



Chronotate: An open-source tool for manual timestamping and quantification of animal behavior

Paul A. Philipsberg^a, Zoé Christenson Wick^a, Keziah S. Diego^a, Nick Vaughan^a, Angelina Galas^{a,b}, Albert Jurkowski^{a,c}, Yu Feng^a, Lauren M. Vetere^a, Lingxuan Chen^a, Iván Soler^a, Denise J. Cai^a, Tristan Shuman^{a,*}

^a Icahn School of Medicine at Mount Sinai, New York NY, United States

^b New York University, New York NY, United States

^c CUNY Hunter College, New York NY, United States

ARTICLE INFO

Keywords:

Open-source
Rodent behavior
Novel object location
Novel object recognition

ABSTRACT

A core necessity to behavioral neuroscience research is the ability to accurately measure performance on behavioral assays, such as the novel object location and novel object recognition tasks. These tasks are widely used in neuroscience research and measure a rodent's instinct for investigating novel features as a proxy to test their memory of a previous experience. Automated tools for scoring behavioral videos can be cost prohibitive and often have difficulty distinguishing between active investigation of an object and simply being in close proximity to an object. As such, many experimenters continue to rely on hand scoring interactions using stopwatches, which makes it difficult to review scoring after-the-fact and results in the loss of temporal information. Here, we introduce Chronotate, a free, open-source tool to aid in manually scoring novel object behavior videos. The software consists of an interactive video player with keyboard integration for marking timestamps of behavioral events during video playback, making it simple to quickly score and review bouts of rodent-object interaction. In addition, Chronotate outputs detailed interaction bout data, allowing for nuanced behavioral performance analyses. Using this detailed temporal information, we demonstrate that novel object location performance peaks within the first 3 s of interaction time and preference for the novel location becomes reduced across the test session. Thus, Chronotate can be used to determine the temporal structure of interactions on this task and can provide new insight into the memory processes that drive this behavior. Chronotate is available for download at: <https://github.com/ShumanLab/Chronotate>.

1. Introduction

The accurate and reliable assessment of behavioral performance in rodents has been of great benefit to virtually every field of neuroscience research from both a basic research and preclinical perspective [1]. Two of the most commonly used rodent behavioral assays are the novel object recognition (NOR) and novel object location (NOL) tasks [2,3]. These two assays test a rodent's ability to learn the layout and features of an environment and identify changes to that environment after a delay [2]. As such, these tasks test both learning and memory, as a rodent must first learn about an environment and then later remember it well enough to determine which object has changed location or identity. Because rodents have an innate tendency to investigate novel over familiar stimuli [4], memory performance can be measured as a ratio of how

much time the rodent spends investigating novel versus familiar stimuli or a new location of familiar stimuli. In fact, this instinct towards investigating novel stimuli or location is so robust that if an equivalent amount of time is spent investigating a novel and a familiar stimulus/location, it is interpreted as the rodent forgetting which is novel and which is familiar [3]. It is therefore critical to the interpretation of this data that the time the rodents spend investigating both novel and familiar stimuli or location be accurately assessed and reported.

A number of machine-learning based tools for automated behavior scoring exist [5–9], but these tools often require specialized video recording setups, large sets of training data, and significant time to fine-tune the underlying models [10]. Moreover, subtle distinctions between nuanced behaviors often require the judgment of the experimenter to resolve ambiguous cases. For example, the difference between an animal

* Corresponding author.

E-mail address: tristan.shuman@mssm.edu (T. Shuman).

<https://doi.org/10.1016/j.neulet.2023.137461>

Received 21 June 2023; Received in revised form 1 August 2023; Accepted 21 August 2023

Available online 23 August 2023

0304-3940/© 2023 Elsevier B.V. All rights reserved.

sniffing an object and an animal grooming while facing an object can be difficult for an automated scoring tool to accurately classify. For these reasons, many experimenters performing the NOR and NOL tasks continue to rely on hand-scoring methods to measure rodent-object investigation times and ratios.

At its simplest, NOR/NOL video scoring requires recording the total time spent interacting with each object. As such, a common method of scoring would consist of using two stopwatches to time the total duration of all bouts of exploration for each of the novel and familiar objects as the video is played back in real-time. This approach has several disadvantages. Recording multiple short bouts of interaction in succession is hindered by the difficulty of pressing the start/stop button of a stopwatch repeatedly, and making even one error can require the scorer to restart from the beginning of the video. Moreover, recording only the total amount of time spent with each object prevents more granular

analysis of how interaction changes over the course of the session. If the experimenter is later interested in reexamining only the first 30 s of exploration, for example, then each video would need to be rescored. Additionally, this approach makes evaluating inter-rater reliability difficult, as scoring of individual bouts cannot be directly compared. Fortunately, however, many of these issues can be addressed by manually noting the duration and timestamp of each bout of interaction, though doing this in a purely manual fashion would be arduous and time-consuming.

Several behavioral scoring software packages support using key-strokes to mark timestamped events on behavioral videos, but the price of these tools can be prohibitive in many applications [2,8]. For this reason, and those discussed above, we developed Chronotate: a free, open-source tool to aid in the hand-scoring of behavior videos. Here we demonstrate Chronotate’s utility for assessing mouse performance on

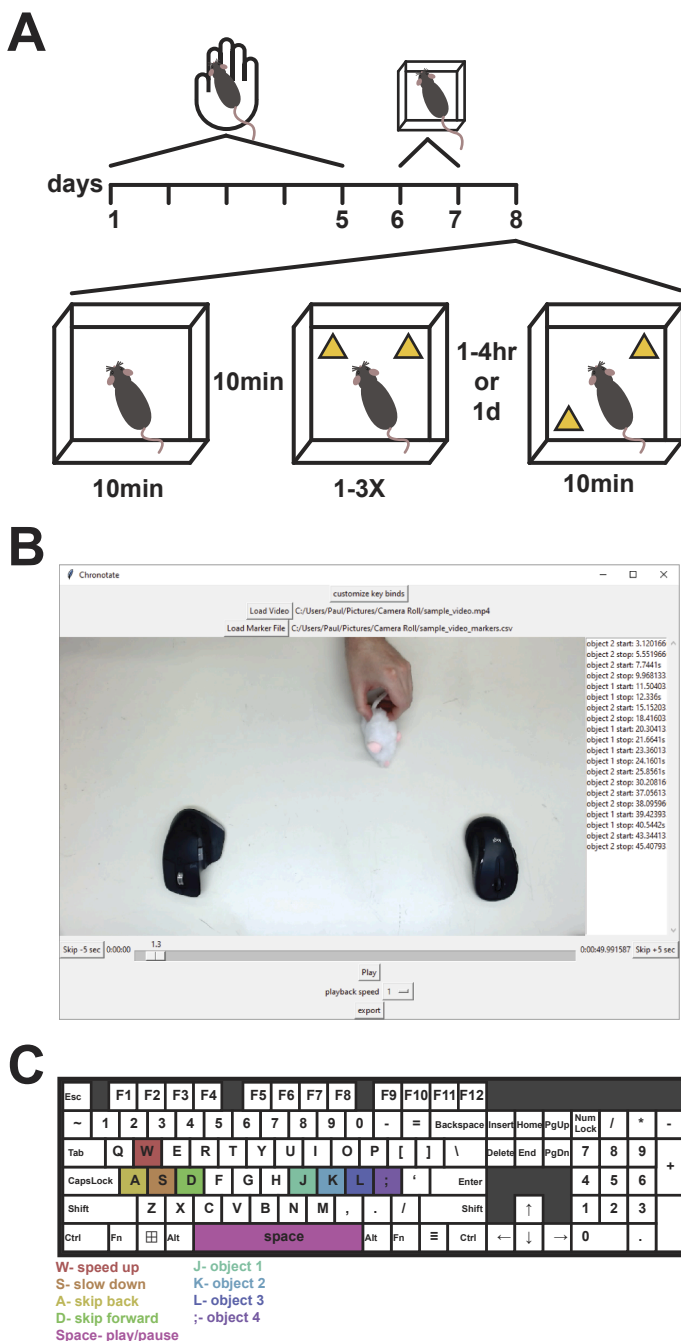


Fig. 1. Chronotate for novel object location test scoring. (A) Schematic of novel object location (NOL) protocol. Mice were handled for 5 days before being habituated to an empty arena for 2 days. On the 8th day of training, mice were re-habituated to the empty arena before a training session, where 2 objects were added to the arena. Following a 1hr, 4hr, or 1 day delay, mice were placed back into the arena with one of the objects relocated. (B) Chronotate graphical interface. A list of time-stamped interaction bouts is displayed on the right and a progress bar along the bottom of the interface allows for seeking to precise time points. (C) Default keyboard controls for scoring object interaction (J,K,L,;-), speeding up or slowing down (W,S), skipping forward or back (D,A), and playing or pausing (space). Key bindings can be customized as desired by activating the ‘customize key binds’ dialog from the top of the main window.

the NOL task and explore object investigation bout-by-bout in addition to surveying cumulative investigation preferences over the duration of testing. We find that not only is Chronotate useful for scoring videos, but that the more granular data analyses it permits can shed light on NOR and NOL behavioral assay design.

2. Materials and methods

2.1. Animals

The use of animals and the experimental protocols were approved by the Icahn School of Medicine's Institutional Animal Care and Use Committee, in accordance with the US National Institutes of Health guidelines.

For comparison of NOL scoring between stopwatches and Chronotate (Figs. 1, 2A), male and female C57BL6 mice (Charles River Lab, strain 027) were ordered directly from the vendor and were tested in the NOL task at 12 weeks of age. For evaluation of inter-rater reliability (Fig. 2B-D), male and female wild-type B6129SF2/J (Jackson Laboratory, RRID: IMSR_JAX: 101045; [11,12]) and 3xTg (Mouse Mutant Resource and Research Center, RRID:MMRRC_034830-JAX; [12]) mice were maintained with in-house breeding in approved breeding facilities at Mount Sinai and were tested in the NOL at 6–8 months of age. For investigating

the time course of object preference (Fig. 3), male C57BL/6J mice (Jackson Laboratory, Bar Harbor, ME) were ordered directly from the vendor and were tested in the NOL at 3–6 months of age, as in Chen et al. 2023 [13]. All animals were housed in a temperature-controlled vivarium and were given food and water *ad libitum*. Lights were kept on a 12-hour light–dark cycle (lights on at 0700).

2.2. Novel object location memory task

For 4–5 days prior to being introduced to the behavioral arena, mice were habituated to transportation from the animal facility to behavior room where they were then gently handled for 5 min (Fig. 1A) [13,14]. On the 6th and 7th days, mice were habituated to the behavioral arena (30 × 30 × 30 cm, white acrylic) for 10 min per day. On these days, the arena was empty except for distinct spatial cues on each of the four walls. On the 8th day, mice were returned to the same arena for one (Fig. 2A, Fig. 3) to three (Fig. 2C-D) training sessions, where two identical objects (plastic toys or door stoppers ~ 2" x 2") were taped down near two corners of the arena (Fig. 1A). Mice were then returned to their home cage for 1 h (Fig. 2A), 4 h (Fig. 2C-D), or 1 day (Fig. 3) before the testing session. During testing, one of the two objects was relocated to a corner of the arena that was previously empty (Fig. 1A).

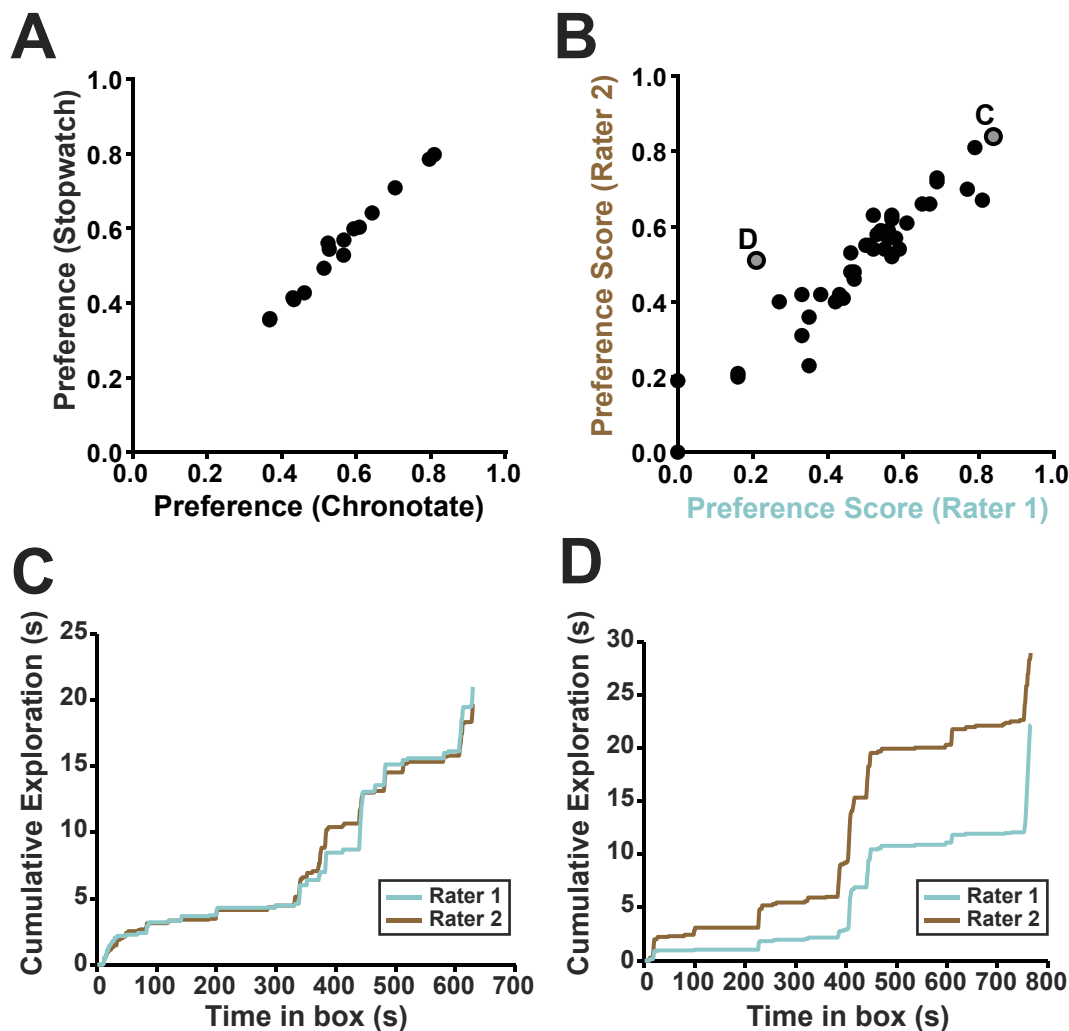


Fig. 2. Time-stamped outputs allow detailed assessment of inter-rater reliability. (A) Preference scores from Chronotate are highly correlated to those scored with a stopwatch (Pearson's $r = 0.99$, $p < 0.0001$, $n = 16$ videos). (B) Preference scores by two independent raters are highly correlated (Pearson's $r = 0.93$, $p < 0.0001$, $n = 42$ animals). (C, D) Cumulative exploration across a video with high inter-rater agreement (C) and a video with lower inter-rater agreement (D). The points corresponding to these two videos are noted in B.

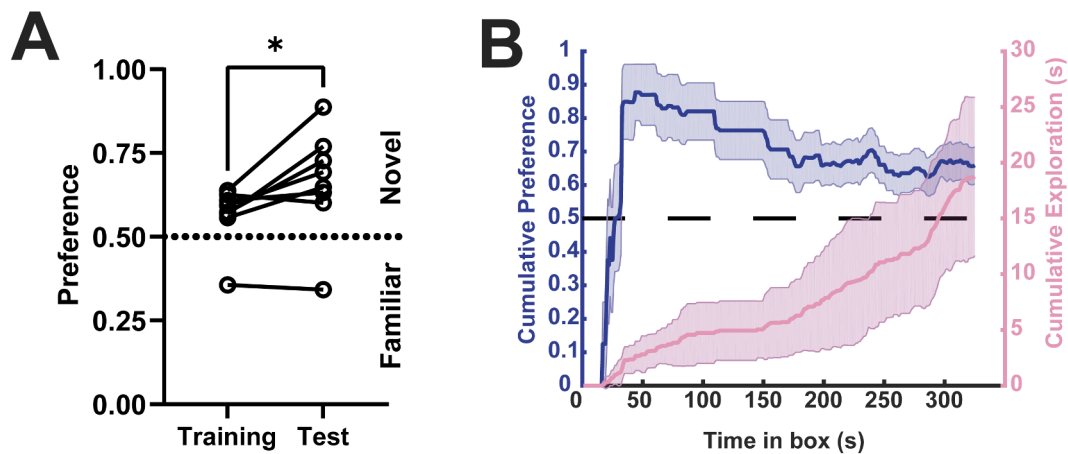


Fig. 3. Determining the time-course of object preference across the testing session. (A) Animals successfully learned to discriminate between the novel and familiar object-location pair during the test session (paired *t*-test $p = 0.028$, $n = 8$ animals), with animals showing a preference for the novel object-location. (B) Mean cumulative preference for the novel object over the course of the session (blue, left axis) and mean cumulative total object exploration time (pink, right axis). Preference for investigating the novel object-location peaks at approximately 45 s after the start of the session, which corresponds to an average combined exploration of both objects of 2.7 s. Shaded error bars represent SEM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Scoring

Mouse-object interactions were scored manually using stopwatches or using Chronotate [13,14]. Interactions were identified as periods in which an animal was sniffing, whisking at the object, or looking at an object while touching it. Time spent sitting on objects, touching them while rearing up, or grooming while facing an object were not counted as exploration. The mouse's preference for investigating the moved (novel) vs the unmoved (familiar) objects was calculated as the ratio of time spent interacting with the novel over the total time:

$$\text{Preference} = \frac{\text{Interaction time with the moved object}}{\text{Total interaction time}}$$

For the analysis in Fig. 3, mice that spent <30 s of exploration or <8 min in the arena during the training session were excluded [13].

For scoring with stopwatches, total time spent with each object was scored with an individual stopwatch for each of the two objects, starting and stopping the timer at the beginning and end of each bout of exploration [13,14]. For scoring with Chronotate, each of the two objects was assigned to a particular keyboard key. Each bout of object interaction was marked by pressing the relevant key at the start of exploration and releasing the key at the end of the bout. Marker files containing time stamped interaction bouts for each video were exported for subsequent analysis.

2.4. Chronotate

Chronotate uses key-down and key-up keyboard events to record video timestamps of the start time and end time of each bout of interaction (Fig. 1C). Simultaneous scoring of up to four different objects is currently supported. Because the recorded timestamps are relative to the actual position in the video and not an external timer, there is no need to watch the video back in real-time while scoring. Chronotate allows the user to speed up video playback to score long videos more efficiently, or slow down playback to more precisely mark the times of rapid events, all with keyboard strokes (Fig. 1C). Additionally, Chronotate allows the user to skip forward or backward in fixed increments, and seek to a specific time using a progress bar integrated into the graphical interface (Fig. 1B). All keyboard shortcuts can be customized by activating the 'customize key binds' dialog from the top of the Chronotate window. Custom key bindings will be saved in a 'keyBinds.json' file in the same directory as the Chronotate application.

Each interaction start/stop event logged is displayed in the side panel for review (Fig. 1B). Following scoring, these markers can be exported and analyzed externally according to the needs of the individual experimenter. The marker file will be saved to the same directory the video is loaded from as a .csv file named for the corresponding video. In addition, a previously exported marker file can be imported alongside its corresponding video in order to review and assess inter-rater reliability. Importantly, clicking on any event recorded in the side panel will seek the video playback to the recorded timestamp, allowing an experimenter to review the scoring of any individual event.

All functionality of Chronotate is implemented in Python. The graphical interface (Fig. 1B) is implemented using the tkinter package. Video playback is supported by a modified version of the open-source tkVideoPlayer package, as well as the Python Imaging Library (Pillow) and pyAV packages [15]. The import and export of csv data is implemented using the pandas package [16]. To run the source files, we recommend setting up a new Conda environment for Chronotate with these dependencies. All source files along with a precompiled executable file (.exe) are available from the Shuman Lab's GitHub page (<https://github.com/ShumanLab/Chronotate>). The precompiled .exe is a stand-alone application compatible with Windows and requires no installation. The .exe has been successfully tested with Windows 10 & 11. (Note that due to GitHub's handling of large files, cloning or downloading the .zip of the repository will result in a broken .exe. Therefore, the .exe must be downloaded separately, which can be done by following the link in the repository's readme page.) The source files require Python 3.6–3.10 and are cross-platform compatible with Mac, Linux and Windows. The source files have been successfully tested on Windows 10, macOS Monterey 12.3, and Ubuntu 18.04, and should also be compatible with other versions of these operating systems.

2.5. Data analysis & statistics

The marker output files from Chronotate were processed by a custom MATLAB script (available on the Shuman Lab's GitHub) to quantify the total exploration time with the novel and familiar objects. The cumulative exploration with each object was calculated on a second-by-second basis to generate a curve of preference for the novel object over the course of the session. All statistical tests were carried out using GraphPad Prism 9.

3. Results

We first show that Chronotate does not introduce a bias into the resulting preference scores. A set of 16 NOL behavior videos were each scored twice by a blinded experimenter, one using stopwatches and a second time using Chronotate. Preference scores were highly correlated (Fig. 2A: Pearson's $r = 0.99$, $p < 0.0001$, $n = 16$ videos), indicating high agreement between the two scoring methods.

One of the major benefits of using Chronotate to score NOR/NOL video data is the ability to review scoring between raters on a level that is more granular than standard approaches. Typically, inter-rater reliability on these tasks is compared by determining how similar final preference, or discrimination index, values are to each other, as we show in Fig. 2B (Pearson's $r = 0.93$, $p < 0.0001$, $n = 42$ animals, comparing preference scores between two independent raters). Unfortunately, however, using this method to troubleshoot videos where two raters have low inter-rater reliability presents the challenge of determining where in the video a difference in scoring emerged. Fortunately, because Chronotate outputs each time-stamped exploration bout, the moments where two raters diverged in their scoring becomes easily identifiable (Fig. 2C, 2D). These portions of the video can then be isolated, reviewed, and re-scored as needed. The NOR/NOL tasks leverage rodent's innate tendency to investigate a novel stimulus or location over a familiar one. As such, our mice trained in the NOL task showed a preference for investigating the novel object-location pair that emerged during the test session (Fig. 3A, paired t -test, $p = 0.028$, $n = 8$ animals; data from Chen et al. 2023 [13]). This indicates that the mice had a memory of the familiar object-location pair from the training session and were therefore able to identify which object had moved. However, novelty is a time-limited phenomenon, and as a rodent spends increasing time in the arena on the testing day, the novelty of the moved object will, in theory, begin to diminish. Using Chronotate to log object exploration and preference on a cumulative bout-by-bout basis, we can see that mice showed a maximal preference for exploring the novel object-location after 2.7 s of total exploration, which occurred 45 s into the test session (Fig. 3B, $n = 8$ animals), and that as the mice begin to spend more time with the objects, this preference is diminished.

4. Discussion

Here, we describe Chronotate and provide examples of its utility for facilitating manual scoring of NOR and NOL behavior videos. In short, Chronotate is an open-source graphical user interface implemented in Python that allows the user to score behavior videos with keyboard strokes and outputs timestamped bouts of interest. We show that this time-stamped data can be used to isolate key events, for example, where two raters diverge in their scoring of the data (Fig. 2D). In addition, this time-stamped data can be used to determine if an object-location preference may have been diminished by the duration of the testing session or the amount of total exploration (Fig. 3).

An additional benefit of Chronotate is the ability to skip forward through portions of behavior videos in which an animal is not in proximity to an object of interest, reducing the amount of time raters must spend reviewing each video. However, a future add-on to this software could integrate open-source location tracking modules, such as the one implemented in ezTrack [10] to automatically direct the user to portions of the video in which the animal is close enough to interact with the object, requiring the rater to review only the timepoints when object interaction is possible.

While we describe Chronotate's utility for NOR/NOL videos, its potential applications far exceed this scope. Not only can Chronotate be adapted for the scoring of any videos requiring manual, time-stamped notation, but future refinements to automated behavior scoring tools will hopefully reduce the need to perform manual scoring of behavior videos. In this scenario, Chronotate may prove useful for generating labeled training data to be passed into supervised machine learning

models in order to improve the identification of active object interaction bouts.

CRedit authorship contribution statement

Paul A. Philipsberg: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Zoé Christenson Wick:** Conceptualization, Writing – original draft. **Keziah S. Diego:** Investigation. **Nick Vaughan:** Investigation. **Angelina Galas:** Investigation. **Albert Jurkowski:** Investigation. **Yu Feng:** Investigation, Methodology. **Lauren M. Vetere:** Investigation, Methodology. **Lingxuan Chen:** Investigation, Methodology. **Iván Soler:** Investigation. **Denise J. Cai:** Supervision, Funding acquisition. **Tristan Shuman:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank all members of the Shuman and Cai labs for their comments and feedback throughout the implementation of Chronotate. This work was supported by NIH grants R01 NS116357 (TS), RF1 AG072497 (TS), R01 MH120162 (DJC), DP2 MH122399 (DJC), R56 MH132959 (DJC), F32 NS116416 (ZCW), F31 AG069496 (LV), and an AES Predoctoral Fellowship (YF).

References

- [1] J.W. Krakauer, A.A. Ghazanfar, A. Gomez-Marin, M.A. MacIver, D. Poeppel, Neuroscience needs behavior: correcting a reductionist bias, *Neuron* 93 (2017) 480–490, <https://doi.org/10.1016/j.neuron.2016.12.041>.
- [2] J.K. Denninger, B.M. Smith, E.D. Kirby, Novel object recognition and object location behavioral testing in mice on a budget, *JoVE*. (2018) 58593, <https://doi.org/10.3791/58593>.
- [3] M. Leger, A. Quiedeville, V. Bouet, B. Haelewyn, M. Boulouard, P. Schumann-Bard, T. Freret, Object recognition test in mice, *Nat. Protoc.* 8 (2013) 2531–2537, <https://doi.org/10.1038/nprot.2013.155>.
- [4] A. Vogel-Ciernia, M.A. Wood, Examining object location and object recognition memory in mice, *Current Protocols in Neuroscience*. 69 (2014) 8.31.1–8.31.17. <https://doi.org/10.1002/0471142301.ns0831s69>.
- [5] A. Mathis, P. Mamidanna, K.M. Cury, T. Abe, V.N. Murthy, M.W. Mathis, M. Bethge, DeepLabCut: markerless pose estimation of user-defined body parts with deep learning, *Nat. Neurosci.* 21 (2018) 1281–1289, <https://doi.org/10.1038/s41593-018-0209-y>.
- [6] A.B. Wiltshko, T. Tsukahara, A. Zeine, R. Anyoha, W.F. Gillis, J.E. Markowitz, R. E. Peterson, J. Katon, M.J. Johnson, S.R. Datta, Revealing the structure of pharmacobehavioral space through motion sequencing, *Nat. Neurosci.* 23 (2020) 1433–1443, <https://doi.org/10.1038/s41593-020-00706-3>.
- [7] T.D. Pereira, N.A. Tabris, A. Matsliah, D.M. Turner, J. Li, S. Ravindranath, E. S. Papadoyannis, E. Normand, D.S. Deutsch, Z.Y. Wang, G.C. McKenzie-Smith, C. C. Mitelut, M.D. Castro, J. D'Uva, M. Kislin, D.H. Sanes, S.D. Kocher, S.-S.-H. Wang, A.L. Falkner, J.W. Shaevitz, M. Murthy, SLEAP: a deep learning system for multi-animal pose tracking, *Nat. Methods* 19 (2022) 486–495, <https://doi.org/10.1038/s41592-022-01426-1>.
- [8] V. Ibañez, L. Bohlen, F. Manuella, I. Mansuy, F. Helmchen, A.-S. Wahl, EXPLORE: a novel deep learning-based analysis method for exploration behaviour in object recognition tests, *Sci. Rep.* 13 (2023) 4249, <https://doi.org/10.1038/s41598-023-31094-w>.
- [9] C.J. Gabriel, Z. Zeidler, B. Jin, C. Guo, C.M. Goodpaster, A.Q. Kashay, A. Wu, M. Delaney, J. Cheung, L.E. DiFazio, M.J. Sharpe, D. Aharoni, S.A. Wilke, L. A. DeNardo, BehaviorDEPOT is a simple, flexible tool for automated behavioral detection based on markerless pose tracking, *Elife* 11 (2022) e74314, <https://doi.org/10.7554/eLife.74314>.
- [10] Z.T. Pennington, Z. Dong, Y. Feng, L.M. Vetere, L. Page-Harley, T. Shuman, D. J. Cai, ezTrack: An open-source video analysis pipeline for the investigation of animal behavior, *Sci. Rep.* 9 (2019) 19979, <https://doi.org/10.1038/s41598-019-56408-9>.

- [11] R. Gstir, S. Schaffner, M. Scheideler, M. Misslinger, M. Griehl, N. Daschil, C. Humpel, G.J. Obermair, C. Schmuckermair, J. Striessnig, B.E. Flucher, A. Hüttenhofer, Generation of a neuro-specific microarray reveals novel differentially expressed noncoding RNAs in mouse models for neurodegenerative diseases, *RNA* 20 (2014) 1929–1943, <https://doi.org/10.1261/rna.047225.114>.
- [12] L.M. Stevens, R.E. Brown, Reference and working memory deficits in the 3xTg-AD mouse between 2 and 15-months of age: a cross-sectional study, *Behav. Brain Res.* 278 (2015) 496–505, <https://doi.org/10.1016/j.bbr.2014.10.033>.
- [13] L. Chen, T.R. Francisco, A.M. Baggetta, Y. Zaki, S. Ramirez, R.L. Clem, T. Shuman, D.J. Cai, Ensemble-specific deficit in neuronal intrinsic excitability in aged mice, *Neurobiol. Aging* 123 (2023) 92–97, <https://doi.org/10.1016/j.neurobiolaging.2022.12.007>.
- [14] L. Chen, Z. Christenson Wick, L.M. Vetere, N. Vaughan, A. Galas, K.S. Diego, P. Philipsberg, D.J. Cai, T. Shuman, Progressive excitability changes in the medial entorhinal cortex in the 3xTg mouse model of Alzheimer's disease pathology, *BioRxiv.* (2023), <https://doi.org/10.1101/2023.05.30.542838>.
- [15] J.A. Clark, Pillow (PIL Fork) Documentation, (2023). <https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf>.
- [16] W. McKinney, Data Structures for Statistical Computing in Python, in: Austin, Texas, 2010: pp. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.