hoc pairwise comparisons of the interaction effect are shown in the figures.

Relative variability
As a measure of relative variability, the coefficient of variation of 23 continuous variables was calculated for each study group by the following formula: (standard deviation/mean) × 100. Values before and after the enrichment phase were compared within each group to detect an increase or decrease.

Habituation response in behavioral paradigms
The Grid Exploratory Paradigm was carried out three times in a row. The explorative behavior seemed to increase over the trials and some animals tried to climb out of the cage in

Table 1. Overview of statistical analysis. Details on the statistical analysis of each variable are provided. If other factors than described in the section on statistical analysis influenced the choice of model, the rationale is given. *[1] The mean of the three test days before and after the enrichment phase was calculated and used for analysis, respectively. *[2] In the Free Exploratory Paradigm, a technical malfunction occurred during the recording of the LE group on day 1 after the enrichment phase. As a result, only the data from day 2 and day 3 were averaged for these four mice. *[3] After the sum of the time moving in the center and in the periphery was calculated for the Free Exploratory Paradigm, the percentage of time moving in each location was determined. If the DLCAnalyzer did not detect a mouse moving in the center or in the periphery on all three days before or after the enrichment phase (i.e., the sum of the time moving in the center and in the periphery was equal 0), the mouse was excluded from the analysis of this variable. It must be noted that the number of mice moving in the arena detected by the DLCAnalyzer may vary from the number of mice that entered the arena with all four paws since we did not track the paws but only the body center using DLC. Abbreviations: CH: conventional housing; EP: enrichment phase; GLMM: generalized linear mixed-effects models; LE: lockbox enrichment; LMM: linear mixed-effects models; SEE: super-environmental enrichment.

Test and variables	n	Statistical analysis	Rationale	Fixed factors	Random factors
Grid Exploratory Paradigm					
Total time of exploration	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group,time, interaction between group and time	animal, cage
Latency to explore	CH: n = 12 SEE: n = 12 LE: n = 12	GLMM with a Gamma distribution	A GLMM was used after fitting of a LMM to the data, inspection of diagnostic plots, and identification of the data distribution using a histogram of the latency to explore. Given the warning regarding singularity in the GLMM with animal and cage as random factors, we opted to proceed with a simpler model that included only the animal as a random factor.	group, time, interaction between group and time	animal
Number of explorations	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time, interaction between group and time	animal, cage
Free Exploratory Paradigm*[1,2]					
Distance travelled in the arena	CH: n = 12 SEE: n = 12 LE: n = 12	LMM	The more complex model was chosen since it allowed for a more nuanced examination of the differential effects of the different groups across time. Due to convergence issues with the LMM including both animal and cage as random factors, only animal was included as random factor.	group, time, interaction between group and time	animal
Percentage of time moving in the center or periphery*[3]	CH: n = 12 SEE: n = 10 before EP and n = 9 after EP LE: n = 10	LMM	If the DLCAnalyzer did not detect a mouse moving in the center or in the periphery on all three days, the mouse was excluded.	group, time	animal, cage
Number of mice entering the arena	CH: n = 12 SEE: n = 12 LE: n = 12	GLMM with a binomial distribution	Given the warning regarding singularity and convergence issues in the GLMM with animal and cage as random factors, only animal was included as a random factor. Due to a warning regarding convergence issues for the more complex model, the simpler model was used.	group, time	animal
Latency to enter the arena	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time, interaction between group and time	animal, cage
Open Field Test*[1]					
Distance travelled	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage

Test and variables	n	Statistical analysis	Rationale	Fixed factors	Random factors
Time spent in the center	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage
Time spent in the periphery	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time, interaction between group and time	animal, cage
Urination	CH: n = 12 SEE: n = 12 LE: n = 12	GLMM with a binomial distribution	Given the warning regarding singularity in the GLMM with animal and cage as random factors, only animal was included as a random factor.	group, time	animal
Defecation	CH: n = 12 SEE: n = 12 LE: n = 12	GLMM with a Poisson distribution	A GLMM was used after fitting of a LMM to the data, inspection of diagnostic plots, and identification of the data distribution using a histogram of the number of fecal boli. Given the warning regarding singularity, only animal was included as a random factor.	group, time	animal
Elevated Plus Maze Test[*1]					
Distance travelled	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage
Number of protected stretch attend:	CH: n = 12 SEE: n = 12 LE: n = 12	LMM	Given the warning regarding singularity only animal was included as a random factor.	group, time	animal
Time spent on open arms or closed arms	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage
Number of transitions to open arms	CH: n = 12 SEE: n = 12 LE: n = 12	GLMM with a Poisson distribution	A GLMM was used after fitting of a LMM to the data, inspection of diagnostic plots, and identification of the data distribution using a histogram of the number of transitions to the open arms.	group, time	animal, cage
Number of transitions to closed arms	CH: n = 12 SEE: n = 12 LE: n = 12	LMM	Given the warning regarding singularity in the LMM, only animal was included as a random factor.	group, time	animal
Ratio of the time spent on the open arms to the time spent on the closed arms	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage
Defecation	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage
Urination	CH: n = 12 SEE: n = 12 LE: n = 12	LMM	Since the residuals of the simpler model deviated from the normal distribution, but the residuals of the more complex model met the normality assumptions, the more complex model was chosen.	group, time, interaction between group and time	animal, cage
Performance in lockboxes in the calorimetric apparatus					
Number of mice solving the lockbox	CH: n = 12 SEE: n = 12 LE: n = 12	Chi-square test, McNemar test	Due to warnings suggesting that the optimization algorithm encountered difficulties, we opted not to use a LMM or GLMM.	–	–

Test and variables	n	Statistical analysis	Rationale	Fixed factors	Random factors
Total interaction time with the lockbox	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, time	animal, cage
Interaction time with the lockbox until solved	Before EP: CH: n = 3 SEE: n = 6 LE: n = 3 After EP CH: n = 8 SEE: n = 6 LE: n = 12	LMM	Only mice that were able to solve the lockbox were included in data analysis. Square root transformation of the values was conducted to meet the assumptions of a normal distribution. Given the warning regarding singularity in the LMM, only animal was included as a random factor.	group, time	animal
Physiological variables					
Resting metabolic rate	CH: n = 12 SEE: n = 12 LE: n = 8	LMM	4 animals of the LE group were excluded because not enough data points of the measurement before the enrichment phase were available for analysis. Values were logarithmized to meet the assumptions of a normal distribution.	group, time	animal, cage
Oxygen consumption during lockbox trial in the calorimetric apparatus	Before EP CH: n = 5 SEE: n = 6 LE: n = 5 After EP CH: n =11 SEE: n = 11 LE: n = 11	LMM	Only mice that interacted with the lockboxes and did not rest during the lockbox trial were included. Comparisons of model fits revealed a significant difference in overall fit statistics between the models. The more complex model had lower AIC, the simpler model had lower BIC values. The more complex model was chosen as it offered a more comprehensive and interpretable analysis of the data.	group, time, interaction between group and time	animal, cage
Fecal corticosterone metabolites	CH: n = 12 SEE: n = 12 LE: n = 12	LMM	Since the residuals of the model including the interaction between group and time were not normally distributed, the simpler model without the interaction as fixed factor was chosen for analysis.	group, time	animal, cage
Adrenal glands relative to the bodyweight	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group	Cage
Bodyweight	CH: n = 12 SEE: n = 12 LE: n = 12	LMM		group, week of experiment	animal, cage

trial 2 and trial 3 which had to be prevented by the experimenter. Due to this disturbance of the animal's behavior in these two trials, only trial 1 was used for further analysis.

The Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test were carried out on three consecutive days. Differences in the distance travelled between the days were analyzed using the Friedman test with Dunn-Bonferroni post-hoc test, as described previously[60] (figure and results given as extended data[61]). As a measure of relative variability of the distance travelled on the three days, the coefficient of variation was calculated. In the Free Exploratory Paradigm, four mice of the LE group had to be excluded from this analysis due to a technical malfunction of the recording on day 1 after the enrichment phase. A repeatability analysis was carried out using "rptR" library for RStudio to determine the source of the variance within the data[62,63], as described previously by Rudeck *et al.*[60]. Briefly, random factors were animal, group, or batch; the fixed factor was the distance travelled. The 95 % confidence interval was based on 500 bootstrap runs and 100 permutations[60]. The likelihood ratio test was used to test for statistical significance of the

repeatability of the distance travelled between the three days[60]. The most important random factor was determined by calculating the repeatability "R" over the three days for all fixed factors including animal, group, and batch as multiple random grouping factors[60].

Interrater reliability

Since the videos of the Grid Exploratory Paradigm and the lockbox trials in the calorimetric apparatus were manually annotated, consistency of the data was determined by analyzing the interrater reliability between the person who labeled all videos and another observer who was asked to label six videos. The interrater reliability was calculated using BORIS[43] (Cohens Kappa) and interpreted according to Landis and Koch[64].

Results

Grid Exploratory Paradigm

The interrater reliability for the video analysis ranged from $\kappa = 0.907$ to $\kappa = 0.966$ which is considered almost perfect according to Landis and Koch[64].

Total time of exploration: In the LMM, the group had no significant effect ($F_{2, 11.09} = 1.20$, $p = 0.338$) on the total time of exploration. Time ($F_{1, 33} = 19.73$, $p < 0.001$) significantly affected the total time of exploration, with higher values before compared to after the enrichment phase. Moreover, the interaction between group and time ($F_{2, 33} = 5.54$, $p = 0.008$) had a significant effect on the total time of exploration.

Post-hoc analysis of the interaction effect revealed that the groups neither differed before (CH versus LE: estimate = 38.75, SE = 28.1, df = 11.1, t.ratio = 1.38, p = 0.326; CH versus SEE: estimate = 36.58, SE = 28.1, df = 11.1, t.ratio = 1.30, p = 0.330; LE versus SEE: estimate = –2.17, SE = 28.1, df = 11.1, t.ratio = –0.077, p = 0.940) nor after the enrichment phase (CH versus LE: estimate = –59.17, SE = 28.1, df = 11.1, t.ratio = –2.103, p = 0.205; CH versus SEE: estimate = –7.92, SE = 28.1, df = 11.1, t.ratio = –0.28, p = 0.919; LE versus SEE: estimate = 51.25, SE = 28.1, df = 11.1, t.ratio = 1.82, p = 0.205). The total time of exploration was higher before compared to after the enrichment phase in the CH group (estimate = 92.50, SE = 20.8, df = 33.0, t = 4.44, p = 0.001), while the values measured before and after the enrichment phase neither differed in the LE (estimate = –5.42, SE = 20.8, df = 33.0, t = –0.260, p = 0.919) nor the SEE group (estimate = 48.00, SE = 20.8, df = 33.0, t = 2.31, p = 0.138) (Figure 5a).

Latency to explore: In the GLMM, the effects of group ($Chi^2 = 15.57$, df = 2, $p < 0.001$), and interaction between group and time ($Chi^2 = 17.35$, df = 2, $p < 0.001$) on the latency to explore were significant. Time had no significant effect ($Chi^2 = 1.04$, df = 1, p = 0.308).

Post-hoc analysis of the group revealed that CH differed compared to LE (estimate = 0.71, SE = 0.30, df = Inf, z.ratio = 2.34, p = 0.03) and SEE (estimate = 0.90, SE = 0.30, df = Inf, z.ratio = 3.03, p = 0.007). LE and SEE did not differ from each other (estimate = 0.20, SE = 0.30, df = Inf, z.ratio = 0.65, p = 0.52).

The post-hoc analysis of the interaction effect demonstrated that some group differences were based on differences between the groups at baseline and were not an effect of the different enrichment conditions. Before the enrichment phase, the latency to explore in mice assigned to SEE was lower than in mice assigned to CH (estimate = 1.41, SE = 0.39, df = Inf, z.ratio = 3.67, p = 0.001) or LE (estimate = 1.22, SE = 0.39, df = Inf, z.ratio = 3.16, p = 0.005) (Figure 5b). There was no difference between CH and LE (estimate = = 0.20, SE = 0.38, df = Inf, z.ratio = 0.52, p = 0.696). After the enrichment phase, the latency to explore differed between CH and LE (estimate = 1.22, SE = 0.38, df = Inf, z.ratio = 3.18, p = 0.005), while there was no significant difference between CH and SEE (estimate = 0.40, SE = 0.38, df = Inf, z.ratio = 1.03, p = 0.385) or LE and SEE (estimate = –0.82, SE = 0.39, df = Inf, z.ratio = –2.09, p = 0.079). When comparing the latency to explore before and after the enrichment phase within the group, the values significantly decreased in the LE group (estimate = 1.35, SE = 0.33, df = Inf, z.ratio = 4.08, $p < 0.001$). There was no significant change over time within the CH group (estimate = 0.33, SE = 0.32, df = Inf, z.ratio = 1.02, p = 0.385) and SEE group (estimate = –0.68, SE = 0.35, df = Inf, z.ratio = –1.93, p = 0.089).

Number of explorations: In the LMM, neither group ($F_{2, 8.63} = 3.32$, $p = 0.085$) nor time ($F_{1, 33} = 0.64$, $p = 0.428$) significantly affected the number of explorations, while the interaction between group and time had a significant effect ($F_{2, 33} = 3.365$, $p = 0.047$). However, the post-hoc analysis of the interaction effect did not identify group differences before (CH versus LE: estimate = –3.67, SE = 2.50, df = 8.63, t.ratio = –1.47, p = 0.53; CH versus SEE: estimate = –6.42, SE = 2.50, df = 8.63, t.ratio = –2.57, p = 0.23; LE versus SEE: estimate = –2.75, SE = 2.50, df = 8.63, t.ratio = –1.10, p = 0.58) or after (CH versus LE: estimate = –1.17, SE = 2.50, df = 8.63, t.ratio = –0.47, p = 0.724; CH versus SEE: estimate = –1.08, SE = 2.50, df = 8.63, t.ratio = –0.43, p = 0.724; LE versus SEE: estimate = 0.08, SE = 2.50, df = 8.63, t.ratio = 0.03, p = 0.974) the enrichment phase were found (Figure 5c). The comparison of the values measured before and after the enrichment phase within the groups did not reveal any significant differences (CH: estimate = –1.17, SE = 1.45, df = 33, t.ratio = –0.80, p = 0.584; SEE: estimate = 4.17, SE = 1.45, df = 33, t.ratio = 2.86, p = 0.108; LE: estimate = 1.33, SE = 1.45, df = 33, t.ratio = 0.92, p = 0.584).

Free Exploratory Paradigm

The variances observed for the distance travelled were explained by the individual animal while group and batch did not significantly contribute to the observed variances (Table 2). The repeatability value of 0.357 after the enrichment phase indicated that 35.7 % of the variance can be explained by the factor animal. Independent of the time point before or after the enrichment phase, the coefficient of variation determined for distance travelled was lowest on day 3 (Table 3).

Distance travelled in the arena: In the LMM, neither group ($F_{2, 54.67}$) = 2.15, p = 0.126) nor the interaction between group and time ($F_{2, 33} = 2.30$, $p = 0.117$) had a significant effect on the distance travelled in the arena (Figure 5d). However, time was found to have a significant effect on the distance travelled

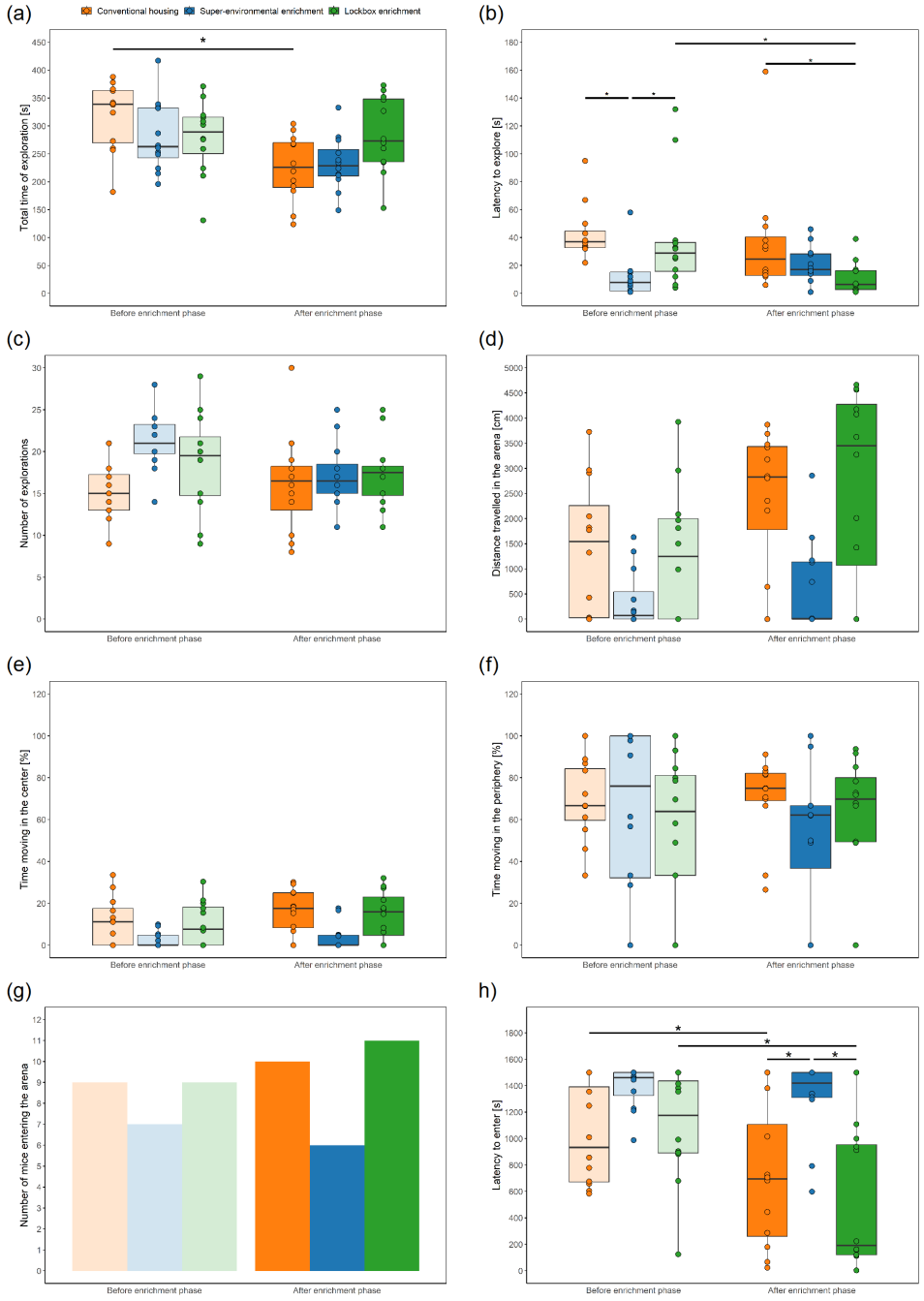


Figure 5. Grid Exploratory Paradigm and Free Exploratory Paradigm before and after exposure to the different types of enrichment. Boxplots: The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile, and the whiskers represent values which are no greater than 1.5 × IQR. Additionally, individual data points representing each animal are overlaid on the boxplot as dots. It's important to note that these dots may overlap due to multiple animals having the same or very similar values. Grid Exploratory Paradigm: (**a**) Total time of exploration, (**b**) latency to explore, (**c**) number of explorations; Free Exploratory Paradigm: (**d**) distance travelled in the arena, (**e**) percentage of time moving in the center, (**f**) percentage of time moving in the periphery, (**g**) number of mice entering the arena (i.e., a mouse entered the arena when all four paws left the tunnel and were placed on the floor of the arena on at least on day out of the three days on which the test was performed), (**h**) latency to enter the arena (i.e., entering the arena was defined as placing all paws on the floor of the arena; if a mouse did not enter the arena, the latency was set to 1500 s). * $p < 0.05$ (linear mixed-effects model or generalized linear mixed-effects model, post-hoc pairwise comparisons of interaction between group and time), n = 12 animals per group. An asterisk (*) indicates a significant difference between two groups before or after the enrichment phase, or a significant difference within a group before and after the enrichment phase (details are given in the results section). For (**e**) and (**f**) applied: if the DLCAnalyzer did not detect a mouse moving in the center or in the periphery on all three days, the mouse was excluded; n = 12 animals both before and after the enrichment phase in the CH group, n = 10 before and n = 9 animals after the enrichment phase in the SEE group, n = 10 animals both before and after the enrichment phase in the LE group were included in this analysis. This number may vary from the number of mice entering the arena with all four paws since we did not track the paws but only the body center using DLC.

Table 2. Repeatability values for distance traveled in the Free Exploratory Paradigm, Open Field, and Elevated Plus Maze test with multiple grouping factors: animal, group, and batch. For the random factors animal, group, and batch, the repeatability R, the 95 % confidence interval (CI) with lower and upper bounds calculated at the 2.5th and 97.5th percentiles, and the p values calculated by the likelihood ratio test were determined over the three days (n = 36 mice). Distribution of Free Exploratory Paradigm data generated before the enrichment phase did not allow to analyze repeatability values. * Due to a technical malfunction of the recording from the lockbox enrichment group on day 1 after the enrichment phase, four mice of this group had to be excluded from the analysis.

Test	Repeatability for animal			Repeatability for group			Repeatability for batch		
	R	CI	p-value	R	CI	p-value	R	CI	p-value
Before enrichment phase									
Open Field Test	0.401	[0.125; 0.559]	< 0.001	0	[0; 0.142]	1	0	[0; 0.134]	1
Elevated Plus Maze Test	0.354	[0.12; 0.531]	< 0.001	0.056	[0; 0.25]	0.192	0	[0; 0.124]	1
After enrichment phase									
Free Exploratory Paradigm*	0.357	[0.144; 0.524]	< 0.001	0	[0; 0.076]	1	0.022	[0; 0.118]	0.316
Open Field Test	0.307	[0.066; 0.468]	0.001	0.021	[0; 0.158]	0.347	0	[0; 0.103]	0.5
Elevated Plus Maze Test	0.298	[0.059; 0.478]	0.001	0.017	[0; 0.172]	0.367	0.044	[0; 0.211]	0.221

Table 3. Coefficient of variation for the distance traveled in the Free Exploratory Paradigm, Open Field, and Elevated Plus Maze test. The data of 36 mice including all study groups are shown as the coefficient of variation, which was calculated using the following formula: (standard deviation/mean) × 100. The lowest value in a row was underlined. For the distance traveled, independent of the time point before or after the enrichment phase, the coefficient of variation was lowest on day 3 in the Free Exploratory Paradigm and the Open Field Test. In the Elevated Plus Maze Test, the coefficient of variation was lowest on day 2 before and on day 1 after the enrichment phase. *Due to a technical malfunction of the recording from the lockbox enrichment group on day 1 after the enrichment phase in the Free Exploratory Paradigm, four mice had to be excluded when calculating mean and standard deviation of the distance traveled on day 1.

Test	Day		
	1	2	3
Before enrichment phase			
Free Exploratory Paradigm	262	149	107
Open Field Test	17	16	13
Elevated Plus Maze Test	19	15	18
After enrichment phase			
Free Exploratory Paradigm*	170	111	79
Open Field Test	26	19	18
Elevated Plus Maze Test	23	27	26

($F_{2, 33}$ = 5.75, p = 0.022), with higher values after than before the enrichment phase.

Percentage of time moving in the center: While the LMM did not reveal group differences in the percentage of time moving in the center ($F_{2, 6}$ = 4.77, p = 0.058) (Figure 5e), time had a significant effect ($F_{1, 35}$ = 4.95, p = 0.033). The mice moved more in the center after than before the enrichment phase (Figure 5e).

Percentage of time moving in the periphery: Neither group ($F_{2, 6}$ = 0.55, p = 0.602) nor time ($F_{1, 35}$ = 0.31, p = 0.582) significantly affected the percentage of time moving in the periphery in the LMM (Figure 5f).

Number of mice entering the arena: In the GLMM, neither group (Chi^2 = 2.83, df = 2, p = 0.243) nor time (Chi^2 = 0.79, df = 1, p = 0.374) significantly affected whether a mouse entered the arena with all four paws on at least one day out of the three days the test was performed (Figure 5g).

Latency to enter the arena: In the LMM, the latency to enter the arena with all four paws was not affected by the group ($F_{2, 51.07}$ = 2.6918, p = 0.077). Time ($F_{1, 33}$ = 7.68, p = 0.009) and the interaction between group and time ($F_{2, 33}$ = 4.542, p = 0.018) had significant effects on the latency. The latency to enter the arena was higher before compared to after the enrichment phase.

Post-hoc analysis of the interaction effect showed that, before the enrichment phase, the groups did not differ (CH versus LE estimate: –71.7, SE = 167, df = 9.95, t.ratio = –0.43, p = 0.676; CH versus SEE: estimate = –365.0, SE = 167, df = 9.95, t.ratio = –2.19, p = 0.089; LE versus SEE: estimate = –293.3, SE = 167, df = 9.95, t.ratio = –1.760, p = 0.164) (Figure 5h). After the enrichment phase the

mice housed under SEE took longer to enter the open arena than mice with access to LE (estimate = –774.8, SE = 167, df = 9.95, t.ratio = –4.65, p = 0.005) and CH mice (estimate = –595.1, SE = 167, df = 9.95, t.ratio = –3.57, p = 0.015). There was no difference between CH and LE (estimate = 179.7, SE = 167, df= 9.95, t.ratio = 1.08, p = 0.354). When comparing the latency to enter the arena within the groups over time, a significant decrease was found in CH (estimate = 313.1, SE = 113, df = 33, t.ratio = 2.77, p = 0.023) and LE (estimate = 564.4, SE = 113, df = 33, p.ratio = 5.00, p < 0.001) but not SEE (estimate = 83.0, SE = 113, df = 33, t.ratio = 0.74, p = 0.50).

Urination and defecation: All mice except from one mouse, that was housed under conventional conditions and excreted one fecal bolus on day 3 after the enrichment phase, neither defecated nor urinated in the arena.

Open Field Test

The variances observed for the distance travelled were explained by the individual animal while group and batch did not significantly contribute to the observed variances (Table 2). The repeatability values of 0.401 (before enrichment phase) and 0.307 (after enrichment phase) indicated that 40.1 % and 30.7 % of the variance can be explained by the factor animal. Independent of the time point before or after the enrichment phase, the coefficient of variation determined for distance travelled was lowest on day 3 (Table 3).

Distance travelled: The LMM revealed that time had a significant effect on the distance travelled ($F_{1, 35}$ = 89.53, p < 0.001), with higher values before when compared to after the enrichment phase. Group ($F_{2, 33}$ = 0.96, p = 0.392) did not significantly influence the distance travelled (Figure 6a).

Time spent in the center: In the LMM, neither group ($F_{2, 6}$ = 1.42, p = 0.312) nor time ($F_{1, 35}$ = 1.03, p = 0.318) significantly affected the time spent in the center (Figure 6b).

Time spent in the periphery: The time spent in the periphery was not affected by group ($F_{2, 11.21}$ = 0.87, p = 0.445), time ($F_{1, 33}$ = 0.61, 0.441), or the interaction between group and time ($F_{2, 33}$ = 0.04, p = 0.965) in the LMM (Figure 6c).

Urination: Given as median and 25th–75th percentiles in brackets, 0 (0–0) urine spots before and 0 (0–0) urine spots after the enrichment phase in mice assigned to CH, 1 (0–1) urine spots before and 0 (0–1) urine spots after the enrichment phase in mice assigned to SEE, as well as 1 (0–1) urine spots before and 0 (0–1) urine spots after the enrichment phase in mice assigned to LE were counted after the Open Field Test (the values were rounded up to whole number). The GLMM revealed that neither group (Chi^2 = 5.64, df = 2, p = 0.060) nor time (Chi^2 = 0.07, df = 1, p = 0.797) significantly affected urination in the Open Field Test.

Defecation: Given as median and 25th–75th percentiles in brackets, mice assigned to CH excreted 1 (1–2) fecal boli before and 2 (2–2) fecal boli after the enrichment phase, mice assigned to SEE excreted 3 (2–4) fecal boli before and 3 (2–4) fecal boli after the enrichment phase, and mice assigned to LE excreted 3 (1–5) fecal boli before and 3 (2–4) fecal boli after the enrichment phase during the Open Field test (the values were rounded up to whole number). The groups (Chi^2 = 6.97, df = 2, p = 0.031) significantly differed in the GLMM, while time (Chi^2 = 0.02 , df = 1, p = 0.882) had no significant effect on the number of fecal boli excreted during the Open Field Test. In the CH group significantly fewer fecal boli were excreted compared to the LE group (estimate = –0.47, SE = 0.20, df = Inf, z.ratio = –2.41, p = 0.030) and the SEE group (estimate –0.46, SE = 0.20, df = Inf, z.ratio = –2.33, p = 0.0298). When inspecting the data, it can be noticed that mice assigned to CH excreted fewer pellets both before and after the enrichment phase, suggesting that this observation was not attributable to the type of enrichment.

Elevated Plus Maze Test

The variances observed for the distance travelled were explained by the individual animal while group and batch did not significantly contribute to the observed variances (Table 2). The repeatability values of 0.354 (before enrichment phase) and 0.298 (after enrichment phase) indicated that 35.4 % and 29.8 % of the variance can be explained by the factor animal. The coefficient of variation calculated for distance travelled was lowest on day 2 before and on day 1 after the enrichment phase (Table 3).

Distance travelled: In the LMM, the distance travelled was significantly affected by the time ($F_{1, 35}$ = 125.87, p < 0.001), with higher values before when compared after the enrichment phase. Group ($F_{2, 6}$ = 1.116, p = 0.387) had no significant effect (Figure 7a).

Number of protected stretch attend: The number of protected stretch attend was significantly affected by group ($F_{2, 33}$ = 7.41, p = 0.002) and time ($F_{1, 35}$ = 19.83, p < 0.001) in the LMM. The number of protected stretch attend was significantly higher before compared to after the enrichment phase. However, post-hoc pairwise comparisons did not reveal any significant group differences (LE versus CH: estimate = –1.54, SE = 0.51, df = 6, t.ratio = -3.02, p = 0.070; CH versus SEE: estimate: –0.63, SE = 0.51, df = 6, t.ratio = –1.23, p = 0.267; LE versus SEE: estimate = 0.92, SE = 0.51, df = 6, t.ratio = 1.80, p = 0.184) (Figure 7b).

Time spent on open arms: While group ($F_{2, 6}$ = 0.18, p = 0.842) had no effect (Figure 7c), time ($F_{1, 35}$ = 149.20, p < 0.001) significantly decreased the time spent on the open arms in the LMM. Values were higher before than after the enrichment phase.

Time spent on closed arms: While the time spent on the closed arms was not affected by group ($F_{2, 6}$ = 0.04, p = 0.957) (Figure 7d), the time had a significantly increasing effect ($F_{1, 35}$ = 147.18, p < 0.001) in the LMM, with higher values after than before the enrichment phase.

Number of transitions to open arms: The GLMM showed that the number of transitions to the open arms was not

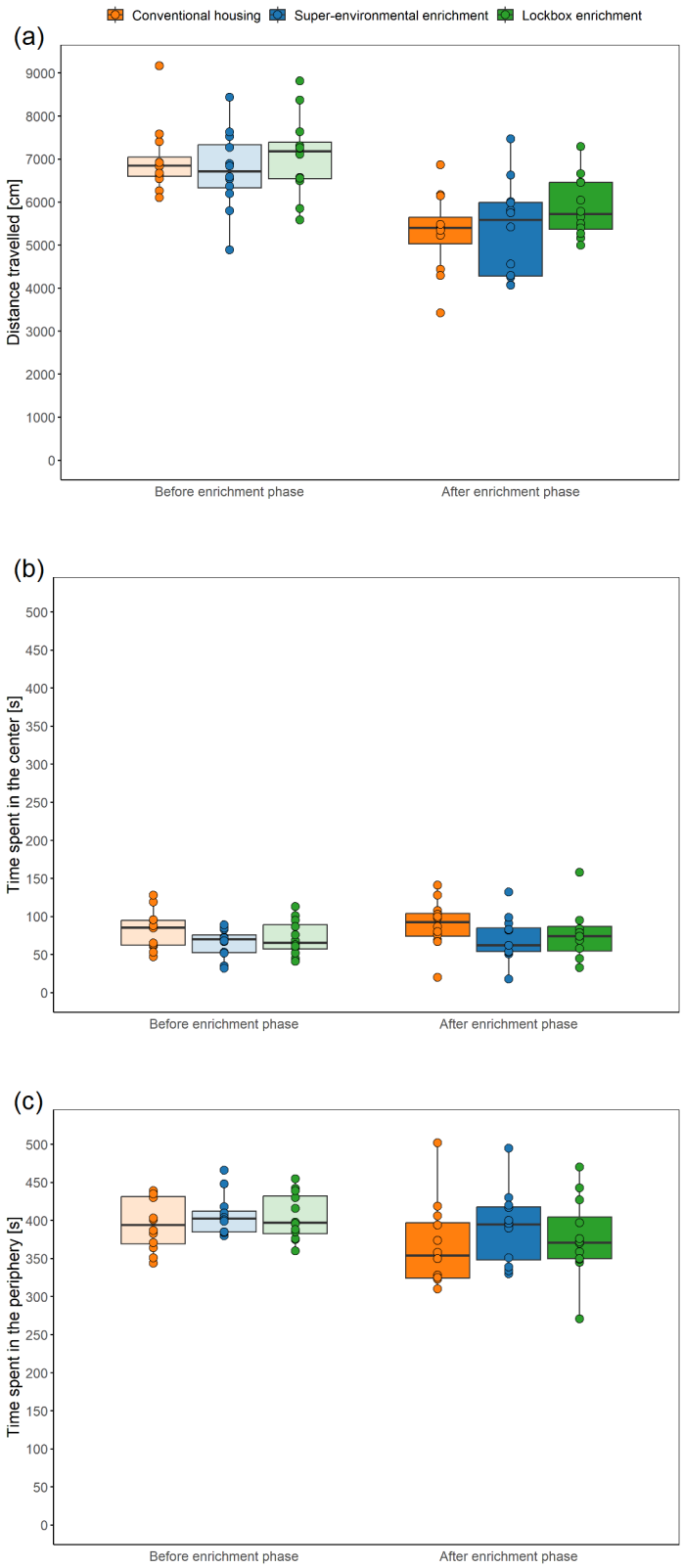


Figure 6. Open Field Test before and after the enrichment phase. Boxplots: The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile, and the whiskers represent values which are no greater than 1.5 × IQR. Additionally, individual data points representing each animal are overlaid on the boxplot as dots. It's important to note that these dots may overlap due to multiple animals having the same or very similar values. (**a**) Distance travelled, (**b**) time spent in the center, (**c**) time spent in the periphery; n = 12 animals per group.

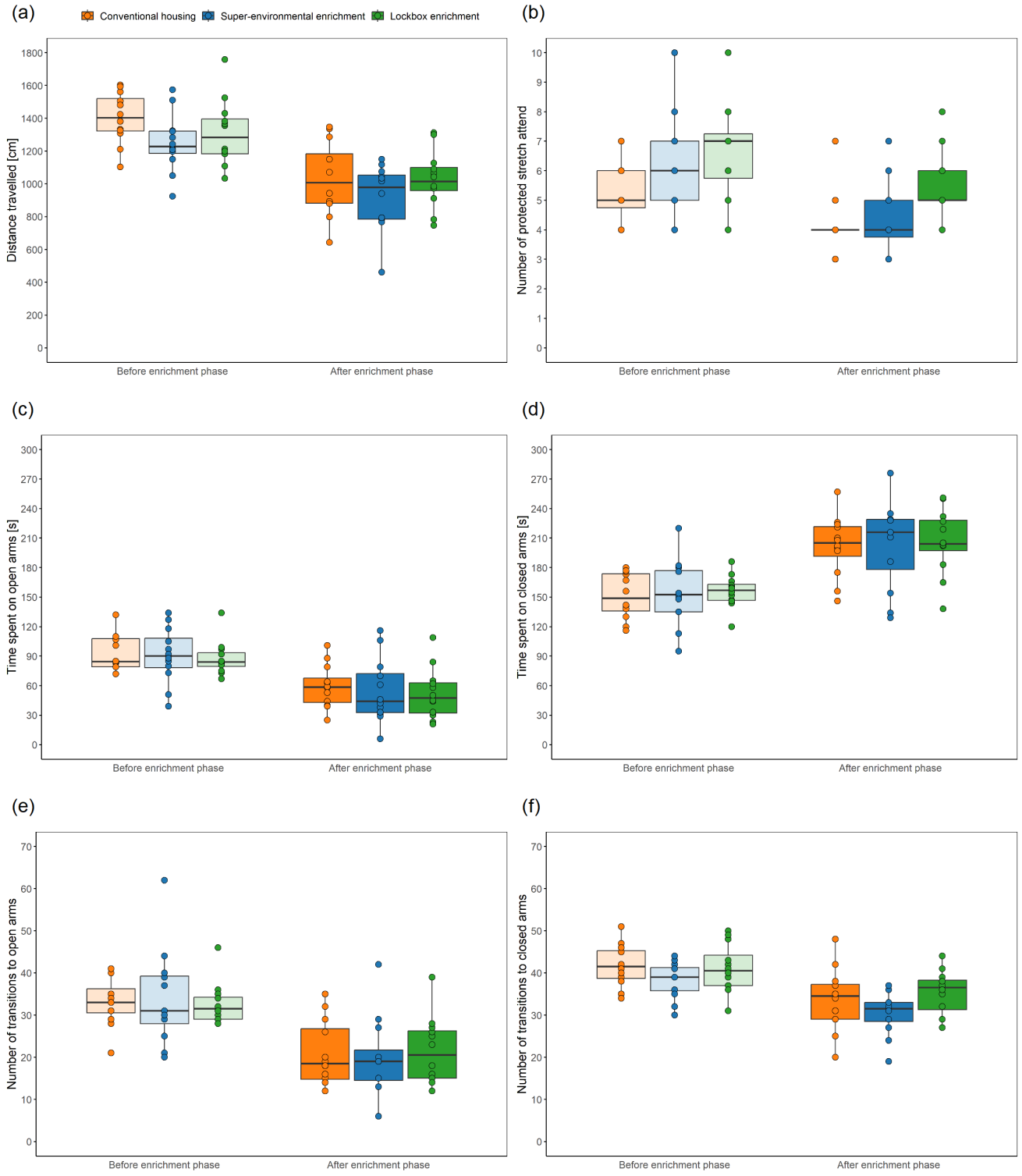


Figure 7. Elevated Plus Maze Test before and after the enrichment phase. Boxplots: The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile, and the whiskers represent values which are no greater than 1.5 × IQR. Additionally, individual data points representing each animal are overlaid on the boxplot as dots. It's important to note that these dots may overlap due to multiple animals having the same or very similar values. (**a**) Distance travelled in the arena, (**b**) number of protected stretch attend, (**c**) time spent on open arms, (**d**) time spent on closed arms, (**e**) number of transitions to open arms, (**f**) number of transitions to closed arms; n = 12 animals per group.

significantly influenced by group (Chi^2 = 0.05, df = 2, p = 0.975) (Figure 7e) but by time (Chi^2 = 104.18, df = 1, p < 0.001). Values were higher before than after the enrichment phase.

Number of transitions to closed arms: Group did not influence the number of the transitions to the closed arms ($F_{2, 33}$ = 3.14, p = 0.057) (Figure 7f), while time had a significantly decreasing effect ($F_{1, 35}$ = 39.44, p < 0.001) in the LMM. The values were higher before than after the enrichment phase.

Ratio of the time spent on the opens arms to the time spent on the closed arms: Given in median and 25th–75th percentiles in brackets, before the exposure to the different types of enrichment, the ratio was 0.55 (0.47–0.83), 0.59 (0.42–0.84), and 0.53 (0.47–0.65) in mice assigned to CH, SEE, and LE, respectively. After the enrichment phase, the ratio was 0.28 (0.18–0.42) in the CH group, 0.20 (0.14–0.48) in the SEE group, and 0.23 (0.14–0.33) in the LE group. Group ($F_{2, 6}$ = 0.17, p = 0.848) had no effect, but time (F1, 35 = 130.33, p < 0.001) significantly decreased the ratio in the LMM, with higher values before compared to after the enrichment phase.

Defecation: Given as median and 25th–75th percentiles in brackets, mice assigned to CH excreted 1 (0–1) fecal boli before and 2 (2–3) fecal boli after the enrichment phase, mice assigned to SEE excreted 2 (1–3) fecal boli before and 2 (1–3) fecal boli after the enrichment phase, mice assigned to LE excreted 1 (0–2) fecal boli before and 3 (3–4) fecal boli after the enrichment phase during the Elevated Plus Maze Test (the values are rounded up to whole number). While the number of fecal boli was not affected by group ($F_{2, 6}$ = 0.18, p = 0.843), it significantly increased over time ($F_{1, 35}$ = 25.97, p < 0.001). The animals excreted more feces during the Elevated Plus Maze Test after compared to before the enrichment phase.

Urination: Given as median and 25th–75th percentiles in brackets, 0 (0–1) urine spots before and 0 (0–1) urine spots after the enrichment phase in mice assigned to CH, 1 (0–1) urine spots before and 0 (0–1) urine spots after the enrichment phase in mice assigned to SEE, 1 (0–1) urine spots before and 1 (1–1) urine spots after the enrichment phase in mice assigned to LE were counted after the Elevated Plus Maze Test (the values are rounded up to whole number). Neither group ($F_{2, 11.13}$ = 0.40, p = 0.680), time ($F_{1, 33}$ = 0.00, p = 1.000) nor interaction between group and time ($F_{2, 33}$ = 2.70, p = 0.08) significantly affected the number of urinary spots counted after the Elevated Plus Maze Test.

Resting metabolic rate

In the LMM, time ($F_{1, 31}$ = 4.49, p = 0.042) significantly affected the RMR, while group had no significant effect ($F_{2, 5}$ = 3.88, p = 0.096) (Figure 8a). The RMR was higher before compared to after the enrichment phase.

Performance in lockbox trials in the calorimetric apparatus

The interrater reliability for the video analysis ranged from κ = 0.648 to κ = 0.885 which is considered substantial to almost perfect according to Landis and Koch[64].

Number of mice solving the lockbox: Before the exposure to the enrichment, the performance in opening the 2-step lockbox did not differ between the groups (Chi^2 = 2.250, df = 2, p = 0.490, Cramer-V = 0.250) (Figure 8b). After the exposure to the enrichment, the three different groups differed from each other (Chi^2 = 7.754, df = 2, p = 0.031, Cramer-V = 0.464): All mice (n = 12) from the LE, 6 mice from the SEE and 8 mice from the CH group succeeded in solving this lockbox. When comparing the number of mice that were able to solve the lockboxes between the lockbox trial before and after the enrichment phase within the groups, the McNemar test revealed that only LE had a significant effect (Chi^2 = 7.11, df = 1, p = 0.004, n = 12; CH: Chi^2 = 3.20, df = 1, p = 0.063, n = 12; SEE: Chi^2 = 0, df = 1, p = 1.000, n = 12).

Total interaction time with the lockbox: In the LMM, the group had no significant effect ($F_{2, 6}$ = 0.17, p = 0.844) (Figure 8c) but time had a significant effect on the interaction time with the lockboxes (F1, 35 = 22.58, p < 0.001). The mice spent more time interacting with the lockboxes after the enrichment phase compared to before it.

Interaction time with the lockbox until solved: Among the mice that were able to solve the lockbox, the group did not significantly affect the interaction time with the lockbox until it was solved in the LMM ($F_{2, 26.38}$ = 0.53, p = 0.594) (Figure 8d). Time had a significant effect ($F_{1, 19.35}$ = 12.67, p = 0.002). The interaction time with the lockbox until it was solved decreased after the enrichment phase compared to before it.

Oxygen consumption during lockbox trial in the calorimetric apparatus

In the LMM, neither group ($F_{2, 16.59}$ = 1.14, p = 0.344), time ($F_{1, 27.38}$ = 2.13, p = 0.156) nor the interaction between group and time ($F_{2, 27.74}$ = 2.84, p = 0.075) significantly affected the oxygen consumption during the lockbox trials (Figure 8e).

Fecal corticosterone metabolites

The LMM revealed that group ($F_{2, 6}$ = 0.12, p = 0.892) had no significant effect on the FCM concentrations (Figure 8f), but the time (F1, 35 =32.89, p < 0.001) significantly affected the values. Independent of the group, the FCM values were higher after the exposure to the different types of enrichment compared to baseline.

Bodyweight

While group ($F_{2, 6}$ = 0.30, p = 0.748) did not significantly affect the bodyweight in the LMM (Figure 9), week of experiment ($F_{1, 683}$) = 3766.77, p < 0.001) had a significant effect, with increasing weight over time.

Adrenal glands

The percentage weight of the adrenal glands relative to the bodyweight measured in week 20 was not affected by the group in the LMM ($F_{2, 6}$ = 0.16, p = 0.858; Table 4).

Pathohistological evaluation

The toxicopathological analysis yielded no evidence of group-specific pathological alterations in the analyzed tissues. A mild, multifocal, chronic, interstitial, lymphocytic nephritis

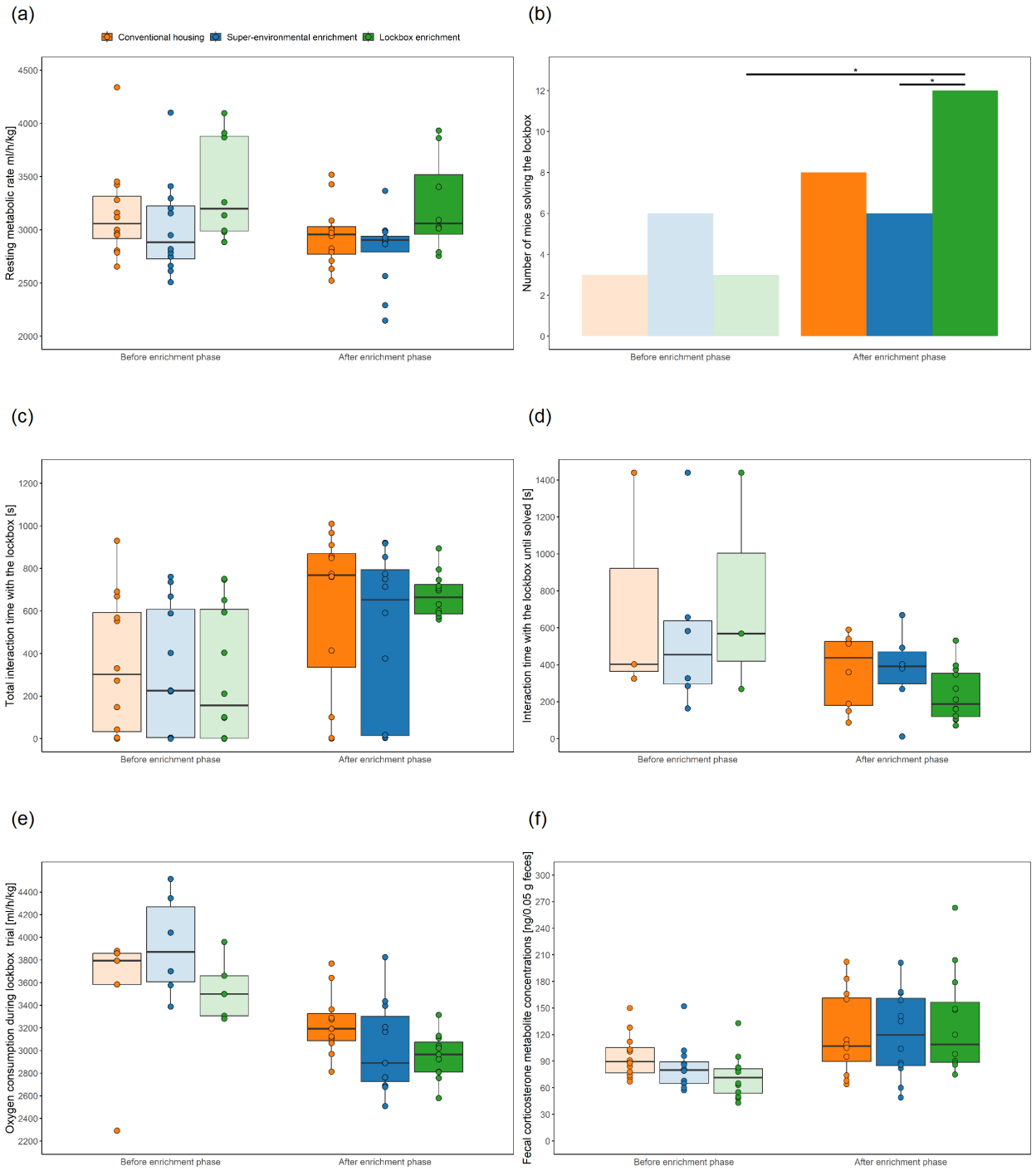


Figure 8. Calorimetric measurement before and after the enrichment phase. Boxplots: The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile, and the whiskers represent values which are no greater than 1.5 × IQR. Additionally, individual data points representing each animal are overlaid on the boxplot as dots. It's important to note that these dots may overlap due to multiple animals having the same or very similar values. (**a**) Resting metabolic rate (convention housing: n = 12 animals, super-environmental enrichment: n = 12 animals, lockbox enrichment: n = 8 animals; 4 animals of the LE group were excluded because not enough data points of the measurement before the enrichment phase were available for analysis), (**b**) number of mice solving the lockbox: n = 12 animals per group (data were analyzed using Chi-square test or McNemar test: * $p < 0.05$), (**c**) total interaction time with the lockbox (n = 12 animals per group), (**d**) interaction time with the lockbox until solved (only mice that were able to solve the lockbox were included; for n number see (**b**)); (**e**) oxygen consumption during lockbox trials in the calorimetric apparatus (only mice that interacted with the lockboxes and did not rest were included; before enrichment phase: n = 5 animals in the CH group, n = 6 animals in the SEE group, and n = 5 animals in the LE group; after the enrichment phase: n = 11 animals per group), (**f**) fecal corticosterone metabolite concentrations (n = 12 animals per group). An asterisk (*) indicates a significant difference between two groups before or after the enrichment phase, or a significant difference within a group before and after the enrichment phase (details are given in the results section).

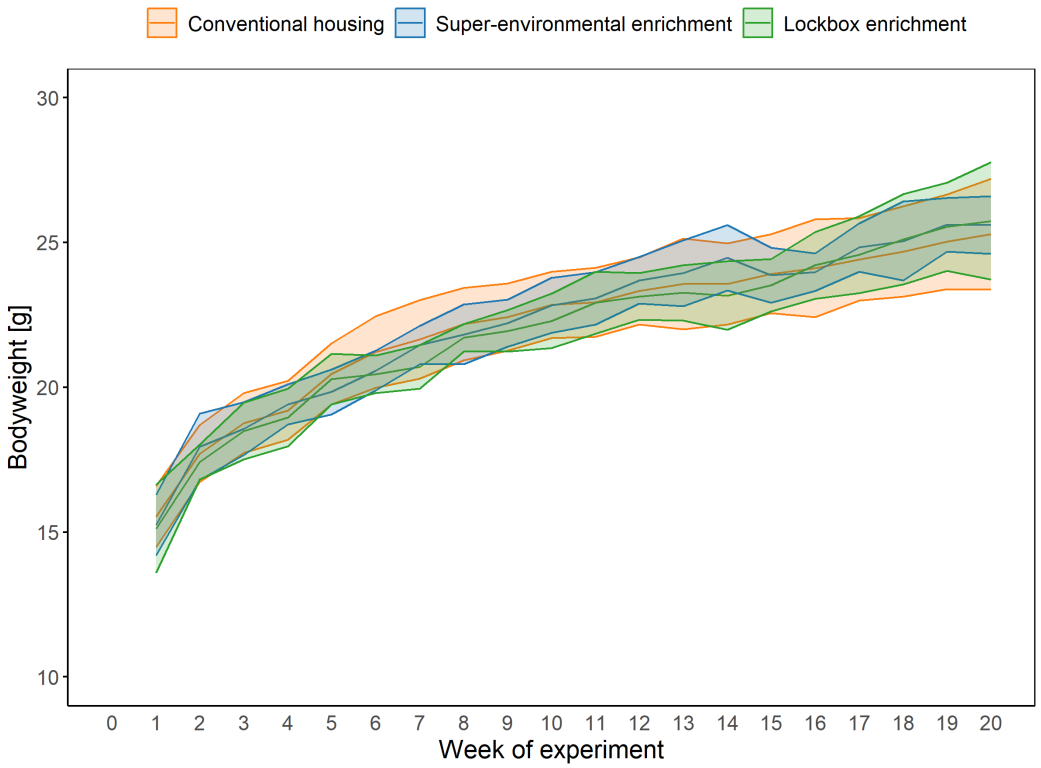


Figure 9. Bodyweight throughout the entire experiment. Data are given as mean ± standard deviation.

Table 4. Percentage weight of the adrenal glands relative to the bodyweight measured in week 20. Data are given as median and 25th–75th percentiles in brackets; the values are rounded up; n = 12 mice per group.

Parameter	Conventional housing	Super-environmental enrichment	Lockbox enrichment
Weight of adrenal glands relative to bodyweight [%]	0.022 (0.021–0.024)	0.021 (0.020–0.026)	0.022 (0.018–0.025)

was noticed in some animals of all groups which was interpreted as a typical background pathology.

Coefficient of variation

The coefficient of variation for 23 continuous variables is shown in Table 5. For most variables, the coefficient of variation increased. However, the coefficient of variation decreased after the enrichment phase when compared to the time before the enrichment phase in 6 (CH), 7 (SEE), or 9 (LE) variables, indicating only slight differences between the study groups.

Discussion

In the scope of animal welfare, we investigated whether LE allowing both manipulative and cognitive activities affected the phenotype and influenced the affective state of young adult female laboratory mice when compared to animals kept under conventional or super-environmentally enriched housing conditions. Main findings of our study were that 2-month exposure to the different types of enrichment affected trait anxiety-related behavior and the performance in solving novel lockbox tasks. No effects were found in standard behavioral tests for state anxiety-related behavior and physiological variables, such as bodyweight, resting metabolic rate, stress hormone metabolite concentrations, and adrenal gland weights. This suggests that the relevant phenotype of the mice was influenced by neither LE nor SEE. The relative variability of most variables increased over time, independent of the housing conditions, while it decreased for some variables, particularly in animals with LE. All lockboxes are available as open-source tools, allowing anyone to use, modify, and distribute them freely.

When interpreting the data of the three study groups, i.e., LE, SEE, and CH, it must be considered that in our study the conventional housing cage may provide the animals with more resources than what is usual in other facilities. All animals,

Table 5. Coefficient of variation for continuous variables examined before and after the enrichment phase. The data of each study group (n = 12 mice per group) are shown as the coefficient of variation, which was calculated using the following formula: (standard deviation/mean) × 100. If the coefficient of variation decreased within a group when compared to baseline (i.e., before the enrichment phase), the value was underlined. Values were rounded up to two decimal places. * The mean of the three testing days was used.

Test	Variable	Conventional housing		Super-environmental enrichment		Lockbox enrichment	
		Before enrichment phase	After enrichment phase	Before enrichment phase	After enrichment phase	Before enrichment phase	After enrichment phase
Grid Exploratory Paradigm	Total duration of exploration [s]	19,42	26,22	22,67	20,80	23,75	24,14
	Latency to explore [s]	48,16	115,66	135,41	72,09	102,22	107,69
	Number of explorations	21,22	36,74	18,56	22,33	31,68	23,42
Free Exploratory Paradigm*	Distance traveled	93,78	59,19	151,74	146,13	104,98	70,73
	Latency to enter the arena	36,74	76,13	11,80	23,00	38,86	98,98
	Time moving in the center [%]	97,77	64,61	144,83	160,40	106,18	79,75
	Time moving in the periphery [%]	27,54	28,59	62,22	67,05	60,09	52,11
Open Field Test	Distance travelled [cm]	17,31	17,31	13,70	20,69	13,29	11,90
	Time spent in the center [s]	30,85	35,03	28,47	41,78	31,92	43,00
	Time spent in the periphery [s]	8,64	15,46	6,58	12,24	7,58	13,85
Elevated Plus Maze Test	Distance travelled [cm]	11,07	21,95	14,21	21,27	15,31	16,85
	Time spent in the center [s]	17,70	28,15	17,78	35,97	13,16	27,81
	Number of transitions to the center	8,95	25,65	16,05	24,50	12,00	18,96
	Time spent on the open arms [s]	21,87	36,78	31,12	59,84	19,75	49,47
	Number of transitions to the open arms	17,58	36,86	33,63	47,04	15,10	36,94
	Time spent on the closed arms [s]	15,23	15,38	21,73	21,78	10,63	16,10
	Number of transitions to the closed arms	33,29	22,50	11,27	16,56	13,77	14,58
	Ratio of the time spent on the opens arms to the time spent on the closed arms	37,88	53,13	53,03	87,50	32,76	71,43
	Number of protected stretch attend	20,00	22,38	33,33	30,73	28,57	22,70
Resting metabolic rate	Oxygen consumption [ml/h/kg]	14,10	10,00	14,73	11,69	14,37	11,99
Performance in opening a 2-step lockbox	Total interaction time with the lockbox [s]	142,83	57,92	94,59	51,71	65,54	53,80

Test	Variable	Conventional housing		Super-environmental enrichment		Lockbox enrichment	
		Before enrichment phase	After enrichment phase	Before enrichment phase	After enrichment phase	Before enrichment phase	After enrichment phase
Oxygen consumption during the first 30-min lockbox trial in the calorimetric apparatus	Oxygen consumption [ml/h/kg]	13,85	8,46	11,88	10,88	8,29	8,24
Analysis of stress hormone metabolite concentrations	Fecal corticosterone metabolites concentrations [ng/0.05g feces]	25,93	38,33	31,17	40,38	34,32	43,95

including the CH group, were provided with more space, than required by law[3]. In addition to nesting material, the "conventional housing" cages were equipped with a shelter and a handling tunnel. This may not accurately reflect the conditions of "conventional housing" in all laboratory animal facilities, which are often more impoverished, but was necessary for the purpose of our study. Otherwise, the differentiation between effects attributed to space and those attributed to enrichment would have been unfeasible.

Neither lockbox nor super-environmental enrichment affected state anxiety-related behavior and physiological variables

The influence of enrichment on the phenotype of laboratory animals is subject to critical discussion because researchers are concerned that even minor changes to the cage environment may impact the data they are generating[65]. In light of this, the absence of differences between our study groups in the Open Field and the Elevated Plus Maze Test, bodyweight, resting metabolic rate, stress hormone metabolite concentrations, and adrenal gland weights may alleviate these concerns. We could only detect two minor differences between the groups (i.e., defecation in the Open Field Test and number of protected stretch attend in the Elevated Plus Maz Test) that are negligible. Differences in defecation appeared to be attributed to baseline variations among the groups and the analysis of the number of protected stretch attend revealed group differences, but post-hoc pairwise comparisons did not identify any significant differences between groups.

Our findings are consistent with Bailoo *et al.* who also did not observe any effect of environmental enrichment on the time spent on the open arms of the Elevated Plus Maze or the time spent in the center of the Open Field in female C57BL/6JRj mice[6]. Additionally, this research group did not find differences in the FCM concentrations or adrenal gland weights between female C57BL/6JRj mice housed under various types of enrichment, ranging from barren to SEE[6]. In our study, all cages, regardless of the group, were equipped with transparent handling tunnels, which can also be considered as an enrichment item. Previous research has demonstrated that the provision of a transparent handling tunnel in a cage does not affect the Open Field Test, Elevated Plus Maze Test, plasma corticosterone concentrations, or adrenal gland weights in female C57BL/6J mice[66]. However, the impact of enrichment on anxiety-related behavior and physiological stress variables of mice may depend on factors such as strain, sex, or age. Contradictory results were obtained in male C3H/eB mice, with environmental enrichment increasing the time spent in the center of the Open Field and elevating corticosterone concentrations[67]. Moreover, environmental enrichment elevated the time spent in the center of the Open Field in male HAB mice[68] and the time spent in the open arms of the Elevated Plus Maze in wild type littermates of 5-HT_{1A}R KO mice on a C57BL/6 background (sex unspecified)[69]. In another study with a relatively small sample size, female C57BL/6J mice appeared to spend more time in the center of the Open Field when they were housed in larger enriched (n = 7) compared to smaller impoverished cages (n = 10)[70]. Dickson *et al.* reported that female C57BL/6J mice housed in groups in a larger, enriched cage spent more time in the closed quadrants in the Elevated Zero Mazes compared to animals individually housed in a small, impoverished cage[71]. In the latter study, it can be assumed that space and social structure played critical roles in addition to physical/inanimate enrichment items. In contrast, in our study we ensured that space and social structure were consistent across all study groups.

The method of FCM analysis used in our study has previously revealed a correlation between elevated FCM levels and a range of stressors[72–76]. The zenith of FCM concentrations typically occurs 8–10 hours post-exposure to a stressor, contingent upon the intestinal transit time[50]. To mitigate the potential impact of the circadian rhythm on FCM excretion[49], all fecal samples excreted over a 24-hour period were collected. The FCMs concentrations measured after the 2-month enrichment phase suggested that the type of enrichment did not significantly affect the hypothalamic-pituitary-adrenal (HPA) axis response during the calorimetric measurement, which was associated with various stress factors for the mice (i.e., social isolation, novel environment, cold stress due to lack of nesting material and enrichment items).

To ensure that potential changes in bodyweight measured during the experimental period were due to the different types of enrichment, all groups received an equal amount of oat flakes. While female C57BL/6JCrl mice (age approximately 2 months to 2 years) housed in a large semi-natural enclosure were previously shown to be heavier than those housed under CH conditions, the animals kept in cages of the same size as the CH group and provided with additional environmental enrichment items did not exhibit significant difference from the CH group[77]. The scientific literature reports varying effects of enrichment on bodyweight, ranging from no effect to an increasing effect. For instance, it was demonstrated that environmental enrichment increased the bodyweight in female C57BL/6JCrl mice (age approximately 8–16 weeks)[4] or in male BALB/c mice (age approximately 8–18 weeks)[78]. This phenomenon may be elucidated by the correlation of bodyweight with the retroperitoneal adipose tissue and bone density[77]. Moreover, it could be argued that heightened physical exercise in enriched, more complex cages could result in a greater muscle weight. However, this hypothesis could previously not be substantiated, potentially due to the advanced age of the animals (approximately 2 years) in the study by Mieske *et al.*[77]. In contrast, no effect of environmental enrichment on the bodyweight was found in female and male C57BL/6NTac mice (age approximately 21 weeks)[79], in C57BL/6JOlaHsd mice, and BALB/cOlaHsd mice (age approximately 5–8 weeks)[80]. Additionally, cognitive tasks were shown to have a decreasing effect on the body weight in young male C57BL/6JCrl mice compared to their counterparts that were not engaged in the learning paradigms[81].

These discrepancies between studies may be attributed to the different types of enrichment provided and the usage of the items by the animals. Moreover, it must be noted that the amount of nesting material also plays a crucial role in bodyweight changes. Large amount of nesting material can lead to a reduction in thermoregulation activity, consequently lowering energy expenditure[77]. A previous study showed that an increased amount of nesting material maintained a more positive energy balance and resulted in greater bodyweight in C57BL/6NCrl mice[82]. Olsson and Dahlborn summarized in a review that mice tended to be heavier and ingested less food when provided with nesting material[5].

In contrast to the present study demonstrating no effect of the different types of enrichment on the resting metabolic rate, environmental enrichment was previously shown to decrease the RMR in female C57BL/6JCrl[77]. The varying study outcomes regarding the RMR may be attributed to age differences among the mice (i.e., mice examined in the study conducted by Mieske *et al.* were 20–22 months old).

Lockbox enrichment decreased trait anxiety-related behavior

Our study demonstrated that trait anxiety-related behavior was decreased by LE over time in the Grid Exploratory Paradigm. Moreover, both LE and CH reduced trait anxiety-related behavior in comparison to SEE in the Free Exploratory Paradigm. Trait anxiety-related behavior describes an enduring personal trait stemming from past experiences of temporary fear[83–85]. Additionally, trait anxiety can also be influenced by genetic factors and breeding conditions[86]. Both test paradigms, the Grid Exploratory Paradigm and the Free Exploratory Paradigm, were based on the voluntary participation of the mice, i.e., the mice voluntarily climbed on the ladder or entered the testing arena. In contrast, in the Open Field or Elevated Plus Maze Test, animals are typically compelled to participate by transferring to the maze. Since tests for trait anxiety-related behavior are generally not conducted as standard, any impact on this variable should not raise concerns among researchers. Instead, it can be viewed as a benefit for animal welfare, potentially positively influencing the affective state of the mice.

In the Grid Exploratory Paradigm, it was notable that CH decreased the total exploration time over time (i.e., when comparing values before and after the enrichment phase), contrasting with the LE and SEE, which seemed to sustain high levels of exploration. Although the latency to explore in the Grid Exploratory Paradigm was initially lower in the SEE group compared to the CH and LE groups before the enrichment phase, these differences ceased to exist after the enrichment phase. Instead, a difference between CH and LE emerged, with LE decreasing the latency to explore, which, however, should not be over-interpreted due to group differences before the enrichment phase. Additionally, only LE decreased the latency to explore over time, suggesting that LE decreased trait-anxiety-related behavior. The effect of LE on exploration may possibly be attributed to the novelty factor introduced by the lockboxes. A study by Bohn *et al.* gave slight hints that mice may prefer complex environments with enrichment items being regularly exchanged over, depending on the study week, complex environments with the same enrichment items and a non-enriched environment[87]. Although LE is not comparable with environmental enrichment items providing complexity, such as houses, lying boards, or other climbing elements, lockboxes provide novel (positive) stimuli to the mice, when regularly exchanged. This may influence the expectation and consequently the behavior of the mice in tests reliant on voluntary participation, where mice have the option to either explore or refrain from exploring the testing paradigm. Since we could not analyze home cage activity and cage preference due to technical malfunctions of the MoPSS devices in our study, it remains unclear whether LE may have increased these variables, potentially supporting the results found.

In the Free Exploratory Paradigm, further interesting observations were made regarding the latency to explore. The latency to explore was reduced in the LE and CH groups compared to the SEE group. This difference may be due to the latter group being accustomed to a higher complexity and presence of hiding places in their home environment, which were absent in the testing arena. In contrast, animals from the LE and CH groups were used to less environmental enrichment items (i.e., house, tunnel, and nesting material only).

Lockbox enrichment improved sequential problem-solving with lockboxes

When a 2-step lockbox was introduced to the mice following the calorimetric measurement, those from the LE group exhibited the best performance after the enrichment phase; specifically, all mice from this group successfully solved the lockbox after the 2-month enrichment phase and the number of mice able to do so increased over time in the LE group only. Additionally, mice from the LE group required the least amount of time to open it, although significance level was not achieved. This enhanced performance is likely attributable to the lockbox training, which facilitated manipulative and cognitive activities. However, the oxygen consumption rate during the lockbox trials did not differ between the study groups.

It could be hypothesized that mice from the LE group experienced less neophobia and more intrinsic interest when faced with the novel 2-step lockbox due to their familiarity with the 3D-printed items and the knowledge that a food reward was hidden inside. However, our data allowed us to reject this hypothesis because the total interaction time with the lockbox did not vary between the study groups. Furthermore, the lockbox, in both states (i.e., open and baited with an oat flake, as well as locked), was initially introduced to mice from all groups in the measurement cage to familiarize them with this new item. Therefore, it can be assumed that all mice were aware of the presence of the oat flake hidden within the lockbox.

Earlier studies have already demonstrated that C57BL/6 mice housed under enriched conditions, which include various stimulating items such as tubes, a running wheel, toys, and treats, exhibited enhanced spatial learning in the Morris Water Maze Test[88]. Moreover, daily transfers to an enriched environment (three hours per day over two months), which provided houses, running wheels, and toys[89] or additionally tunnels, platforms, and pictures[90], have been shown to improve recognition memory in a novel-object recognition-memory task and emotional memory in a contextual fear-conditioning task in another mouse strain[89,90]. Given these findings, it is conceivable that LE, which encourages manipulative and cognitive activities, could similarly enhance the cognitive performance of mice. In the aforementioned studies, toys provided to the mice in enriched environments served as stimuli. If these toys were baited with treats and required specific actions to access the food, they would function on a similar principle to lockboxes.

For future studies, it would be of interest to investigate whether LE also has the potential to enhance the performance of mice in other learning and memory tasks.

Lockbox enrichment decreased relative variability in a higher number of variables

Independent of the type of enrichment, the relative variability increased in most variables over time, i.e., values of measured parameters were higher after compared to before the enrichment phase. However, for some variables, the coefficient of variation did not increase but decreased. This trend was most often observed with LE and least often with CH, suggesting that neither lockbox nor SEE had a stronger impact on phenotypic variation. This is in line with a systematic review, demonstrating that there were no differences in the coefficient of variation across variables between environmental enrichment and control groups[91]. Even a more complex, semi-natural environment does not increase the relative variability[77]. In a comprehensive study, the provision of a shelter and nesting material had no consistent effect on the relative variability of various physiological parameters, leading to the conclusion that the impact was negligible[79]. These observations were confirmed in further studies[6,92–94]. It must also be considered that the mice of our study were tested at different developmental stages in life, i.e., in late adolescence and adulthood, which may have also contributed to differences in the relative variability measured before and after the enrichment phase.

Repeatability analysis revealed impact of personality on data variances

To determine the primary contributor to variability in our data, we performed a repeatability analysis across the three days of testing for distance travelled measured in the Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test. In all three tests, the variances observed for locomotor activity were primarily explained by the individual animal, while group and batch did not significantly contribute to the observed variances. Our findings indicated that the individual personality of the mice played a critical role, as previously discussed as an important factor in explaining variances of data measured in mice[60,95,96]. This underscored the significance of repeated testing, as explaining the variability observed in the data is a crucial step in addressing the reproducibility crisis[60]. The emphasis on the personality of laboratory animals also opens up new avenues for enhancing animal welfare and identifying opportunities to improve the well-being of mice on an individual level[97].

Habituation response in Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test

The Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test were conducted over three consecutive days to habituate the animals to the test situation, thereby circumventing an increase in variance due to novelty. A habituation response to the testing apparatus is characterized by lower levels of exploration across the days of testing. This procedure is considered to improve the reliability of the measurements[60].

In our study, the distance travelled in the Open Field Test decreased over the three days before but increased after the enrichment phase, suggesting a habituation response at the first time point of testing. The data generated at the second time point leave uncertainty regarding whether habituation failed or additional factors were accountable. When the mice were tested for the first and second time, they were at an

age of 7–8 weeks and 17–18 weeks, respectively. It is interesting to note that other studies repeating the Open Field Test utilized mice aged no older than 10 weeks. In these studies, it was reported that locomotor activity in the Open Field Test decreased over the days of testing in male C57BL/6JCrl mice at the age of 6–7[60] and female C57BL/6JRj mice at the age of approximately 10 weeks[6]. Additionally, Bailoo *et al.* demonstrated a decrease in locomotor activity in the Open Field Test over testing days in C57BL/6 mice aged approximately 9–10 weeks[98]. Hence, the different observations made before and after the enrichment phase may be attributed to the age of the animals. A similar observation was made in the Elevated Plus Maze Test. The distance travelled did not alter before but increased over the three days after the mice were exposed to the enrichment.

In the Free Exploratory Paradigm, the distance travelled in the testing arena increased over the three consecutive days on both time points, i.e., before and after the enrichment phase. In contrast to the Open Field Test, the mice can voluntarily choose to enter and explore the testing arena in this paradigm. Therefore, habituation is associated with a higher locomotor activity in this test situation.

The decreased coefficient of variation of distance travelled on day 3 in the Free Exploratory Paradigm and the Open Field Test suggested a successful habitation in these tests. In contrast, in the Elevated Plus Maze Test, the coefficient of variation did not decrease over the three days, which may indicate that the animals did not habituate to this testing apparatus.

On the one hand, habituation can reduce data variability in certain behavioral paradigms, thereby improving data quality and the reproducibility of research results. However, on the other hand, habituation to a test situation can influence exploration and, consequently, anxiety-related behavior in the tests discussed above. Therefore, reduced exploration due to increasing familiarity is also an expected outcome in some behavioral assays, such as the novel-context recognition test in the Open Field[99]. In this context, repeated exposure to the Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test may limit the interpretability of the results.

Implementation of lockbox enrichment in laboratory animal facilities

All lockboxes are available as open-source tools, enabling any establishment to use and modify them freely. When enrichment strategies at establishments are reviewed and updated, the implementation of lockboxes may be considered. A feasible implementation approach tailored to the specific circumstances of each establishment must be developed. An advantage of lockboxes over other enrichment strategies that facilitate cognitive and/or manipulative activities outside the home cage, such as mazes, is that they can be presented to the mice in their home cages. It is also possible to connect a playpen[100], including one or more lockboxes, with the home cage. However, it is crucial that the mice are given "a degree of control and choice over their environment to reduce stress-induced behavior"[3]. If short individual lockbox sessions in the home cage or playpen, similar to the temporary lockbox enrichment introduced above, are not feasible, an alternative approach could be to provide several lockboxes to the animal group (e.g., a platform with several single mechanisms). These mechanisms can be rebaited or changed during the daily inspection, similar to the permanent lockbox enrichment described above. We are also working on an automated lockbox. If the box automatically locks the mechanisms and rebaits itself after being solved, it would benefit both the animals and the caretakers.

In the future, lockboxes may also serve as a novel approach in behavioral neuroscience, as discussed by Lang *et al.*[18]. Initial results from the mice's performance during the lockbox training sessions indicated that lockboxes may be beneficial for investigating sequential decision-making and motor learning[34].

To validate the safety of the PLA material used for printing the lockboxes, a pathohistological evaluation was conducted at the end of the study, revealing no evidence of toxicopathological effects associated with PLA. As a result, exposure to lockboxes made from PLA during the temporary enrichment phase over two months (i.e., lockbox training sessions in the morning) and permanent LE phase over one month (i.e., 24 hours/day before sacrificing) did not cause any harmful effects. However, in some cases, such as studies conducted under "Good Laboratory Practice" (GLP) conditions, specific material requirements must be met. Moreover, the dimensions of the lockboxes may not be suitable for all cage sizes. In such scenarios, researcher or companies can utilize and modify our freely available lockboxes designs[38] to construct lockboxes in other sizes, forms, or materials.

One challenge encountered when introducing lockboxes into animal facilities is adhering to hygiene restrictions. When lockboxes are printed outside the barrier and then transported inside, they must undergo disinfection. In our study, we disinfected the lockboxes using 70% ethanol and allowed them to dry overnight. Autoclaving PLA is not feasible due to its sensitivity to heat. Instead, alternatives such as peracetic acid or hydrogen peroxide airlocks could be considered. Before implementation, compatibility tests should be conducted on small PLA samples to ensure no adverse effects occur.

Personal subjective observation

Before the lockbox training sessions, the doors of the gate system were closed to install the lockbox for the first mouse to be trained. During this preparation time, all four mice usually attempted to access cage B and consequently queued inside or in front of the tube. The same behavior was observed when a mouse was in cage B working on the lockbox; the other mice tried to get through the door of the gate. This observation was only made as a side note and was not systematically documented. However, it may suggest that the mice were eager to gain access to cage B and subsequently

the LE and/or the food reward. Unfortunately, we could not analyze home cage activity and cage preference due to technical malfunctions of the MoPSS devices in batch 2 and batch 3. Therefore, it remains unclear whether the mice may have developed a place preference for cage B as a consequence of the LE and must be investigated in future studies.

Conclusion

All lockboxes are available as open-source tool, allowing everyone to use them as enrichment strategy or as a novel approach in behavioral neuroscience. We demonstrated that the lockboxes served as an effective enrichment strategy facilitating manipulative and cognitive activities. The utilization of LE positively affected the affective state and improved sequential problem-solving with novel lockboxes in female C57BL/6JCrl mice. Our findings demonstrated that implementing LE, much like other enrichment strategies, will neither influence the relevant phenotype of the mice nor exacerbate the reproducibility crisis.

Ethics and consent

Research involving animals was performed according to the guidelines of the German Animal Welfare Act and the Directive 2010/63/EU for the protection of animals used for scientific purposes. Approval for animal experimentation and maintenance of the animals was granted by the Berlin State Authority (referred to as "Landesamt für Gesundheit und Soziales"; permit number: G0249/19; date of approval: February 14, 2020). The pre-registration of the study can be found in the animal study registry (animalstudyregistry.org, doi: 10.17590/asr.0000237).

Data availability

Zenodo: Data and ARRIVE 2.0 checklist for the original article "Lockbox enrichment facilitates manipulative and cognitive activities for mice". https://zenodo.org/doi/10.5281/zenodo.10907380[101].

An xlsx file containing data for all tests presented in this article and the "ARRIVE guidelines 2.0: author checklist" are provided.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Zenodo: Mouse Lockboxes - 3D printing files and videos (v1.0.0). https://doi.org/10.5281/zenodo.11143666[38].

STL files of the lockboxes as well as door systems and information on 3D printing details are given. Moreover, videos of the mice solving the lockboxes are provided.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Zenodo: Extended data of the article "Lockbox enrichment facilitates manipulative and cognitive activities for mice": Supplementary Figure. https://doi.org/10.5281/zenodo.13269372[61].

Figure and results of the distance traveled in the Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test during habituation are shown.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements

We thank the animal care takers for their diligent care of the mice, as well as Clara Bekemeier and Leonie-Iva Pawlak for their invaluable support during the study. This article is based upon work from COST Action "Improving biomedical research by automated behaviour monitoring in the animal home-cage" (TEATIME; CA20135; cost-teatime.org) supported by COST (European Cooperation in Science and Technology). We acknowledge the use of ChatGPT for proofreading and grammar checking of some sentences in the manuscript, as well as for refining the code used to create the figures.

References

1. Mieske P, Hobbiesiefken U, Fischer-Tenhagen C, *et al.*: **Bored at home?-A systematic review on the effect of environmental enrichment on the welfare of laboratory rats and mice.** *Front Vet Sci.* 2022; **9**: 899219.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Cait J, Cait A, Scott RW, *et al.*: **Conventional laboratory housing increases morbidity and mortality in research rodents: results of a meta-analysis.** *BMC Biol.* 2022; **20**(1): 15.
PubMed Abstract | Publisher Full Text | Free Full Text

3. **Directive 2010/63/EU of the European Parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes.** Official Journal of the European Union ABI. L276/33-L276/79, 2010.
Reference Source

4. Hobbiesiefken U, Mieske P, Lewejohann L, *et al.*: **Evaluation of different types of enrichment - their usage and effect on home cage behavior in female mice.** *PLoS One.* 2021; **16**(12): e0261876.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Olsson IA, Dahlborn K: **Improving housing conditions for laboratory mice: a review of "environmental enrichment".** *Lab Anim.* 2002; **36**(3): 243–70.
PubMed Abstract | Publisher Full Text

6. Bailoo JD, Murphy E, Boada-Saña M, *et al.*: **Effects of cage enrichment on behavior, welfare and outcome variability in female mice.** *Front Behav Neurosci.* 2018; **12**(2): 232.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Mieske P, Diederich K, Lewejohann L: **Roaming in a land of milk and honey: life trajectories and metabolic rate of female inbred mice living in a semi naturalistic environment.** *Animals (Basel).* 2021; **11**(10): 3002.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Gui L, Luo Z, Shan W, *et al.*: **Role of Sox2 in learning, memory, and postoperative cognitive dysfunction in mice.** *Cells.* 2021; **10**(4): 727.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Robison LS, Francis N, Popescu DL, *et al.*: **Environmental Enrichment:**

disentangling the influence of novelty, social, and physical activity on Cerebral Amyloid Angiopathy in a transgenic mouse model. *Int J Mol Sci.* 2020; **21**(3): 843.
PubMed Abstract | Publisher Full Text | Free Full Text

10. Leidinger C, Herrmann F, Thöne-Reineke C, *et al.*: **Introducing clicker training as a cognitive enrichment for laboratory mice.** *J Vis Exp.* 2017; (121): 55415.
PubMed Abstract | Free Full Text

11. Krakenberg V, Wewer M, Palme R, *et al.*: **Regular touchscreen training affects Faecal Corticosterone Metabolites and anxiety-like behaviour in mice.** *Behav Brain Res.* 2021; **401**: 113080.
PubMed Abstract | Publisher Full Text

12. Quante SM, Siewert V, Palme R, *et al.*: **The power of a touch: regular touchscreen training but not its termination affects hormones and behavior in mice.** *Front Behav Neurosci.* 2023; **17**: 1112780.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Krakenberg V, Siestrup S, Palme R, *et al.*: **Effects of different social experiences on emotional state in mice.** *Sci Rep.* 2020; **10**(1): 15255.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Steventon JJ, Harrison DJ, Trueman RC, *et al.*: ***In vivo* MRI evidence that neuropathology is attenuated by cognitive enrichment in the Yac128 Huntington's Disease mouse model.** *J Huntingtons Dis.* 2015; **4**(2): 149–60.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Martinez-Coria H, Yeung ST, Ager RR, *et al.*: **Repeated cognitive stimulation alleviates memory impairments in an Alzheimer's Disease mouse model.** *Brain Res Bull.* 2015; **117**: 10–5.
PubMed Abstract | Publisher Full Text

16. Burnham JC: **Thorndike's puzzle boxes.** *J Hist Behav Sci.* 1972; **8**(2): 159–67.
PubMed Abstract | Publisher Full Text

17. Auersperg AMI, Kacelnik A, von Bayern AMP: **Explorative learning and functional inferences on a five-step means-means-end problem in Goffin's cockatoos (*Cacatuagoffini*).** *PLoS One.* 2013; **8**(7): e68979.
PubMed Abstract | Publisher Full Text | Free Full Text

18. Lang B, Kahnau P, Hohlbaum K, *et al.*: **Challenges and advanced concepts for the assessment of learning and memory function in mice.** *Front Behav Neurosci.* 2023; **17**: 1230082.
PubMed Abstract | Publisher Full Text | Free Full Text

19. Auersperg AMI, von Bayern AM, Gajdon GK, *et al.*: **Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm.** *PLoS One.* 2011; **6**(6): e20231.
PubMed Abstract | Publisher Full Text | Free Full Text

20. Epstein R, Kirshnit CE, Lanza RP, *et al.*: **'Insight' in the pigeon: antecedents and determinants of an intelligent performance.** *Nature.* 1984; **308**(5954): 61–2.
PubMed Abstract | Publisher Full Text

21. Huber L, Rechberger S, Taborsky M: **Social learning affects object exploration and manipulation in keas, *Nestor notabilis*.** *Anim Behav.* 2001; **62**(5): 945–54.
Publisher Full Text

22. Suwandschieff E, Huber L, Bugnyar T, *et al.*: **Kea, bird of versatility. Kea parrots (*Nestor notabilis*) show high behavioural flexibility in solving a demonstrated sequence task.** *J Ornithol.* 2024; **165**(1): 49–55.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Miyata H, Gajdon GK, Huber L, *et al.*: **How do keas (*Nestor notabilis*) solve artificial-fruit problems with multiple locks?** *Anim Cogn.* 2011; **14**(1): 45–58.
PubMed Abstract | Publisher Full Text

24. Hunt GR: **Manufacture and use of hook-tools by New Caledonian crows.** *Nature.* 1996; **379**: 249–51.
Publisher Full Text

25. Taylor AH, Hunt GR, Holzhaider JC, *et al.*: **Spontaneous metatool use by New Caledonian crows.** *Curr Biol.* 2007; **17**(17): 1504–7.
PubMed Abstract | Publisher Full Text

26. Bird CD, Emery NJ: **Insightful problem solving and creative tool modification by captive nontool-using rooks.** *Proc Natl Acad Sci U S A.* 2009; **106**(25): 10370–5.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Wimpenny JH, Weir AAS, Clayton L, *et al.*: **Cognitive processes associated with sequential tool use in New Caledonian crows.** *PLoS One.* 2009; **4**(8): e6471.
PubMed Abstract | Publisher Full Text | Free Full Text

28. Cheke LG, Bird CD, Clayton NS: **Tool-use and instrumental learning in the Eurasian jay (*Garrulus glandarius*).** *Anim Cogn.* 2011; **14**(3): 441–55.
PubMed Abstract | Publisher Full Text

29. Whiten A, Custance DM, Gomez JC, *et al.*: **Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*).** *J Comp Psychol.* 1996; **110**(1): 3–14.
PubMed Abstract | Publisher Full Text

30. Carvalho S, Cunha E, Sousa C, *et al.*: **Chaînes opératoires and resource-exploitation strategies in chimpanzee (*Pan troglodytes*) nut cracking.** *J Hum Evol.* 2008; **55**(1): 148–63.
PubMed Abstract | Publisher Full Text

31. Carvalho S, Biro D, McGrew WC, *et al.*: **Tool-composite reuse in wild chimpanzees (*Pan troglodytes*): archaeologically invisible steps in the technological evolution of early hominins?** *Anim Cogn.* 2009; **12 Suppl 1**: S103–14.
PubMed Abstract | Publisher Full Text

32. Vrbanec L, Matijević V, Guenther A: **Enhanced problem-solving ability as an adaptation to urban environments in house mice.** *Proc Biol Sci.* 2021; **288**(1945): 20202504.
PubMed Abstract | Publisher Full Text | Free Full Text

33. Mazza V, Guenther A: **City mice and country mice: innovative problem solving in rural and urban noncommensal rodents.** *Anim Behav.* 2021; **172**: 197–210.
Publisher Full Text

34. Boon MN, Andresen N, Traverso S, *et al.*: **Mechanical problem solving in mice.** *bioRxiv.* 2024.
Publisher Full Text

35. FELASA working group on revision of guidelines for health monitoring of rodents and rabbits, Mähler Convenor M, Berard M, *et al.*: **FELASA recommendations for the Health Monitoring of mouse, rat, hamster, guinea pig and rabbit colonies in breeding and experimental units?** *Lab Anim.* 2014; **48**(3): 178–92.
PubMed Abstract | Publisher Full Text

36. Habedank A, Urmersbach B, Kahnau P, *et al.*: **O mouse, where art thou? The Mouse Position Surveillance System (MoPSS)-an RFID-based tracking system.** *Behav Res Methods.* 2022; **54**(2): 676–89.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Habedank A, Urmersbach B, Kahnau P, *et al.*: **O mouse, where art thou? The Mouse Position Surveillance System (MoPSS) - an RFID based tracking system - supplementary material.** [Data set]. *Zenodo.* 2021.
Publisher Full Text

38. Hohlbaum K, Andresen N, Boon MN, *et al.*: **Mouse Lockboxes - 3D printing files and videos (v1.0.0).** [Dataset]. Zenodo. 2024.
http://www.doi.org/10.5281/zenodo.11143666

39. Kahnau P, Habedank A, Schwarck C, *et al.*: **Protocol for voluntary tunnel handling of mice.** (last accessed on 3 March, 2024).
Reference Source

40. Mathis A, Mamidanna P, Cury KM, *et al.*: **DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.** *Nat Neurosci.* 2018; **21**(9): 1281–9.
PubMed Abstract | Publisher Full Text

41. Nath T, Mathis A, Chen AC, *et al.*: **Using DeepLabCut for 3D markerless pose estimation across species and behaviors.** *Nat Protoc.* 2019; **14**(7): 2152–76.
PubMed Abstract | Publisher Full Text

42. Sturman O, von Ziegler L, Schläppi C, *et al.*: **Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions.** *Neuropsychopharmacology.* 2020; **45**(11): 1942–52.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Friard O, Gamba M: **BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations.** *Methods Ecol Evol.* 2016; **7**(11): 1325–30.
Publisher Full Text

44. He K, Zhang X, Ren S, *et al.*: **Deep residual learning for image recognition.** In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2016; 770–778.
Publisher Full Text

45. Insafutdinov E, Pishchulin L, Andres B, *et al.*: **Deepercut: a deeper, stronger, and faster multi-person pose estimation model.** In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VI 14.* Springer International Publishing, 2016; **9910**: 34–50.
Publisher Full Text

46. Nickerson DM, Facey DE, Grossman GD: **Estimating physiological thresholds with continuous two-phase regression.** *Physiol Zool.* 1989; **62**(4): 866–87.
Publisher Full Text

47. Garner J, Gaskill B, Rodda C, *et al.*: **The mouse ethogram.** (last accessed on 3 March, 2024).
Reference Source

48. Palme R, Touma C, Arias N, *et al.*: **Steroid extraction: get the best out of faecal samples.** *Wien Tierarztl Monat.* 2013; **100**: 238–46.
Reference Source

49. Touma C, Palme R, Sachser N: **Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones.** *Horm Behav.* 2004; **45**(1): 10–22.
PubMed Abstract | Publisher Full Text

50. Touma C, Sachser N, Möstl 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**(3): 267–78.
PubMed Abstract | Publisher Full Text

51. Bert B, Schmidt N, Voigt JP, *et al.*: **Evaluation of cage leaving behaviour in rats as a free choice paradigm.** *J Pharmacol Toxicol Methods.* 2013; **68**(2): 240–9.
PubMed Abstract | Publisher Full Text

52. R Core Team: **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria, 2022.
Reference Source

53. Bates D, Mächler M, Bolker B, *et al.*: **Fitting linear mixed-effects models using lme4.** *J Stat Softw.* 2015; **67**(1): 1–48.
Publisher Full Text

54. Kuznetsova A, Brockhoff PB, Christensen RHB: **lmerTest package: tests in**

linear mixed effects models. *J Stat Softw.* 2017; **82**(13): 1–26.
Publisher Full Text

55. Wickham H: **ggplot2: elegant graphics for data analysis.** Springer-Verlag New York, 2016.
Reference Source

56. Xiao N: **ggsci: scientific journal and Sci-Fi themed color palettes for 'ggplot2'.** R package version 3.0.1, 2024.
Reference Source

57. Kassambara A: **ggpubr: 'ggplot2' based publication ready plots.** R package version 0.6.0, 2023.
Reference Source

58. Lenth R: **emmeans: estimated marginal means, aka least-squares means.** R package version 1.10.0, 2024.
Reference Source

59. Fox J, Weisberg S, Price B: **car: companion to applied regression.** R package version. 3.1-2, 2023.
Reference Source

60. Rudeck J, Vogl S, Banneke S, *et al.*: **Repeatability analysis improves the reliability of behavioral data.** *PLoS One.* 2020; **15**(4): e0230900.
PubMed Abstract | Publisher Full Text | Free Full Text

61. Hohlbaum K, Andresen N, Mieske P, *et al.*: **Extended data of the article "Lockbox enrichment facilitates manipulative and cognitive activities for mice": Supplementary Figure**. [Extended data]. Zenodo. 2024.
http://www.doi.org/10.5281/zenodo.13269372

62. Nakagawa S, Schielzeth H: **Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists.** *Biol Rev Camb Philos Soc.* 2010; **85**(4): 935–56.
PubMed Abstract | Publisher Full Text

63. Stoeffel M, Nakagawa S, Schielzeth H: **An introduction to repeatability estimation with rptR.** 2018; [updated 2018/01/042018/06/12].
Reference Source

64. Landis JR, Koch GG: **The measurement of observer agreement for categorical data.** *Biometrics.* 1977; **33**(1): 159–74.
PubMed Abstract | Publisher Full Text

65. Toth LA: **The influence of the cage environment on rodent physiology and behavior: implications for reproducibility of pre-clinical rodent research.** *Exp Neurol.* 2015; **270**: 72–7.
PubMed Abstract | Publisher Full Text

66. Oatess TL, Harrison FE, Himmel LE, *et al.*: **Effects of acrylic tunnel enrichment on anxiety-like behavior, neurogenesis, and physiology of C57BL/6J mice.** *J Am Assoc Lab Anim Sci.* 2021; **60**(1): 44–53.
PubMed Abstract | Publisher Full Text | Free Full Text

67. Benaroya-Milshtein N, Hollander N, Apter A, *et al.*: **Environmental Enrichment in mice decreases anxiety, attenuates stress responses and enhances Natural Killer cell activity.** *Eur J Neurosci.* 2004; **20**(5): 1341–7.
PubMed Abstract | Publisher Full Text

68. Sah A, Rooney S, Kharitonova M, *et al.*: **Enriched environment attenuates enhanced trait anxiety in association with normalization of aberrant neuro-inflammatory events.** *Int J Mol Sci.* 2022; **23**(21): 13052.
PubMed Abstract | Publisher Full Text | Free Full Text

69. Rogers J, Vo U, Buret LS, *et al.*: **Dissociating the therapeutic effects of Environmental Enrichment and exercise in a mouse model of anxiety with cognitive impairment.** *Transl Psychiatry.* 2016; **6**(4): e794.
PubMed Abstract | Publisher Full Text | Free Full Text

70. Hüttenrauch M, Salinas G, Wirths O: **Effects of long-term Environmental Enrichment on anxiety, memory, hippocampal plasticity and overall brain gene expression in C57BL6 mice.** *Front Mol Neurosci.* 2016; **9**: 62.
PubMed Abstract | Publisher Full Text | Free Full Text

71. Dickson PE, Mittleman G: **Environmental Enrichment influences novelty reactivity, novelty preference, and anxiety via distinct genetic mechanisms in C57BL/6J and DBA/2J mice.** *Sci Rep.* 2021; **11**(1): 3928.
PubMed Abstract | Publisher Full Text | Free Full Text

72. Atanasov NA, Sargent JL, Parmigiani JP, *et al.*: **Characterization of train-induced vibration and its effect on Fecal Corticosterone Metabolites in mice.** *J Am Assoc Lab Anim Sci.* 2015; **54**(6): 737–44.
PubMed Abstract | Free Full Text

73. Voigt CC, Klöckner P, Touma C, *et al.*: **Hormonal stress response of laboratory mice to conventional and minimally invasive bleeding techniques.** *Animal Welfare.* 2013; **22**(4): 449–55.
Publisher Full Text

74. Walker MK, Boberg JR, Walsh MT, *et al.*: **A less stressful alternative to oral gavage for pharmacological and toxicological studies in mice.** *Toxicol Appl Pharmacol.* 2012; **260**(1): 65–9.
PubMed Abstract | Publisher Full Text | Free Full Text

75. Hohlbaum K, Bert B, Dietze S, *et al.*: **Impact of repeated anesthesia with Ketamine and Xylazine on the well-being of C57BL/6JRj mice.** *PLoS One.* 2018; **13**(9): e0203559.
PubMed Abstract | Publisher Full Text | Free Full Text

76. Hohlbaum K, Leidinger C, Palme R, *et al.*: **Castration of adult male C57BL/6JRj mice allows for resocialization and social housing of previously single-housed males: a harm-benefit analysis.** *Berliner und Münchener Tierärztliche Wochenschrift.* 2020; **133**.
Publisher Full Text

77. Mieske P, Scheinpflug J, Yorgan TA, *et al.*: **Effects of more natural housing conditions on the muscular and skeletal characteristics of female C57BL/6J mice.** *Lab Anim Res.* 2023; **39**(1): 9.
PubMed Abstract | Publisher Full Text | Free Full Text

78. Gurfein BT, Stamm AW, Bacchetti P, *et al.*: **The calm mouse: an animal model of stress reduction.** *Mol Med.* 2012; **18**(1): 606–17.
PubMed Abstract | Publisher Full Text | Free Full Text

79. André V, Gau C, Scheideler A, *et al.*: **Laboratory mouse housing conditions can be improved using common environmental enrichment without compromising data.** *PLoS Biol.* 2018; **16**(4): e2005019.
PubMed Abstract | Publisher Full Text | Free Full Text

80. Wirz A, Mandillo S, D'Amato FR, *et al.*: **Response, use and habituation to a mouse house in C57BL/6J and BALB/c mice.** *Exp Anim.* 2015; **64**(3): 281–93.
PubMed Abstract | Publisher Full Text | Free Full Text

81. Kahnau P, Guenther A, Boon MN, *et al.*: **Lifetime observation of cognition and physiological parameters in male mice.** *Front Behav Neurosci.* 2021; **15**: 709775.
PubMed Abstract | Publisher Full Text | Free Full Text

82. Johnson JS, Taylor DJ, Green AR, *et al.*: **Effects of nesting material on energy homeostasis in BALB/cAnNCrl, C57BL/6NCrl, and Crl:CD1(ICR) mice housed at 20 °C.** *J Am Assoc Lab Anim Sci.* 2017; **56**(3): 254–9.
PubMed Abstract | Free Full Text

83. Belzung C, Berton F: **Further pharmacological validation of the BALB/c neophobia in the free exploratory paradigm as an animal model of trait anxiety.** *Behav Pharmacol.* 1997; **8**(6–7): 541–8.
PubMed Abstract | Publisher Full Text

84. Reiss S: **Trait anxiety: it's not what you think it is.** *J Anxiety Disord.* 1997; **11**(2): 201–14.
PubMed Abstract | Publisher Full Text

85. Lister RG: **Ethologically-based animal models of anxiety disorders.** *Pharmacol Ther.* 1990; **46**(3): 321–40.
PubMed Abstract | Publisher Full Text

86. Rex A, Sondern U, Voigt JP, *et al.*: **Strain differences in fear-motivated behavior of rats.** *Pharmacol Biochem Behav.* 1996; **54**(1): 107–11.
PubMed Abstract | Publisher Full Text

87. Bohn L, Bierbaum L, Kästner N, *et al.*: **Structural enrichment for laboratory mice: exploring the effects of novelty and complexity.** *Front Vet Sci.* 2023; **10**: 1207332.
PubMed Abstract | Publisher Full Text | Free Full Text

88. Kempermann G, Kuhn HG, Gage FH: **More hippocampal neurons in adult mice living in an enriched environment.** *Nature.* 1997; **386**(6624): 493–5.
PubMed Abstract | Publisher Full Text

89. Rampon C, Tang YP, Goodhouse J, *et al.*: **Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice.** *Nat Neurosci.* 2000; **3**(3): 238–44.
PubMed Abstract | Publisher Full Text

90. Tang YP, Wang H, Feng R, *et al.*: **Differential effects of enrichment on learning and memory function in NR2B transgenic mice.** *Neuropharmacology.* 2001; **41**(6): 779–90.
PubMed Abstract | Publisher Full Text

91. Kentner AC, Speno AV, Doucette J, *et al.*: **The contribution of Environmental Enrichment to phenotypic variation in mice and rats.** *eNeuro.* 2021; **8**(2): ENEURO.0539-20.2021.
PubMed Abstract | Publisher Full Text | Free Full Text

92. Vogt MA, Mertens S, Serba S, *et al.*: **The 'Cage Climber' - a new enrichment for use in large-dimensioned mouse facilities.** *Appl Anim Behav Sci.* 2020; **230**: 105078.
Publisher Full Text

93. Lewejohann L, Reinhard C, Schrewe A, *et al.*: **Environmental bias? Effects of housing conditions, laboratory environment and experimenter on behavioral tests.** *Genes Brain Behav.* 2006; **5**(1): 64–72.
PubMed Abstract | Publisher Full Text

94. Wolfer DP, Litvin O, Morf S, *et al.*: **Laboratory animal welfare: cage enrichment and mouse behaviour.** *Nature.* 2004; **432**(7019): 821–2.
PubMed Abstract | Publisher Full Text

95. Lewejohann L, Zipser B, Sachser N: **"Personality" in laboratory mice used for biomedical research: a way of understanding variability?** *Dev Psychobiol.* 2011; **53**(6): 624–30.
PubMed Abstract | Publisher Full Text

96. Roche DG, Careau V, Binning SA: **Demystifying animal 'personality' (or not): why individual variation matters to experimental biologists.** *J Exp Biol.* 2016; **219**(Pt 24): 3832–43.
PubMed Abstract | Publisher Full Text

97. Richter SH, Hintze S: **From the individual to the population–and back again? Emphasising the role of the individual in animal welfare science.** *Appl Anim Behav Sci.* 2019; **212**: 1–8.
Publisher Full Text

98. Bailoo JD, Jordan RL, Garza XJ, *et al.*: **Brief and long periods of maternal separation affect maternal behavior and offspring behavioral development in C57BL/6 mice.** *Dev Psychobiol.* 2014; **56**(4): 674–85.
PubMed Abstract | Publisher Full Text

99. Wagatsuma A, Okuyama T, Sun C, *et al.*: **Locus Coeruleus input to hippocampal CA3 drives single-trial learning of a novel context.** *Proc Natl Acad Sci U S A.* 2018; **115**(2): E310–E316.
PubMed Abstract | Publisher Full Text | Free Full Text

100. Ratuski AS, Makowska IJ, Dvorack KR, *et al.*: **Using approach latency and anticipatory behaviour to assess whether voluntary playpen access is rewarding to laboratory mice.** *Sci Rep.* 2021; **11**(1): 18683.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

101. Hohlbaum K, Andresen N, Mieske P, *et al.*: **Data and ARRIVE 2.0 checklist for the original article "Lockbox enrichment facilitates manipulative and cognitive activities for mice".** [Dataset]. Zenodo. 2024.
http://www.doi.org/10.5281/zenodo.10926604

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 09 September 2024

https://doi.org/10.21956/openreseurope.19899.r43585

© **2024 Cushman J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

✔ **Jesse D Cushman**
National Institutes of Health, Durham, NC, USA

The authors have substantially improved the manuscript and I have no further concerns. I look forward to seeing their future work on this subject.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: behavioral neuroscience, hippocampus-dependent behavior and physiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 05 August 2024

https://doi.org/10.21956/openreseurope.19043.r41985

© **2024 Cushman J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

? **Jesse D Cushman**
[1] National Institutes of Health, Durham, NC, USA

[2] National Institutes of Health, Durham, NC, USA

In this manuscript examines the impact of an enrichment approach that involves repeated experience with lock boxes on behavioral, physiological and metabolic endpoints. The goal was to assess whether this could be a viable option for enrichment that could improve the overall wellbeing of the mice and what impact it might have on reproducibility. Overall, I think that the lock-boxes are an incredibly creative solution and the fact that they've made them publicly available so that anyone can print and construct them is a huge contribution to the field. The videos showing the mice interacting with and solving the lock-boxes are a great resource as well. The additional resources for gates and various other tools are very useful, too. The study was conducted with a clear effort for rigor and reproducibility. Unfortunately, there are several aspects of how the study was conducted, analyzed, presented, and interpreted that raise some concerns for me. I will outline those below. I want to emphasize that I appreciate the large amount of work that was put into this and I think there is a great deal of value to the scientific community that this effort has produced. As outlined in my final comment, I think a greater emphasis should be placed on the contribution to the field that will come from the open source tools themselves.

Major concern:

The within-subjects design for tasks employed is highly problematic. The exploration-based tasks such as open field, grid exploration and elevated plus maze are not designed to be run in a within-subjects repeated measures design. Exploration decreases with increasing familiarity with the environment. Reduction in exploration with increasing familiarity plays a huge role in behavior and is a direct part of many assays, such as Tonegawa's "novel context recognition" [2] and object recognition assays. Familiarity greatly alters the "anxiety-like" measures of open arm time and visits in the elevated plus maze. As the environment becomes more familiar and the exploratory drive diminishes this leads to less need to explore the more threatening open arms (see Rogers and Dalvi, 1997, for example)[Ref 1]. The results in this study confirm all of this. Figure 7a, for example, shows the expected reduction in distance travelled that would occur after averaging together the first three days of exposure to an open field vs. the second three (novel context recognition). Figure 8C,D and E shows the expected reduction in open arm transitions that would be expected to occur in a familiar plus maze. A better design would have been to conduct this in a between-subjects design since all three groups were run. Little was gained from a within-subjects design and a lot was lost. The data comparing the effects of the different housing conditions is unfortunately rather uninterpretable because of this. The authors need to address the limitations of repeat testing due to the within-subject design limiting the interpretability of the findings. In the discussion there is reference to the "novelty factor", but clearly this concept needed to have been a part of the initial study design planning.

In addition, there are some measures where there were differences prior to the enrichment conditions, which should not be the case, and makes any interpretations highly problematic. For example, the conclusion that the lock-box enrichment reduces trait anxiety, which I think refers to the reduced latency to explore the grid, shown in the right side of Fig 5B. The differences in the groups before enrichment make this difficult to interpret, which is problematic because this reduction is highlighted as a major finding in the Abstract.

Minor:

The descriptions of the statistical approaches are overly complex. I think it would be better to outline the overall statistical strategy, the decision work-flow that is being employed of how or why one analysis approach was chosen over another and then provide a table or overall

summary. For most of the sections the analysis approach was the same and does not need to be repeated multiple times. Although the statistical approach appears to have been intended to be quite rigorous, the variability in how each task was analyzed relative to the others creates an impression of inconsistency. Please provide more detail on the overall strategy and less detail on each analysis performed. The details could be better placed in a table. The write up of the stats also makes reading the results difficult. It is hard to read through the statistical details in the parenthesis and still follow what is being said. This could also be helped with a stats table. Despite all this detail I found it hard to understand what the n's were for each analysis, which did vary due to various technical malfunctions. This could be included in the stats table to make it more clear.
I found the lines in the figures indicating statistical significance a bit difficult to follow. Perhaps details could be put in the figure legend?
It is unfortunate that males could not be included. What is the plan to integrate the analysis of the males in the future. Will sex differences be examine directly in any way?
Figure 6 doesn't really help show anything useful (other than the clear reductions in exploration that occur with increasing familiarity). Data needs to be separated by group to be useful.
Stating that training with the lockboxes improved sequential problem solving might be misleading, as the training involved the same task. Wording it this way makes it sound like a different cognitive assay was run and the lock box enrichment improved performance on a different task. The results indicate that training and experience with lock boxes improved performance with solving lock boxes. This should be made more clear.

General question: How would lock boxes be implemented at a vivarium level? They would need to be re-baited and each mouse would need the highly labor intensive individualized exposure that was used here. Would experimenters or animal techs re-bait the lock boxes? Could the lock boxes be designed to be repeatable somehow? It's not clear to me how these lock boxes (as amazing as they are) could really be implemented at scale.
The final conclusion of the paper is quite conservative and, I think, defensible given the data: "*Our findings demonstrated that implementing LE, much like other enrichment strategies, will not exacerbate the reproducibility crisis.*" To my mind, however, what I find most exciting about the paper is the open source tools it makes available to the community. I would have liked to see more emphasis on solving the lock boxes as a task in its own right. I could see this as being very useful in cognitive assessments in many modelling situations. Perhaps the emphasis of the paper could highlight this more? Otherwise, I think the greatest strength of the paper – the introduction of an innovative open source tool, is rather lost in large amount of data that is difficult to interpret.

References

1. Rodgers RJ, Dalvi A: Anxiety, defence and the elevated plus-maze.*Neurosci Biobehav Rev*. 1997; **21** (6): 801-10 PubMed Abstract | Publisher Full Text
2. Wagatsuma A, Okuyama T, Sun C, Smith LM, et al.: Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context.*Proc Natl Acad Sci U S A*. 2018; **115** (2): E310-E316 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: behavioral neuroscience, hippocampus-dependent behavior and physiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 15 Aug 2024

Katharina Hohlbaum

We thank the reviewer for taking the time to read and comment on our work, which greatly enhanced the value of our manuscript. Below we answered point-by-point to all major and minor concerns:

Major concerns:

1.**Within-subjects design:** There is a significant reproducibility crisis in research, and it is imperative for all researchers to consider ways to enhance data reliability. One strategy is to identify methods to reduce data variability. A high variability in the data can obscure group/treatment effects, potentially leading to a misinterpretation of the results. If animals are habituated to the test situation and the variability decreases, the data will be more reliable, which was previously proposed by Rudeck at al. (2020). Therefore, the Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test were conducted over three consecutive days. Table 2 shows that the coefficient of variation decreased in the Free Exploratory Paradigm and Open Field Test over the three days, demonstrating that repeated testing in these two paradigms was an efficient strategy to reduce data variability. However, we recognize the reviewer's criticism about repeated testing, which should be taken into account. We added an additional paragraph to the discussion section on "Habituation response in Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test" and pointed out that the repeated testing in the Free Exploratory Paradigm, Open Field Test, and Elevated Plus Maze Test can limit the interpretability of the results. Rudeck J, Vogl S, Banneke S, *et al.*: Repeatability analysis improves the reliability of

behavioral data. *PLoS One.* 2020; **15**(4): e0230900.

2.**Baseline values:** We hope that the reviewer acknowledged that we examined and reported “baseline” values (i.e., before enrichment phase). Unfortunately, this is still not standard in all animal behavior studies. We would like to address a possible misunderstanding regarding the statement that “there are some measures where there were differences prior to the enrichment conditions”. There is only one variable with significant baseline differences between the groups, which was transparently pointed out in the results section on “Latency to explore” and in the discussion section on “Lockbox enrichment decreased trait anxiety-related behavior". In brief, mice assigned to CH and LE showed a higher latency to explore compared to mice assigned to SEE before the enrichment phase. After the enrichment phase, these differences ceased to exist and only a difference between CH and LE was detected (i.e., the two groups that did not differ at baseline). When comparing the latency to explore before and after the enrichment phase within the groups, the values significantly decreased in the LE group only. In response to the reviewer's concerns, we will focus our analysis on the within-group comparison of the Grid Exploratory Paradigm to draw the conclusion that lockbox enrichment reduces trait anxiety-related behavior. Additionally, this conclusion is further supported by the results from the Free Exploratory Paradigm. The abstract and discussion section on “Lockbox enrichment decreased trait anxiety-related behavior” were revised accordingly.

Minor concerns:
1. Statistical analysis: As suggested by the reviewer, more general information on the “overall statistical strategy” was added to the paragraph on “Statistical analysis” and an additional table providing an overview of the statistical strategy (Table 1) was created. Table 1 provides details on the statistical analysis of each variable. If other factors than described in the section on statistical analysis influenced the choice of model, the rationale is given in Table 1. In the scope of transparency and reproducibility, the authors would prefer not to remove any further details on the statistical analysis form the manuscript. For researchers not familiar with LMM/GLMM it may not be clear that it is important to compare model fits and chose the most suitable model for the dataset. Depending on the fixed/random factors, distribution of data/residuals, and model fit comparisons, it may occur that not the same model can be used for analyzing the different datasets of the study. This may be the reason why the reviewer had the impression that “the variability in how each task was analyzed relative to the others creates an impression of inconsistency”. If the reviewer is more specific about this criticism and suggests how to improve the statistical analysis, the authors will be happy to discuss his suggestions.

2. Symbols indicating significant differences in figures: We are sorry that the meaning of lines and asterisks were not clear. To provide further clarification, we added an explanation to the legends of figure 5 and 8: An asterisk (*) indicates a significant difference between two groups before or after the enrichment phase, or a significant difference within a group before and after the enrichment phase (details are given in the results section).

3.Male animals: As explained in the manuscript, the study planning included both sexes but so far were have only been able to examine females due to the time-consuming nature of the study. We emphasized this limitation in the article: “The study planning included both

sexes. Due to the time-consuming nature of the study, thus far, we have only been able to examine female animals, which must be considered as limitation of the present article."

4.Figure 6: As suggested by the other reviewer, Figure 6 was moved to the supplementary material: Hohlbaum, K., Andresen, N., Mieske, P., et al.: Extended data of the article "Lockbox enrichment facilitates manipulative and cognitive activities for mice": Supplementary Figure [Extended data]. Zenodo. 2024. https://doi.org/10.5281/zenodo.13269372https://doi.org/10.5281/zenodo.13269372 Data were not separated by groups in this figure since differences in the distance traveled between the days were of interest. To compare the data from the present study with those generated by Rudeck et al. (2020), the same statistical analysis was applied.

5. Wording "sequential problem solving with lockboxes": We apologize for the misleading wording in the abstract and conclusion. After reading the discussion section of the manuscript, it should be clear that the authors were referring to sequential problem-solving with lockboxes. To clarify this, the corresponding sentences in the abstract and conclusion were rephrased.
However, the authors also would like to point out that the training of the LE group did not involve the same lockbox as used during the calorimetric measurement when a lockbox was presented to mice of all groups (i.e., CH, SEE, and LE). Mice of the LE group were trained on the lockboxes shown in Figure 3a and 3b. At the end of the calorimetric measurement, a novel lockbox was presented to the mice (Figure 3c).

6. General question on implementation of lockboxes a vivarium level: First, it is essential to recognize our moral responsibility and duty to continuously refine the experimental and housing conditions of laboratory animals. This obligation applies to everyone working with laboratory animals. The Directive 2010/63/EU stipulates that the enrichment strategies in establishments shall be regularly reviewed and updated. We can consider various strategies for implementing lockboxes in everyday work of an animal facility. However, since the strategies depend on several factors (such as space, housing conditions, hygiene levels, etc.), each establishment must develop a feasible approach tailored to their specific circumstances. As described in the discussion section, lockboxes can either be added to the home cages or playpens which are connected to the home cage. If short individual lockbox sessions in the home cage, similar to the temporary lockbox enrichment introduced in this manuscript, are not feasible, an alternative could be to provide several single mechanisms in the group cage, which can be rebaited or changed during the daily inspection, similar to the permanent lockbox enrichment described here. Our lab is also working on an automated lockbox. If the box automatically locks the mechanisms and is rebaited after being solved, it would benefit both the animals and the caretakers. We added these thoughts to the discussion section on "Implementation of lockbox enrichment in laboratory animal facilities".

7.Conclusion of the paper: The present manuscript is part of an extensive study divided into several substudies. In the substudy presented in this article, we focused on phenotyping and animal welfare; another substudy analyzes the problem-solving strategies of mice in lockbox sessions using computer vision methods, which may be of particular interest to researchers in the field of behavioral neuroscience. The preprint of this substudy

can be found here (Boon et al. 2024): https://doi.org/10.1101/2024.07.29.605658. We agree that lockbox tasks can be useful in cognitive assessment, as highlighted in a review by Lang et al. (2023). However, when introducing lockboxes as a novel type of enrichment or task for cognitive assessment, understanding their impact on the phenotype of mice is crucial. Discussions on the implementation of enrichment strategies often include concerns that data variability will increase and/or that the animals' phenotype will be significantly affected. Consequently, animals that would naturally live in complex habitats are still housed in small, barren cages, which negatively impacts their well-being. Therefore, the data presented in this manuscript are of utmost importance to this discussion and need to be shared with the scientific community in their current detail. We hope this will contribute to improve the welfare of laboratory mice in the future. According to your advice, we highlighted the points mentioned above (i.e., open-source tool, novel approach in behavioral neuroscience) in the abstract, discussion, and conclusion. Marcus N. Boon, et al. Mechanical problem solving in mice. bioRxiv 2024.07.29.605658; doi: https://doi.org/10.1101/2024.07.29.605658. Lang B, Kahnau P, Hohlbaum K, *et al.*: Challenges and advanced concepts for the assessment of learning and memory function in mice. *Front Behav Neurosci.* 2023; **17**: 1230082.

Competing Interests: No competing interests were disclosed.

Reviewer Report 12 July 2024

https://doi.org/10.21956/openreseurope.19043.r41692

© 2024 Bains R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rasneer Sonia Bains
[1] Mary Lyon Centre at MRC Harwell, Oxfordshire OX11 0RD, UK
[2] Mary Lyon Centre at MRC Harwell, Oxfordshire OX11 0RD, UK

The manuscript is well thought out, and the experimental design, analysis and reporting has been done in line with best practices and reproducibility in mind. I congratulate the authors on this exemplary effort. If I may, I would like to recommend a few minor changes that may help the readers along as it is a very comprehensive manuscript.

1. Of the two main conclusions: adding enrichment won't alter the phenotype and that the lock box enrichment facilitates manipulative and cognitive activities, only the latter is clear from the outset. It would be good to state both of these clearly at the beginning.
2. To improve the legibility of the Box and Whisker plots, may I suggest using a single legend explain the colours, it will really help the readers.
3. Figure 6 could reasonably be moved to supplementary data.
4. When describing the results the authors refer to comparisons that are not shown in the figure, just before mentioning the figure, this is a bit confusing, e.g. "**Total interaction time with the lockbox:** While the group did not affect the interaction time with the

lockboxes ($F_{2, 6}$ = 0.17, p = 0.844), time had a significant effect ($F_{1, 35}$ = 22.58, p < 0.001) in the LMM (Figure 9c). The mice interacted for a longer time duration with the lockbox after compared to before the enrichment phase." Please either show these data or move "(Figure9c)" to the place in the text where it relates to the data shown. This is also true of most text references to Figure 8.

5. Legend for Figure 5, please replace "und" with "and" :)
6. When describing the results from DLC the authors mention that data from some animals may not be included, would it be possible to say how many animals per dataset were not included?
7. Thank you for sharing the interrater reliability data. It is refreshing to see such attention to best practices.
8. The two statements below are a bit confusion, please could you consider rephrasing them:
    1. **Total interaction time with the lockbox:** While the group did not affect the interaction time with the lockboxes ($F_{2, 6}$ = 0.17, p = 0.844), time had a significant effect ($F_{1, 35}$ = 22.58, p < 0.001) in the LMM (Figure 9c). The mice interacted for a longer time duration with the lockbox after compared to before the enrichment phase.
    2. **Interaction time with the lockbox until solved:** Among the mice that were able to solve the lockbox, time significantly affected the interaction time with the lockbox until it was solved ($F_{1, 19.35}$ = 12.67, p = 0.002) in the LMM (Figure 9d). Group had no significant effect ($F_{2, 26.38}$ = 0.53, p = 0.594). The interaction time with the lockbox until solved was decreased after compared to before the enrichment phase.
9. It is not entirely clear from statement "The FCM values were higher after the exposure to the different types of enrichment compared to baseline." on Page 26 if this is back up with stats, please could you rephrase.

While the authors explain why the data only relates to one sex, it would have been good to see the male data. I take it on good faith that the authors will be publishing these data soon. I very much look forward to reading that manuscript too.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and does the work have academic merit?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Behavioural Phenotyping. Experimental design and Analysis.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 15 Aug 2024

Katharina Hohlbaum

We thank the reviewer for taking the time to read and evaluate our article and for appreciating our work. The comments provided were highly constructive and have helped to improve the manuscript significantly. Below we answered point-by-point to all minor concerns.

1. **Stating main conclusions clearly at the beginning:** You are absolutely right. We revised the abstract, conclusion, and first paragraph of the discussion section to emphasize this aspect.
2. **Color legends in figures:** The figures were revised accordingly.
3. **Figure 6:** Figure 6 was moved to the supplementary data: Hohlbaum, K., Andresen, N., Mieske, P., et al.: Extended data of the article "Lockbox enrichment facilitates manipulative and cognitive activities for mice": Supplementary Figure [Extended data]. Zenodo. 2024. https://doi.org/10.5281/zenodo.13269372
4. **Order of figure citations:** This does seem to be confusing. As suggested, the "figure names" were moved to locations in the text where they more closely correspond to the data being presented.
5. **Typo:** Thanks for spotting this typo :)
6. **Numbers of animals:** The number of animals per dataset is provided in the figure and table legends. To make this information more accessible to the reader, we have created an additional table (Table 1) that includes the number of animals per dataset, along with an overview of the statistical analyses. The section on the statistical analysis had to be shortened due to concerns of the other reviewer. He stated that the “descriptions of the statistical approaches are overly complex” and suggested creating this table.
7. **Interrater reliability:** Thanks a lot for appreciating!
8. **Rephrasing:**
   1. In the LMM, group had no significant effect ($F2, 6 = 0.17$, $p = 0.844$) (Figure 9c) but time had a significant effect on the interaction time with the lockboxes ($F1, 35 = 22.58$, $p < 0.001$). The mice spent more time interacting with the lockboxes after the enrichment phase compared to before it.
   2. Among the mice that were able to solve the lockbox, group did not significantly affect the interaction time with the lockbox until it was solved in the LMM ($F_{2, 26.38} = 0.53$, $p = 0.594$) (Figure 9d). Time had a significant effect ($F_{1, 19.35} = 12.67$, $p = 0.002$). The interaction time with the lockbox until it was solved decreased after the enrichment phase compared to before it.
9. **Rephrasing:** The sentence referred to the finding that the fixed factor “time” significantly affected the values ($F_{1, 35} = 32.89$, $p < 0.001$). The sentence was rephrased

to clarify its meaning: “Independent of the group, the FCM values were higher after the exposure to the different types of enrichment compared to baseline.”

Competing Interests: No competing interests were disclosed.