Mechanical problem solving in mice

Marcus N. Boon[1,2,*] Niek Andresen[1,2,4,*] Soledad Traverso[1,2,*]
Sophia Meier[1,2] Friedrich Schuessler[1,2] Olaf Hellwich[1,2]
Lars Lewejohann[1,3,4] Christa Thöne-Reineke[1,4] Henning Sprekeler[1,2,†]
Katharina Hohlbaum[1,3,†]

[1]Cluster of Excellence "Science of Intelligence", Berlin, Germany
[2]Technische Universität Berlin, Germany
[3]Bundesinstitut für Risikobewertung, Berlin, Germany
[4]Freie Universität Berlin, Germany
∗, † Equal contribution
† Corresponding authors
(katharina.hohlbaum@bfr.bund.de, h.sprekeler@tu-berlin.de)

Abstract

Recent advances in automated tracking tools have sparked a growing interest in studying naturalistic behavior. Yet, traditional decision-making tasks remain the norm for assessing learning behavior in neuroscience. We introduce an alternative sequential decision-making task for studying mouse behavior. It consists of an open-source, 3D-printed "lockbox", a mechanical riddle that requires four different mechanisms to be solved in sequence to obtain a reward. During the task, the mice move around freely, allowing the expression of complex behavioral patterns. We observed that mice willingly engage in the task and learn to solve it in only a few trials. To analyze how the mice solved the task, we recorded their behavior in a multi-camera setup and developed a custom data analysis pipeline to automatically detect the interactions of the mice with the different lockbox mechanisms for a large corpus of video footage (> 300h, 12 mice). The pipeline allows us to further delineate why mouse performance increases over trials. Our analyses suggest that this is not due to an increased interaction time with the task or the acquisition of a smart solution strategy, but primarily due to habituation to the lockbox. Lockboxes may hence be a promising approach to study both abstract sequential decision making and low-level motor learning in a single task that can be rapidly learned by mice.

Keywords: mice, lockbox, mechanical problem solving, decision making, sequential problem, automatic video analysis

1 Introduction

Systems neuroscience has traditionally relied on simplified experimental setups, in which animals – often mice – are trained to react to relatively simple stimuli, often with constraints in their physical movements. This approach has been very successful because it simplifies the analysis of both behavior and neural data, but it also comes with challenges. The simplicity of the task constrains the rich and high-dimensional repertoire of neural activity that is observed during natural behavior [7, 24], and despite its apparent simplicity, the animals often need many trials to learn the task. Unfortunately, the latter is often true for mice, which are one of the most widely used models systems in neuroscience due to the broad availability of genetic tools.

Driven by the ability to record from many neurons simultaneously and the development of new automated tracking tools [17, 22, 23, 25], there has been a recent trend towards more complex experiments with freely moving animals [4, 5, 13–15]. Deep learning-based software now allows researchers to analyze animal behavior in more naturalistic settings, capturing the full spectrum of ways an animal might engage with its environment in an automated way. However, the increased freedom in these experiments comes with the challenge that it can be difficult to define appropriate task variables for analysis, because

the internal goals of the animal are often unknown.

One method to gain control over the intentions of a freely moving animal is to use foraging tasks. Inspired by Thorndike's [27] work and by research on birds [1, 2, 10, 19, 26] and other species [3, 12], we designed a so-called "lockbox" for mice. In contrast to most mouse tasks that involve a single decision [6, 18, 20, 28], the lockbox requires mice to solve a sequence of four mechanical problems to reach the goal, without intermediate rewards (Figure 1a). The task thereby enables an anchoring of analyses to well-defined points within the task, without otherwise constraining behavior. In a cohort of 12 mice, we find that the animals voluntarily engage in the task and perform it successfully within very few trials.

To understand how the mice solved the problem, we developed a computational pipeline to analyze multi-perspective video data and quantify the interactions of the mice with the lockbox. Our data suggest that the rapid improvement in performance is not due to an increasing interaction time with the task but to a habituation to the lockbox, and that the mice appear to solve the task using a random exploration strategy. The lockbox can also be used as enrichment in laboratory mouse housing [9] and the design for 3D-printing is open-source [8].

2 Results

2.1 The lockbox task

To study complex problem-solving in mice, we exposed them to a sequential lockbox problem. The problem consists of four interlocked mechanical problems, or "mechanisms", which must be solved sequentially. The mouse can collect a reward upon opening a sliding door, but the door is blocked by a ball that cannot be removed before a stick is pulled out, which in turn cannot be pulled out before a lever is raised (Figure 1a). All four mechanisms are designed to be easily movable by the mice. The movements required to solve all different mechanisms are sufficiently large to distinguish intentional actions from accidental touches. The mechanical interconnections between the mechanisms are visible to the mice, allowing them to observe the effects of their actions and to, in principle, infer which mechanism needs to be solved next. The animals move freely and can decide to engage with or disengage from the task at any time voluntarily. We often observed anticipatory queuing of the mice at the entrance to the cage where they could perform the task (an exemplary recording of this behavior can be found on GitHub[1]).

We used a three-stage curriculum to train a cohort of mice (n=12). Each mouse was first exposed to the complete lockbox problem once. This was followed by a single mechanism training (SMT) phase of eleven trials, during each of which the mouse could solve all individual mechanisms, rewarded independently, in a randomized order (Figure 1b). Finally, the full lockbox was presented again five times on subsequent days (stage 3, Figure 1c). In each full lockbox trial, the mouse had 30 minutes to open the lockbox and collect a reward; otherwise, the trial was deemed unsuccessful. For more details of the training paradigm, see subsection 5.6 and Hohlbaum et al. [9].

All mice solved the full lockbox task within very few exposures (Figure 1d). When the mice encountered the task for the first time, only three out of twelve could solve it. After the SMT phase, there was a sharp rise in the fraction of animals that succeeded (9/12), and this fraction continued to rise slowly across further trials except for a decrease by one count in the last trial. After the first trial, all mechanisms (lever, stick, ball, door) were opened with high probability once they were unlocked by opening the previous mechanisms (lever: 93%, stick 91%, ball: 100%, door: 98%, the lever being unlocked already at the start of the trial).

The low performance in the first trial is unlikely due to a lack of motor skills, because almost all mice were able to solve all four individual mechanisms in the first trial in the SMT phase. During this phase, the mice learned to solve the individual mechanisms more reliably (Figure 1e) and faster (Figure 1f). Notably, the mice consistently needed more time for the first mechanism of each SMT trial, although they were presented in randomized order (Figure 1g). Hence, the decreasing time for the single mechanisms may in part be explained by habituation. Consistent with this hypothesis, the single mechanism training phase also strongly reduced the latency at which the mice first approached the full lockbox (Figure 1h).

[1] `https://github.com/RefinementReferenceCenter/MouseLockBox/blob/main/Door_system/Door_system.mp4`

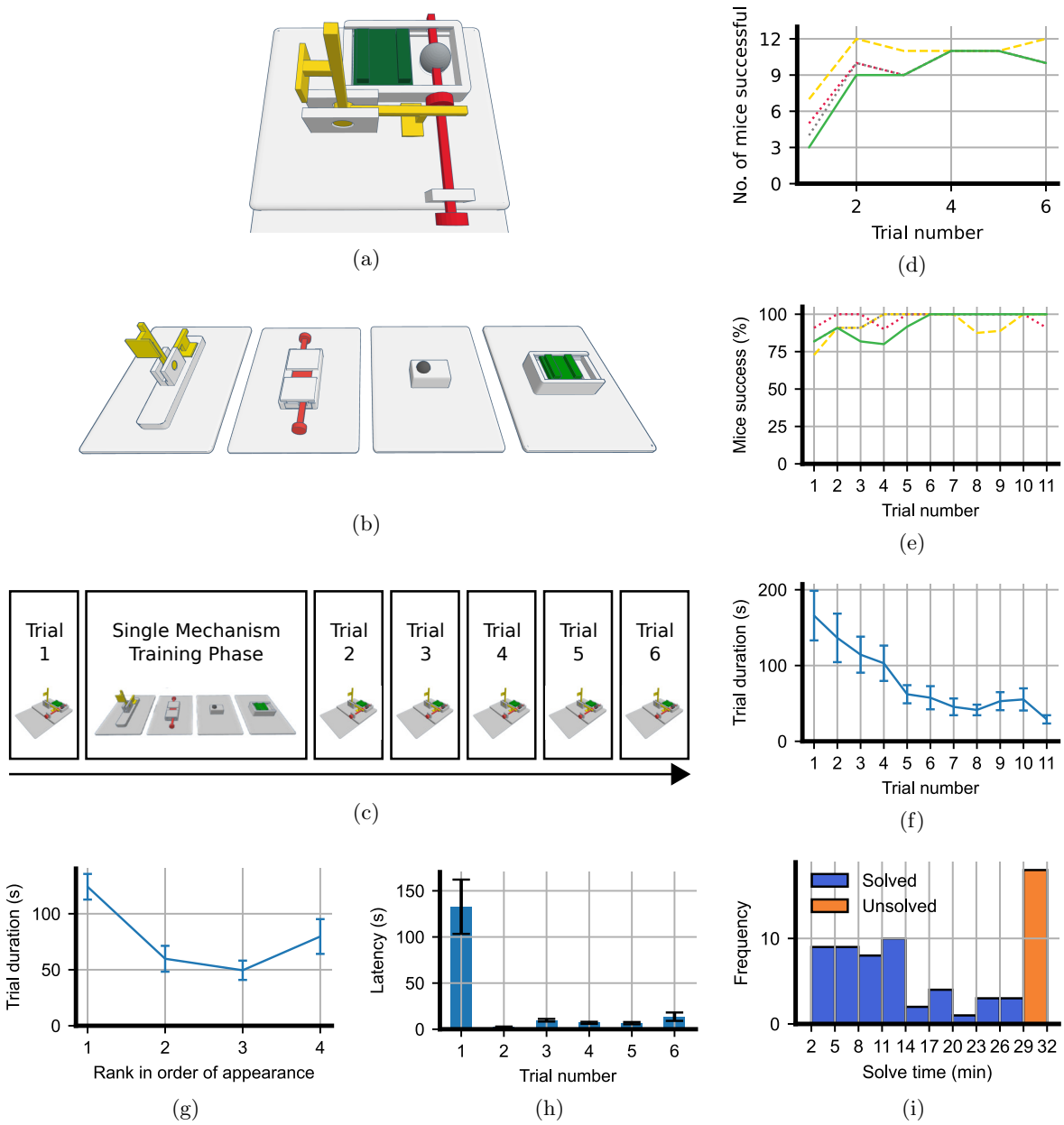

Figure 1: **The lockbox task.** **(a)** Design of the combined lockbox. The yellow lever blocks the red stick blocking the gray ball which blocks the green door concealing a food reward underneath. **(b)** Design of the individual mechanisms in isolation. Each mechanism reveals the reward after a single successful action. **(c)** Mouse training curriculum. Each individual full lockbox trial was performed on one separate day. The single mechanism training took two weeks. **(d)** Learning curves for the mechanisms in the combined lockbox. The number of successful mice was counted for each trial. **(e)** Learning curves for the mechanisms in the single mechanism training phase. The proportion of successful mice was computed for each trial. **(f)** Trial duration of single mechanism trials in seconds. The mean and standard error of the mean are taken over all mice and mechanisms for each trial. **(g)** Mean trial duration in order of appearance. The mean trial duration was computed conditioned on the order in which a mechanism was presented in a trial. **(h)** Start latency (in seconds) for each trial, indicating the time taken for mice to initiate the task. **(i)** Histogram of the time taken by mice (in minutes) to solve the task and their frequencies.

Hence, the substantial performance increase in the second trial may not primarily be caused by skill acquisition, but also by habituation to the task.

The probability of a mouse solving the puzzle after having done so previously was high (≈ 0.9).

Specifically, 10 out of 12 mice consistently solved the puzzle after the first success; one mouse failed only once afterwards, and one mouse increasingly disengaged from the task after succeeding in the second trial.

Once mice learned to solve the lockbox, they typically completed the task in less than half the allocated time. The 16-minute mark was exceeded only by a minority (Figure 1i). This indicates qualitative behavioral differences in failure vs. success trials.

2.2 Automated tracking

To better understand how the mice solved the problem, we recorded the entire training process on video (>300h) and implemented an automated video analysis pipeline (Figure 2a). We recorded mouse behavior from three perspectives using infrared-sensitive cameras. In each video, we tracked key bodyparts of the mouse and using DeepLabCut (DLC, [17]). The keypoint positions in the three perspectives were combined into a 3D reconstruction using standard triangulation methods and nonlinear temporal filtering with a skeletal model of the mouse. From the resulting tracked positions of the mouse, we classified each video frame based on whether the mouse interacted with the lockbox and which mechanisms it was interacting with. To be able to correlate mouse behavior with the state of the lockbox, we also annotated which lockbox mechanisms were open.

To benchmark the accuracy of the video tracking, we compared the results of the tracking pipeline with manual annotations for two videos (Figure 2b). For each video, two independent annotators marked the interactions between the mouse and the lockbox as well as the changes in the state of the lockbox mechanisms throughout the video footage. The automated pipeline results and the manual annotations are highly correlated (Figure 2c).

2.3 Interactions with the lockbox

To identify behavioral differences between success and failure trials, we used the tracking pipeline to quantify how much the mice interacted with the lockbox. We find that failures are not explained by a lack of interactions with the lockbox. On average, the mice spent significantly more time interacting with the lockbox during failure than during success trials (727.1 s vs. 347.5 s, $p<.001$). However, the fraction of interaction time (persistence) was significantly smaller (40.7% vs. 54.2%, $p<.001$), suggesting that the mice were less persistent [11] and potentially less motivated in failure trials.

To further investigate if the increased performance is related to task learning or merely due to habituation, we quantified different behavior metrics across trials and found no evidence of behavioral changes. Given that mice appeared to have different levels of motivation during success and failure trials, we analyzed them separately by plotting the behavioral metrics as a function of how often the mouse had succeeded or failed, respectively. We find that the total interaction time (Figure 2d) does not change significantly as a function of the number of success and failure trials (linear regression to reject the null hypothesis of no change over time; success: $p = .787$; failure: $p = .923$), and neither does the persistence (Figure 2e, success: $p = .385$; failure: $p = .924$). The mice also barely change their preferences for the different lockbox mechanisms (Figure 2f, lever $p=.597$, stick $p=.009$ and slope=-.009, ball $p=.671$, door $p=.0171$ and slope=.009).

To investigate whether mice learned the task sequence, we transformed the data into a sequence of state-interaction pairs. To this end, we defined an interaction with a given mechanism of the lockbox as an uninterrupted period during which the mouse was close to the mechanism in question. For each interaction, we recorded the state of the lockbox by storing which mechanisms needed to be solved next. A "state-advancing interaction" is an interaction with a mechanism that, if successfully manipulated, would advance the lockbox to its next state. We then compared the fraction of state-advancing interactions in a given lockbox state to the overall frequency at which the mouse interacted with this mechanism. In solved trials, mice had significantly more state-advancing interactions with the stick, ball, and door than suggested by their overall preferences (Figure 2g, lever: $p = .923$, stick: $p = .02$, ball: $p < .01$, door: $p = .033$). This finding suggests that the actions of the mice with the stick, ball and door might be intentional and could be goal-directed or memorized rather than just random.

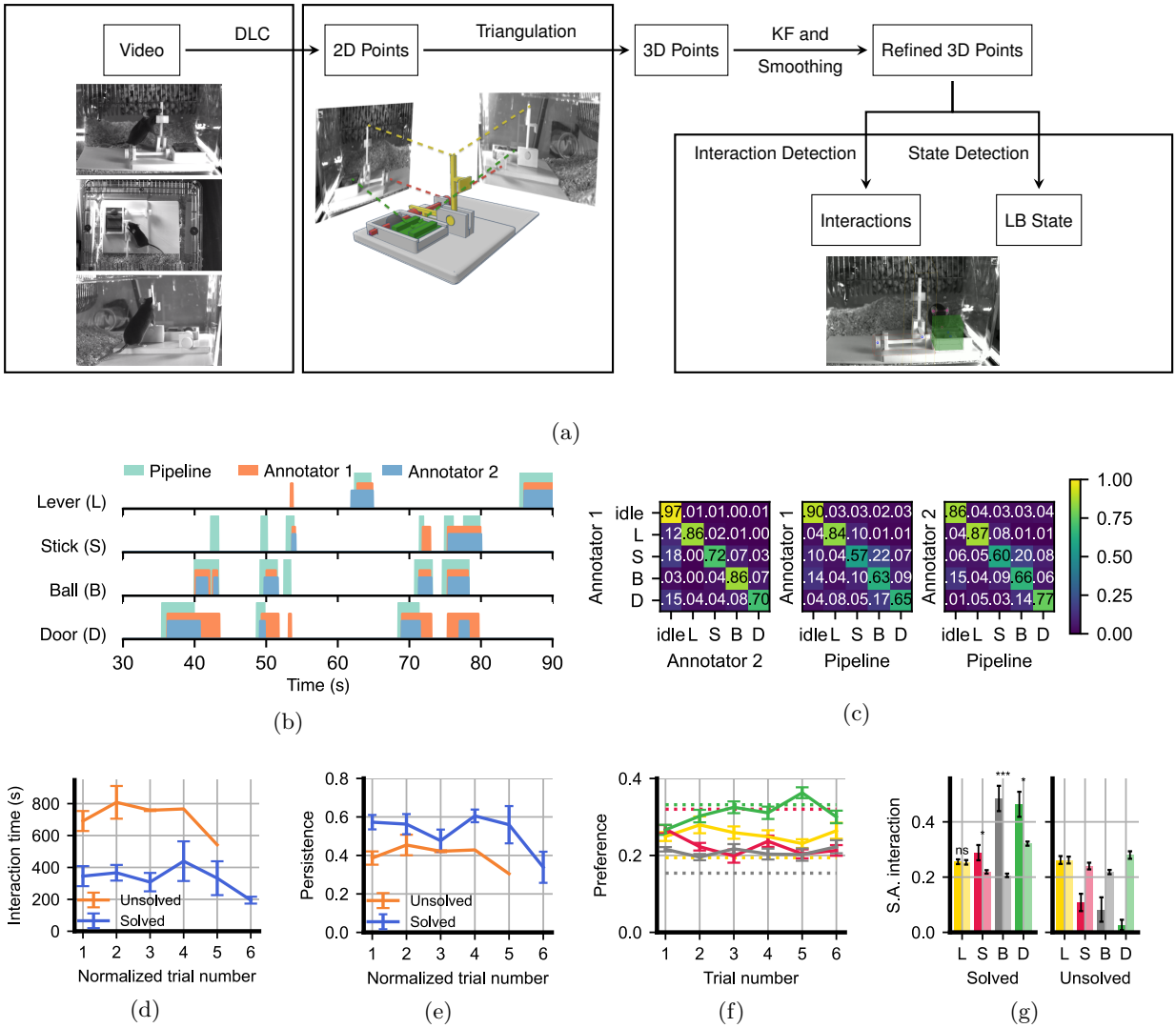

Figure 2: **The data processing pipeline.** **(a)** Schematic of the video analysis pipeline. 2D tracks are extracted using DeepLabCut. The scene is then reconstructed in 3D through triangulation. 3D tracks are refined with an (Extended) Kalman filter and smoother. Finally, mouse interactions and lockbox states are detected. **(b)** Comparison of the pipeline-detected interactions with two human annotators. An excerpt of a mouse's first 1.5-minute interaction from a trial is shown. Colored boxes indicate detected interactions between the mouse and the respective mechanism by the annotator or pipeline. **(c)** Confusion matrices comparing the detected interactions of the pipeline with two human annotators. Interaction data from two trials are used for the comparison. The labels "L", "S", "B", and "D" represent interactions with the lever, stick, ball, and door, respectively. The "idle" label represents the inactive time of the mouse. **(d)** Total required interaction time for solved and unsolved trials. **(e)** Persistence for solved and unsolved trials, calculated as the ratio of the interaction time to the total trial length. **(f)** Preferences over trials (solved and unsolved). The dotted-lines correspond to the preferences of a random model based on the areas of the lockbox mechanisms. **(g)** State-advancing interactions vs. stateless overall preferences. Fully opaque bars represent state-advancing interactions with each mechanism, while semi-transparent bars show the overall distribution of interactions with each mechanism across all mice and trials.

2.4 Behavioral strategies

To assess whether mice develop a strategy for solving the lockbox task, we performed a Bayesian analysis to infer which of two pre-defined strategic pattern best fits the behavioral data of the mice. We consider a "random strategy", in which the mice interact with the lockbox mechanisms independently of the state of the lockbox, while taking into account an initial prior preference (Figure 3a). The "smart"

strategy follows an ϵ-greedy approach in which the mouse interacts in a goal-directed manner, i.e., with the state-advancing lockbox mechanism with a probability of 1 - ϵ and interacts in a random manner with a probability of epsilon, where $\epsilon \in [0, 1]$. For $\epsilon = 1$, the "smart" strategy changes into the "random" strategy.

Each of these strategy models define conditional probabilities for the interactions with the different mechanisms, given the lockbox state and the strategy. This allows us to compute the likelihood of both the strategy (Figure 3c) and the exploration parameter ϵ (Figure 3e) given the interaction data, for each mouse and trial. For this inference, we only considered trials in which the animal solved the task, such that interaction data for all lockbox states are available. Again, we validated the tracking pipeline by comparing the inference results to those of the manually annotated trials (Figure 3b).

The random strategy is most consistent with the observed interaction data (Figure 3c). This may at first glance appear inconsistent with the over-representation of state-advancing interactions (Figure 2g), but it can be explained by the very different numbers of interactions in the different lockbox states. The mice often take a long time to open the first mechanism – the lever – but then quickly solve the remaining mechanisms, suggesting a switch from exploratory to exploitative behaviour. The number of interactions before the lever is opened outnumbers the overall number of interactions in later states (LB closed: 2995, lever open: 1154, stick open: 551, ball open: 308). In contrast, the frequency of state-advancing interactions in Figure 2g weights the different lockbox states equally. To test if this explains the discrepancy, we resampled the same number of interactions for each mouse and trial according to the empirical interaction frequencies in the different lockbox states, but distributed them evenly across the lockbox states. Performing the strategy inference on this surrogate data indeed increasingly favored the "smart" strategy (Figure 3d) over trials, albeit with relatively high exploration parameter ϵ (Figure 3e). In summary, the Bayesian analysis therefore suggests that the mice follow a largely random solution strategy to solve the problem, although we cannot rule out a transition from exploratory to strategic behavior.

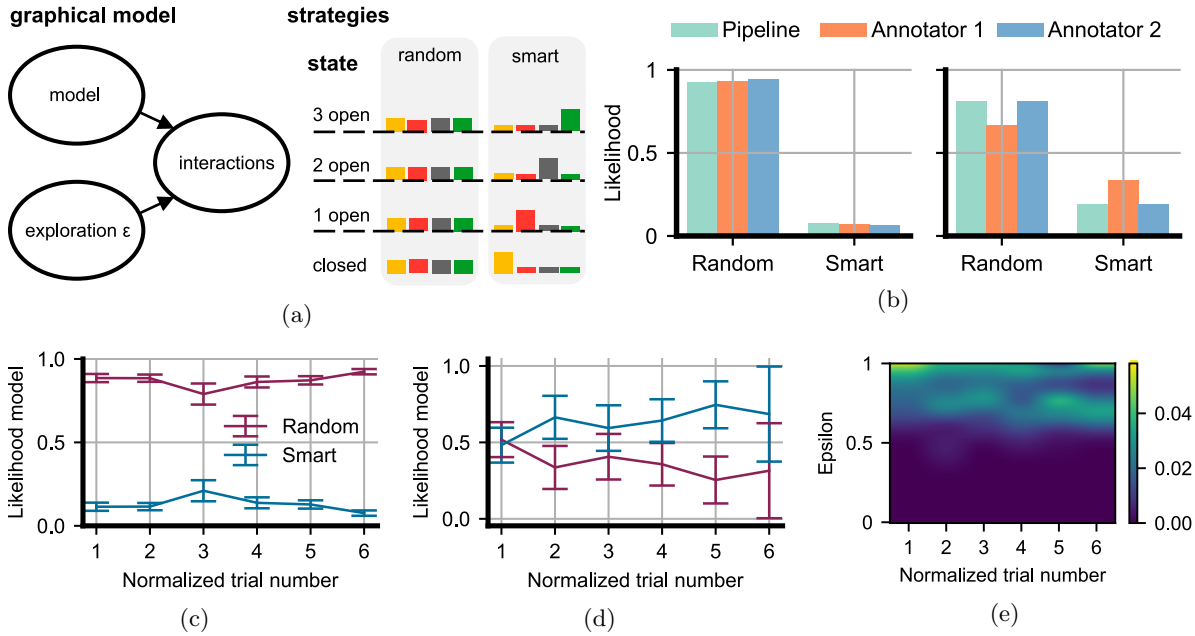

Figure 3: **Bayesian inference of mouse strategies.** **(a)** Graphical model for the Bayesian inference and visualization of the behavioral policies for the two strategies. **(b)** Comparison of model likelihoods for human-annotated data and the pipeline-generated interaction sequences for the two manually annotated trials. **(c, d)** Model likelihoods for the solved trials generated using c) the original interactions data and d) surrogate data with the same number of interactions per lockbox state. **(e)** Heat map of the likelihood of the exploration parameter ϵ for the solved trials, averaged across mice.

3 Discussion

In summary, we presented the lockbox as a novel sequential decision making paradigm for mice. After a short habituation period, the mice learned to solve it within a few trials and required little time to open it, outpacing the learning speed typically observed in conventional binary decision-making tasks.

Mice were more persistent in solved trials and performed more state-advancing interactions than in unsolved trials. However, once we factored out the influence of solved vs. unsolved trials, none of the studied behavioural metrics showed a significant change throughout the experiment. This suggests that most of the performance increase in the first and second trial may be accredited to the habituation to the lockbox rather than to learning.

Unlike many controlled experiments, the task does not confine the mice to limited choices or movements. This freedom complicates the analysis of their behavior. We therefore developed a computer vision-based processing pipeline that transforms multi-camera video footage into discrete sequences of interactions with the different lockbox mechanisms.

Bayesian inference on the resulting state-action sequences indicates that mice follow a random strategy to solve the problem, and there was little evidence that their strategy changes across trials. This observation is consistent with the interpretation that the increased performance arises primarily from habituation. However, the over-representation of state-advancing interactions in solved trials suggest that mice may have acquired knowledge about the task and switch from an exploratory to an exploitative behavior within the trial. Unfortunately, a statistical test of this hypothesis requires a larger dataset and is therefore beyond the scope of this study.

We were surprised both by the fact that the mice solved this complex task so effectively and by the fact that they appeared to show very little learning. However, from an ecological perspective, a random strategy may be a very effective heuristic. Learning is beneficial only when the same problem setting occurs consistently over time, such that the effort of learning a task-specific behavior is amortized later [16]. If the problems encountered during foraging are highly variable, a random strategy may in fact be a good choice, albeit probably inferior to the more demanding variant of mental simulation. Yet, random strategies are only effective in decision problems with a relatively small number of actions. This suggests that the mice have a very effective prior that allows them to selectively try out a small set of promising candidate actions on the lockbox, despite the fact that they had limited exposure to levers, sticks, balls or doors in the past. Our present analysis cannot rule out that the mice learned such candidate actions during the single mechanism training (SMT) phase, and that the increase in performance in the second trial is due to narrowing down the number of actions that are then explored randomly. Neither can we rule out that the SMT phase induced motor learning and merely increased the success rate of individual interactions during random exploration.

A potential inroad to understanding whether the mice learn motor skills or an abstract strategy for sequential problems would be to study how they generalize to new variants of the lockbox. If the increased performance was due to motor learning or the identification of promising candidate actions, they should generalize to lockboxes composed of the same mechanisms in a different order. If they acquired an abstract understanding of sequential problems, their performance on entirely new sequential problems should improve with experience.

Many of the analyses we performed were limited by the relatively small number of trials and mice. In particular, we cannot exclude that the mice eventually acquire a smarter, greedy strategy after many trials. Therefore, a promising avenue to pursue would be an automated setup, in which the mice have perpetual access to a robotized and sensorized lockbox that is automatically rebaited once it has been solved. This would enable the acquisition of larger datasets and more conclusive analyses. It would also be very interesting to study the performance of other species on the task, to identify differences in solution strategy.

4 Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project number 390523135.

The authors thank Patrik Reiske and Leonie Pawlak for their help with manual data annotation.

5 Methods

5.1 Ethics statement

Animal experimentation was performed according to the guidelines of the German Animal Welfare Act and the Directive 2010/63/EU for the protection of animals used for scientific purposes. Maintenance of mice and all animal experimentation were approved by the Berlin State Authority ("Landesamt für Gesundheit und Soziales", permit number: G0249/19). The experiment was pre-registered in the animal study registry (animalstudyregistry.org, doi: 10.17590/asr.0000237), comprising several substudies, alongside the analysis of lockboxes as behavioral enrichment, which is published in an accompanying paper [9].

5.2 Animals

12 female C57BL/6J mice were obtained from Charles River Laboratories (Sulzfeld, Germany). They were free of pathogens listed in the FELASA recommendation [21]. The mice arrived at the age of four weeks at the animal facility, where they were housed in filter top cages under specifically pathogen free barrier conditions. The mice were maintained under a room temperature of 22 ± 1 °C, a relative humidity of 55 ± 10%, and a 12/12-hour light/dark cycle. Lights turned on at 7 AM; 30 min before, a simulated sunrise started and smoothly increased the light intensity (Philips HF 3510, 100-240 vac, 50-60 Hz, Philips GmbH Market DACH, Hamburg, Germany). The mice were housed in groups of four in two Makrolon type III cages (39 × 23 × 15 cm) that were connected to each other by a gate. The floor of both cages was covered with fine wooden bedding material (JRS Lignocel FS14, spruce/fir, 2,5-4 mm). One of the cages contained a red triangular plastic house with two exits (long side: 14.5, short sides: 11.5, height: 5cm, TheMouseHouse, Tecniplast), a transparent handling tunnel (length: 11.5, outer diameter: 4cm, custom-built from GEHR PMMA XT® ACRYL, Mannheim, Germany), and three thin (cellulose, unbleached, layers, 20x20cm, Lohmann & Rauscher International GmbH & CO KG) as well as two thick (23x24,8cm folded, Essity ZZ Towel) paper tissues for nesting; the other cage was not enriched with any material and was considered as training cage. Food pellets (LASvendi, LAS QCDiet, Rod 16, autoclavable) and tap water were provided *ad libitum* in both cages. The animals were handled using a tunnel. In the first two weeks, they were habituated to tunnel handling by the animal technicians and experimenter, who were all female. After the mice had habituated to the animal facility for a week, radio frequency identification (RFID) transponders were subcutaneously implanted into the neck region using an application cannula (ISO-Compliant Transponder, Peddymark Limited, Redhill, Surrey, United Kingdom; cannula: 2.6 × 0.15 × 40 mm) under a short isoflurane anesthesia (induction: 4% isoflurane, maintenance: 1-2% isoflurane; carrier gas: 100% oxygen, flow: 1 liter/min; Isofluran CP 1ml/ml, CP-Pharma Handelsgesellschaft mbH, Burgdorf, Germany). Before the transponder implantation, 1 mg/kg meloxicam (Loxicom 0.5 mg/ml, Elanco Animal Health, Bad Homburg, Germany or Metacam 0.5 mg/ml, Boehringer Ingelheim, Ingelheim am Rhein, Germany) was orally administered. Besides the RFID transponders, color marking of the tail was used for rapid visual identification of the individuals (edding 750 paint marker, edding International GmbH, Ahrensburg, Germany). Before the lockbox training period, the mice were tested in the open field, elevated plus maze, free exploratory paradigm, grid exploratory paradigm, and calorimetry for another study[9].

5.3 Home cage set-up

The home cage of the mice consisted of two Makrolon type III cages connected to each other by a gate. The gate was made of three transparent tubes (11.5 × 4 cm, custom-built from GEHR PMMA XT® ACRYL, Mannheim, Germany) and two doors (custom-built, 3D printed, STL files can be found on GitHub: github.com/RefinementReferenceCenter/MouseLockBox). The doors could be opened and locked by the experimenter. Except during the training periods, the mice had access to both cages. During the training periods, the doors of the gate were closed and the mice could only individually enter

the training cage. The cage lid of the training cage was modified for video monitoring and the grid was replaced with transparent acrylic glass in an area of 22 × 15 cm.

5.4 Video monitoring

Three Cameras acA1920-40um (Basler AG, Ahrensburg, Germany; Kowa Lens LM25HC F1.4 f25mm, Kowa Optimed Deutschland GmbH, Düsseldorf, Germany) were used to record the mice during the training periods from the top, front and side (1936 × 1216, 30 frames/s, AVI format). The cage was illuminated with two infrared lights (Synergy 21 IR-Strahler 60W, ALLNET GmbH Computersysteme, Germering, Germany).

5.5 Lockbox designs

The lockbox were designed using Tinkercad (www.tinkercad.com). The STL files were exported from Tinkercad and the G-code was generated with Cura_SteamEngine 4.4.0. The lockboxes were printed using an Ultimaker 3 Extended and an Ultimaker S3 (0.4 mm nozzles). Polylactic acid (PLA) of different colors was used as material. The lockbox consisted of a white platform, a yellow lever, a red stick, a gray ball, and a green sliding door. Information on 3D printing details and the STL files of the lockbox can be found on GitHub [8].

5.6 Lockbox training

After the mice had been kept for four weeks in the animal facility, they were habituated to the food reward, i.e., oat flakes (kernige Haferflocken, Wurzener Nahrungsmittel GmbH, Wurzen, Germany). On four consecutive days, eight oat flakes were scattered on the floor of the training cage in the morning. On the following day, the lockbox training started. For this, the lockboxes were baited with an oat flake. There was one lockbox for each mouse to exclude odor cues from conspecifics. The lockbox training was carried out, as follows: When all mice left the training cage, the doors of the gate were locked and the closed lockbox was placed into the training cage. Then the first mouse was allowed to enter the training cage and explore the lockbox for a maximum of 30 min. A training trial ended after the mouse opened the lockbox or after 30 min. At the end of the trial, the lockbox was removed from the training cage when the mouse no more touched the lockbox with any paw or 1 min after the end of the trial. Finally, the doors of the gate were opened and the mouse could leave the training cage. Then another lockbox was placed into the training cage and the procedure was repeated for all mice over six trials. In all trials, the mice entered the training cage in a randomized order. In trial 1, the mice explored the closed lockbox for the first time. Between trial 1 and trial 2, the mice were trained on the single components of the lockbox, i.e., the lever, the stick, the ball, and the sliding door—over a period of 20 days, including a few days without lockbox training. In trial 2, the open lockbox baited with an oat flake was presented to the mice for 1 min. After that, it was re-baited with a food reward and closed. The lockboxes were cleaned with ethanol 70 % after all mice completed their training trial.

5.7 Tracking

The behavior of the mice was tracked using DeepLabCut (DLC, Mathis et al. [17], version 2.2.1.1). Two individual trackers were trained to localize different key points on the lockbox and the mouse, respectively. Among the full dataset of 3934 images, 3554 (90%) were randomly assigned for training and 380 (10%) for testing both trackers. To extract the images from the videos and to train the networks, methods provided by DLC were used with default settings (see Appendix **??** for further information). The three 2D coordinates obtained by applying the trackers to the videos from the three cameras were concatenated into a 6D vector $x(t)$. To infer a 3D position for the various key points, we used an extended Kalman filter & smoother for which (part of) the observation matrix consisted of the 2D to 3D mapping created by triangulating (using RANSAC triangulation) on the closed lockbox key points. The underlying generative model maps the degrees of freedom of a skeletal model of the mouse and the lockbox to the 6-dimensional coordinates of all key points that are obtained from the DLC tracker. To clarify the terminology and notation: We will refer to the degrees of freedom of the mouse/lockbox model as the hidden causes θ of the system, the inferred positions of body/lockbox part i in 3D as 3D coordinates r_i and the combined 6D coordinates of keypoint i obtained from the DLC tracker as tracks x_i.

The Kalman filter combines a model of the dynamics of the hidden causes θ with a nonlinear emission model for the different tracks x_i. To obtain smooth tracks, we modeled the dynamics of the lockbox and the mouse in a state space consisting of the hidden causes θ and their first, second and third temporal derivatives (velocity v, acceleration a and jerk j):

$$\theta_{t+1} = \theta_t + v_t \Delta t - \nabla E(\theta) \tag{1}$$

$$\dot{v}_{t+1} = v_t + a_t \Delta t \tag{2}$$

$$\dot{a}_{t+1} = a_t + j_t \Delta t \tag{3}$$

$$\dot{j}_{t+1} = j_t + \sqrt{\Delta t} \Sigma_\theta^{1/2} \xi_t \,. \tag{4}$$

Here, Δt denotes the discretization time step and ξ is N-dimensional Gaussian noise with unit covariance. Σ_θ denotes a diagonal covariance matrix that characterizes the amount of noise in the dynamics that can be accounted for by the observations x. $E(\theta)$ is an energy function and the term $-\nabla E(\theta)$ biases the state estimate towards regions of small energy. For some variables this was used to incorporate prior knowledge. For example, the mouse is unlikely to lie on its back, so the angle that captures how far it rears from a sitting position should be small unless there are convincing observations to the contrary.

The hidden causes θ comprise the degrees of freedom of the lockbox and the skeletal model of the mouse. Different lockbox components have different degrees of freedom. The state of the lever can be characterized by a single tilting angle ranging from 0 to $\pi/2$. The state of the stick and the sliding door can be described by by a 1-dimensional variable describing how far it has been shifted from an arbitrary zero position. The ball is free to move in 3D space, but we included an energy function

$$E(\theta) = A \left[\exp(-|\theta_{ball} - \theta_0|/2\sigma_1^2) - \exp(-|\theta_{ball} - \theta_0|/2\sigma_2^2) \right] \,,$$

which favours finding it at its original position θ_{ball} in the horizontal plane, unless it had previously been removed.

The hidden causes of the mouse comprise a center of mass r_{cm} and three angles ϕ, β, γ that describe its orientation in the horizontal plane, its rearing angle, and the curvature of the back. These variables are used to translate the geometry of the skeletal model to 3D coordinate of the nose, the ears and the tailbase of the mouse: $r_i = r_{cm} + O_i(\phi, \beta, \gamma) d_i$. Here, d_i are distances from the center of mass in the mouse-centered coordinate system, which define the geometry of the skeletal model. These distances were chosen once in line with a typical mouse, but not fine-tuned to the individual mice. The matrices $O_i(\phi, \beta, \gamma)$ translate mouse-centered into global coordinates by rotating the relevant body parts suitably. Analogous models are used to translate the hidden causes of the different parts of the lockbox into global 3D coordinates.

Finally, the 3D coordinates r_i of the different keypoints are converted into 6D tracks x_i using a linear transformation:

$$x_i = H r_i + x^{(0)} + \Sigma_x^{1/2} \zeta_i \,. \tag{5}$$

Here, H denotes a 6×3 matrix that transforms 3D positions in millimeters into 6D positions in pixel space, and $x^{(0)}$ is the origin of the coordinate system in 6D pixel space. The Gaussian random vector ζ_i has unit variance and accounts for observation noise. The scale of this noise was defined by the matrix Σ_x. We included the confidence $c \in [0, 1]$ of the DLC tracker by scaling up the observation noise when keypoints were with a low confidence:

$$\Sigma_x = \Sigma_{x,0} \times \frac{1}{c^2} \,.$$

This scaling is applied individually for each tracked keypoint and camera. The effect of this rescaling is to ignore unreliable observations and infer the respective keypoint from either the other camera views or from the assumed continuity in time. Such unreliable observations are very frequent due to occlusions. For a reliably tracked point ($x \approx 1$), the observation noise is given by $\Sigma_{x,0}$.

The scaling matrix H and the origin $x^{(0)}$ in pixel space were constrained by tracking four immobile keypoints on the lockbox, for which the positions in the (lockbox-centered) 3D coordinate system are known from the construction plan of the lockbox. Using the tracked pixel positions of these points, H and $x^{(0)}$ can be obtained from a linear regression.

The extended Kalman filter used a local linearization of this model to infer the hidden causes of the skeletal model and the lockbox. Their positions in 3D space, in turn, can then be computed using the skeletal model. Because the 3D coordinate system is anchored on the lockbox, this allowed us to detect interactions of the mouse with the different lockbox parts.

5.8 Interaction analysis

To detect interactions of the mouse with the different parts of the lockbox, we defined bounding boxes around each lockbox part. Each video frame was then classified as an interaction of the mouse with a given lockbox part when the nose of the mouse was in the associated bounding box. This classification was used to calculate interaction times with the individual lockbox parts and the lockbox as a whole. We are aware that the position of the nose is a rather simple proxy for an interaction, but the paws were too poorly tracked to serve as a reliable indicator for interactions, because of frequent occlusions (paws hidden below the mouse).

5.9 Bayesian analysis of behavioral strategies

For a more fine-grained quantitative analysis of the strategic behavior of the mice, we performed a Bayesian analysis that tested whether the interaction pattern of the mice was consistent with different strategies. To this end, mouse behavior was modeled as a discrete sequence of interactions with the four different lockbox mechanisms (lever - 1, stick - 2, ball - 3, door - 4). The model also considers four different lockbox states, which correspond to which mechanism of the lockbox had to be opened next to proceed (lever - 1, stick - 2, ball - 3, door - 4). The model assigns a conditional probability $\pi(i|j) = \pi(\text{interaction } i|\text{lockbox state } j)$ to interact with each mechanism of the lockbox, given the overall state of the lockbox.

We compared two different strategies. The first strategy was random behavior, in which the mouse interacted with the different lockbox mechanisms with the same probability, irrespective of the state of the lockbox:

$$\pi(i|j) = P(i)\,,$$

where $P(i)$ denotes the prior preference distribution of the mouse, which is determined by calculating the fraction of interactions per mechanism across the entire trial.

The second strategy describes goal-directed behavior in the sense that the mouse interacts most with the lockbox mechanism that had to be solved next to progress:

$$\pi(i|j) = (1-\epsilon)\delta_{ij} + \epsilon P(i)\,.$$

Here, δ_{ij} denotes the Kronecker symbol, which is 1 if $i = j$ and zero otherwise. The parameter $\epsilon \in [0, 1]$ controls the amount of randomness (or exploratory behavior) of the mouse. For $\epsilon = 1$, the behavior reverts to the above random strategy, for $\epsilon = 0$, the mouse interacts exclusively with the lockbox mechanism that needs to be opened next.

Assuming that the interactions are statistically independent samples from the distribution π_{ij}, this model allows to compute the likelihood of the randomness parameter ϵ for any time sequence (I_t, s_t) of interactions I_t and states s_t:

$$L(\epsilon) = \prod_{t=1}^{(I,s)} \pi(I_t, s_t). \tag{6}$$

The model also allows to perform model comparison, i.e., the likelihood of the two different strategies given the observed state-interaction sequence.

To perform this inference on the data, a definition of a single interaction is required. To this end, we defined a single interaction to be one uninterrupted sequence of interaction-detected frames with one of the lockbox mechanisms. We set a duration threshold of 1.5 seconds for the interactions. This avoids labeling the walking of the mouse over the mechanisms as an interaction and leads to interaction sequences that correlate well with the manual annotations. In rare cases where the mouse solved a given state of the lockbox very quickly, this led to the situation that the lockbox proceeded to the next state without a single interaction. Because this is obviously wrong, we detected these cases and manually added a single interaction with the appropriate mechanism.

References

[1] Alice MI Auersperg. Exploration technique and technical innovations in corvids and parrots. In *Animal creativity and innovation*, pages 45–72. Elsevier, 2015.

[2] Alice MI Auersperg, Auguste MP Von Bayern, Gyula K Gajdon, Ludwig Huber, and Alex Kacelnik. Flexibility in problem solving and tool use of kea and new caledonian crows in a multi access box paradigm. *PLoS One*, 6(6):e20231, 2011.

[3] Sarah E Daniels, Rachel E Fanelli, Amy Gilbert, and Sarah Benson-Amram. Behavioral flexibility of a generalist carnivore. *Animal cognition*, 22:387–396, 2019.

[4] Sandeep Robert Datta, David J Anderson, Kristin Branson, Pietro Perona, and Andrew Leifer. Computational neuroethology: a call to action. *Neuron*, 104(1):11–24, 2019.

[5] Emily Jane Dennis, Ahmed El Hady, Angie Michaiel, Ann Clemens, Dougal R Gowan Tervo, Jakob Voigts, and Sandeep Robert Datta. Systems neuroscience of natural behaviors in rodents. *Journal of Neuroscience*, 41(5):911–919, 2021.

[6] Michael J Galsworthy, Jose L Paya-Cano, Lin Liu, Santiago Monleón, Gregory Gregoryan, Cathy Fernandes, Leonard C Schalkwyk, and Robert Plomin. Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavior genetics*, 35:675–692, 2005.

[7] Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv*, page 214262, 2017.

[8] Katharina Hohlbaum, Niek Andresen, Marcus N. Boon, Pia Kahnau, Paul Mieske, Henning Sprekeler, Olaf Hellwich, Christa Thöne-Reineke, and Lars Lewejohann. Mouselockbox, 2023. URL `https://github.com/RefinementReferenceCenter/MouseLockBox`.

[9] Katharina Hohlbaum, Niek Andresen, Paul Mieske, Pia Kahnau, Benjamin Lang, Kai Diederich, Rupert Palme, Lars Mundhenk, Henning Sprekeler, Olaf Hellwich, et al. Lockbox enrichment facilitates manipulative and cognitive activities for mice [version 1; peer review: awaiting peer review]. *Open Research Europe*, 4:108, 2024.

[10] Ludwig Huber, Sabine Rechberger, and Michael Taborsky. Social learning affects object exploration and manipulation in keas, nestor notabilis. *Animal Behaviour*, 62(5):945–954, 2001.

[11] Sarah L Jacobson, Amanda Puitiza, Rebecca J Snyder, Ashley Sheppard, and Joshua M Plotnik. Persistence is key: investigating innovative problem solving by asian elephants using a novel multi-access box. *Animal Cognition*, pages 1–13.

[12] Sarah L Jacobson, Juthapathra Dechanupong, Wantida Horpiencharoen, Marnoch Yindee, and Joshua M Plotnik. Innovating to solve a novel puzzle: Wild asian elephants vary in their ability to problem solve. *Animal Behaviour*, 205:227–239, 2023.

[13] Ashley L Juavinett, Jeffrey C Erlich, and Anne K Churchland. Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility. *Current opinion in neurobiology*, 49:42–50, 2018.

[14] Ann Kennedy. The what, how, and why of naturalistic behavior. *Current opinion in neurobiology*, 74:102549, 2022.

[15] John W Krakauer, Asif A Ghazanfar, Alex Gomez-Marin, Malcolm A MacIver, and David Poeppel. Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, 93(3):480–490, 2017.

[16] Robert Tjarko Lange and Henning Sprekeler. Learning not to learn: Nature versus nurture in silico. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7290–7299, 2022.

[17] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.

[18] Valeria Mazza and Anja Guenther. City mice and country mice: innovative problem solving in rural and urban noncommensal rodents. *Animal Behaviour*, 172:197–210, 2021. ISSN 0003-3472. doi:https://doi.org/10.1016/j.anbehav.2020.12.007. URL https://www.sciencedirect.com/science/article/pii/S0003347220303584.

[19] Hiromitsu Miyata, Gyula K Gajdon, Ludwig Huber, and Kazuo Fujita. How do keas (nestor notabilis) solve artificial-fruit problems with multiple locks? *Animal Cognition*, 14:45–58, 2011.

[20] M-B Nada, Johannes Fuss, Massimo Trusel, Michael J Galsworthy, Kristin Bobsin, Giovanni Colacicco, Robert MJ Deacon, Marco A Riva, Christoph Kellendonk, Rolf Sprengel, et al. The puzzle box as a simple and efficient behavioral test for exploring impairments of general cognition and executive functions in mouse models of schizophrenia. *Experimental neurology*, 227(1):42–52, 2011.

[21] FELASA Working Group on Revision of Guidelines for Health Monitoring of Rodents, Rabbits, M Mähler, M Berard, R Feinstein, A Gallagher, B Illgen-Wilcke, K Pritchett-Corning, and M Raspa. Felasa recommendations for the health monitoring of mouse, rat, hamster, guinea pig and rabbit colonies in breeding and experimental units. *Laboratory animals*, 48(3):178–192, 2014.

[22] Talmo D Pereira, Nathaniel Tabris, Junyu Li, Shruthi Ravindranath, Eleni S Papadoyannis, Z Yan Wang, David M Turner, Grace McKenzie-Smith, Sarah D Kocher, Annegret L Falkner, et al. Sleap: Multi-animal pose tracking. *BioRxiv*, pages 2020–08, 2020.

[23] Cristina Segalin, Jalani Williams, Tomomi Karigo, May Hui, Moriel Zelikowsky, Jennifer J Sun, Pietro Perona, David J Anderson, and Ann Kennedy. The mouse action recognition system (mars) software pipeline for automated analysis of social behaviors in mice. *Elife*, 10:e63720, 2021.

[24] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364 (6437):eaav7893, 2019.

[25] Jennifer J Sun, Serim Ryou, Roni H Goldshmid, Brandon Weissbourd, John O Dabiri, David J Anderson, Ann Kennedy, Yisong Yue, and Pietro Perona. Self-supervised keypoint discovery in behavioral videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2171–2180, 2022.

[26] Elisabeth Suwandschieff, Ludwig Huber, Thomas Bugnyar, and Raoul Schwing. Kea, bird of versatility. kea parrots (nestor notabilis) show high behavioural flexibility in solving a demonstrated sequence task. *Journal of Ornithology*, 165(1):49–55, 2024.

[27] Edward L. Thorndike. *Animal intelligence: Experimental studies.* Macmillan Press, 1st edition, 1911. doi:10.5962/bhl.title.55072.

[28] Lara Vrbanec, Vanja Matijević, and Anja Guenther. Enhanced problem-solving ability as an adaptation to urban environments in house mice. *Proceedings of the Royal Society B*, 288(1945):20202504, 2021.

A Appendix: 2D tracking

The dataset used to develop the trackers for mice and lockbox consists of a total of 79 videos showing the two different lockbox training conditions, i.e. training on single mechanisms of the lockbox and on the complete lockbox, and recorded from the three different perspectives (front, top, and side). From those 79 video recordings, 3934 images were extracted using the frame extraction method provided by DeepLabCut (DLC, Mathis et al. [17], version 2.2.1.1), keeping the default settings, that is, automatic extraction mode, kmeans algorithm for frame selection, no frame cropping, and a clustering step of 1 to include each frame for clustering. Among the full dataset, 3554 images (90%) were randomly assigned for training and 380 (10%) for testing both trackers. The composition ratio of the images of the two lockbox training conditions and the three perspectives, respectively, was maintained in the training dataset, while for the testing dataset the proportion of images for each case was evenly distributed. The detailed composition of the entire dataset as well as the training and testing sets is listed in Table 1 and Table 2. Each image was manually annotated with various keypoints on the mouse, such as nose, ears, paws, and tail, and on the lockbox (lever, stick, ball, and sliding door) using DLC's interactive graphical user interface as labeling tool.

	Front	Top	Side	Total (%)
Single mechanisms	380	490	327	1197 (30%)
Lockbox	780	1117	840	2737 (70%)
Total (%)	1160 (30%)	1607 (40%)	1167 (30%)	3934

Table 1: The numbers of images taken of the two lockbox training conditions and the three different perspectives, respectively, used in the full dataset.

	Single mechanisms	Lockbox	Front	Top	Side	Total (%)
Training set	1007	2547	1030	1487	1037	3554 (90%)
Testing set	190	190	130	120	130	380 (10%)

Table 2: The numbers of images taken of the two lockbox training conditions and the three camera perspectives, respectively, used for training and testing sets.

Two individual trackers were developed to localize different keypoints on the lockbox and the mouse, respectively. The Lockbox Tracker and the Mouse Tracker were trained and tested independently, both following the same training and testing process and using the aforementioned image sets. The network configurations of each model were set to the default settings of DeepLabCut, employing the ResNet-50, a deep convolutional neural network architecture, pre-trained on the ImageNet dataset for pose estimation. The training and testing of the trackers was performed on a Linux system with 183-Ubuntu SMP kernel version and NVIDIA Quadro RTX 4000 GPU 8GB VRAM. Each training session had 1,030,000 training iterations, a batch size of 1, and a learning rate of 5×10^{-3}, retaining the multi-step learning rate decay as defined by DLC. The maximum input size of images was changed from 1500 by default to 2500 to allow larger frame sizes during training.

To evaluate the performance of the trackers, the accuracy, sensitivity, and precision analyses were performed and the pixel error between label and prediction was computed according to following equations:

$$\text{Accuracy} = \frac{true\,positive + true\,negative}{labels + predictions} \tag{7}$$

$$\text{Precision} = \frac{true\,positive}{predictions} \tag{8}$$

$$\text{Sensitivity} = \frac{true\,positive}{labels} \tag{9}$$

$$\text{Pixel Error} = \sqrt{(prediction - label)^2} \tag{10}$$

where *true positive* and *true negative* are the numbers of times, the trackers correctly predicted that a keypoint was present and not present, respectively. A point was considered correctly predicted with

a tracker confidence ≥ 0.6. The pixel error describes the distance between the predicted point and the labeled point for one keypoint in pixels. Note, that the above equations refer to the tracker predictions for one image and one keypoint. The overall performances of the trackers were calculated as the averages of all images and the number of keypoints. The results of the evaluation can be found in Table 3. The Lockbox Tracker performs slightly better than the Mouse Tracker, especially in terms of precision. This is probably due to the fact that the lockbox was always placed in the same position, whereas the mouse could move freely in the cage. As a result, the possible coordinates of the lockbox parts were limited in space and vary less, even if they were moved.

Tracker	Mean Accuracy	Mean Precision	Mean Sensitivity	Mean Pixel Error
Lockbox	87.94%	96.68%	84.93%	27.68 px
Mouse	84.95%	89.83%	67.92%	17.07 px

Table 3: Overall performance of the Mouse Tracker and the Lockbox Tracker.