Check for updates

METHOD ARTICLE

REVISED Development of an IntelliCage-based cognitive bias test for mice

[version 2; peer review: 2 approved]

Pia Kahnau[1], Anne Jaap[1], Birk Urmersbach[1], Kai Diederich[1], Lars Lewejohann[1,2]

[1]Laboratory Animal Science, German Federal Institute for Risk Assessment, Berlin, 10589, Germany
[2]Insitute of Animal Welfare, Animal Behavior and Laboratory Animal Science, Freie Universität Berlin, Berlin, 14163, Germany

v2 **First published:** 25 Nov 2022, **2**:128
https://doi.org/10.12688/openreseurope.15294.1
Latest published: 08 Aug 2023, **2**:128
https://doi.org/10.12688/openreseurope.15294.2

Abstract

The cognitive bias test is used to measure the emotional state of animals with regard to future expectations. Thus, the test offers a unique possibility to assess animal welfare with regard to housing and testing conditions of laboratory animals. So far, however, performing such a test is time-consuming and requires the presence of an experimenter. Therefore, we developed an automated and home-cage based cognitive bias test based on the IntelliCage system. We present several developmental steps to improve the experimental design leading to a successful measurement of cognitive bias in group-housed female C57BL/6J mice. The automated and home-cage based test design allows to obtain individual data from group-housed mice, to test the mice in their familiar environment, and during their active phase. By connecting the test-cage to the home-cage via a gating system, the mice participated in the test on a self-chosen schedule, indicating high motivation to actively participate in the experiment. We propose that this should have a positive effect on the animals themselves as well as on the data. Unexpectedly, the mice showed an optimistic cognitive bias after enrichment was removed and additional restraining. An optimistic expectation of the future as a consequence of worsening environmental conditions, however, can also be interpreted as an active coping strategy in which a potential profit is sought to be maximized through a higher willingness to take risks.

Keywords

cognitive bias, judgment bias, home-cage based, IntelliCage, conditioning, learning behavior, mice

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 08 Aug 2023	✓ view	✓ view
version 1 25 Nov 2022	? view	? view

1. **David P. Wolfer**, University of Zurich, Zurich, Switzerland
ETH Zurich, Zurich, Switzerland

2. **Ewelina Knapska**, Nencki Institute of Experimental Biology, Warsaw, Poland

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



This article is included in the Neuroscience gateway.



This article is included in the Psychology gateway.



This article is included in the COST Actions gateway.

Corresponding author: Pia Kahnau (pia.kahnau@bfr.bund.de)

Author roles: Kahnau P: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Jaap A**: Conceptualization, Writing – Review & Editing; **Urmersbach B**: Writing – Review & Editing; **Diederich K**: Project Administration, Writing – Review & Editing; **Lewejohann L**: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was financially supported by the European Union through the COST Action grant CA20135 and by the European Union's Horizon 2020 research and innovation programme under the grant agreement No. 861952. This work was funded by the DFG (FOR 2591; LE 2356/5-1). The authors declare no competing interests.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Kahnau P *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: Kahnau P, Jaap A, Urmersbach B *et al.* **Development of an IntelliCage-based cognitive bias test for mice [version 2; peer review: 2 approved]** Open Research Europe 2023, **2**:128 https://doi.org/10.12688/openreseurope.15294.2

First published: 25 Nov 2022, **2**:128 https://doi.org/10.12688/openreseurope.15294.1

REVISED **Amendments from Version 1**

In addition to minor linguistic and orthographic improvements, some major changes have been made to the manuscript. In the introduction, a scientific article was added as a reference, which also describes a cognitive bias test in the IntelliCage. The methods have been expanded to include a description of the habituation steps to the home-cage based set-up. It was also explained in more detail why only females were used, and that it is necessary to test this set-up as well as the experiment with male mice. The discussion of developmental step 2 was expanded to include information on the IntelliCage entries (the data for this can be found in the supplements). This information is important to better describe learning and non-learning behavior. In the discussion of developmental step 3, the statement "cognitive bias could be manipulated" was weakened and described and discussed in more detail.

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

Introduction

It has been shown that in both humans and animals, past experiences influence future expectations (Harding *et al.*, 2004; Mendl *et al.*, 2009; Paul *et al.*, 2005). Individuals with negative experiences or in a bad mood are more likely to be "pessimistic" about future events and, vice versa, individuals with positive experiences or in a good mood are more likely to be "optimistic". In the past, many tests for various species have been developed to investigate the emotional state of animals (Jirkof *et al.*, 2019). To examine the influence of emotional or affective states on expectations of future events, a number of cognitive bias tests have been developed (Boleij *et al.*, 2012; Harding *et al.*, 2004; Hintze *et al.*, 2018; Schlüns *et al.*, 2017; Verbeek *et al.*, 2014).

The common feature of these tests is the need for conditioning the subjects to scalable stimuli, *e.g.*, odors, tones, or spatial positions. The animals learn that they will receive a reward for the stimulus at one end of the scale and a punishment for the second stimulus at the other end of the scale. After successful conditioning, the actual test follows, in which ambiguous stimuli are presented to the animals. These ambiguous stimuli are located on the scale between the already known stimuli. The reaction towards these ambiguous stimuli is then measured and analyzed: It is assumed that if the response to the ambiguous stimulus is similar to the positively conditioned stimulus, the animals seem to expect a reward. In this case, they had a positive expectation of the future event, or in other words, they appear to be "optimistic". However, if the response resembles the response of the negatively conditioned stimulus, the animals seem to have a negative expectation or seem to be "pessimistic".

The first cognitive bias test was developed by Harding and colleagues in 2004 (Harding *et al.*, 2004). Rats were conditioned to press a lever in response to hearing the positively-associated tone-frequency to receive a reward or not to press a lever to avoid a punishment after hearing the negatively-associated tone-frequency. The cognitive bias test revealed that rats kept under unpredictable housing conditions were less likely to press the lever for a reward in response to ambiguous tone-frequencies than rats kept under normal housing conditions. It was thus concluded that the negative experience rendered them 'pessimistic'.

Although mice are the most commonly used experimental animals (Lewejohann *et al.*, 2020), it took eight years before the first results of a cognitive bias test for mice were published (Boleij *et al.*, 2012). Boleij and colleagues conditioned mice to various odor stimuli, which predicted either a palatable or an unpalatable food reward. First, it was shown that BALB/cJ mice were able to discriminate between odor stimuli, whereas 129P3/J mice were not. Second, it was shown that BALB/cJ mice tested under more aversive white light conditions had a higher latency in response to the ambiguous stimulus than mice tested under less aversive red-light conditions.

Further cognitive bias test methods followed in which mice were conditioned to spatial positions (Bailoo *et al.*, 2018; Kloke *et al.*, 2014; Novak *et al.*, 2015; Verjat *et al.*, 2021), to tactile stimuli (Novak *et al.*, 2016), to different tunnel lengths (Krakenberg *et al.*, 2019), to auditory stimuli (Jones *et al.*, 2017), to light stimuli (Alboni *et al.*, 2017), to olfactory stimuli (Resasco *et al.*, 2021), or in an automated touchscreen-based set-up presenting different patterns on a screen (Krakenberg *et al.*, 2019). Most studies showed that mice could be conditioned to the different stimuli and that the data plotted on the axis of stimuli increasing from negative to positive result in a sigmoidal curve (increasing s-shape slopes from the negative to the positive stimulus). These sigmoidal curves indicate that ambiguous stimuli are perceived differently compared to the conditioned stimuli, which is an important criterion for the validity of cognitive bias tests (Gygax, 2014; Hintze *et al.*, 2018; Krakenberg *et al.*, 2019).

In most of the studies, it is necessary for both the conditioning and the test itself to remove the mice from their home-cages and manually place them in the respective test set-ups. As a consequence, the animals have to be handled, taken out of their familiar environment, separated from their group members (if kept in groups) and forced to participate in the test irrespective of their current state of motivation. In fact, this may have a negative effect on the animals' state of mind during the conditioning phase and as a result the cognitive bias test might also be influenced. This implies that in order to minimize external influence on the cognitive bias, the best handling method has to be chosen (*e.g.*, known influence of tail handling compared to cup and tunnel handling on anxiety-like behavior (Hurst & West, 2010) and that the animals have to be very well-habituated to the test set-ups. Nevertheless, even with the best handling and habituation, a possibly negative influence of the separation from the home-cage and/or the group (Krohn *et al.*, 2006; Manouze *et al.*, 2019) as well as the experimenter's immediate influence on the mice, and thereby the test results, must be taken into account. Home-cage based experiments are a good way to address this shortcoming and are becoming

increasingly common (see https://www.cost-teatime.org/ for more on home-based cage-based systems + (Kahnau *et al.*, 2023)).

For example, Alboni and colleagues used the IntelliCage system (TSE-Systems, Germany) to perform a cognitive bias test on male mice (Alboni *et al.*, 2017).

The IntelliCage is a home-cage based test system that allows automated data acquisition, which can improve the reproducibility of the data (reviewed in Voikar & Gaburro, 2020). Depending on size and weight of the animals, it is possible to keep up to 16 mice in the IntelliCage as one social group. The IntelliCage has four conditioning corners, each containing an radio frequency identification (RFID) antenna. Since the mice also carry an RFID transponder, it is possible to obtain individual data from mice living in groups and to study activity and learning behavior in social groups (Endo *et al.*, 2011; Kahnau *et al.*, 2021; Krackow *et al.*, 2010; Voikar *et al.*, 2018).

In the study of Alboni and colleagues, different male mice groups were conditioned to two lights (blue and yellow) that were associated with either a reward or a punishment within the IntelliCage. As an ambiguous stimuli both lights were presented simultaneously. When presented with this ambiguous condition a nosepoke scored as optimistic behavior and no nosepoke scored as pessimistic behavior. The results showed that mice treated with the anti-depressant fluoxetine appeared to be more optimistic in an enriched environment than mice not treated with fluoxetine. In this study, several mice were housed together in one (or more) IntelliCages. The system allows individual data to be collected using RFID sensors, but mutual interference between the mice cannot be ruled out. It is possible that individual conditioned corners were preferred and therefore visited more frequently and/or that mice pushed each other out of corners to gain access, preventing individual mice from performing their task undisturbed. In addition only one ambiguous condition (both lights) could be presented sparing near positive and near negative ambiguous stimuli.

To overcome this shortcoming, we added a gate to the IntelliCage, the AnimalGate (AnimalGate, TSE-Systems, Germany), in order to test the mice individually and undisturbed by group members. Our test set-up consisted of a home-cage, a gate and an IntelliCage (test-cage). Through the gate, it is possible to separate the mice and let them individually enter the test-cage. This is especially important to allow all individuals within the group to be conditioned and tested without disturbance by group members. Another advantage is that the mice can individually decide when to enter the test-cage and participate in the experiment, rather than being coerced by an experimenter-imposed schedule. As a result, the influence of the experimenter and the influence on the wake/sleep rhythm is reduced to a minimum, except for daily visual inspection and weekly cleaning of the set-up. It has already been shown that rats and mice can independently transfer themselves from their home-cages to test-cages individually to perform tasks within test-cages (Kahnau *et al.*, 2022A; Kaupert *et al.*, 2017; Mei *et al.*, 2020; Rivalan *et al.*, 2017; Winter & Schaefers, 2011). A slight disadvantage is that since only one mouse can be within the test-cage at a time and other motivated mice have to wait until this mouse has left the test-cage. However, we could show in a recent experiment with a comparable set-up that no single mouse was constantly blocking others from getting access (Kahnau *et al.*, 2022A).

Within our automated and home-cage based test set-up, we conditioned female C57BL/6J mice to different tones. De Hoz and Nelken as well as Francis and colleagues already showed that mice were able to differentiate between different tones (De Hoz & Nelken, 2014; Francis & Kanold, 2017). Here, we present our different developmental steps and results of the cognitive bias tests. Our first hypothesis was that it is possible to condition mice within the IntelliCage based set-up and that the cognitive bias is influenced by the removal of enrichment and by repeated restraining. Here we present the individual developmental steps of our automated and home-cage based cognitive bias test, which were based on each other and the optimizations we implemented through previous experience. We show that it is possible to successfully condition mice in a relatively short time and measure the cognitive bias of mice, with minimal intervention and time investment by the experimenter.

Methods

Animals and housing conditions

In this study, three developmental steps with three different mouse groups (one developmental step per group) are presented in which different conditioning methods are described (Table 1). As this is an exploratory study, it was not possible to estimate the effect size and power in advance. All three groups served as their own controls as before and after comparisons were made. All 36 female C57BL/6J mice were purchased from Charles River Sulzfeld, Germany. For each developmental step, the three groups consisted of 12 mice. This group size was chosen due to the size of the entire IC based set-up. It is possible to keep 12 mice in this set-up with additional enrichment and still providing enough space for exhibiting species specific behaviors. RFID technology also makes it possible to obtain individual data from mice living in a social group. All mice were four weeks old upon arrival but were bought at different time points. All efforts were undertaken to minimize animal suffering. No medical treatment was required at any time for the mice due to pain, suffering, or harm. Further details on the mouse groups are given at the respective developmental steps.

For the establishment of the home-cage based cognitive bias test, females were used exclusively since they show little agonistic behavior and can be easily kept in groups. In addition, females do not show pronounced territorial behavior that would excludes others from certain resources (Mieske *et al.*, 2021) However, it is worth noting that in a previously published study, 48 male mice were kept in groups of 12 in four ICs without such problems (Kahnau *et al.*, 2021). These mice were grouped at the age of 21 days and by the end of the long-term study (up to 2 years) there were no injury-related failures. The early grouping of the mice, but also the group size (Lidster *et al.*, 2019), seems to have a positive effect on the agonistic

Table 1. Experimental procedure. IC = IntelliCage.

	Group one	Group two	Group three
Developmental step	1	2	3
Year	2019	2019	2020
Conditioning protocol	Gate: passing the gate	Corner: visiting the IC corner	Corner: visiting the IC corner
Tone	Sequences	Frequencies	Frequencies
Tone length	6.6 sec.	0.5 and 1 sec.	2 sec.
Airpuff length	1 sec.	1 sec.	2 sec.

behavior of the male mice. However, as this was an explorative study to establish a new test design, there was the concern that individual male mice might occupy the gate and thereby prevent others from entering the test cage. This would reduce the number of visits for the other mice and possibly have a negative effect on learning behavior. To circumvent this from the beginning we used female mice in this study. However, it is important to emphasize that follow-up studies with male mice are needed. As this is the only way to prove that the method can be used regardless of sex.

We deliberately used an inbred strain to minimize genetic variability. However, despite all efforts of standardization, minimal genetic drift and varying epigenetic influences can occur during breeding. In order to randomize the factors that could not be controlled for, all mice in each experiment were born and raised by different mothers and foster mothers to ensure maximum genetic and epigenetic independence between individuals. Immediately after arrival a health inspection was performed and the mice were weighed and color-marked (edding 750, colors: black, white, red, yellow, silver) on the tail for visual identification. The mice were housed within the home-cage based set-up, and no data was recorded for the first two weeks. The day after arrival, tunnel handling training to reduce handling stress (Gouveia & Hurst, 2013; Hurst & West, 2010) was started and conducted for three weeks (see video tutorial).

One week after arrival, all mice received RFID transponders (Euro ID, FDX-B, ISO 11784/85). The evening before the transponder transplantation, an analgesic (meloxicam 1mg/kg, Meloxidyl by CEVA) was given orally by fixing the mice in the experimenter's hand, to reduce possible pain caused by implantation. The transponders were implanted under isoflurane anesthesia (induction of anesthesia: 4l/min 4%; maintenance of anesthesia: 1l/min 1–2%) subcutaneously in the neck region about 1cm behind the ears. Out of 36 mice, two mice lost their transponders by the morning after transponder implantation and the procedure had to be repeated. None of the 36 mice needed medical treatment after transponder implantation.

One week after transponder implantation, the mice moved to the housing room where also the home-cage based experiments were conducted. The room temperature and humidity were 22°C +/- 3°C and 55% +/- 15%. The light/dark cycle was set to 12/12 hours with light off at 7 pm in winter months and at 8 pm in summer because of the switch from winter to summertime. The sunrise was simulated with a wake-up light (Philips HF 3510, 100–240 vac, 50–60 Hz, Philips Consumer Lifestyle B.V. Netherlands) half an hour before the room-light was switched on. The wake-up light was placed on the ground in a corner of the housing room with the light directed towards the animals. The light intensity increased gradually and reached the full intensity at 7/8 am (depending on season). The daily visual health inspection was performed between 7/8 am to 10 am (depending on season). The home-cage set-up was cleaned once a week. Bedding, nesting material, and enrichment items were replaced. A small handful of old bedding was transferred to the new home-cage. On the same day, the mice were weighed and re-color-marked.

Home-cage based set-up

In all developmental steps, the same home-cage based set-up was used. This set-up (Figure 1) consisted of three compartments: a home-cage, a gate (AnimalGate), and a test-cage (IntelliCage, IC). As the gate had doors, an RFID reader, and infrared barriers, it was possible to allow only one mouse at a time to pass through the gate from the home-cage into the IC. This individual transfer was necessary to allow the mice to perform the tests individually and undisturbed by group members. All other mice of the social group had to wait until the one mouse within the IC moved back through the gate into the home-cage. However, getting used to the setup was a gradual process.

Step 1:

First, all doors in the gate and conditioning corners were left open at all times. This allowed all mice to move freely, explore and gain access to the water bottles. During this phase, all mice were able to remain in the test cage at the same time.

Step 2:

The center gate door and the conditioning corner doors were closed. Gate: Each time a mouse entered the gate, the door opened, the mouse could pass through the gate, and the door closed again. Then another mouse could enter the gate. During

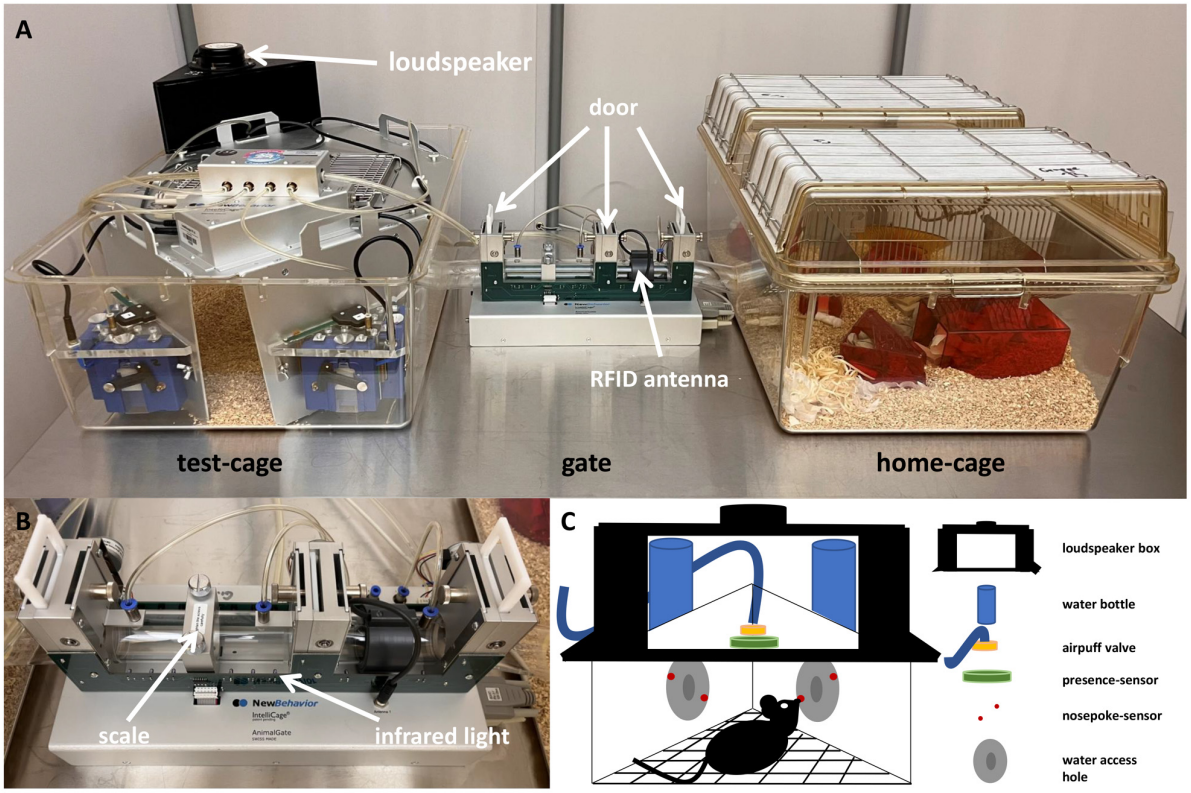

Figure 1. Home-cage based set-up based on the IntelliCage system. A: The set-up consisted of the IntelliCage used as the test-cage, which is connected through the AnimalGate to the home-cage. The IntelliCage was equipped with four conditioning corners and bedding. The home-cage was equipped with bedding, nesting, enrichment and food *ad libitum* (not shown here). The AnimalGate had three doors, one radio frequency identification (RFID) antenna. **B**: In addition, the AnimalGate had eight infrared barriers and one scale to measure the animal's weight during each gate passage. **C**: Within the IntelliCage corners, water could be provided. In addition, each corner had one radio frequency identification antenna, one presence-sensor, one airpuff-valve, two water dispensers and two doors.

this phase, all mice were able to move between the home-cage and the test-cage. IC: On entering the conditioning corners, the doors opened and access to the water bottles was free. The doors closed when the mouse left the corner.

Step 3:

Only one mouse was allowed to pass through the gate and stay in the test cage. Gate: The middle door was closed. It opened as soon as a mouse entered the gate from the home cage. At the same time, the rear (first AG door) and front (third gate door) doors closed. As soon as the mouse entered the rear part of the gate and passed the middle door, the middle door also closed and the weight of the mice was measured. After seven seconds, the front and middle doors opened and the mouse could leave the gate and enter the test cage. The first door of the gate remained closed until the mouse had left the test cage through the gate again. IC: the doors inside the conditioning corners were closed and opened when the mouse did a nosepoke on the nosepoke-senor.

The home-cage was a Makrolon type IV cage (floor space 2065 cm^2) with a filtertop equipped with 3-4 cm bedding (Poplar Granulate 2-3 mm, Altromin, Germany), two red triangle-shaped houses ("TheMouseHouse", Tecniplast, Italy), nesting material (eight papers, six paper nesting stripes and six cotton rolls), four wooden bars to chew on, and food *ad libitum* (autoclaved pellet diet, LAS QCDiet, Rod 16, Lasvendi, Germany). Within the home-cage was also an acrylic tube (4 cm diameter, 17.5 cm long), which was used for tunnel handling to reduce handling stress. Mouse group three additionally received nesting materials upon weekly changing: folded paper stripes, mid coarse wood wool and square hemp pads. Also, one resting platform and a running disk (InnoDome with InnoWheel, Bio-Serv) was placed within the home-cage and the mice received weekly changing toys filled with millet (organic peeled golden millet, Bohlsener Mühle) once per week.

The IC is a computer and RFID technology-based test system with four conditioning corners. Each corner contained an RFID antenna at the corner entrance, a presence sensor, which detected differences in temperature, two nosepoke infrared sensors, two doors through which the water access can be regulated, two water dispensers, and an airpuff valve for the possibility of a mild punishment (0.5 bar). Depending on the conditioning

method, one or four IC corners were active, in which water was provided. The IC contained only bedding material.

In order to perform experiments within the IC system, it is necessary to habituate the mice to the system first. The mice had to learn how to pass through the AnimalGate and where to access water within the IC. For this purpose, the mice were habituated gradually to the AnimalGate and IC doors. Initially, all AnimalGate and IC doors were permanently open (phase: 'all doors open'). Thus, it was possible for all mice to move freely within the system. As a next step, the doors of the AnimalGate were closed, and opened only when a mouse entered the AnimalGate, which is similar to the next IC habituation step when the corner doors were closed and opened due to a visit (phase: 'visit open doors'). In the final phase of habituation, only one mouse could stay in the IC, and the IC doors could only be opened with a nosepoke.

Conditioning protocol

The basic requirement for performing a cognitive bias test is to condition the animals to scalable stimuli. In our study, the mice were conditioned to auditory stimuli. Three different conditioning protocols were performed with each of the different mouse groups. Common to all protocols was that the mice had to learn that for one presented tone (positive tone); they received water as a reward; if they made a nosepoke within the IC corner (correct behavior). For another tone (negative tone), they received an airpuff as a punishment, if they made a nosepoke (incorrect behavior). If the mice did not make a nosepoke after hearing the positive tone (incorrect behavior), they received no water. If the mice did not make a nosepoke after hearing the negative tone (correct behavior), they did not receive an airpuff (Table 2). All tones were created by using the online tool onlinetonegenerator.com and Audacity (AudacityCross-Platform Sound Editor).

Since the mice only had the opportunity to drink water in the IC, it was necessary to monitor whether all mice drank daily. If a mouse did not drink for 24 h, the mouse was offered water in a separate cage for 15 minutes. After these 15 minutes, they were placed back in the home-cage. If drinking did not occur in the IC for three consecutive days, these mice were taken out of the experiment by allowing them access to water within the IC corner without tones. These mice were no longer participating in the conditioning phase and cognitive bias test, but were still left in the group, leaving the social structure unchanged throughout the experiment.

Table 2. Description of the possible events during the conditioning within the IntelliCage corner.

Nosepoke	Positive Tone	Negative Tone
Yes	water	airpuff
No	nothing	nothing

For more clarity, the individual development steps are described individually below. The respective results and conclusions follow the method description of the individual development steps.

Analysis

Data analysis was done with the open-source statistical software R (version 4.0.3, RCoreTeam, 2020). For data visualization the R package ggplot2 (Wickham, 2016) was used. Model assumptions were inspected visually first by Q-Q plots, and secondly by visualizing variance homogeneity of the residuals versus fitted values. Individual animals served as the experimental unit, as only one mouse was in the test cage at a time. A total of 36 mice were used, which were divided into three groups (12 mice per group and developmental step). Since data were collected automatically, blinding was not necessary.

Analysis of data from gate conditioning protocol. For the gate conditioning protocol (detailed description below), the mice first had to learn which corner was the active corner. Therefore, the number of visits to the active corner was compared to the number of visits to the inactive corners for each mouse per day during the first 14 days (when only the positive tone was presented). A visit was recorded by the IC-system each time a mouse entered a corner, and both the RFID transponder number was detected and the presence-sensor was activated. The visit number was used as the outcome in a linear mixed-effects model (R package nlme [Pinheiro *et al.*, 2020]). The experimental days were used as a fixed effect (factor with 14 levels). The type of visit (factor with two levels: visits in active corners versus visits in inactive corners) and the interaction of type of visits and day were used also as fixed effects. The variable 'days nested in animals' (n = 12) were set as a random effect. Sum-contrasts were used for days and type of visits.

For the evaluation of the two gate conditioning runs (run 1 n = 11, run 2 n = 12), the frequency with which the mice passed the AnimalGate was first determined for each mouse for each day, *i.e.*, how often mice were presented with tone-sequences. The duration from entering to leaving the IC was defined as IC-session. From this, we determined how often the positive and negative tone-sequences were played (per animal, per day). Next, we determined how often the mice visited the active corner and made nosepokes on the nosepoke-sensor during the positive and negative tone-sequence IC-sessions. The number of nosepokes was used as the outcome in a linear mixed-effects model (R package nlme). In this model, the experimental days were defined as days and used as a fixed effect (factor with nine levels in AnimalGate conditioning run 1, factor with 14 levels in AnimalGate conditioning run 2). Within the statistical model, the type of tone-sequence (two-level factor: positive versus negative tone-sequence) and the interaction of type of tone-sequence and day were also used as fixed effects. Sum-contrasts

were used for day and type of tone-sequence. The variable 'experimental days nested in animals' was set as a random effect.

Analysis of data from corner conditioning protocol. For the evaluation of the corner conditioning protocol (detailed description below), the frequency with which the mice (group 2 n = 12, group 3 n = 12) visited the active corner within the IC was first determined for each mouse for each day, *i.e.*, how often mice were presented with tone-frequencies (inactive corners were blocked with a plug). From this, we determined how often the positive and negative tones were played (per animal, per day). Next, we determined how often the mice visited the active corner and made nosepokes at the nosepoke-sensor during the positive and negative tone. The number of nosepokes was used as the outcome in a linear mixed-effects model (R package nlme). In this model, the experimental days were defined as days and used as a fixed effect (factor with 48 levels). Within the statistical model, the type of tone-frequency (two level factor: positive tone-frequency versus negative tone-frequency) and the interaction of type of tone-frequency and day were also used as fixed effects. Sum-contrasts were used for day and type of tone-frequency. The variable 'experimental days nested in animals' was set as a random effect. To test for effects of interaction of day and tone-frequency, *post hoc* comparison was conducted (R package emmeans [Lenth, 2020]).

Learning success for visit conditioning. Descriptive statistics were used to assess individual learning success by observing correct nosepoke behavior. Correct nosepoke behavior at the positive tone was defined as a corner visit during which at least one nosepoke was made. Correct nosepoke behavior for the negative tone was defined as a corner visit without a nosepoke. For each mouse, we first determined how many positive tone trials and negative tone trials had occurred. Then, the numbers of positive tone trials with nosepokes and the number of negative tone trials without nosepokes were determined. Since the probabilities for the positive and negative tone trials were different, percentage values were calculated. From this, the corrected nosepoke behavior was plotted for each animal individually. The learning criterion was set as follows: First, we checked whether the values for the positive and negative tone were above the 50% chance level. Then, on 75% of the conditioning days, the correct nosepoke behavior had to be above the chance level in order to reach the learning criterion.

Cognitive bias test. All mice reaching the learning criterion were used in the cognitive bias test (test 1 and 2 n = 9). All other mice remained in the group, but no tones were presented when they entered the IC corner. For the cognitive bias test, the mice were presented with three additional (ambiguous) tones. First, for each mouse we determined how many nosepokes they made in response to the five different tones. The number of nosepokes was used as the outcome in a linear mixed-effects model (R package nlme). In this model, the tones (factor with five levels) and measurement (cognitive bias test 1: factor with three levels (baseline measurement 1, negative conditions and baseline measurement 2), cognitive bias test 2: factor with four levels (baseline measurement 1 and 2, negative conditions and baseline measurement 3) and the interaction were used as fixed effects. The variable 'treatment nested in animals' was set as a random effect. If the model indicated a significant effect of treatment or tone, we conducted a pairwise *post hoc* analysis (R package emmeans).

Body weight and IntelliCage behavior. For the evaluation of body weight, number of nosepokes and visits, the corresponding values were determined for each animal for each day (group 2 n = 12, group 3 n = 12). These three variables were used as the outcome in three different linear mixed-effects models (R package nlme). Treatment (group two: factor with eight levels (0%, 5%, 10%, 16%, 20%, 33% and 50% probability of negative tone and visit open doors), group three: factor with seven levels (0%, 20% and 50% probability of negative tone, nosepoke open doors, baseline measurement and negative conditions)), day (group two: factor with 75 levels, group three: factor with 100 levels) and the interaction of treatment and day was used as a fixed effect. The variable 'experimental days nested in animals' were set as a random effect.

Developmental Step 1

Methods

Animals. The 12 female mice of group one arrived at the institute in February 2019. At the start of the first developmental step, the mice were seven weeks old. After the experiment presented here, the mice were 18 weeks old and used in home-cage based learning tasks (data not published) and in a consumer demand test, which was also performed within the home-cage based set-up presented here (Kahnau *et al.*, 2022A). The mice started barbering behavior at the age of 18 weeks and immediately following the experiment presented here. Barbering behavior is commonly found in C57BL/6J mice (Kahnau *et al.*, 2022B; Sarna *et al.*, 2000). The reason for this behavior is not yet understood.

Gate conditioning protocol. The gate conditioning protocol was pre-registered in the Animal Study Registry (doi: 10.17590/asr.0000121). The mice were conditioned to tone-sequences. These sequences had a play time of 6.6 seconds at a frequency of 8 kHz and comprised either short tone-sequences with long breaks or long tone-sequences with short breaks (Figure 2).

Each mouse was randomly assigned one of two tone-sequences; thus six out of twelve mice had tone-sequence A and the other six had tone-sequence B as the positive tone stimulus. The other tone-sequence was consequently the negative stimulus. One loudspeaker was placed on top of the IC (on the grid) facing in the direction of the IC inside, allowing the mice to hear the tone-sequences. The tone-sequences were played when entering the IC after passing through the gate.

Within the IC, each mouse was randomly assigned one active corner (three mice per corner), in which the mice received either the water reward or an airpuff punishment depending on the tone-sequence. Visiting the other three corners had no consequences.

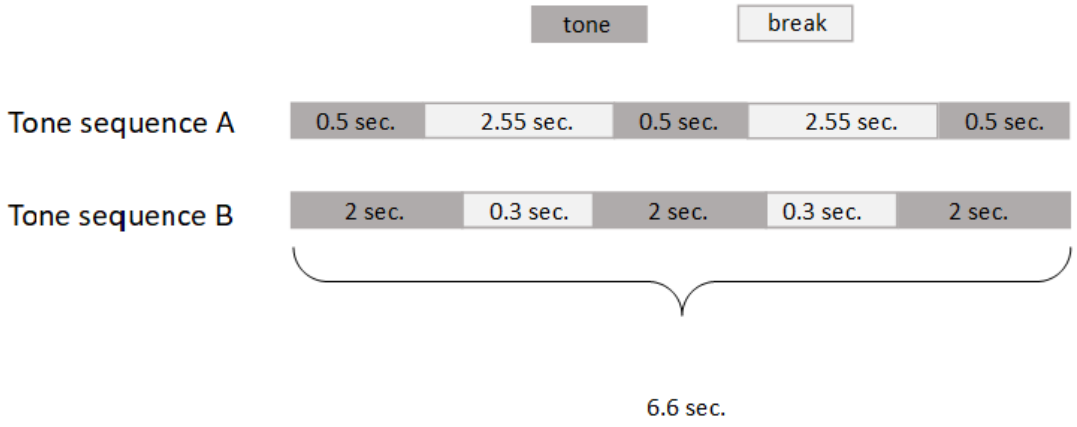

Figure 2. Tone-sequences used for AnimalGate conditioning.

The mice had to learn first which corner their active corner was (one out of four) and second that a tone was played every time they entered the IC through the gate. This corner and positive tone conditioning ran for 14 days. When visiting the active corner and activating the nosepoke-sensor, the doors were opened for five seconds. To prevent the mice from staying too long inside the corner, an airpuff was released after another five seconds. To open the doors within the IC corner again, the IC had to be left through the gate (end of IC-Session). By re-entering the IC, a new trial was initiated.

After corner and positive tone conditioning, the negative tone-sequence was added. To prevent the mice from having too many negative experiences directly at the beginning of the conditioning phase, the probability of the negative tone being played was increased successively. Therefore, two runs were carried out. For gate conditioning run 1, the probability of playing the negative tone was 33%. For gate conditioning run 2, the probability of playing the negative tone was 50%. To initiate a new trial, the IC had to be re-entered through the gate, *i.e.*, mice that could not drink after a negative tone-sequence or did not drink after a positive tone-sequence had to leave and re-enter the IC for the next chance to drink.

Results

Corner and positive tone-sequence conditioning. The mice first had to learn which corner was the assigned active corner. Over a period of 14 days, the animals were successfully conditioned to the active corner (main effect visits: $F_{1,154} = 225.44$, $p < 0.0001$). The overall number of visits decreased over the experimental days (interaction: $F_{13,154} = 6.63$, $p < 0.0001$, Figure 3).

Gate conditioning protocol. The mice had to learn to make nosepokes after hearing positive tone-sequences and refrain from making nosepokes after hearing negative tone-sequences. In gate conditioning run 1 with 33% chance of hearing a negative tone sequence (Figure 4), the mice did not make more or less nosepokes after hearing positive or negative tone-sequences on average (main effect tone-sequence: $F_{1,90} = 0.22$; $p = 0.64$). The mice did not learn to differentiate between tone-sequences over time (interaction: $F_{8,90} = 0.82$; $p = 0.59$). However, the mice made fewer nosepokes regardless of tone-sequences over time (main effect day: $F_{8,80} = 4.58$; $p = 0.0001$).

In gate conditioning run 2 with the chance of hearing a negative tone sequence increase to 50% (Figure 5), the mice made, on average, more nosepokes for the positive tone-sequence (main effect tone-sequence: $F_{1,77} = 18.9$; $p < 0.0001$) but did not learn to differentiate between the tone-sequences (interaction: $F_{6,77} = 0.62$; $p = 0.71$). During run 2 the mice made more nosepokes over time regardless of tone-sequences (main effect day: $F_{6,66} = 2.45$; $p = 0.03$).

Discussion

The first developmental step was described as ‘gate conditioning protocol’, where tone-sequences were played whenever a mouse passed the gate and entered the IC. The initial idea of using tone-sequences was to easily create ambiguous sequences once the positive and negative sequences were successfully conditioned. Although it was possible to condition the mice to their respective randomly assigned IC corner, the mice were not able to distinguish between two tone-sequences. The mice were unable to associate a water reward with one tone-sequence and a mild airpuff punishment with another tone-sequence. The unsuccessful conditioning could have different reasons.

First, mouse-specific ultrasonic vocalization series can have a length of two seconds. They are variable in their sequence but are released at a more or less constant frequency. There are also short sequences (a few milliseconds long) that vary in both sequence and frequency (Ehret, 2018). Our artificially created, very static tone sequences at constant frequency had a length of 6.6 seconds, which may be too long to be perceived as relevant for the mice. The tone-sequences might have shown better results if shortened. To the best of our knowledge, there have been no experiments to condition mice to artificially

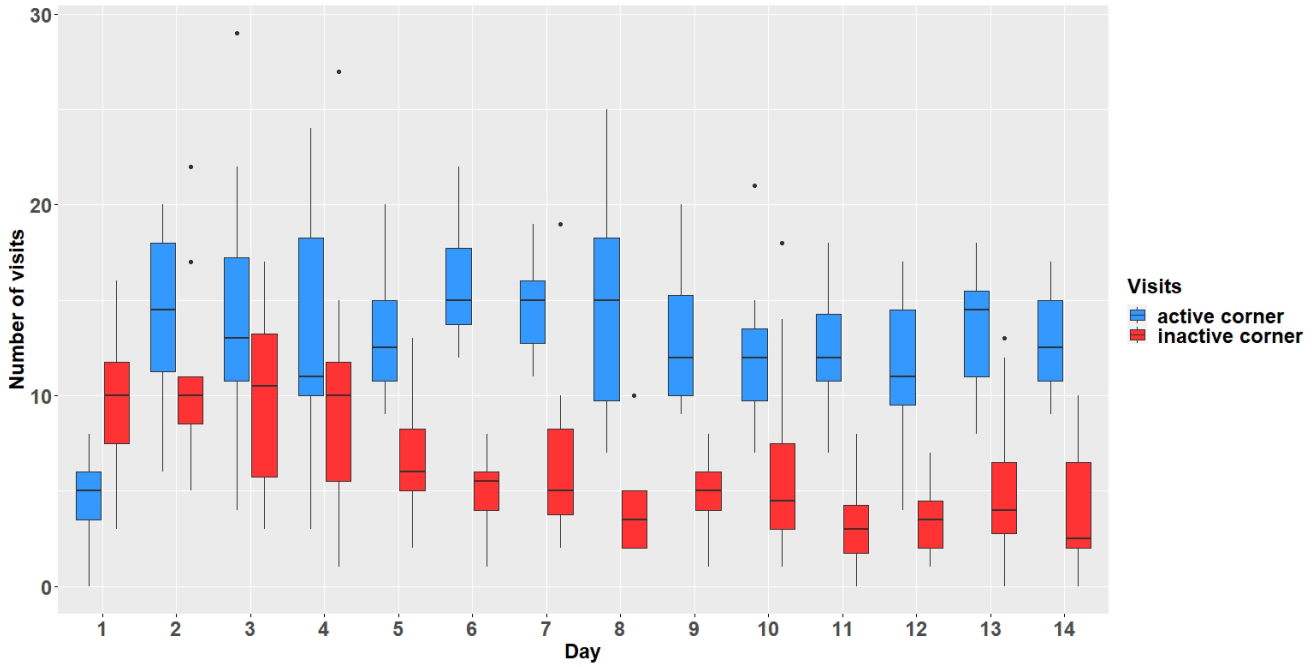

Figure 3. Comparison of visit numbers in active and inactive corners. The y-axis shows the number of visits which were made within the active and inactive corners. The x-axis shows the experimental days. n = 12.

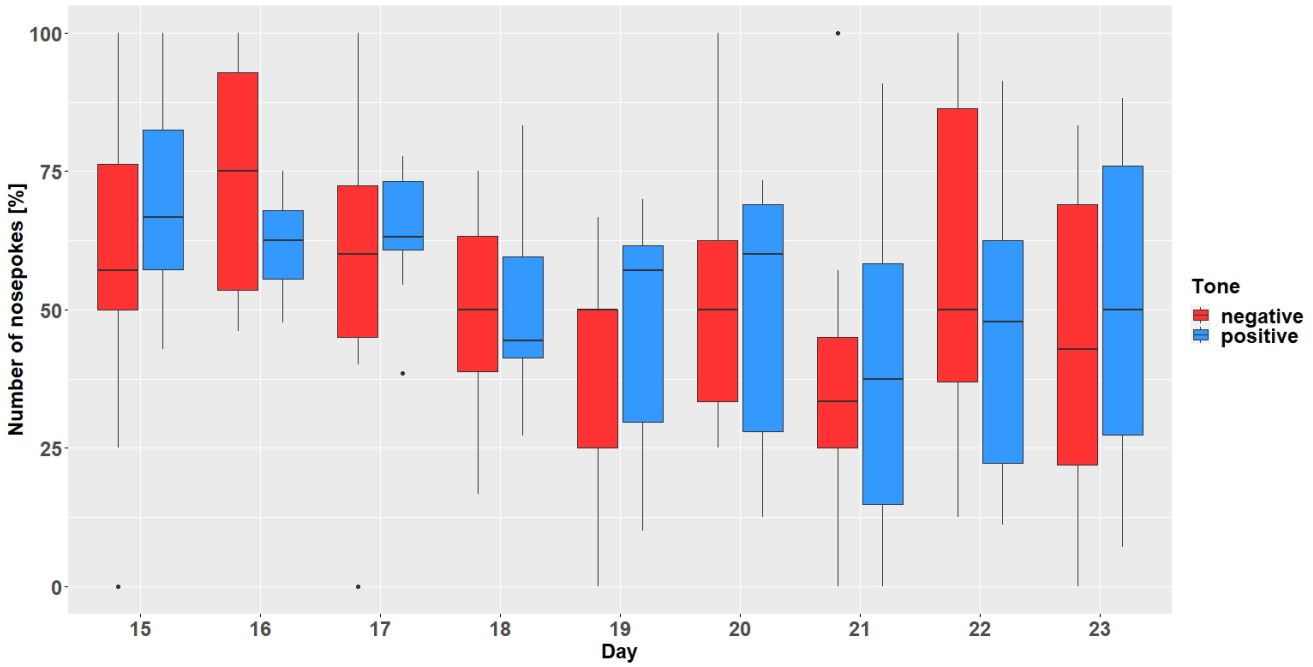

Figure 4. Gate conditioning run 1. The y-axis shows the number of nosepokes which were made in response to the presented tone-frequencies. Number of nosepokes are given in percent since the probability of the two tone-frequencies being played was different (positive = 67%, negative = 33%). After hearing a positive tone-frequency, a nosepoke had to be made, but not after hearing a negative tone-frequency. The x-axis shows the experimental days. n = 11.

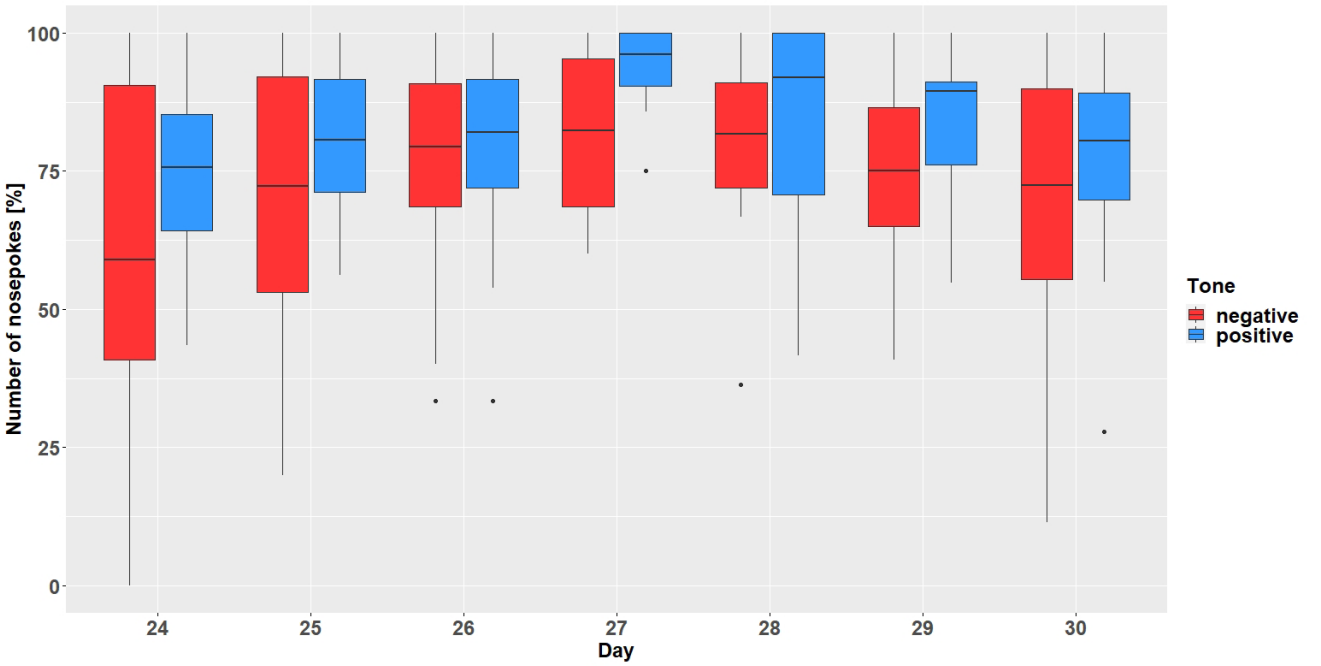

Figure 5. Gate conditioning run 2. The y-axis shows the number of nosepokes which were made in response to the presented tone-frequencies. The probability of the two tone-frequencies being played was 50:50. After hearing a positive tone-frequency, a nosepoke had to be made, but not after hearing a negative tone-frequency. The x-axis shows the experimental days. n = 12.

created tone-sequences like the ones we used during developmental step 1. However, past studies showed the possibility to condition mice to tones, namely tone-frequencies (De Hoz & Nelken, 2014; Jones *et al.*, 2017). Therefore, we decided to use tone-frequencies instead for the next developmental step.

Second, the timing at which the tone-sequences during gate conditioning were presented was not optimal. Tones were initiated by each pass through the gate and played when the IC was entered. Whether the mouse then also directly visited the IC corner was probably dependent on how strong the motivation to drink was. Therefore, it might be possible that too much time passed between the tone and the actual corner visit, and thus, no association was established between these two events. The timing between stimulus presentation and event onset is important for successful conditioning, as shown, for example, by clicker training (Lattal, 2010).

Therefore, we decided to change the time point of tone presentation and relocated the conditioning completely to the IC corner. From then on, the sound was played when the mouse entered the IC corner. This improvement reduced the time span from the presentation of the stimulus to the corresponding nosepoke behavior to a minimum. To prevent a possible overlap effect of the unsuccessful conditioning on the next developmental step, we continued to work with a naïve mouse group.

Developmental Step 2

Methods

Animals. The twelve female mice of group two arrived at the institute in October 2019. At the start of the second developmental step presented here, the mice were 14 weeks old. The mice started barbering behavior at an age of 20 weeks, during the conditioning phase. At the end of the experiment, the mice were 26 weeks old and used in another experiment to develop a conditioned place preference test to assess severity of experimental procedures (publication in preparation).

Corner conditioning protocol. Since the gate conditioning protocol was not successful in group one, we improved the conditioning protocol and decided to no longer condition to tone-sequences but to tone-frequencies.

The hearing range of mice is between 2 kHz and 70 kHz (Heffner & Heffner, 2007). To find different frequencies with equal sound pressure levels (SPL) in the corner, a measuring microphone (miniDSP Umik-1 calibrated USB microphone) and the software Room EQ Wizard were used. In a study by de Hoz and Nelken, mice were successfully conditioned to tone-frequencies between 6 kHz and 13 kHz (De Hoz & Nelken, 2014). The same frequency range was used for our study. With a digital signal processor (miniDSP 2x4), the SPL of the played tone was optimized, to ensure that all tones were played at the

same volume within the corner. This was done to ensure that variations in SPL stemming from the speaker and confined space in which they are played were as small as possible. However, it must be emphasized that the perception of mice differs from that of humans and that there possibly are influences, which we are unable to detect and/or assess.

In addition to different tones, we decided to change the time when the tones were played. For the corner conditioning protocol, tone-frequencies (positive or negative tone) were played when entering an IC corner instead of when leaving the gate and entering the IC. One single corner within the IC was chosen as the active corner for all mice to set the focus of the mice to this corner and to ensure that the tone quality was the same for all mice. All other corners were made unreachable by 3D printed plugs made from gray polylactic acid (PLA). In order to initiate a new trial, the mice had to re-enter the active corner. During one IC session, multiple trials could be initiated by the mouse re-entering the active corner without having to leave the IC again (as it was the case for gate conditioning protocol). Within the active corner and after hearing the positive tone-frequency, the IC doors could be opened by a nosepoke for seven seconds.

To play the tone-frequencies, one loudspeaker was placed on top of the active corner directed towards the inside of the corner, so the mice were able to hear the tones. In order to be able to position the loudspeaker, it was integrated into a black 3D printed box (Figure 1C).

The tone-frequency at one end of the scale was 6.814 kHz at 70 decibel (dB), the other tone-frequency on the other end of the scale was 13.629 kHz at 70 dB. At the beginning of the conditioning phase, only the positive tone frequency was played during a visit in the active IC corner. The probability of the negative tone-frequency was increased progressively to avoid too many negative experiences at the beginning of the experiment (*Extended data* [Kahnau *et al.*, 2022C]).

The tone-frequencies had at first a length of 0.5 seconds. The tone length was extended to one second on experimental day 20. During the experimental phase, there were several technical problems and therefore, data for some days were lost. On several occasions, the body weight of the animals could not be recorded due to the AnimalGate being blocked by bedding material. Removing the bedding from the AnimalGate solved this problem. An unexpected failure of the control unit led to missing data recording on days 23, 99, 104, 105. The whole IC system had to be restarted to resolve these failures.

Results

Corner conditioning protocol. After visiting the active corner, one out of two tone-frequencies was randomly presented. In total (Figure 6), the mice made more nosepokes at the positive

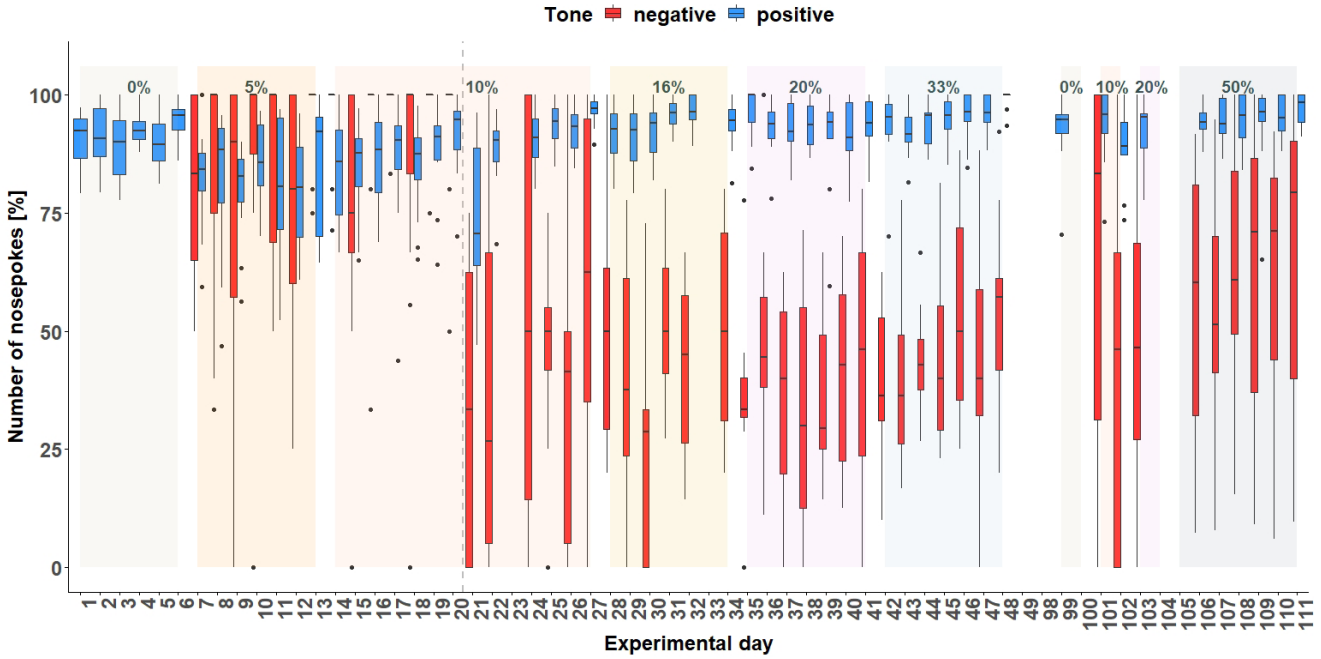

Figure 6. Corner conditioning group two. Number of nosepokes in percent made in response to two different tone-frequencies. The data for experimental day 23, 99, 104, 105 is missing due to technical problems with the IntelliCage system. No data from day 49 to 84 is available, because the mice were not in the home-cage based set-up as the set-up had to be maintained. From experimental day 85 the mice were kept in the set-up again. In order to habituate the mice to the set-up again, no sounds were played on days 85 to 98. On the y-axis, the number of nosepokes in percent is shown. The x-axis shows the experimental days. The dashed line marks the time point when the tone length was increased to one second. 0% = no negative tone, 5% = 5% negative tone probability, 10% = 10 percent negative tone probability, 16% = 16% negative tone probability, 20% = 20% negative tone probability, 33% = 33% negative tone probability, 50% = 50% negative tone probability.

tone compared to the number of nosepokes made at the negative tone-frequency (main effect tone: $F_{1,452} = 795$, $p < 0.0001$). The mice differentiated between the two-tone frequencies after the tone length was increased to 1 second on experimental day 20 (interaction: $F_{47,452} = 9.51$, $p < 0.0001$, table S3 *Extended data* [Kahnau *et al.*, 2022C]). In addition, the mice made less nosepokes in total after day 20 (main effect experimental day: $F_{47,473} = 5.39$, $p < 0.0001$).

Individual learning success. Since the results are considered for each mouse, the results are evaluated descriptively. The individual learning success was considered during the time period when the negative tone was played with a probability of 33% (Figure 7) and 50% (Figure 8). These were chosen because the negative tone was played enough times to allow a meaningful comparison of the nosepoke behavior.

At the time when the negative tone was played with a probability of 33%, seven mice (ro_si_2, ro_sw_2, sw_ge_2, sw_si_2, we_ro_2, we_si_2, and we_sw_2) out of 12 mice reached the learning criterion. Mouse ro_ge-2 stopped to drink before the negative tone was played with a probability of 33%.

Increasing the probability of the negative tone to 50% resulted in more incorrect nosepoke behavior in response to the negative tone. Only four mice (ro_sw_2, sw_si_2, we_ro_2 and we_si_2) out of 12 mice reached the learning criterion (75% of correct nosepoke behavior over 50%). The mice ro_ge_2 and ro_si_2 did not drink and were taken out of the experiment.

Body weight and IntelliCage behavior. Body weight, number of licks, and number of visits were recorded throughout the experimental period (Figure 9). Body weight was influenced by the treatment ($F_{7,803} = 2.33$, $p = 0.02$) as well as by the experimental day ($F_{1,803} = 211$, $p < 0.0001$). Also, the interaction treatment and day had an influence on body weight ($F_{7,803} = 2,36$, $p = 0.02$). Body weight increases over time, but there are days when the mice lose weight. This is influenced by the tasks they have to perform in the IC. In addition, the number of licks over time was influenced by treatment ($F_{7,813} = 20.71$, $p < 0.0001$) as well as experimental day ($F_{1,813} = 14,3$, $p > 0.0001$) and seems to decrease with the presentation of the tones. This influence seems to be particularly strong on individual experimental days (interaction: $F_{17,813} = 7,99$, $p < 0.0001$), which is also reflected in the number of visits (interaction: $F_{7,813} = 22.31$, $p < 0.0001$). These were also influenced by the treatment ($F_{7,814} = 47$, $p < 0.0001$) but not influenced by the experimental day ($F_{1,815} = 0.05$, $p = 0.83$).

Discussion

The second developmental step was described as 'corner conditioning protocol', where tone-frequencies were played whenever a mouse visited the active IC corner. With this protocol, it was possible for the first time for single mice to distinguish between two different tone-frequencies within the set-up presented here. Two mice ceased drinking in the IC during the conditioning phase. Therefore, these mice were excluded from the experiment, *i.e.*, for them the tone presentation was turned off and they were able to open the doors by a nosepoke at each visit. Dropouts also occurred in other studies, where individual animals did not reach the learning criterion and thus the actual test phase (*e.g.*, Bračić *et al.*, 2022; Hintze *et al.*, 2018; Kloke *et al.*, 2014; Krakenberg *et al.*, 2019). Since only one mouse could be in the IC at a time, the question arises whether individual mice could not learn because they could enter the IC less often. The data show that all mice were able to enter

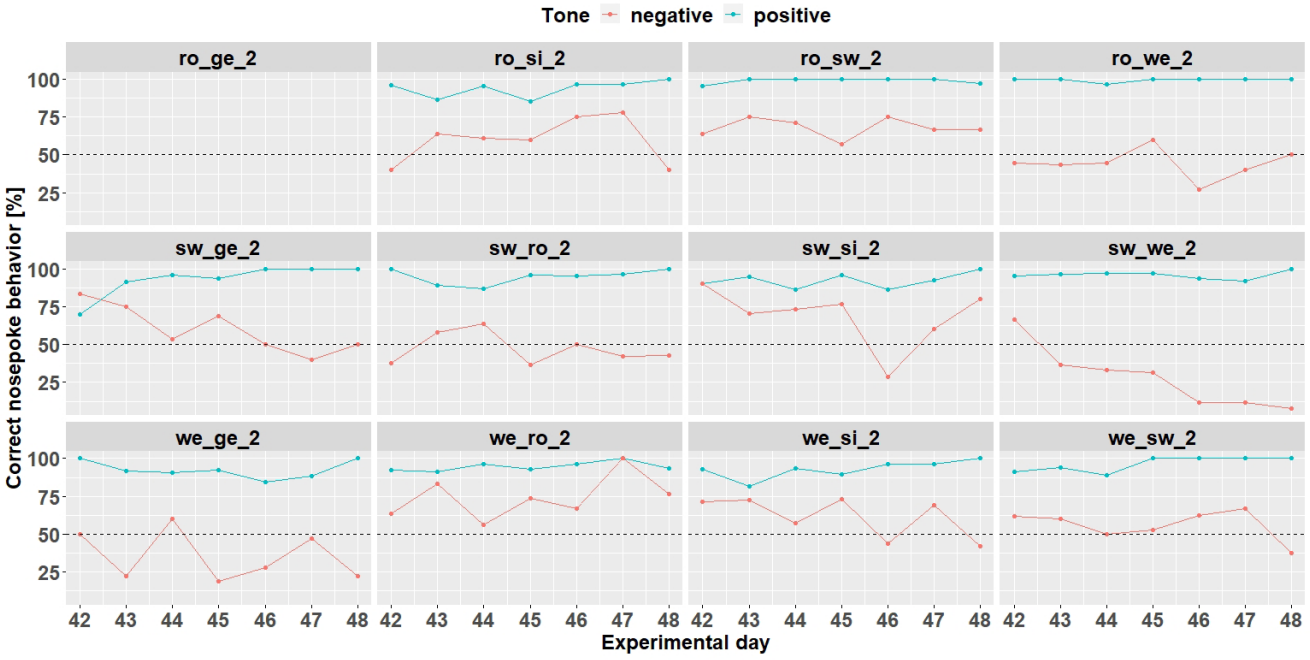

Figure 7. Individual learning success during conditioning when the negative tone-frequency was presented with a probability of 33%. Mouse ro_ge_2 was taken out of the experiment. Learning criterion 75% of correct nosepoke behavior over 50%.

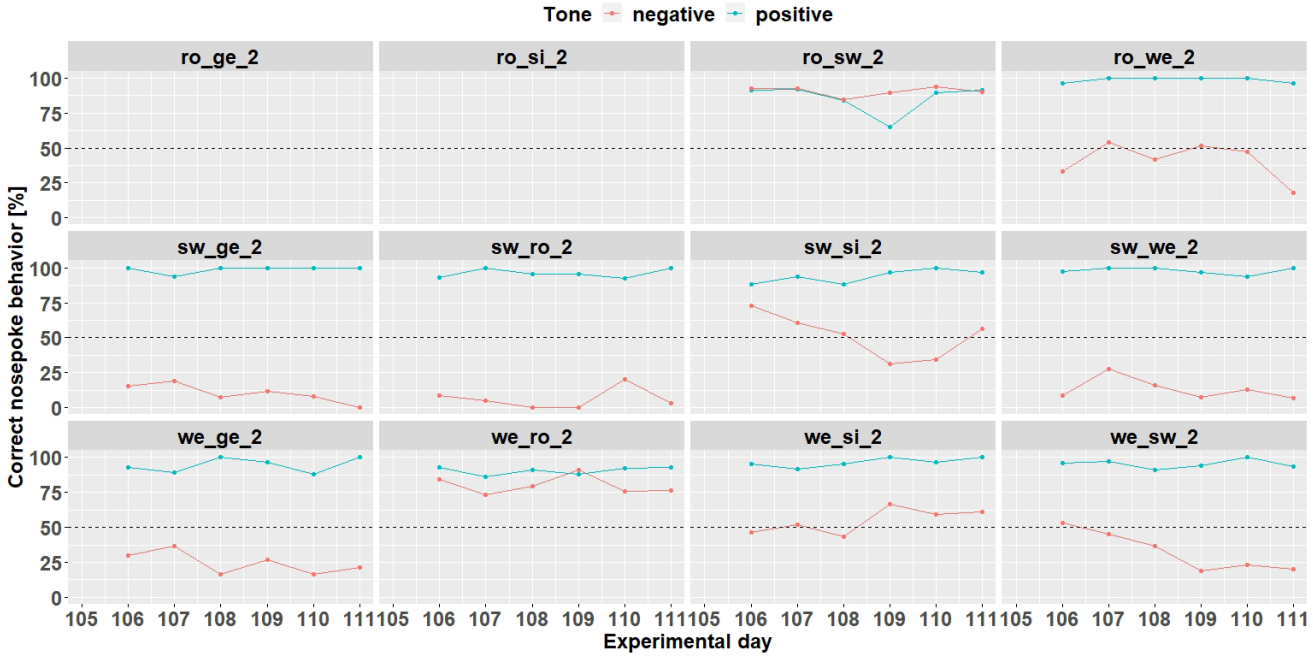

Figure 8. Individual learning success during conditioning when the negative tone-frequency was presented with a probability of 50%. The data of day 105 is missing due to technical problems with the IntelliCage system. Ro_ge_2 and ro_si_2 did not participate any longer in the experiment. Learning criterion 9 trials over 50% out of 12.

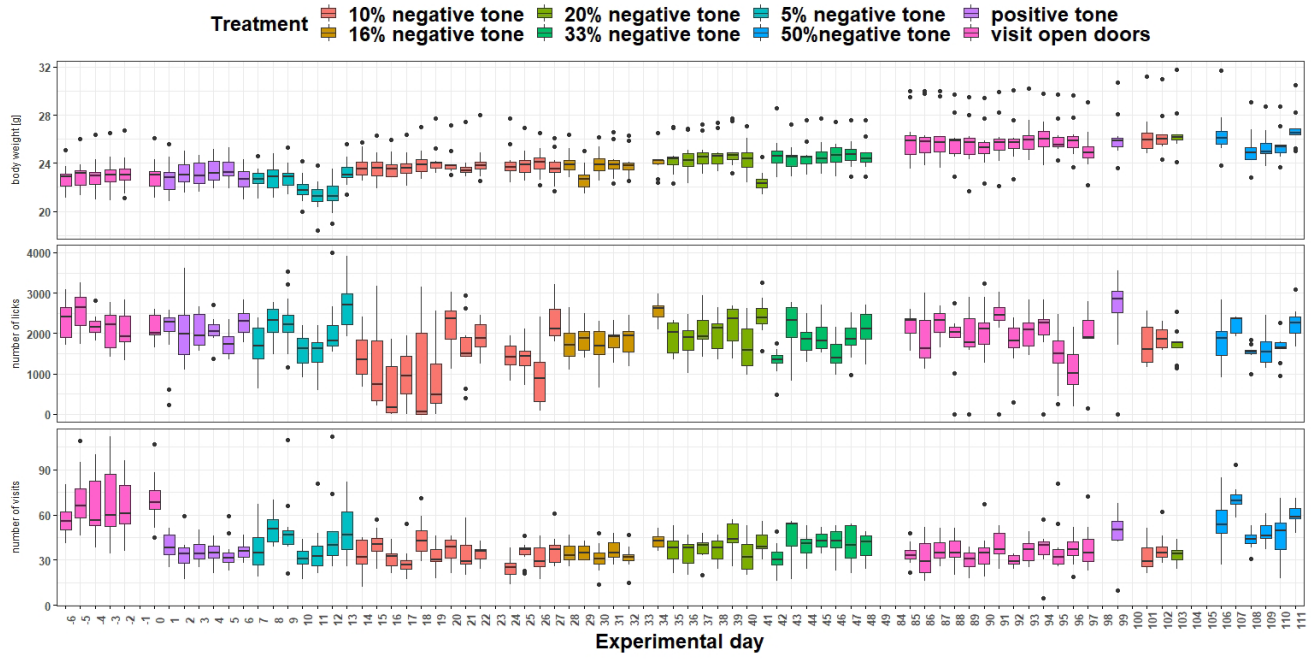

Figure 9. Measurement of body weight, IntelliCage corner visits and lick number over time. The x-axis shows the experimental days. On the y-axis first the body weight, second the lick number and third the visit number is shown. Different tones with different playback probabilities were presented throughout the experimental period (treatment). The data of experimental day -1, 23, 33, 98, 100, 104 and 105 are missing, due to technical issues. During experimental days 49 to 84 no tones were played.

the IC several times a day. Once they were in the IC, they were also able to repeatedly enter the active corner of the IC. We therefore assume that individual mice did not fail to learn because they were able to enter the IC less often. For the mice that stopped drinking, it seems more likely that the risk of getting an airpuff was too high and they stopped drinking. When the tones were turned off for these mice, they started drinking again.

The other ten mice of the group continued to drink within the IC but they did not initially distinguish between the two different tones. After changing the tone length (from half a second to one second), significant increase in the number of nosepoke behavior could be detected. The mice did more nosepokes in response to the positive tone compared to the nosepoke number for the negative tone. However, the number of nosepokes for the negative tone increased when the probability of it occurring was increased (up to 50%).

This was particularly evident in the examination of individual learning performance, when nosepoking was barely supressed by the negative tone. Overall, mice made many correct responses for the positive tone, but markedly fewer correct responses for the negative tone. Accordingly, the mice seemed to have a high motivation to perform nosepokes regardless of the outcome.

There was also an increase in the number of visits over the course of the experiment. However, the number of licks per day hardly changed. The explanation might be that the possibility to drink was reduced by increasing the number of trials with the negative tone. Thus, to get the same amount of liquid, more visits had to be made. It may be that the motivation to interact with the nosepoke sensor was so strong that the risk of punishment was accepted. This would be in line with literature data showing that mice continue to operate a lever although it was associated with a stimulation of 'aversive brain regions' (Cazala, 1986).

By giving many incorrect responses to the negative tone, the mice also received a correspondingly high number of airpuffs, which in turn could have led to habituation to the airpuff. The punishment would therefore no longer be perceived as a valid punishment (Kahnau *et al.*, 2021). Another explanation could be that the permanent presentation of the tones caused them to no longer be perceived as relevant but rather as a kind of background noise, and nosepokes were made independently of the tones.

In many previous cognitive bias tests, mice were placed in a designed test apparatus for a defined test period and were exposed to the stimuli for that defined time (*e.g.*, Bailoo *et al.*, 2018; Boleij *et al.*, 2012; Kloke *et al.*, 2014; Krakenberg *et al.*, 2019; Richter *et al.*, 2012). After the test phase, the mice were transferred back to their home-cages, where they spent their time undisturbed until the next test phase. On the contrary, in our system, which also served as the home-cage, no such breaks occurred. Thus, the stimulus might have had none or little relevance and the focus might be on opening the doors, driven by the motivation to drink.

Our results suggest that rest periods should be included in order to maintain the concentration and/or motivation of the mice. Therefore, for the next developmental step, we decided to schedule breaks, while the mice had access to the water without presentation of the tones, between the individual conditioning and testing phases. To exclude possible influences from previous conditioning phases, we again worked with another naïve mouse group in the next developmental step.

Developmental Step 3

Methods

Animals. The twelve female mice of group three arrived at the institute in September 2020. At the start of the third developmental step, the mice were six weeks old. At the end of this experiment, the mice were 21 weeks and used in various cognitive experiments (data not published) and in an experiment to develop a home-cage based consumer demand test based on the mouse positioning surveillance system (data not published yet). The mice started barbering behavior at the age of 31 weeks, 10 weeks after the experiment presented here.

Corner conditioning protocol. In order to successfully condition the mice of group three to tone-frequencies, further modifications were made to the corner conditioning protocol described earlier. This experiment was pre-registered in the Animal Study Registry (doi: 10.17590/asr.0000228). In the active corner and after hearing the positive tone-frequency, the IC doors could be opened by a nosepoke for ten seconds. In addition, the tone length as well as the airpuff length was extended to two seconds. The tone-frequencies for the first conditioning phase of group three were the same as for group two (6.814 kHz at 70 dB and 13.629 kHz at 70 dB). For the second conditioning phase, tone-frequencies between 6.814 kHz at 70 dB and 9.636 kHz at 70 dB were used. Also, for group three, the probability of the negative tone was increased step by step (*Extended data*).

Cognitive bias test. After the conditioning phase, the cognitive bias test followed. This was done by adding ambiguous tone-frequencies, which were calibrated between the positive and negative tone-frequencies (first cognitive bias test: 8.103 kHz, 9.636 kHz, 11.459 Hz, second cognitive bias test: 7.431 kHz, 8.103 kHz, 8.836 kHz). For the determination of these ambiguous tones, the geometric mean, which is the perceived middle between two tones, was used. To determine the geometric mean (GM), the square of the product of the two chosen tone frequencies is calculated.

$$GM = \sqrt{f_1 \cdot f_2}$$

The tritone of the original low and high frequency is then used as the respective high and low frequency to calculate two additional tritones, generating a scale of five tones, each perceptibly equidistant to their neighbors. The SPL was checked with a measuring microphone and the Room Acoustics Software.

The probability for each of the three ambiguous tone frequencies to be played was 5%. By entering the active corner, one of

the five different tone-frequencies was randomly presented. The mice received water by performing a nosepoke at the positive tone, and received an airpuff by performing a nosepoke at the negative tone. The mice received neither a reward nor a punishment for the ambiguous tones. For data evaluation, the nosepoke behavior toward the ambiguous tone-frequencies was measured.

During baseline measurement, the housing conditions were as described in section "Home-cage based set-up". To manipulate the cognitive bias, the housing conditions were changed. In accordance with studies in rats, we expect the cognitive bias of the mice to be negatively affected by the removal. The mice had less bedding (2cm high), less nesting material (four papers), less housing (one mouse house), no running disk, one handling tube, two wooden gnawing sticks, no active enrichment and no resting platform. For further treatment effect, the mice were additionally restrained. For this purpose, the mice were handled by tail and placed in a tube. In the tube, the mice were unable to move and had to remain in the tube for three minutes. This procedure was performed on four consecutive days at 08:00 to 9:30 o'clock during the cognitive bias measurement. The order in which the mice were restrained was randomized for each day using the R statistical program.

Results

Corner conditioning protocol. From day 57 (Figure 10), the tone-frequencies were changed. Over the entire period, the mice made more nosepokes in response to the positive tone compared to the number of nosepokes made in response to the negative tone-frequency ($F_{1,429}$ = 3578, p < 0.0001). The experimental days seem to have an influence on the nosepoke number (main effect experimental day: $F_{48,418}$ = 4.77, p < 0.0001) as well as the interaction of day and tone ($F_{48,429}$ = 6.13, p < 0.0001). This shows that the mice can discriminate between the sounds, but on some days there were fewer nosepokes overall.

Individual learning success conditioning phase 1. Conditioning phase 1 run for 11 days (Figure 11). Experimental day 16 was quite noticeable, where all mice performed worse. It was found that a technical problem occurred during the tone playback. Therefore, for the evaluation of the learning success only 10 days were used.

Nine mice (ro_ge_3, ro_sw_3, ro_we_3, sw_ge_3,sw_si_3, sw_we_3, we_ge_3, we_ro_3 and we_si_3) out of 12 mice reached the learning criterion (75% of correct nosepoke behavior over 50%). The mice ro_si_3, sw_ro_3 and we_sw_3 stopped to drink and were taken out of the experiment at day 18.

Cognitive bias test 1. During the first CB test (Figure 12), the tone-frequencies influenced the number of nosepokes, which were made after hearing the tone-frequencies ($F_{4,96}$ = 28.55, p < 0.0001). The more similar the tone becomes to the positive tone, the more nosepokes the mice made. A *post hoc* comparison showed that, except for the negative and near-negative tone (tone-frequency which is close to the negative tone-frequency), the mice discriminated between the different frequencies (Table 3). Also, the treatment (baseline measurement and

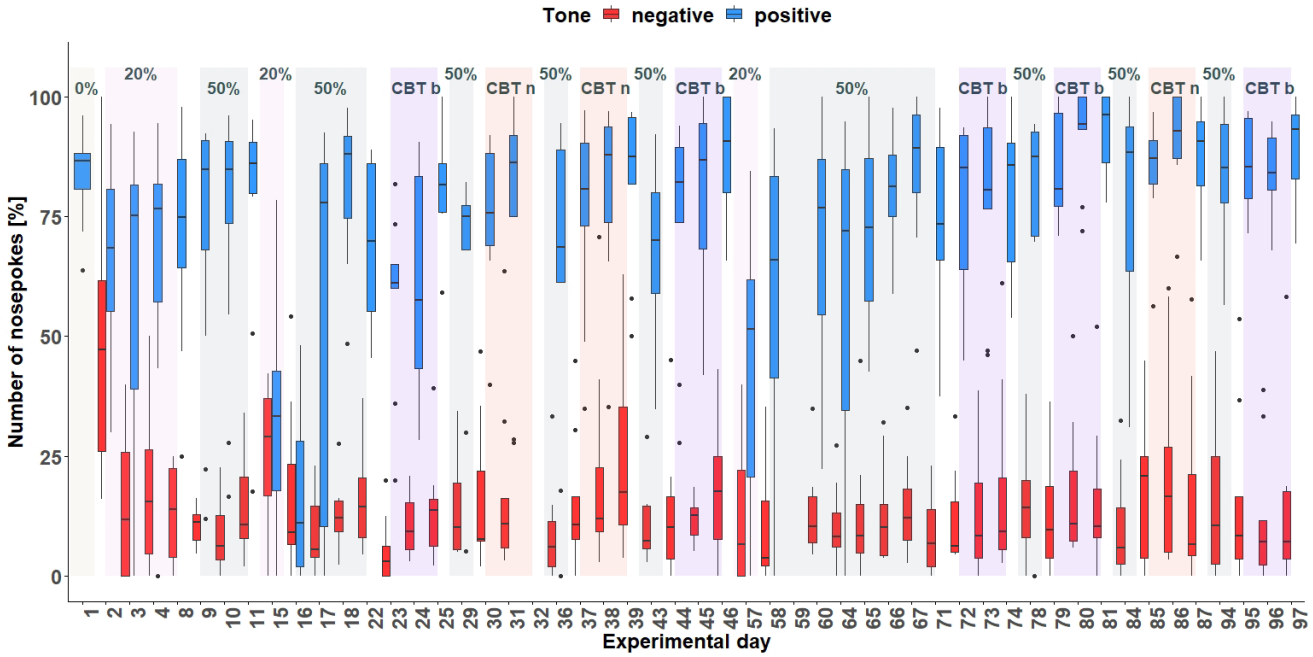

Figure 10. Corner conditioning group three. Number of nosepokes in percent made in response to two tone-frequencies. The data for experimental day 32 and 59 are missing due to technical problems with the IntelliCage system. There was an experimental break between day 47 and 56. After each treatment, no tones were presented. On the y-axis, the number of nosepokes in percent are shown. The x-axis shows the experimental days. During the experimental time period, the tones were presented with different probabilities. 0% = no negative tone, 20% = 20% negative tone probability, 50% = 50% negative tone probability, CBT b = cognitive bias measurement baseline, CBT n = cognitive bias measurement under negative conditions with less bedding and nesting, no enrichment and daily restraining.

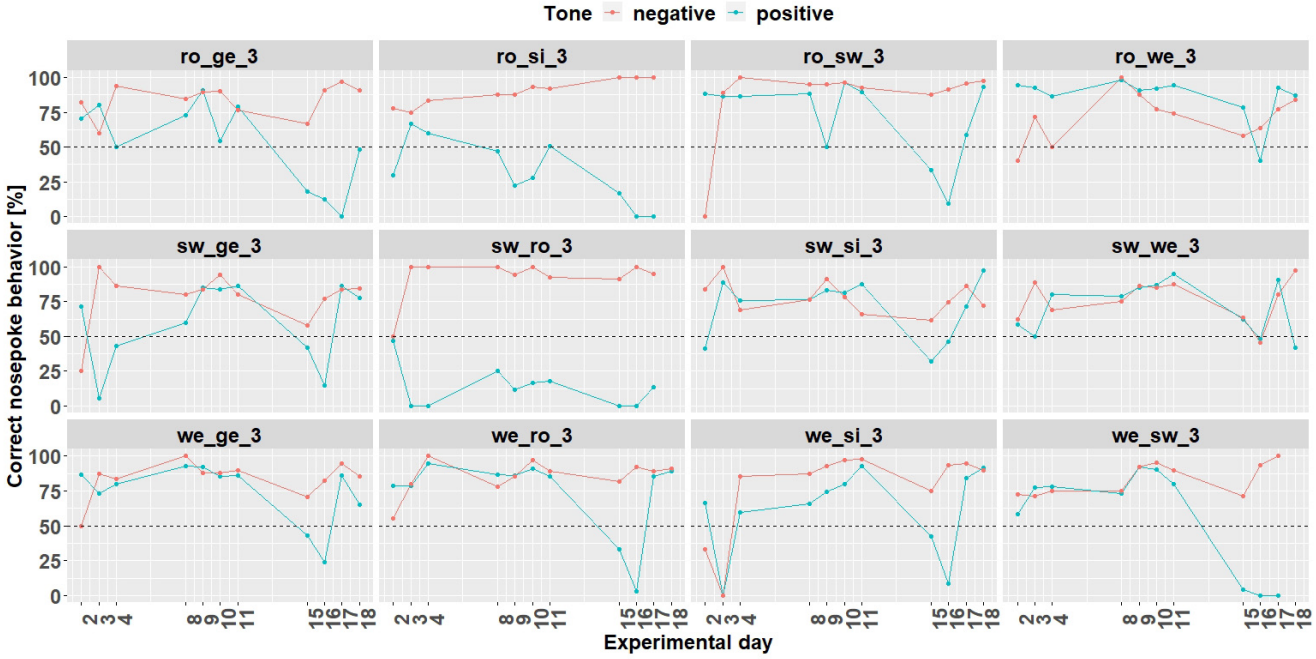

Figure 11. Individual learning success during conditioning phase 1 of mouse group three. On day 16, due to technical problems, the tones were not played correctly The mice ro_si_3, sw_ro_3 and we_sw_3 were excluded from the experiment from day 18 onwards. Learning criterion: 75% of correct nosepoke behavior over 50%.

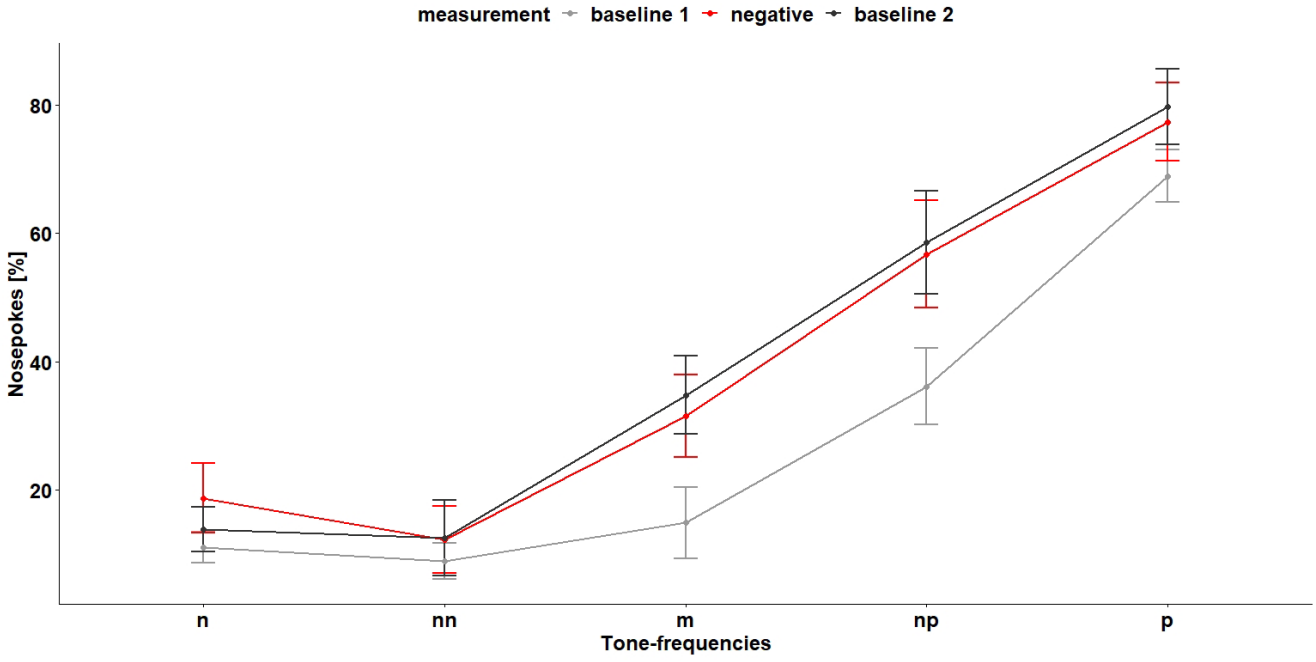

Figure 12. Cognitive bias test 1. The x-axis shows the tone-frequencies with n = negative tone, nn = near-negative tone, m = middle tone, np = near-positive tone and p = positive tone. The y-axis shows the number of nosepokes in percent made in response to the tone-frequencies. During negative measurement the housing conditions were changed compared (less bedding and nesting and no enrichment) to baseline measurement and the mice were restrained daily. n = 9.

Table 3. Results of the *post hoc* comparison of the performed nosepokes in response to the tone-frequencies for the first cognitive bias test. n = negative tone, nn = near-negative tone, m = middle tone, np = near-positive tone and p = positive tone.

Comparison	Estimate	SE	df	t.Ratio	p-Value
m – n	12.52	3.87	96	3.24	<0.001
m – nn	15.79	3.87	96	4.08	<0.001
m – np	-23.41	3.87	96	-6.06	<0.0001
m – p	-48.27	3.87	96	-12.48	<0.0001
n – nn	3.27	3.87	96	0.85	0.4
n – np	-35.93	3.87	96	-9.29	<0.0001
n – p	-60.79	3.87	96	-15.72	<0.0001
nn – np	-39.2	3.87	96	-10.14	<0.0001
nn – p	-64.06	3.87	96	-16.57	<0.0001
np – p	-24.86	3.87	96	-6.43	<0.0001

negative treatment (less bedding and nesting, no enrichment and daily restraining) had an influence on the nosepoke behavior of the mice ($F_{2,16}$ = 5.08, p = 0.02). A *post hoc* comparison showed that the mice made less nosepokes during baseline 1 measurement compared to baseline 2 measurement and negative treatment (Table 4). The interaction of tone-frequency and treatment had no influence on the nosepoke behavior ($F_{8,96}$= 1.05, p = 0.4). Which suggests that the mice do not evaluate the different tone frequencies differently during the different treatments.

Individual learning success conditioning phase 2. Due to technical issues, the data of day 59 are missing and was excluded for learning success evaluation. During conditioning phase 2 (Figure 13) 8 (ro_sw_3, ro_we_3, sw_ge_3, sw_si_3, sw_we_3, we_ge_3, we_ro_3 and we_si_3) out of 12 mice reached the learning criterion. The mice ro_si_3 and we_sw_3 stopped to drink and were taken out of the experiment at day 59. The mouse sw_ro_3 stopped to drink, too, and was taken out of the experiment at day 65.

Also, during the second CB test (Figure 14), the tone-frequencies influenced the number of nosepokes ($F_{4,112}$ = 27.27, p < 0.0001). Again, the mice did not differentiate between the negative and near-negative tone but between all other tone-frequencies (Table 5). The measurement and the interaction of tone-frequency and treatment had no influence on the nosepoke number (main effect treatment: $F_{3,21}$ = 1.67, p = 0.2, interaction: $F_{13,112}$ = 0.62, p = 0.8).

Body weight and IntelliCage behavior. Body weight (Figure 15) was influenced by the treatment ($F_{5,963}$ = 17.4, p < 0.0001) as well as by the experimental day ($F_{1,963}$ = 196, p < 0.0001). Over time, body weight increased continuously.

Table 4. Results of the *post hoc* comparison of the performed nosepokes during baseline measurement and negative treatment in response to the tone-frequencies. b = baseline, n = negative treatment (less bedding and nesting, no enrichment and daily restraining).

Comparison	Estimate	SE	df	t.Ratio	p-Value
b1 – b2	11.917	3	16	-3.979	<0.01
b1 – n	-11.345	3	16	-3.788	<0.01
b2 – n	0.573	3	16	0.191	0.85

Also, the interaction of experimental day and treatment influenced body weight ($F_{5,963}$ = 12.52, p < 0.0001). There are a few days when the mice lose a noticeable amount of weight. The number of licks (Figure 15) over time were influenced by treatment ($F_{5,963}$ = 30.79, p < 0.0001) but not by experimental day ($F_{1,963}$ = 0.03, p = 0.9) or the interaction of experimental day and treatment ($F_{5,963}$= 1.8, p = 0.1). The number of visits (Figure 15) were influenced by treatment ($F_{5,963}$ = 50.29, p < 0.0001). The analysis showed a tendency towards influence of the experimental day on the visit numbers ($F_{1,963}$ = 3.5, p = 0.06). However, the interaction of experimental day and treatment had an influence on the visit number ($F_{5,963}$ = 6.8, p < 0.0001).

Discussion

The third developmental step was also described as 'corner conditioning protocol', where tone-frequencies were played whenever a mouse visited the active IC corner. The tone length was changed again (from one to two seconds) compared to developmental step two. The assumption was that this change would allow the mice to discriminate the tone-frequencies more easily. In the study by de Hoz and Nelken, the tone-frequencies were played throughout the complete time of a corner visit. The playing of the tone was stopped only after the mouse left the corner and was re-initiated by a new corner visit (De Hoz & Nelken, 2014). This extreme adjustment of playback length was not considered for our experiment, since it is not known how the individual tone presentation length influences the nosepoke behavior, and thus, the cognitive bias of the mice. There was a potential for individual visit durations to have an influence on the individual mouse assessment of ambiguous tone, making the results difficult to interpret and thus reducing the validity of the data.

Like in developmental step two, some mice in group three could not be conditioned to the tone-frequencies. However, the remaining mice learned effectively and made more nosepokes in response to the positive tone compared to the negative tone. In addition, the mice seemed to be more hesitant in nosepoke behavior compared to the mice in developmental step two. This becomes evident when examining individual learning performance: There were slightly fewer correct responses for the positive tone and more correct responses for the negative tone. This implies that they performed less nosepokes overall,

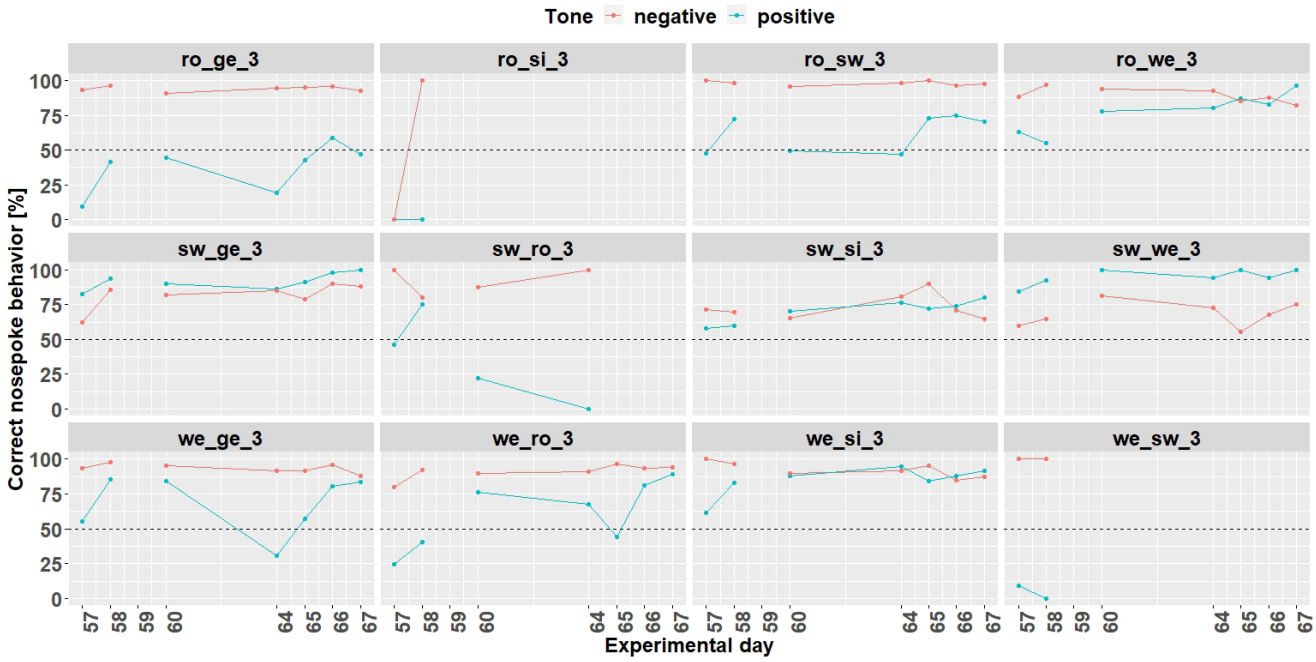

Figure 13. Individual learning success during conditioning phase 2 of mouse group three. Data of day 59 is missing due to technical problems.

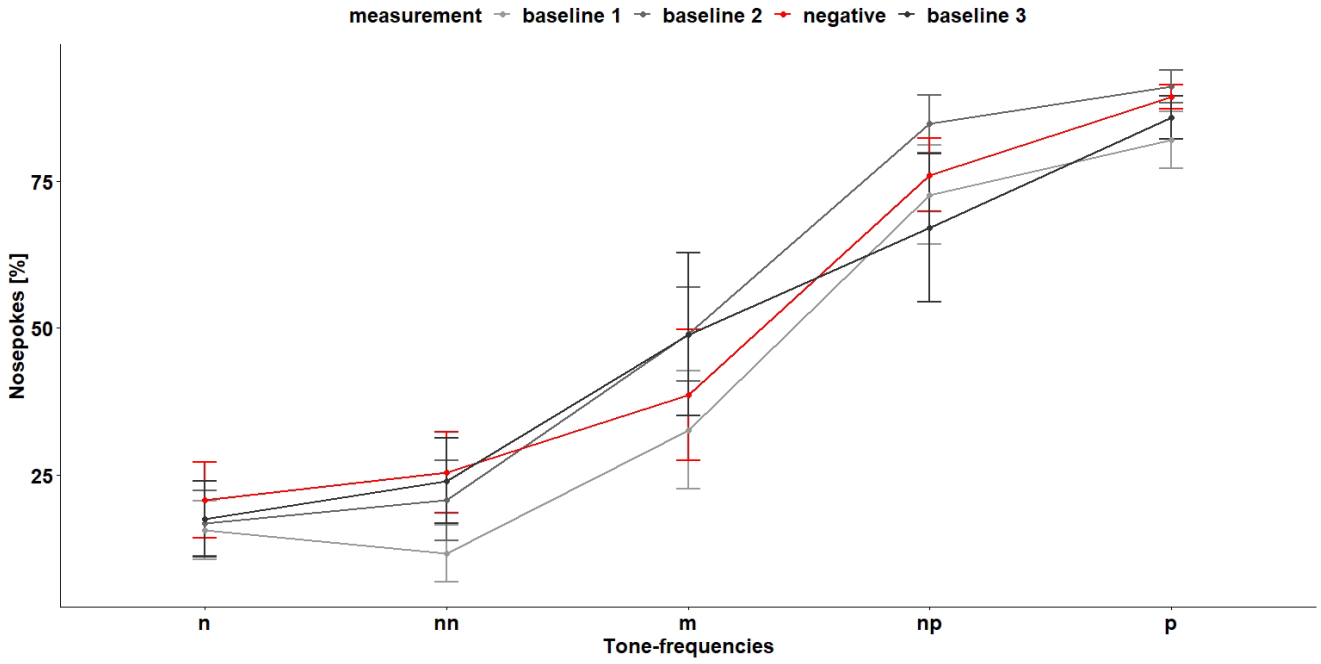

Figure 14. Cognitive bias test 2. The x-axis shows the tone-frequencies with n = negative tone, nn = near-negative tone, m = middle tone, np = near-positive tone and p = positive tone. The y-axis shows the number of nosepokes in percent made in response to the tone-frequencies. During negative treatment the housing conditions were changed compared to baseline measurement and the mice were restrained daily. n = 9.

Table 5. Results of the *post hoc* comparison of the performed nosepokes in response to the tone-frequencies for the first cognitive bias test. n = negative tone, nn = near-negative tone, m = middle tone, np = near-positive tone and p = positive tone.

Comparison	Estimate	SE	df	t.Ratio	p-Value
m – n	24.72	4.42	112	5.59	<0.0001
m – nn	21.92	4.42	112	5.0	<0.0001
m – np	-32.89	4.42	112	-7.44	<0.0001
m – p	-44.87	4.42	112	-10.15	<0.0001
n – nn	-2.79	4.42	112	-0.63	0.53
n – np	-57.61	4.42	112	-13.03	<0.0001
n – p	-69.59	4.42	112	-15.74	<0.0001
nn – np	-54.82	4.42	112	-12.4	<0.0001
nn – p	-66.80	4.42	112	-15.1	<0.0001
np – p	-11.98	4.42	112	-2.71	<0.001

which has a positive effect on the number of correct answers for the negative tone but a negative effect on the answers for the positive tone. The airpuff seems to be perceived as negative. However, since some mice had to be excluded in this and in the previous developmental step because they stopped drinking, it should be considered whether the airpuff of 0.5 bar is too intense and might be reduced which could reduce the drop-out rate.

In other studies, punishment is not used at all (Graulich *et al.*, 2016; Hintze *et al.*, 2018; Novak *et al.*, 2016; Verjat *et al.*, 2021), as it is discussed that punishment during conditioning and in the test itself may already have an influence on the cognitive bias (Roelofs *et al.*, 2016). Also in the study of Alboni and colleagues punishment was not used within the IC system (Alboni *et al.*, 2017). It is quite conceivable that individual mice, motivated by their everyday exploratory behavior, enter individual corners without drinking motivation, so that corner visits also occur without nosepoke. In order to be able to interpret the mouse behavior clearly, it is recommended to use mild airpuffs as punishment in IC experiments. In addition, conditioning with punishment seems to be easier to learn and thus seems to succeed faster (Lagisz *et al.*, 2020). In our system,

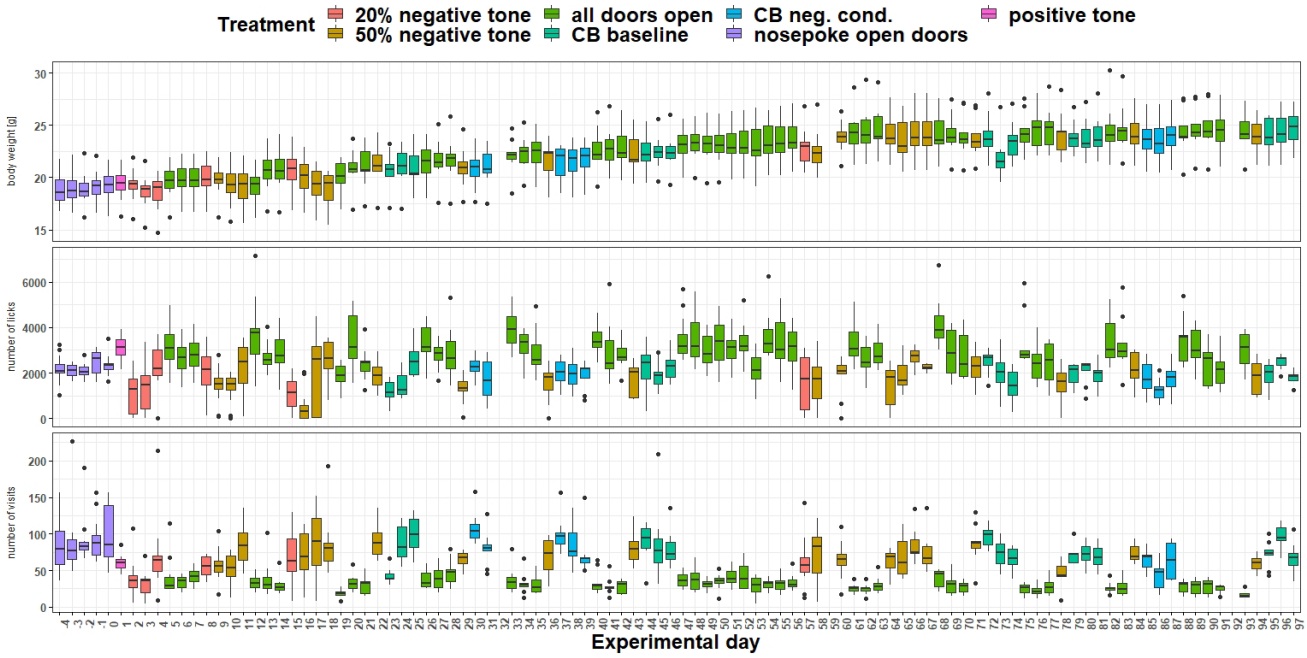

Figure 15. Measurement of body weight, IntelliCage corner visits and lick number over time. The x-axis shows the experimental days. On the y-axis, first the body weight, second the lick number and third the visit number is shown. Different tones with different playback probabilities were presented throughout the experimental period (treatment). The data of experimental day 32, 59 and 92 is missing, due to technical issues.

we chose to use a punishment because the behavior of the mice can be interpreted clearly. The mice want to avoid the airpuff and therefore do not poke when the negative tone is presented. We come to this conclusion based on our experience of mice immediately performing nosepokes upon entering the IC corners if no airpuffs are included in an experimental design.

Lagisz and colleagues identified in their systematic review and meta-analysis that a go/go active choice paradigm (go to receive a reward and go to avoid a punishment) leads to the most sensitive set-up (Lagisz *et al.*, 2020). It is discussed whether in a go/no-go paradigm the no-go behavior could be related to reduced activity or motivation and not to negative expectation of the future event (Enkel *et al.*, 2010; Matheson *et al.*, 2008). Nevertheless, we chose in our system a go/no-go paradigm. The mice had to nosepoke (go) to receive the reward (water) and not to nosepoke (no-go) to avoid the punishment (airpuff). In addition, the mice had to leave the IC corner and re-enter it (go) to initiate a new trial. We chose a go/no-go paradigm for the same reason that we used the airpuff as a punishment. The behavior in response to the tones is more easily distinguished and interpreted. In addition, by self-initiating the trial, there are no waiting times and the mice have the possibility to complete the trial in a self-determined manner (Hintze *et al.*, 2018; Krakenberg *et al.*, 2019). This choice of experimental design allows us to assume that the mice are highly motivated and facilitates the derivation of a conclusive interpretation of the mice's behavior.

In the third development step we also analyzed the visit and lick behavior. Both seem to be influenced by the treatment (breaks, conditioning or cognitive bias measurement). By starting conditioning, fewer visits and licks were made. It can be assumed that the lick number is also influenced by the circumstance that the IC doors were permanently open during the breaks. This allowed the mice to drink more per visit during the breaks, which consequently reduced the number of visits and increased the number of licks. The data suggest that the mice need more time to drink, as weight was also affected by the treatment. It would therefore be reasonable to increase IC open-door-time. However, the open-door-time should not be so long that the number of visits is reduced because more licks might be made per visit and thus fewer visits are needed and made overall. This in turn would lead to a reduced number of trials for evaluation.

Because we assumed successful conditioning in developmental step three, the cognitive bias test followed. With the automated and home-cage based set-up presented here, it was possible to measure the cognitive bias of female C57BL/6J mice. Our data showed a sigmoidal curve of data points decreasing from positive tone-frequency to negative tone-frequency. Our result suggests that the ambiguous tone-frequencies are perceived and interpreted differently with respect to the previously conditioned tone-frequencies, which is a basic requirement of a valid cognitive bias test (Gygax, 2014; Hintze *et al.*, 2018; Krakenberg *et al.*, 2019).

We hypothesized that mice living in enriched housing conditions (from 28 days of age) would be affected in their emotional state by removal of enrichment and additional restraining. In accordance with the literature on known effects of worsening housing and living conditions in rodents, we expected a pessimistic cognitive bias as a result. In fact, we were able to detect a change in the cognitive bias. The mice showed more nosepoke behavior while kept under negative conditions compared to the time of the first baseline measurement, indicating a positive, optimistic cognitive bias. This increased nosepoke behavior was still evident during the second baseline measurement, when the negative conditions had been eliminated. This result is surprising because studies in rats showed that rats housed under negative housing conditions showed a negative cognitive bias (Burman *et al.*, 2009; Harding *et al.*, 2004) and a transfer from standard to enriched housing conditions led to a shift from pessimistic to optimistic cognitive bias (Brydges *et al.*, 2011; Richter *et al.*, 2012). So far, only Alboni and colleagues and Resasco and colleagues were able to measure an influence of housing conditions on cognitive bias in mice. Unlike to our study, enriched housed mice seemed to have a positive expectancy related to the ambiguous stimulus compared to standard housed mice (Alboni *et al*., 2017; Resasco *et al*., 2021).

The question arises why the mice in our experiment seem to have a more optimistic cognitive bias after removing enrichment and with restraining. It is theoretically possible that the removal of enrichment has no negative effect on the mice. However, the mice were additionally restrained daily during this developmental stage. Based on the behavior of the mice, we assume that the mice had a negative experience at least concerning the restraining. Anecdotally, we can report that the mice showed a reduced interaction with the experimenter's hand after the first trial and seemed to no longer want to enter the restrain tube freely. The mice also tried to leave the tube, as evidenced by frantic turning inside the tube. Another possible explanation could be that the mice experienced boredom due to the removal of enrichment, since most stimulating objects had been removed. According to optimal arousal theory, individuals strive for an optimal arousal state. If an individual does not have this arousal state and/or experiences boredom, it would seek something arousing/stimulating. However, if the arousal state is too strong, the individual would seek less arousing stimuli (Mitchell *et al.*, 1984).

In our experiment, this could indicate that the mice did not have an optimal arousal state due to the removal of enrichment and that this is targeted by an increased willingness to take risks to receive an airpuff. However, the mice were also additionally restrained. Thus, it is not possible to identify which factor (removal of enrichment or restraining) or both factors had an influence on the cognitive bias. The influence also seems to be so strong that an increased nosepoke behavior (compared to the first baseline measurement) could also be detected for the second baseline measurement. This raised the question of whether the mice really had a more optimistic cognitive bias or whether the tones were too "easy" to distinguish. Therefore, we decided to reduce the tone scalar.

The mice also learned to discriminate between tones which were closer to each other, and learned when they received water and when they received an airpuff. Therefore, another cognitive bias test was performed.

During the second test phase, we again observed a sigmoidal curve in the data, but no change in cognitive bias. This result is consistent with other studies (Bailoo *et al.*, 2018; Bračić *et al.*, 2022). It should be noted that during the second test phase, the period of negative conditions was significantly shorter. It is possible that one week has no influence on the cognitive bias of mice or that the experiences already made have led to a kind of habituation. Also, the number of animals with nine mice might be too small to statistically detect an existing difference of small effect size. It is also possible that the test systems developed so far, including the system presented here, were not sensitive enough. In addition, the possible change in cognitive bias might not last long enough to be measured or is masked by positive stimulation due to cognition training (Krakenberg *et al.*, 2019).

Another reason why we could not measure a change could be that a group of mice serving as their own control is not informative enough, as we cannot rule out a temporal carrying over effect for the second baseline. However, the study of Bracic and colleagues showed that the cognitive bias was repeatable over multiple measurements (Bračić *et al.*, 2022). Further experiments are necessary to better interpret the results presented here. For example, it is necessary to test whether a mouse group can serve as its own control group in our test set-up. We therefore conducted a follow-up study in which two groups of mice with different experiences are tested simultaneously and repeatedly. The analysis of this data will allow us to compare two independent groups with each other, but also to compare each group with itself.

It should be noted that a too-frequent repetition of presenting the ambiguous stimuli could also lead to mice learning that neither reward nor punishment occurs with ambiguous stimuli (Roelofs *et al.*, 2016). This might in fact have been the case here, as the mice never returned to baseline 1 level. It seems that the mice make more nosepokes over time. This could indicate that the mice have learned that there is no reward or punishment for the ambiguous tones. However, it is positive to note that the sigmoidal curve is present in all measurements, suggesting that the mice evaluate the ambiguous tones differently from the conditioned tones. If the mice unambiguously could differentiate all frequencies, we would expect a much higher nosepoke behavior (close to 100%) in response to the ambiguous tone frequencies. This assumption is based on experience from previous IC experiments (Kahnau *et al.*, 2021). To prevent the animals in the cognitive bias test from learning that there is neither reward nor punishment for the ambiguous stimuli, it is necessary to ensure that the ambiguous stimuli are distributed in an appropriately high experimental number of positive and negative stimuli (Krakenberg *et al.*, 2019).

In some mouse test systems, the trial number per session is relatively low, *i.e.*1-32 trials (Boleij *et al.*, 2012; Kloke *et al.*, 2014; Novak *et al.*, 2016). In contrast, in the set-up of Hintze and colleagues and in the automated touch-screen system of Krakenberg and colleagues, up to 54 trials per day could be performed. However, it was always necessary to remove the mice from their familiar environment, thus separating them from their group members (with the exception of individual housing) and determining the time of the test, which could have an influence on the motivation to participate in the test.

In our set-up, the number of trials varied depending on how frequently the IC corners were visited (group three: 4 - 214 visits = trials), but were distributed over the entire day. The animals decided independently from the experimenter when to enter the IC (if the IC was not already occupied by another mouse), which makes a high motivation to participate in the test plausible. Even though only one mouse could be in the IC at a time, it was possible for all mice to enter the IC several times a day, and thus, initiate trials in the IC itself. This was also shown in an automated and home-cage based consumer demand test, for which a similar test setup was used as described here (Kahnau *et al.*, 2022A). It is also not necessary to manipulate the night/day rhythm (as *e.g.*, in Krakenberg *et al.*, 2019), as in experiments in which the presence of an experimenter during data acquisition is required. This, in turn, drastically reduces the time required (daily control of animals and set-up of about 30 minutes).

Conclusions

The cognitive bias test seems to be a suitable test method to measure the affective state of animals (Lagisz *et al.*, 2020). So far, however, these tests are very labor intensive and require animals to be tested outside of their home cages, which has implications for the animals and thus the data. (*e.g.*, Bračić *et al.*, 2022; Hintze *et al.*, 2018; Kloke *et al.*, 2014; Krakenberg *et al.*, 2019). Therefore, we aimed to develop an automated and home-cage based cognitive bias test for mice in which individual mice can perform the task undisturbed by group members.

In the study presented here, we describe the developmental steps for such a test concluding in a method that allows measuring the cognitive bias in mice. By presenting the various stages of development, we intended to provide a better understanding of the structure of the test method. We also contribute to providing comprehensive information to the scientific community that can be used to develop further automated and home-cage based systems.

Automation and home-cage based testing offers the advantages of testing the mice in their familiar environment and during their active phase. The influence of the animals on each other is reduced, as only one mouse can be in the test-cage at a time. Also, the influence of the experimenter is reduced to a minimum. The fact that the mice can choose the time of the experiment and initiate trials themselves gives them control over what occurs and suggests that the mice are highly motivated. All this, in turn, might have a positive impact on the validity of the data.

All three groups of mice showed barbering behavior over their lifespan. This behavior occurred at different ages in the respective groups. Group one showed barbering behavior immediately after the experiments presented here, group two during the experiments and group three a few weeks after the experiments presented here. However, it is likely that the behavior was present earlier, as it was only visible through fur lesions. Barbering is a common behavior in female C57 mice (Garner, 2005; Kahnau *et al.*, 2022B). The reasons for the occurrence of this behavior are still unknown. To gain a better understanding of the behavioral course of barbering, we have developed a score sheet (Kahnau *et al.*, 2022B). Whether and what influence barbering has on the mice and thus on the experimental data is unclear. We assume that the influence on the data presented here is rather low, as we were able to condition mice and measure the cognitive bias. Nevertheless, it is necessary to investigate this behavior further and to report it if it occurs.

We were able to measure the cognitive bias of the mice although further research is needed for a better understanding of the mice's cognitive bias measured in the IntelliCage based system. We will continue to develop our test system and use it to assess the burden of commonly used behavioral tests such as the Water Maze Test, and to include the perspective of the mouse in this assessment.

Ethical approval

All experiments were approved by the Berlin state authority, Landesamt für Gesundheit und Soziales (LAGeSo), under license No. G 0182/17 and were in accordance with the German Animal Protection Law (TierSchG, TierSchVersV).

Data availability

Underlying data

Zenodo: Data and supplementary material for the paper "Development of an IntelliCage based Cognitive Bias Test for Mice.", 10.5281/zenodo.8203397 (Kahnau *et al.*, 2022C)

Extended data

Zenodo: Data and supplementary material for the paper "Development of an IntelliCage based Cognitive Bias Test for Mice.", 10.5281/zenodo.8203397 (Kahnau *et al.*, 2022C)

One .pdf file contains supplementary material, the other .pdf file the ARRIVE checklist. In addition, there is one .xlsx file for each developmental step that contains all the data for that step. Each of these three .xlsx files contains a ReadMe sheet describing all variables. The different .txt files contain the R scripts that were used to analyze the data.

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

Authors' contribution

Conceptualization: Pia Kahnau, Lars Lewejohann, Anne Jaap

Investigation/Methodology: Pia Kahnau

Technical support: Birk Urmersbach

Formal Analysis: Pia Kahnau

Project Administration: Lars Lewejohann, Kai Diederich, Pia Kahnau

Funding acquisition: Lars Lewejohann

Supervision: Lars Lewejohann

Writing – original draft preparation: Pia Kahnau

Writing – review and editing: Pia Kahnau, Anne Jaap, Birk Urmersbach, Kai Diederich, Lars Lewejohann

Acknowledgements

The authors thank the animal caretakers of the German Federal Institute for Risk Assessment, especially Carola Schwarck, Lisa Gordijenko and Heiderose Camus, for their support in the animal husbandry.

Lars Lewejohann and Pia Kahnau are members of the COST Action CA20135 – Improving biomedical research by automated behaviour monitoring in the animal home-cage (TEATIME).

References

Alboni S, Van DIjk RM, Poggini S, *et al.*: **Fluoxetine effects on molecular, cellular and behavioral endophenotypes of depression are driven by the living environment.** *Mol Psychiatry.* 2017; **22**(4): 552–561.
PubMed Abstract | Publisher Full Text | Free Full Text

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**: 232.
PubMed Abstract | Publisher Full Text | Free Full Text

Boleij H, van't Klooster J, Lavrijsen M, *et al.*: **A test to identify judgement bias in mice.** *Behav Brain Res.* 2012; **233**(1): 45–54.
PubMed Abstract | Publisher Full Text

Bračić M, Bohn L, Siewert V, *et al.*: **Once an optimist, always an optimist? Studying cognitive judgment bias in mice.** *Behav Ecol.* 2022; **33**(4): 775–788.
PubMed Abstract | Publisher Full Text | Free Full Text

Brydges NM, Leach M, Nicol K, *et al.*: **Environmental enrichment induces optimistic cognitive bias in rats.** *Anim Behav.* 2011; **81**(1): 169–175.
Publisher Full Text

Burman OHP, Parker RMA, Paul ES, *et al.*: **Anxiety-induced cognitive bias in non-human animals.** *Physiol Behav.* 2009; **98**(3): 345–350.
PubMed Abstract | Publisher Full Text

Cazala P: **Self-stimulation behavior can be elicited from various 'aversive' brain structures.** *Behav Brain Res.* 1986; **22**(2): 163–171.
PubMed Abstract | Publisher Full Text

de Hoz L, Nelken I: **Frequency tuning in the behaving mouse: Different**

bandwidths for discrimination and generalization. *PLoS One.* 2014; **9**(3): e91676.
PubMed Abstract | Publisher Full Text | Free Full Text

Ehret G: **Characteristics of Vocalization in Adult Mice.** In: *Handbook of Behavioral Neuroscience.* 2018; **25**: 187–195.
Publisher Full Text

Endo T, Maekawa F, Võikar V, *et al.*: **Automated test of behavioral flexibility in mice using a behavioral sequencing task in IntelliCage.** *Behav Brain Res.* 2011; **221**(1): 172–181.
PubMed Abstract | Publisher Full Text

Enkel T, Gholizadeh D, Von Bohlen Und Halbach O, *et al.*: **Ambiguous-cue interpretation is biased under stress- and depression-like states in rats.** *Neuropsychopharmacology.* 2010; **35**(4): 1008–1015.
PubMed Abstract | Publisher Full Text | Free Full Text

Francis NA, Kanold PO: **Automated operant conditioning in the mouse home cage.** *Front Neural Circuits.* 2017; **11**: 10.
PubMed Abstract | Publisher Full Text | Free Full Text

Garner JP: **Stereotypies and other abnormal repetitive behaviors: Potential impact on validity, reliability, and replicability of scientific outcomes.** *ILAR J.* 2005; **46**(2): 106–117.
PubMed Abstract | Publisher Full Text

Gouveia K, Hurst JL: **Reducing Mouse Anxiety during Handling: Effect of Experience with Handling Tunnels.** *PLoS One.* 2013; **8**(6): e66401.
PubMed Abstract | Publisher Full Text | Free Full Text

Graulich DM, Kaiser S, Sachser N, *et al.*: **Looking on the bright side of bias—Validation of an affective bias test for laboratory mice.** *Appl Anim Behav Sci.* 2016; **181**: 173–181.
Publisher Full Text

Gygax L: **The A to Z of statistics for testing cognitive judgement bias.** *Anim Behav.* 2014; **95**: 59–69.
Publisher Full Text

Harding EJ, Paul ES, Mendl M: **Animal behaviour: cognitive bias and affective state.** *Nature.* 2004; **427**(6972): 312.
PubMed Abstract | Publisher Full Text

Heffner HE, Heffner RS: **Hearing ranges of laboratory animals.** *J Am Assoc Lab Anim Sci.* 2007; **46**(1): 20–22.
PubMed Abstract

Hintze S, Melotti L, Colosio S, *et al.*: **A cross-species judgement bias task: Integrating active trial initiation into a spatial Go/No-go task.** *Sci Rep.* 2018; **8**(1): 5104.
PubMed Abstract | Publisher Full Text | Free Full Text

Hurst JL, West RS: **Taming anxiety in laboratory mice.** *Nat Methods.* 2010; **7**(10): 825–826.
PubMed Abstract | Publisher Full Text

Jirkof P, Rudeck J, Lewejohann L: **Assessing Affective State in Laboratory Rodents to Promote Animal Welfare—What Is the Progress in Applied Refinement Research?** *Animals.* 2019; **9**(12): 1026.
PubMed Abstract | Publisher Full Text | Free Full Text

Jones S, Paul ES, Dayan P, *et al.*: **Pavlovian influences on learning differ between rats and mice in a counter-balanced Go/NoGo judgement bias task.** *Behav Brain Res.* 2017; **331**: 214–224.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Kahnau P, Jaap A, Diederich K, *et al.*: **Determining the value of preferred goods based on consumer demand in a home-cage based test for mice.** *Behav Res Methods.* 2022A; **55**(2): 751–766.
PubMed Abstract | Publisher Full Text | Free Full Text

Kahnau P, Jaap A, Hobbiesiefken U, *et al.*: **A preliminary survey on the occurrence of barbering in laboratory mice in Germany.** *Animal Welfare.* 2022B; **31**(4): 433–436.
Publisher Full Text

Kahnau P, Jaap A, Urmersbach B, *et al.*: **Data and supplementary material for the paper "Development of an IntelliCage based Cognitive Bias Test for Mice".** [Data set]. Zenodo. 2022C.
http://www.doi.org/10.5281/zenodo.7310180

Kahnau P, Mieske P, Wilzopolski J, *et al.*: **Development and Application of Home Cage Monitoring in Laboratory Mice and Rats : a Systematic Review Short title : Systematic Review of Home Cage Monitoring in Rodents.** *BioRxiv.* 2023.
Publisher Full Text

Kaupert U, Thurley K, Frei K, *et al.*: **Spatial cognition in a virtual reality home-cage extension for freely moving rodents.** *J Neurophysiol.* 2017; **117**(4): 1736–1748.
PubMed Abstract | Publisher Full Text | Free Full Text

Kloke V, Schreiber RS, Bodden C, *et al.*: **Hope for the best or prepare for the worst? Towards a spatial cognitive bias test for mice.** *PLoS One.* 2014; **9**(8): e105431.
PubMed Abstract | Publisher Full Text | Free Full Text

Krackow S, Vannoni E, Codita A, *et al.*: **Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage.** *Genes Brain Behav.* 2010; **9**(7): 722–731.
PubMed Abstract | Publisher Full Text

Krakenberg V, Woigk I, Garcia Rodriguez L, *et al.*: **Technology or ecology? New tools to assess cognitive judgement bias in mice.** *Behav Brain Res.* 2019; **362**: 279–287.
PubMed Abstract | Publisher Full Text

Krohn TC, Sørensen DB, Ottesen JL, *et al.*: **The effects of individual housing on mice and rats: A review.** *Animal Welfare.* 2006; **15**(4): 343–352.
Reference Source

Lagisz M, Zidar J, Nakagawa S, *et al.*: **Optimism, pessimism and judgement bias in animals: A systematic review and meta-analysis.** *Neurosci Biobehav Rev.* 2020; **118**: 3–17.
PubMed Abstract | Publisher Full Text

Lattal KA: **Delayed Reinforcement of Operant Behavior.** *J Exp Anal Behav.* 2010; **93**(1): 129–139.
PubMed Abstract | Publisher Full Text | Free Full Text

Lenth R: **Estimated Marginal Means, aka Least-Squares Means.** 2020.
Reference Source

Lewejohann L, Schwabe K, Häger C, *et al.*: **Impulse for animal welfare outside the experiment.** *Lab Anim.* 2020; **54**(2): 150–158.
PubMed Abstract | Publisher Full Text | Free Full Text

Lidster K, Owen K, Browne WJ, *et al.*: **Cage aggression in group-housed laboratory male mice: an international data crowdsourcing project.** *Sci Rep.* 2019; **9**(1): 15211.
PubMed Abstract | Publisher Full Text | Free Full Text

Manouze H, Ghestem A, Poillerat V, *et al.*: **Effects of single cage housing on stress, cognitive, and seizure parameters in the rat and mouse pilocarpine models of epilepsy.** *eNeuro.* 2019; **6**(4): ENEURO.0179-18.2019.
PubMed Abstract | Publisher Full Text | Free Full Text

Matheson SM, Asher L, Bateson M: **Larger, enriched cages are associated with 'optimistic' response biases in captive European starlings (*Sturnus vulgaris*).** *Appl Anim Behav Sci.* 2008; **109**(2–4): 374–383.
Publisher Full Text

Mei J, Kohler J, Winter Y, *et al.*: **Automated radial 8-arm maze: A voluntary and stress-free behavior test to assess spatial learning and memory in mice.** *Behav Brain Res.* 2020; **381**: 112352.
PubMed Abstract | Publisher Full Text

Mendl M, Burman OHP, Parker RMA, *et al.*: **Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms.** *Appl Anim Behav Sci.* 2009; **118**(3–4): 161–181.
Publisher Full Text

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.* 2021; **11**(10): 3002.
PubMed Abstract | Publisher Full Text | Free Full Text

Mitchell D, Koleszar AS, Scopatz RA: **Arousal and T-maze choice behavior in mice: A convergent paradigm for neophobia constructs and optimal arousal theory.** *Learn Motiv.* 1984; **15**(3): 287–301.
Publisher Full Text

Novak J, Bailoo JD, Melotti L, *et al.*: **An exploration based cognitive bias test for mice: Effects of handling method and stereotypic behaviour.** *PLoS One.* 2015; **10**(7): e0130718.
PubMed Abstract | Publisher Full Text | Free Full Text

Novak J, Stojanovski K, Melotti L, *et al.*: **Corrigendum to "Effects of stereotypic behaviour and chronic mild stress on judgement bias in laboratory mice" [Appl. Anim. Behav. Sci. 174 (2016) 162–172].** *Appl Anim Behav Sci.* 2016; **179**: 104.
Publisher Full Text

Paul ES, Harding EJ, Mendl M: **Measuring emotional processes in animals: The utility of a cognitive approach.** *Neurosci Biobehav Rev.* 2005; **29**(3): 469–491.
PubMed Abstract | Publisher Full Text

Pinheiro J, Bates D, DebRoy S, *et al.*: **Linear and Nonlinear Mixed Effects Models**. 2020.
Reference Source

RCoreTeam: **A language and environment for statistical computing**. 2020.
Reference Source

Resasco A, MacLellan A, Ayala MA, *et al.*: **Cancer blues? A promising judgment bias task indicates pessimism in nude mice with tumors.** *Physiol Behav.* 2021; **238**: 113465.
PubMed Abstract | Publisher Full Text

Richter SH, Schick A, Hoyer C, *et al.*: **A glass full of optimism: Enrichment effects on cognitive bias in a rat model of depression.** *Cogn Affect Behav Neurosci.* 2012; **12**(3): 527–542.
PubMed Abstract | Publisher Full Text

Rivalan M, Munawar H, Fuchs A, *et al.*: **An automated, experimenter-free method for the standardised, operant cognitive testing of rats.** *PLoS One.* 2017; **12**(1): e0169476.
PubMed Abstract | Publisher Full Text | Free Full Text

Roelofs S, Boleij H, Nordquist RE, *et al.*: **Making decisions under ambiguity: Judgment bias tasks for assessing emotional state in animals.** *Front Behav Neurosci.* 2016; **10**: 119.
PubMed Abstract | Publisher Full Text | Free Full Text

Sarna JR, Dyck RH, Whishaw IQ: **The Dalila effect: C57BL6 mice barber whiskers by plucking.** *Behav Brain Res.* 2000; **108**(1): 39–45.
PubMed Abstract | **Publisher Full Text**

Schlüns H, Welling H, Federici JR, *et al.*: **The glass is not yet half empty: agitation but not *Varroa* treatment causes cognitive bias in honey bees.** *Anim Cogn.* 2017; **20**(2): 233–241.
PubMed Abstract | **Publisher Full Text**

Verbeek E, Ferguson D, Lee C: **Are hungry sheep more pessimistic? The effects of food restriction on cognitive bias and the involvement of ghrelin in its regulation.** *Physiol Behav.* 2014; **123**: 67–75.
PubMed Abstract | **Publisher Full Text**

Verjat A, Devienne P, Rödel HG, *et al.*: **More exploratory house mice judge an ambiguous situation more negatively.** *Anim Cogn.* 2021; **24**(1): 53–64.
PubMed Abstract | **Publisher Full Text**

Voikar V, Gaburro S: **Three Pillars of Automated Home-Cage Phenotyping of Mice: Novel Findings, Refinement, and Reproducibility Based on Literature and Experience.** *Front Behav Neurosci.* 2020; **14**: 575434.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Voikar V, Krackow S, Lipp HP, *et al.*: **Automated dissection of permanent effects of hippocampal or prefrontal lesions on performance at spatial, working memory and circadian timing tasks of C57BL/6 mice in IntelliCage.** *Behav Brain Res.* 2018; **352**: 8–22.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Wickham H: **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016.
Reference Source

Winter Y, Schaefers ATU: **A sorting system with automated gates permits individual operant experiments with mice from a social home cage.** *J Neurosci Methods.* 2011; **196**(2): 276–280.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 03 October 2023

https://doi.org/10.21956/openreseurope.17783.r34338

© **2023 Knapska E.** 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.

✔ **Ewelina Knapska**
Laboratory of Emotions' Neurobiology, Center of Excellence for Neural Plasticity and Brain Disorders BRAINCITY, Nencki Institute of Experimental Biology, Warsaw, Poland

The authors successfully addressed all my concerns.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Behavioral Neuroscience, working with mice and rats. I have been working with the IntelliCage system since 2004.

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.

Reviewer Report 15 August 2023

https://doi.org/10.21956/openreseurope.17783.r34337

© **2023 Wolfer D.** 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.

✔ **David P. Wolfer**
[1] Institute of Anatomy, Division Functional Neuroanatomy, University of Zurich, Zurich, Switzerland
[2] Institute of Human Movement Sciences and Sport, D-HEST, ETH Zurich, Zurich, Switzerland

The authors have addressed all my concerns.

Competing Interests: Conflict of interest statement. For the sake of transparency, I would like to

mention that last author Lars Lewejohann and the reviewer both participate in the COST action CA20135 TEATIME. However, we are not actively collaborating or working on common publications. In 2011, were both authors of the publication of a multi-lab study (PLoS One. 2011 Jan 31;6(1):e16461.doi: 10.1371/journal.pone.0016461.), but this did not involve direct interactions between us. Further, the reviewer participated in the development of the original idea of the IntelliCage system and acted as a scientific advisor of NewBehavior, the company which commercialized IntelliCage until it was sold to the company TSE Systems. I never had nor do I currently have any commercial interest in the IntelliCage system or in any competing product. In conclusion, my review of this work was not hindered by a conflict of interest.

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 14 February 2023

https://doi.org/10.21956/openreseurope.16536.r30671

© **2023 Knapska E.** 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.

?

Ewelina Knapska
[1] Laboratory of Emotions' Neurobiology, Center of Excellence for Neural Plasticity and Brain Disorders BRAINCITY, Nencki Institute of Experimental Biology, Warsaw, Poland
[2] Laboratory of Emotions' Neurobiology, Center of Excellence for Neural Plasticity and Brain Disorders BRAINCITY, Nencki Institute of Experimental Biology, Warsaw, Poland

The manuscript by Pia Kahnau *et al.* describes an attempt to develop a method of measuring cognitive bias in mice using the IntelliCage. The cognitive bias test measures how emotional states affect cognitive processes. The test uses classical or operant conditioning to measure responses to ambiguous cues. It can be used to measure animals' welfare state, as those in better welfare make more optimistic judgments. The cognitive bias test is well-described in many species, including mice. The novelty of the studies described in the manuscript is using the IntelliCage, an automated cage for assessing mice behavior and learning, to develop a new protocol for measuring cognitive bias. Developing a reliable protocol would be useful as it would be a valuable tool to assess the welfare of the laboratory mice and study mechanisms of cognitive bias. The authors present three experiments through which they came closer to designing optimal test conditions. The methods are well described, and all the necessary experimental details are provided. However, the aim of that study, i.e., the development of a reliable cognitive bias test for mice in the IntelliCage, has not been achieved. Thus, additional experimental work is needed.

Major comments:

1. The mice were only tested for cognitive bias in the third experiment, in which a shift toward more optimistic decisions was observed after a supposedly stressful procedure. This effect was maintained even when the animals were not stressed, i.e., their responses did not return to the baseline. This result should be carefully considered. One of the possible scenarios the authors did not discuss is that the procedure they used to stress the mice was actually not aversive. Removing some elements of the environment which was already enriched (the mice had access to a large territory, had to learn a new behavior, etc.) and adding very mild restrain changed the behavior of mice in the direction opposite to the predicted one. It could be the effect of positive cognitive bias. Still, the change conditions could also affect learning as the response curve did not return to baseline. I recommend testing the protocol using conditions already known to induce cognitive bias.

2. To show the reliability of the cognitive bias protocol, it should be tested on more than one cohort of mice.

3. As during the training and testing only one mouse could enter the IntelliCage at the same time (when another animal did not already occupy it), it would be good to show, for each animal, how often it entered the training/testing chamber. Such data may explain the relatively high drop-out rate of animals that were unable to learn (if some mice waited for another trial too long, it might have affected memory consolidation).

4. The test has been developed for females only. The authors should discuss how to develop such a test for male mice.

Minor comments:

1. It would be helpful for the reader to get information that mice are individually recognized in the IntelliCage system thanks to the transponders they carry early in the text (e.g., in the Introduction).

2. It needs to be clarified whether, during habituation to nose-poking, all mice had access to the IntelliCage or whether they went there one by one as during the following stages.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the

findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Behavioral Neuroscience, working with mice and rats. I have been working with the IntelliCage system since 2004.

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 01 Aug 2023

Pia Kahnau

First of all, we would like to thank you for taking the time to review our manuscript in the open review process. The summary is concise and the clearly formulated comments will help to improve the manuscript.

Major comment 1: It is true that the data suggest a more optimistic emotional state due to the supposedly more negative experiences (removal of enrichment and repeated confinement). In rats and also in mice, housing conditions have been shown to influence the cognitive biases of animals. However, in our study, the group of mice used was their own control. Based on the results, one has to ask whether a group of mice can be the control at the same time, whether it is even possible to measure a before-after effect with our experimental design (see discussion on page 28). The extent to which the removal of the enrichment actually has a negative effect on the mice has indeed not been conclusively clarified. However, restraining is a paradigm regularly used to inflict negative experience and we assume that it is not too bold to claim that the restrain test was a negative experience for the mice, at least for the duration of the restraining. We observed that after the first trial on the first day, the mice did not enter the restrain tube voluntarily and showed clear avoidance behavior in the form of reduced interaction with the experimenter's hand. Unfortunately, we can only report on this behavior anecdotally. However, it could be that the end of the restraint test had a positive effect on the cognitive bias, along the lines of "now everything is fine again", and may even have overridden the influence of the removal of the enrichment. We have added this consideration in the discussion (page 27). It is also possible that the mice learned that nosepoking in response to the ambiguous tones had no consequences. This point was elaborated in the discussion (page 28).

Major comment 2: We agree. Therefore, another study is planned in which two groups (a treatment and a control group) will be tested simultaneously. Nevertheless, in the presented study we were already able to measure the cognitive bias of the mice in an automated and home cage based set-up, although, it was not possible to clearly manipulate the cognitive bias. Nevertheless, we believe it is important to describe the steps involved in measuring the cognitive bias, as this information could potentially be helpful in the development of more home cage-based experiments. We consider this paper as a methods

paper rather than a results paper.

Major comment 3: Thank you for this comment. It is theoretically possible that some mice were not able to enter the IntelliCage or were able to enter less frequently and were therefore presented with the tones less often during the conditioning phase. We noticed that all mice were able to enter the IntelliCage several times per night. Once a mouse was in the IntelliCage, it had the ability to repeatedly enter the corner of the IntelliCage and may have been able to compensate for fewer IntelliCage entries. We have added two figures to the supplements showing that the mice were able to enter both the IntelliCage and the IntelliCage corner several times. We assume that the mice that had not learned to discriminate between the tones had an aversion to the possible punishment and not that they had not learned because they could enter the IntelliCage less often (discussion page 18).

Major comment 4: Thank you for mentioning this, it is true that we have only used females so far. The results may not be transferable to males in a one-to-one manner, so it is useful and necessary to test both the set-up and the experiment with males. In the methods section (page 6), we have explained in more detail the need to pay attention to behavior in experiments with males. On the one hand, that individual males do not occupy the IntelliCage and prevent other mice from entering, and on the other hand, agonistic behavior needs to be monitored in order to intervene as quickly as possible in case of injury.

Minor comment 1: We agree that the information about the individual recognizability of the IntelliCage system through RFID antennas and transponders should already be mentioned in the introduction and helps the reader to understand the functionality of the system. Therefore, we have included this information in the introduction (page 4).

Minor comment 2: That is an important hint, thank you very much. Especially the habituation of the set-up is important and should be described in the manuscript. That is why we have added the information in the methods section (page 8).

Competing Interests: No competing interests were disclosed.

Reviewer Report 09 February 2023

https://doi.org/10.21956/openreseurope.16536.r30672

© 2023 Wolfer D. 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.

?

David P. Wolfer
[1] Institute of Anatomy, Division Functional Neuroanatomy, University of Zurich, Zurich, Switzerland
[2]

Institute of Human Movement Sciences and Sport, D-HEST, ETH Zurich, Zurich, Switzerland
[3] Institute of Anatomy, Division Functional Neuroanatomy, University of Zurich, Zurich, Switzerland
[4] Institute of Human Movement Sciences and Sport, D-HEST, ETH Zurich, Zurich, Switzerland

Pia Kahnau and co-authors describe the development and validation of a test protocol which allows to assess emotional state in mice using the IntelliCage system. The concept of the test is that in a training stage the mice first learn to discriminate two acoustic stimuli and to associate one of them with a reward (delivery of water) and the other with a punishment (delivery of an air puff). In the subsequent test stage, the mice are confronted with ambiguous stimuli and their bias toward interpreting them as predicting reward or punishment is taken as a readout of their emotional state. This concept of testing emotional state by measuring cognitive bias is well established. New is the presentation of a protocol that works in the automated IntelliCage system which allows to test the mice with minimal handling by humans and without social deprivation. In the first two steps of development, the authors established the training stage of the protocol in which mice reliably and measurably learned to discriminate two acoustic stimuli. In the third step, the test stage of the protocol was developed and tested by confronting mice with ambiguous stimuli after they had been stressed by repeated restraint and sensory deprivation.

The development of this new protocol is well described and documented. The presented data demonstrate that it has the potential to become a new test of cognitive bias which is as animal-friendly as possible, but a final convincing proof that it works reliably and is indeed suitable for this purpose requires additional experimental work.

Major points:

1. My principal concern with this paper is that the conclusion that using the presented protocol the authors “were able to measure and manipulate the cognitive bias” is overstated. As the authors correctly note, further research is needed, but not only to “for a better understanding of the mice’s cognitive bias” but to establish convincingly that that the protocol is indeed suitable for the intended purpose. As already pointed out in the Discussion, there are two major problems with the data. First, there was uncertainty what direction of bias to expect after the stress manipulation used in the experiment. Second, it could not be demonstrated that the mice returned to baseline after recovery. Thus, the possibility remains that the mice learned during establishment of the first baseline that the new stimuli do not predict punishment and therefore showed a positive bias in subsequent tests. So, the observed cognitive bias may reflect learning rather than emotional state. I would suggest to re-test the protocol while (i) using a well-established way to change emotional state such as protocols used in depression models and (ii) using a cross-sectional design in which an experimental and a control group can be directly compared while they are still naïve to the ambiguous stimuli.

2. Only female mice were tested. While testing male subjects is certainly beyond the scope of the present study, it needs to be acknowledged that follow up studies will be needed to demonstrate that the method works independently of sex.

Minor points:

1. There was at least one previous attempt to measure cognitive bias in the IntelliCage system, published in Alboni *et al.* (2017[1]). I would appreciate it if the authors could critically comment on these attempts in the Introduction.

2. I could not find a formal power analysis of the experimental designs used in this work, but I would assume that with the given numbers of subjects and drop-outs it would be difficult to statistically demonstrate absence of effect. So, negative test outcomes essentially remain inconclusive. Therefore, the authors should tone down their interpretation of negative test outcomes and state that there was no evidence for an effect or that time points/stimuli were indistinguishable instead of claiming that there was no effect.

3. In several instances, the authors merely state that there was a significant effect. They should also give a brief description of what the effect actually was, so that its meaning can be appreciated by the reader. Especially with interaction effects this is not always self-evident.

4. Typos: p20 “It is discussed wheather”, p21 “the period of negative conditions was significantly lower” shorter?

5. I assume that the specific IntelliCage protocols would be made available upon request.

References

1. Alboni S, van Dijk RM, Poggini S, Milior G, et al.: Fluoxetine effects on molecular, cellular and behavioral endophenotypes of depression are driven by the living environment.*Mol Psychiatry*. 2017; **22** (4): 552-561 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: Conflict of interest statement. For the sake of transparency, I would like to mention that last author Lars Lewejohann and the reviewer both participate in the COST action CA20135 TEATIME. However, we are not actively collaborating or working on common publications. In 2011, were both authors of the publication of a multi-lab study (PLoS One. 2011

Jan 31;6(1):e16461.doi: 10.1371/journal.pone.0016461.), but this did not involve direct interactions between us. Further, the reviewer participated in the development of the original idea of the IntelliCage system and acted as a scientific advisor of NewBehavior, the company which commercialized IntelliCage until it was sold to the company TSE Systems. I never had nor do I currently have any commercial interest in the IntelliCage system or in any competing product. In conclusion, my review of this work was not hindered by a conflict of interest.

Reviewer Expertise: Behavioral Neuroscience, mostly using mice as disease models. Have been working with the IntelliCage system for almost 2 decades now.

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 01 Aug 2023

Pia Kahnau

Thank you for the clear summary and for agreeing to review our manuscript in the open peer review process. Your comments address important aspects of the limitations of our work and the editing will help to improve our manuscript. We understand the objection that our claim that we "were able to measure and manipulate the cognitive bias" is overstated and we adjusted our interpretation of the data accordingly.

Major point 1: Regarding the "uncertainty what direction of bias to expect": In the discussion of developmental step 3 we state that we expect a negative cognitive bias, but we agree that this expectation can already be mentioned earlier. Therefore, we added the sentence "In accordance with studies in rats, we expect the cognitive bias of the mice to be negatively affected by removal" on page 20. Regarding the concern that "it could not be demonstrated that the mice returned to baseline after recovery": It is true, that the mice in the cognitive bias test 1 do not return to the baseline level. This could indeed be an indication of learning. It appears that the mice have learned that they will not be rewarded or punished for the ambiguous tone frequencies. The nosepoke behavior also appears to be higher in the second cognitive bias measurement compared to baseline 1 in the cognitive bias test 1. The mice seem to be nosepoke more overall over time. Whether this is due to a more optimistic emotional state, or whether the mice have learned that ambiguous tones will not be punished, cannot be determined from the data presented here. Rather, it seems that a before and after comparison is not possible in this set-up. The conclusion that one mouse group seems to be unsuitable as its own control was added to the discussion of developmental step 3 (page 28). In a follow-up study, two groups with different experiences will be tested simultaneously and the cognitive bias will be compared. However, the sigmoidal curve is maintained in each measurement and the mice nosepoke more to the near positive tones than to the mid tones. This suggests that the mice are evaluating whether or not to make nosepokes. If the mice had learned that they would not be punished on the ambiguous tones, I would expect a much higher rate of nosepokes (close to 100%) in response to the ambiguous frequencies, based on experience from other IntelliCage experiments (Kahnau et al 2021). Nevertheless, learning cannot be completely

ruled out here, so the claim that we were able to manipulate the cognitive bias was significantly weakened. *Kahnau P, Guenther A, Boon MN, et al.: Lifetime Observation of Cognition and Physiological Parameters in Male Mice. Front Behav Neurosci. 2021;15:709775. 34539359 10.3389/fnbeh.2021.709775 8442583*

Major point 2: We agree that it is necessary to test the set-up and method additionally with males. Therefore, we have added to the methods section on page 6 that male mice should also be tested in this set-up to measure their Cognitive Bias. In addition, we pointed out possible experimental difficulties in tests with male mice in such a set-up.

Minor point 1: Thank you for pointing out this paper, which we have included in our introduction (page 3) and discussion (page 26, 27). It is true that in Alboni et al. (2017) (among other results) the result of an IntelliCage-based test of cognitive bias are shown. For this test, mice were conditioned to light (blue and yellow). The stimuli were presented every time a corner was visited, as in our study, which had the advantage that the mice could initiate each trial themselves and thus experienced more self-determination, leading to an improvement in the experimental conditions. The focus of this paper was certainly not to describe the development of the method. Therefore, some parameters that would have been helpful for the evaluation of the study are unfortunately missing. For example, neither the paper nor the supplements provide more detailed information on the number of animals that could not be successfully conditioned. Unfortunately, it is also not clear how many IntelliCages were used in total and how many animals were kept as a social group in an IntelliCage. However, the density of a cage could have an effect on learning, as mice can pull, push or shove each other out of corners. A disadvantage of the cognitive bias test shown in this study is that only one ambiguous stimulus is presented, probably due to the fact that the IntelliCage has only three different lights (blue, yellow and red). Because only one ambiguous stimulus is presented, it is not possible to plot the data on a sigmoid curve. In such a plot of the data, it is not possible to see how well the mice discriminate between the conditioned stimuli and how they interpret the ambiguous stimulus. According to Lagisz and colleagues, even a single ambiguous stimulus seems to reduce the sensitivity of the test (Lagisz et al. 2020). Unfortunately, it is also not stated from how many sessions the cognitive bias test results were obtained. Nevertheless, a difference in emotional state was found between two groups of mice. An enriched environment combined with fluoxetine treatment seems to have a positive effect on the emotional state of depressed mice compared to mice that lived in the same environment but were not treated with fluoxetine. We have not yet been able to clearly demonstrate a manipulation of cognitive bias with our method. *Lagisz M, Zidar J, Nakagawa S, et al.: Optimism, pessimism and judgement bias in animals: A systematic review and meta-analysis. Neurosci Biobehav Rev. 2020;118:3–17. 32682742 10.1016/j.neubiorev.2020.07.012*

Minor point 2: As this was an exploratory study, it was not possible to estimate the effect size and power in advance. Especially at the beginning, it was not clear how the mice would behave in this experimental design. We decided on 12 animals per group (and developmental stage) because it is possible to keep 12 mice in a type IV cage and still provide different enrichments. RFID technology also allows individual data to be obtained. We agree that the number of animals used might seem small and that the statements on negative results regarding the results of the cognitive bias tests should be toned down. We

did so and we have added some paragraphs under “Animals and housing conditions” (page 5), “discussion” (page 28) elaborating on the group size.

Minor point 3: We thank you for pointing this out. We have described the results more clearly in the results section.

Minor point 4: Thank you for these hints. We have corrected the typos.

Minor point 5: Of course, the IntelliCage protocols will be made available if there is interest.

Competing Interests: No competing interests were disclosed.