

Novel Tools and Methods

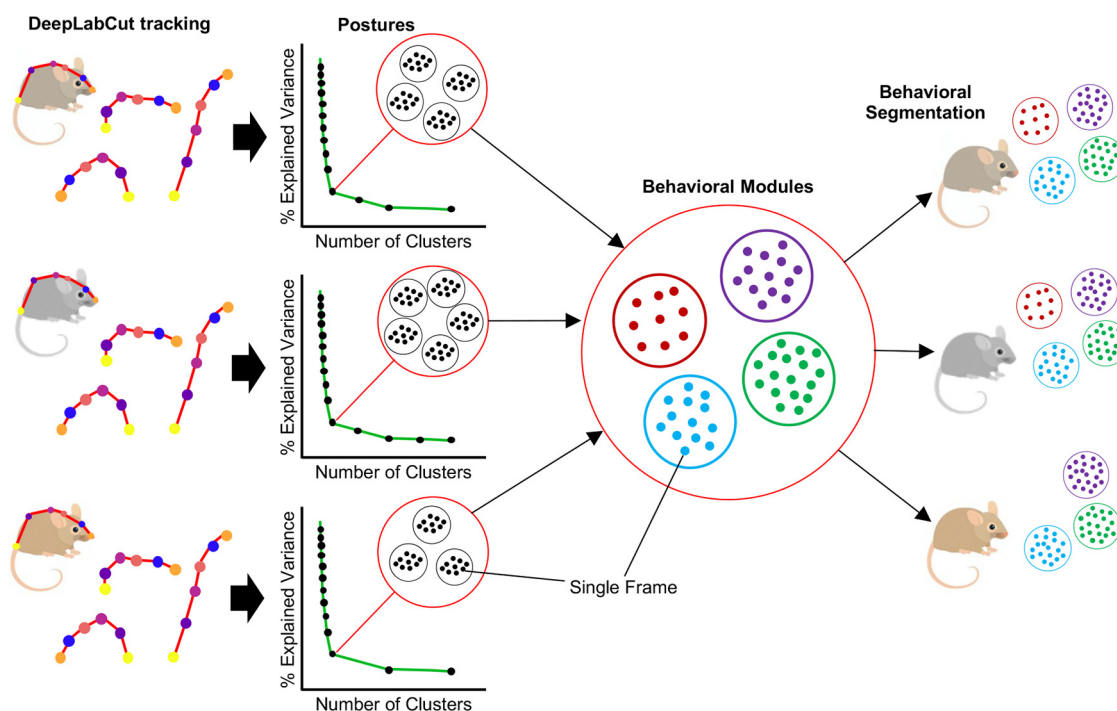
Automated Segmentation of the Mouse Body Language to Study Stimulus-Evoked Emotional Behaviors

Gabriele Chelini,¹  Enrico Maria Trombetta,¹ Tommaso Fortunato-Asquini,² Ottavia Ollari,¹ Tommaso Pecchia,¹ and  Yuri Bozzi^{1,3}

<https://doi.org/10.1523/ENEURO.0514-22.2023>

¹Center for Mind/Brain Sciences (CIMEC), University of Trento, Rovereto 38068, Italy, ²Department of Cellular, Computational, and Integrative Biology (CIBIO), University of Trento, Trento 38123, Italy, and ³Consiglio Nazionale delle Ricerche (National Council of Research) Neuroscience Institute, Pisa 56124, Italy

Visual Abstract



Understanding the neural basis of emotions is a critical step to uncover the biological substrates of neuropsychiatric disorders. To study this aspect in freely behaving mice, neuroscientists have relied on the observation of ethologically relevant bodily cues to infer the affective content of the subject, both in neutral conditions or in response to a stimulus. The best example of that is the widespread assessment of freezing in experiments testing both conditioned and unconditioned fear responses. While robust and powerful, these approaches come at a cost: they are usually confined within selected time windows, accounting for only a limited portion of the complexity of emotional fluctuation. Moreover, they often rely on visual inspection and subjective judgment, resulting in inconsistency across experiments and questionable result interpretations. To overcome these limitations, novel tools are arising, fostering a new avenue in the study of the mouse naturalistic behavior. In this work we developed a computational tool [stimulus-evoked behavioral tracking in 3D for

Significance Statement

We combined novel technical advancements of system neuroscience with a traditional ethology-based behavioral observation to design a simple computational tool for behavior detection in freely moving mice. Using machine learning-based behavioral tracking and unsupervised cluster analysis, we identified statistically meaningful postures on a subsecond scale. We validated this tool in the whisker nuisance (WN) task, where mice are challenged with a prolonged and invasive whisker stimulation, showing that identified postures can be used as a proxy for stimulus-driven fearful and explorative behaviors. With this tool we aim to automate, accelerate and standardize data collection across research laboratories, improving the quality and reproducibility of behavioral studies in mice.

rodents (SEB3R)] to automate and standardize an ethologically driven observation of freely moving mice. Using a combination of machine learning-based behavioral tracking and unsupervised cluster analysis, we identified statistically meaningful postures that could be used for empirical inference on a subsecond scale. We validated the efficacy of this tool in a stimulus-driven test, the whisker nuisance (WN) task, where mice are challenged with a prolonged and invasive whisker stimulation, showing that identified postures can be reliably used as a proxy for stimulus-driven fearful and explorative behaviors.

Key words: behavioral segmentation; DeepLabCut; mouse behavior; mouse body language; mouse emotion; whisker nuisance

Introduction

As with any behaving animal, naturalistic mouse behavior incorporates innate strategies to balance explorative and aversive responses to the surrounding environment. Neuroscientists often exploit this dichotomy to design behavioral tests assessing mouse emotional behavior. Many of these tests put the animal in a forced choice between a putatively safe location versus a potentially threatening alternative (Takao and Miyakawa, 2006; Komada et al., 2008; Bailey and Crawley, 2009; Campos et al., 2013; Kraeuter et al., 2019). Other approaches use stimulus-driven tasks to assess the positive or negative valence of the emotional response to a reward or fearful cue respectively, using one specific behavioral output as a proxy to infer emotional state (such as self-administration of a

positive stimulus or freezing in response to a threat; Panlilio and Goldberg, 2007; Campos et al., 2013; Gafford and Ressler, 2016). While extremely valuable, none of these approaches address the dynamic change in the subjects' affective state over extended periods, limiting results interpretation to circumscribed context rather than interrogating the emotional fluctuation occurring during ecological behavioral flow. Thanks to the advent of advanced tools for behavioral tracking and unbiased computer-driven classification, novel ethologically relevant methods to classify mouse emotions are arising (Wiltshchko et al., 2015; Mathis et al., 2018; Nath et al., 2019; Dolensek et al., 2020; Bohoslav et al., 2021; Hsu and Yttri, 2021; Karashchuk et al., 2021; Luxem et al., 2022), promising a new era in the use of mouse behavior as an optimal translational tool for neuropsychiatric research. To fully exploit the potential of these methods, the field craves for open-access resources to break-down and analyze the output. For instance, using the pose-estimation software DeepLabCut (DLC), it is possible to reliably track several bodily hotspots on a behaving animal, allowing for virtually limitless identification of animal behavioral components. However, every new tracking configuration will require dedicated analysis pipeline to swiftly transform DLC output files into interpretable results. Generate and sharing tools to perform this transformation is an integral step to standardize behavioral assessment across research laboratories, thus improving interpretation and replicability. In this work, we conceptualized an automated procedure to discriminate mouse body postures on a subsecond scale that we named SEB3R (stimulus-evoked behavioral tracking in 3D for rodents), using a combination of machine-learning based behavioral tracking and unsupervised cluster analysis. After visualizing ~100 videos of freely moving mice responding to a nonpainful whisker stimulation, we concluded that a good predictor of mice affective response could be simplified as the spatial location, along the vertical (z)-axis,

Received December 27, 2022; accepted July 24, 2023; First published August 30, 2023.

The authors declare no competing financial interests.

Author contributions: G.C., T.P., and Y.B. designed research; G.C., T.F.-A., and O.O. performed research; G.C., E.M.T., and T.P. contributed unpublished reagents/analytic tools; G.C., E.M.T., and T.F.-A. analyzed data; G.C. and Y.B. wrote the paper.

G.C. was supported by the CARITRO Postdoctoral Fellowship, funded by Fondazione Cassa di Risparmio di Trento e Rovereto. Y.B. is supported by the Strategic Project TRAIN-Trentino Autism Initiative (<https://projects.unitn.it/train/index.html>) from the University of Trento (Grant 2018-2022).

Acknowledgments: We thank all the administrative and technical staff of CIMeC for support. We also thank Michela Maffei, animal caretaker of the CIMeC animal facility, for her endless care and support provided in managing the animal colony used in the study. We thank the carpenter Marco Chelini for donating materials and equipment used in the study.

Correspondence should be addressed to Gabriele Chelini at gabriele.chelini@unitn.it or Yuri Bozzi at yuri.bozzi@unitn.it.

<https://doi.org/10.1523/ENEURO.0514-22.2023>

Copyright © 2023 Chelini et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

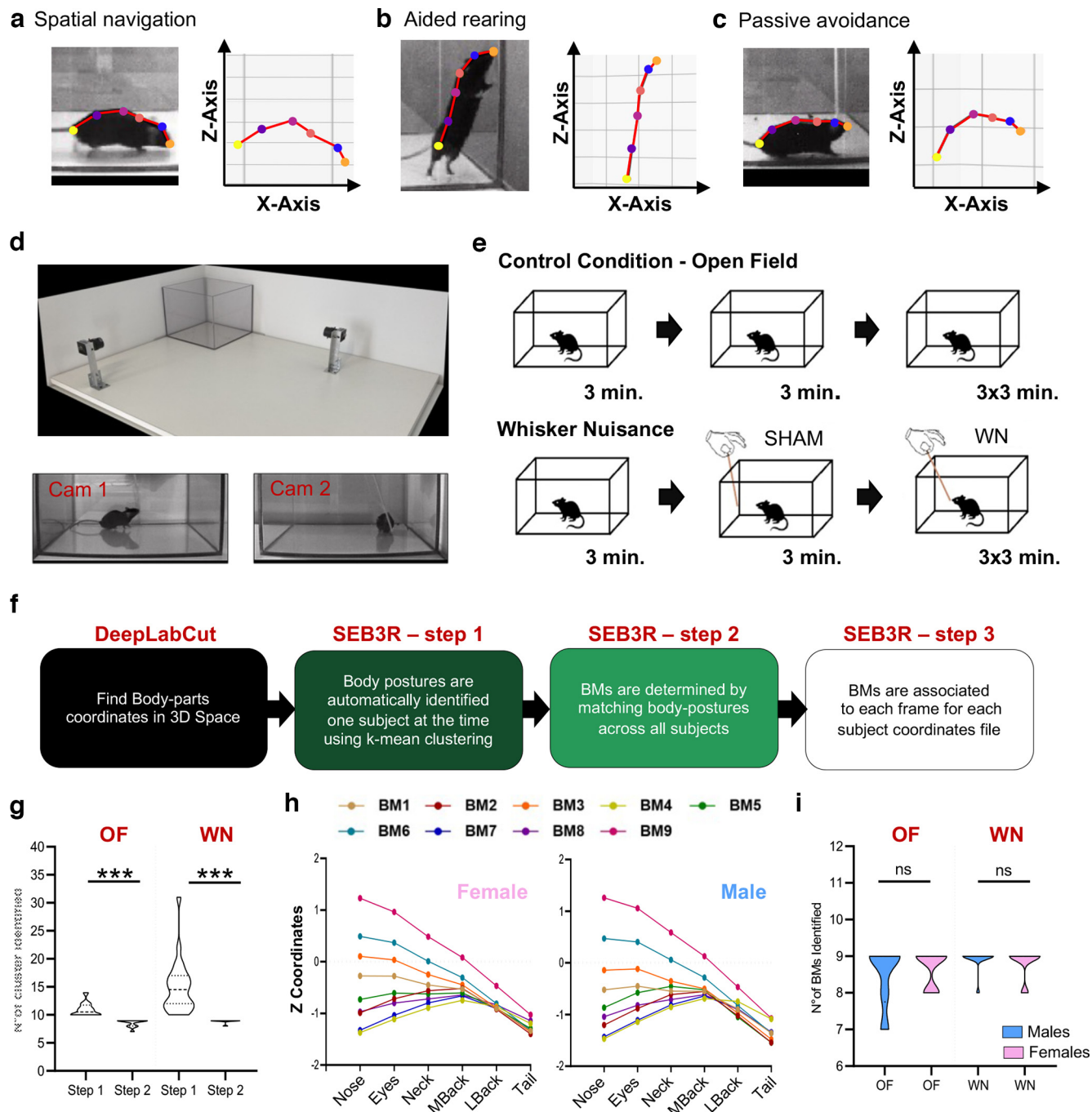


Figure 1. SEB3R pipeline. **a–c**, Examples of body postures determined by visual inspection (left) and corresponding coordinates detected by DLC-3D. Bodily hotspots are indicated by colored dots (orange, nose; blue, eyes; pale red, neck; magenta, mid back; purple, lower back; yellow, tail attachment). **d**, Top, Angular view of the arena for 3D motion capture. Bottom, Orthogonal view of the same mouse during the WN. **e**, Design of the validation experiment. Top, Open field (OF). Bottom, Whisker nuisance task. **f**, Schematic workflow of the SEB3R pipeline. **g**, A two-step cluster analysis eliminates interindividual variability. **h**, Schematic view of BMs identified during the open field session, for both sexes. **i**, The final number of BMs is insensitive to size-biases, thus resulting in analogous identification in both sexes (Extended Data Figs. 1-1, 1-2, 1-3, 1-4, 1-5, 1-6 show screenshots for the critical outputs of the SEB3R pipeline). *** $p < 0.0001$; ns, not significant.

of specific “bodily hotspots”: nose, eyes, neck, mid back, lower back, tail attachment (Fig. 1a–c). More specifically, the relative position of one hotspot compared with the others changes drastically according to specific behavioral states such as spatial navigation, explorative

rearing, or passive avoidance (Fig. 1a–c). Taking advantage of the new-generation software for pose estimation DeepLabCut (DLC; Mathis et al., 2018; Nath et al., 2019), we tracked these hotspots in freely moving mice and used the output coordinates to read mice body language. To

validate this method, we compared behavior in mice navigating undisturbed an open field (OF) arena against a group of littermates challenged with an invasive whisker stimulation [whisker nuisance task (WN); McNamara et al., 2010; Fontes-Dutra et al., 2018; Chelini et al., 2019; Balasco et al., 2022a; Fig. 1e].

Materials and Methods

Animals

All experimental procedures were performed in accordance with Italian and European directives (DL 26/2014, EU 63/2010) and were reviewed and approved by the University of Trento animal care committee and Italian Ministry of Health. Animals were housed in a 12/12 h light/dark cycle with unrestricted access to food and water. All killings for brain explant were performed under anesthesia and all efforts were made to minimize suffering. A total of 48 age-matched adult wild-type littermates of both sexes (weight 25–35 g) were used for the study. All mice were generated from our inbred CNTNAP2 colony with C57BL/6 background (Balasco et al., 2022b). Thirty-six mice were assigned to the whisker stimulation test, while 12 were assigned to the open field condition.

3D motion capture

A basic requirement for DLC tracking in 3D (DLC-3D) is to capture the subject motion using (at least) two video cameras in a stereo-configuration (Mathis et al., 2018; Nath et al., 2019). To this goal, two video cameras (ImageSource DMK27BUR0135 USB3-monochrome, equipped with TAMRON 13VM308ASIRII objectives) were secured, using a metal pedestal, on a board of laminated wood, in orthogonal positioning respect to each other (Fig. 1d). At the corner of the base, two aluminum guides were placed to ensure the stability of the cubic cage during the test. This allowed us to replace the experimental cage at every given animal, thus eliminating experimental biases because of the odor of other mice, that could potentially alter the emotional state of the test subject. Videos were acquired at a rate of 25 frames per second (fps) using the OBS-studio software, and edited using a commercially available software (OpenShot Video Editor), before being analyzed with DLC. For detail about how to use DLC-3D we recommend this link: <https://github.com/DeepLabCut/DeepLabCut/blob/master/docs/Overviewof3D.md>.

Behavioral testing

For all experiments, mice were previously acclimated with the experimental setup for 4 d. For the first 3 d, mice were placed into the experimental cage (a fiberglass cubic box with open top; 25 × 25 × 20 cm) for 2 min. This time was extended to 10 min on day 4 (day before the test). This habituation scheme will be referred as heavy habituation (HH in figures and tables) for the rest of the manuscript. On the day of the test, mice were acclimated with the experimental environment for 5 min, before the beginning of the recordings. To assess the efficacy of our pipeline in discriminating between experimental conditions, a group of

12 mice underwent WN after only 2 d of habituation [referred to as mild habituation (MH) in figures and tables]. All experiments were performed in dim light.

Open field

Mice were placed and left free to navigate the experimental cage for a total of 20 min. Video recordings followed the same chronological scheme of the WN to be able to compare the two conditions with each other (Fig. 1e, top).

Whisker nuisance task (WN)

WN was performed with some variation compared with what previously described (Chelini et al., 2019; Balasco et al., 2022a). Before the beginning of the actual test, 3 min of the animal freely moving in the open field (OF) arena were videotaped as a baseline activity session. The testing phase was composed of four sessions, lasting 3 min each. In the first session (SHAM), a wood stick was introduced in the experimental cage, avoiding direct contact with the animal. The following three sessions (T1–T2–T3) consisted in stimulating mice's whiskers by continuously deflecting vibrissae using the wooden stick (at a frequency of approximately three strokes per second; Fig. 1e, bottom).

To dissect the complexity of behavioral response, multiple behavioral categories were independently quantified by a trained observer. The identified categories, re-adapted from previous version of the test (McNamara et al., 2010; Fontes-Dutra et al., 2018; Chelini et al., 2019; Balasco et al., 2022a), included fearful and curious behaviors. Fearful behaviors were divided into active avoidance (time the animal spends actively moving away from the stick) and passive avoidance (time the animal spends in a defensive posture consisting in curved back, protracted neck, and stretched limbs or retracted in fully hunched posture; this measure also included the time in freezing). Curious behaviors were divided in aided rearing (when the animals, during a rearing action, leans on the arena's walls investigating the surrounding environment) and unaided rearing (when the rearing is not supported by walls; this behavior usually occurs toward the center of the arena and is associated with stick exploration during WN).

Tissue processing and immunofluorescence

Ninety minutes after the end of WN, mice were deeply anesthetized with isoflurane and killed by decapitation. Brains were excised, washed in 0.1% PBS and postfixed overnight in 4% paraformaldehyde (PFA), switched to a cryoprotectant solution (80% PBS, 20% glycerol with 0.1% sodium azide) and stored at 4°C. Cryoprotected brains were sectioned on a vibratome (Leica, VT1200) at 40- μ m thickness. Serial sections were collected in 24 separate compartments and stored at 4°C in cryoprotectant solution.

Free-floating slices were rinsed three times in PBS (10 min each), then washed in PBS containing 0.2% detergent (Triton X-100, Fisher, AC215680010) for 30 min. Tissue sections were then incubated in blocking solution [2% bovine serum albumin (BSA), 1% fetal bovine serum (FBS) in PBS] for 3 h and then transferred to primary

antibody solution [2% BSA, 1% FBS, 1:1000 dilution of primary antibody (rabbit anti-Arc/Arg 3.1; Proteintech, catalog 316290-1-AP)] and incubated at room temperature for 24 h. Then, sections were rinsed three times in PBS (5 min each) and placed in a fluorophore-conjugated (Alexa Fluor Plus 488) secondary antibody solution [1:300 dilution of donkey anti-rabbit secondary antibody (Thermo Fisher Scientific, AB_2762833) in PBS] for 24 h. Sections were then washed 5 min in PB, mounted on superfrost slides, dried for 1 h and coverslipped with fluorescent mounting medium (Southern biotech 0100-01). Slides were stored at 4°C in the dark until use.

Confocal microscopy and image analysis

A confocal laser scanning microscope Leica TCS-SP8, equipped with a HC PL APO 20× objective and interfaced with Leica LAS-X software was used to detect ARC immunolabeling in the lateral (LA) and basolateral (BLA) nuclei of the amygdala. Images were recorded at a resolution of 1024 pixels square, 400-Hz scan speed. Excitation/emission wavelengths were: 490/520 for Alexa-488 fluorophore. Acquisition parameters were set during the first acquisition and kept consistent for all the images. Corrected fluorescence intensity (CFI) was quantified using ImageJ software according to the formula: $CFI = \text{integrated density} - (\text{area size} \times \text{mean fluorescence of background readings})$.

Data collection and analysis

Data collection for all studies was conducted by investigators blind to experimental conditions. All statistical analyses were conducted using JMP-pro software (JMP Statistical Discovery LLC, 2023). For the purpose of this study, we have used three different experimental settings (open field, WN mild habituation and WN heavy habituation) to assess SEB3R ability to discriminate between experimental conditions. By consequence, the absolute time spent in BMs was not a reliably comparable measure, because of noticeable changes in the animal's behavior during the SHAM session. These discrepancies are a direct consequence of the animal familiarity with the experimental setting (habituation regimen) or the lack of the stimulus in the control condition. To overcome this technical limitation, two separate normalization strategies were used. Time spent in BMs was normalized using the habituation session (first 3 min of videorecording) as a reference baseline using the formula: $(\text{number of frames during each test session} - \text{number of frames during habituation session}) / (\text{number of frames during each test session} + \text{number of frames during habituation session})$, to demonstrate SEB3R ability to discriminate between stimulus-independent and stimulus-driven changes. Similarly, to demonstrate SEB3R ability to discriminate between whisker-independent and whisker-dependent behaviors we normalized the time spent in BMs using the SHAM session as a reference with the formula: $(\text{number of frames during whisker stimulation session} - \text{number of frames during SHAM session}) / (\text{number of frames during whisker stimulation session} + \text{number of frames during SHAM session})$.

SEB3R workflow

The SEB3R pipeline is summarized in Figure 1f. To run the MATLAB scripts, body-parts coordinates need to be first identified using DLC-3D. For instruction about DLC usage we suggest this link (<https://github.com/DeepLabCut>) and the relative literature (Mathis et al., 2018; Nath et al., 2019). As indicated on DLC-3D instructions, no major camera requirements are needed, as long as the videos can be reliably labeled to train the network. We recommend users to choose carefully the cameras acquisition rate depending on their experimental needs, as the final output of SEB3R will provide posture identification for each individual frame. Similarly, it is important to ensure both devices run at the same acquisition rate and are temporally synchronized. As for the neural network, for the validation experiment, we used ResNet-50, but we see no reasons the pipeline would not work with other options included in DLC. The body parts required by the analysis are (shown in Fig. 1a–c): nose, eyes, neck, mid back, lower back, and tail attachment.

After training your 3D network, analyze your videos and save the output tracking file in .CSV format. For each subject, place all the tracking files into a folder labeled with the animal identification (ID) number. Files should also contain the same ID in the name. Then place all subjects' folders into the same mother-folder. It is recommended, for experiments composed of multiple sessions, to analyze videos from each session separately. Doing so, the pipeline will provide a separate output for each of the session, facilitating data analysis. In this case it is important for all DLC-3D files to contain the same number of frames analyzed.

SEB3R step 1

Identify meaningful postures in a single subject. Launch the script PoseExtraction by selecting your mother-folder. Assuming you have multiple tracking files for each subject (like in the case of an experiment with multiple sessions on the same mouse), the first step of SEB3R imports and combines the columns containing the z-coordinates of each subject. This step will ensure that postures identification is uniformed across various experimental conditions, limiting the risk of obtaining false-positive results. Then, from the z-coordinates, and within each frame, the linear distance between all the identified hotspot is calculated subtracting the values of each individual coordinate with the value of all the others. This will result in a 15-by- N matrix, with $N = \text{sum of the number of frames contained in the original coordinate files}$. This simple step eliminates most of the interindividual variability because of discrepancies in the animal dimensions, transforming the row coordinates into relative distances scaled by the animal body size. Using the relative distances matrix, the algorithm runs a k -means clustering using a customized version of the MATLAB function `kmean: kmean_opt` (De Landtsheer, 2019). This function uses the Elbow method, a statistical strategy to determine the number of k -mean clusters to choose (Thorndike, 1953). In our context, the advantage of this methods is to identify a “ K ” number of clusters preventing any input from the human user, thus achieving complete unsupervised identification. Each subject is processed independently, but the

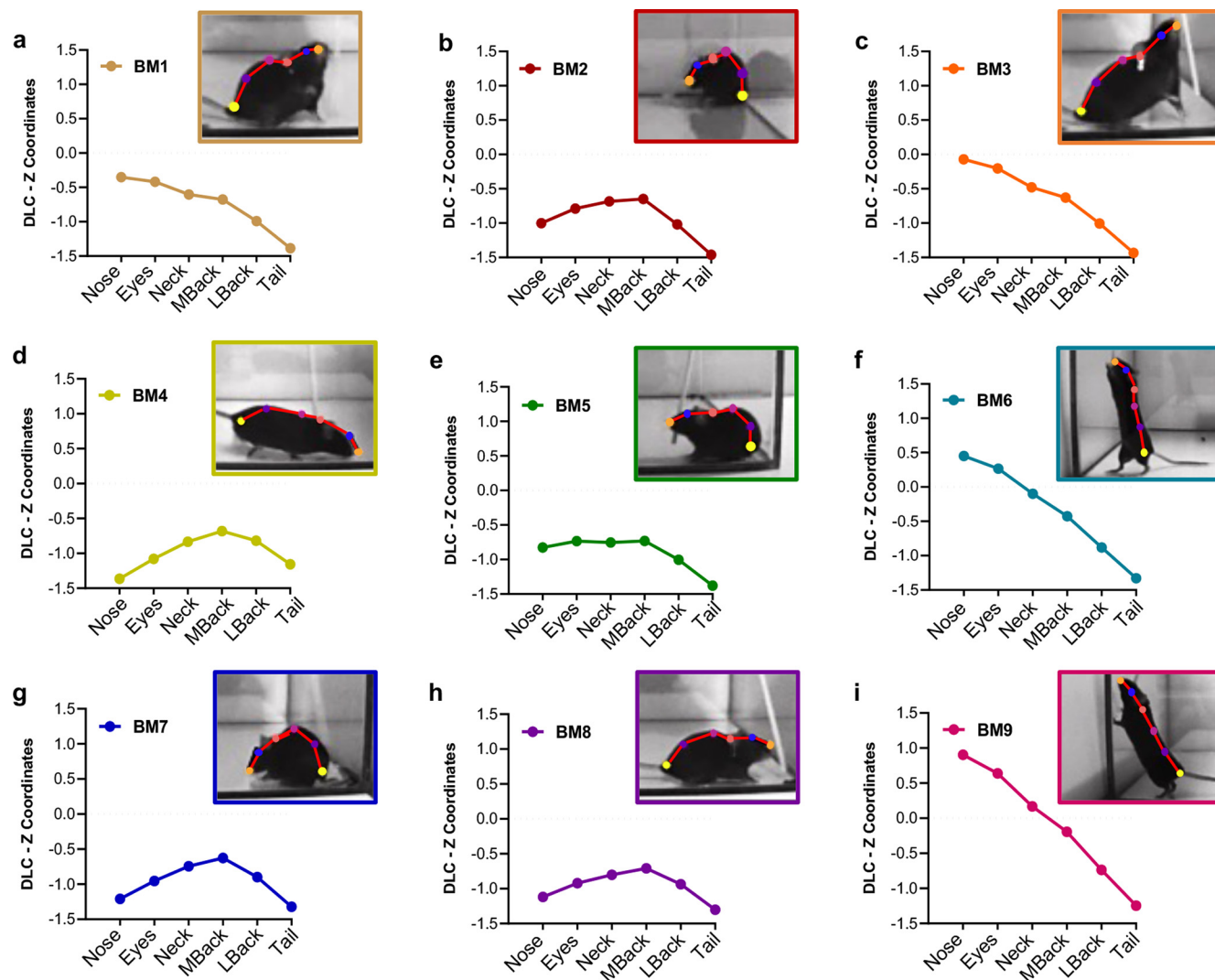


Figure 2. Behavioral modules (BMs) identified by SEB3R. **a–i**, The plots report the DLC z-coordinates for each of the six bodily hotspots (nose, eyes, neck, mid back, lower back, and tail attachment) that allow to identify each of the nine BMs. A representative frame is reported for each BM.

algorithm will sequentially run on all the files included in the mother folder, providing the results of *k-mean* clustering for each individual animal. The script runs for 5–10 min per subject, depending on your device, and outputs two folders containing a summary of the results of the clustering. Clusters identified in this step correspond to statistically relevant postures and will be recapitulated in an output file (MouseIDDistMean.CSV) containing (1) the animal ID in the first column; (2) the averaged distances between hotspots sorted by individual postures (columns 2–16); (3) the posture identification number for each row, on column 17 (Extended Data Fig. 1-1). A copy of all the MouseIDDistMean.CSV files for each subject will be saved in the folder named Mean Distances, within the mother folder. These are the files that will be used in step 2.

SEB3R step 2

Match postures identified within each subject to determine behavioral modules (BMs) replicated across the entire

experimental group. Launch the script ModulesExtraction and select all the folder Mean Distances (containing the output files named MouseIDDistMean.CSV) within the mother folder. This script imports and combines all the averaged distance matrices for each subject and runs a second *k-mean* clustering, automatically matching postures identified in different animals. In this case, the user needs to choose the *k-number* of clusters to select to maximize BM representation of in the cohort. A window will pop-up asking the user to specify the *k-number* desired, that will correspond to the final number of BM (Extended Data Fig. 1-2). To this goal, it is recommended to use a $k = \text{MIN}(\text{poses}) + 1$, where $\text{MIN}(\text{poses})$ is the minimum number of postures identified in a single subject. As an example, in our validation experiment the lowest number of postures identified in a single subject was 8, hence we chose a $k = 9$. This crucial step converges highly similar poses (within and between subjects) into uniform BMs replicated across all (or most) animals, which can be used for comparisons, under the

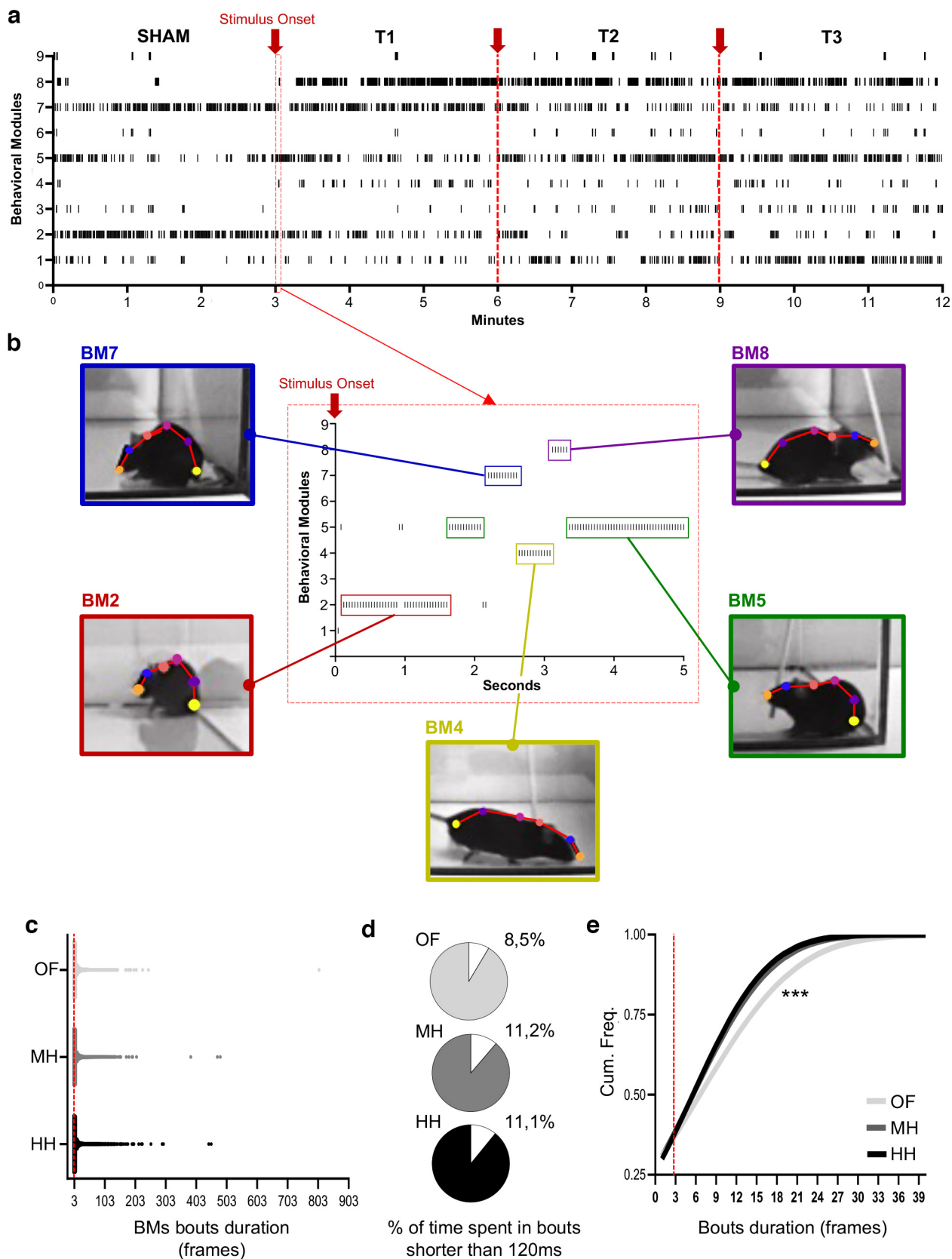


Figure 3. Behavioral recognition on a frame-by-frame resolution in the WN. **a**, Raster plot showing the occurrence of BMs in a prototypical mouse for the entire duration of the WN. Each vertical bar (|) indicates a BM detected in the corresponding period (sham session and trial sessions T1–T3). Stimulus onset at the beginning of each stimulation trial is indicated by a red arrow. **b**, Example

continued

of BMs detected during the first 5 s of whisker stimulation in the WN; the *x*-axis reports the seconds after the beginning of whisker stimulation in T1. **c**, Violin plot showing the spread in the duration of BMs sequences. **d**, Percentage of frames belonging to BMs sequences shorter than three frames (120 ms) is lower in the OF condition compared with WN. **e**, Cumulative frequencies chart shows that BMs sequences are longer in the OF compared with WN (*x*-axis is cropped at a duration of 39 frames to better appreciate the differences; all 3 curves perfectly overlap after this point). *** $p < 0.0001$.

assumption that they are equally informative of the same behavioral state. The final number of BMs obtained with this method has minimal interindividual variability (Fig. 1g). Furthermore, we found no differences in the number of BMs identified in males and female subject (Fig. 1h,i), suggesting that the animal body size does not influence the detection of BMs. For all subjects, the matched pairs (posture-BMs) are recapitulated in the output file named ModulesAssigned.CSV (Extended Data Fig. 1-3) that will be directly saved in the mother folder. This file will be required to proceed with the third and last step.

SEB3R step 3

Assign BMs to each frame. Launch the script AssignModulesAndQuantify and select the mother folder and the ModulesAssigned.CSV file. The algorithm will import, one-by-one, the Clustered.CSV files generated in step 1. These files originate from the initial DLC-3D tracking and contain: the animal ID in column 1, the *z*-coordinates for each frame in columns 2–7, and the posture identifier (i.e., clusters identified in step 1) assigned to each frame in column 8 (Extended Data Fig. 1-4). Using the output of step 2 (ModulesAssigned.CSV), the postures contained in column 8 of Clustered.CSV are matched with the corresponding BMs, generating an output file identical to the Clustered.CSV, with the addition of a column 9 containing BMs identifying number for each frame (MouseID.CSV; Extended Data Fig. 1-5). The output files of this step will be saved in the folder Modules located within each subject subfolder (one file will be generated for each original DLC-3D tracking files input at the beginning of step 1). Thanks to this step, each frame is labeled with the appropriate BM, providing behavioral recognition on a timescale that depends on the original cameras acquisition frame-rate, allowing absolute flexibility in analyzing time-dependent changes in the behavioral flow. Figure 2 illustrates the nine BMs detected in our validation experiment. Finally, for each subject, the pipeline prints a file recapitulating the total number of frames assigned to each independent BM (FrequenciesMouseX.CSV; Extended Data Fig. 1-6), which can be used as an outcome measure to quantify the expression of different behaviors over consecutive sessions of the same test. In these files, each row corresponds to a BM and each column contains the number of frames assigned to that BM for each separate session of the test, according to the number of DLC-3D tracking files originally provided.

Code accessibility

This code can run on any personal laptop on MATLAB-2020 and subsequent versions. The MATLAB code for SEB3R with the relative instructions are available on

public repository GitHub, at the following address: <https://github.com/gchelini87/SEB3R>.

We also share the data resulting from our validation experiments on g-node platform as an aid to familiarize with the use of SEB3R: https://gin.g-node.org/gchelini/SEB3R_Validation_Data/settings.

Results

Previous studies show that during the consecutive sessions of the WN the mouse attitude toward the stimulus (wooden stick) shifts drastically from a complex aversive response to relaxed and explorative (Balasco et al., 2022a). For this reason, this test was chosen to assess whether the BMs detected with SEB3R (Fig. 2a–i) could be used to reliably identify changes in stimulus-driven fearful and curious behaviors over an extended experimental session.

SEB3R provides subsecond behavioral detection

Figure 3 shows that BMs can be reliably identified on a frame-by-frame resolution during the WN. All nine BMs are detected throughout the four sessions (sham, T1–T3) of the whisker stimulation task.

We found that BMs duration is highly variable, and the majority of BMs events are composed of sequences of one to ~20 frames, with a corresponding median of three frames (120 ms) that is shared by all three experimental conditions (Fig. 3c). Exceptionally we observed some longer bouts spanning from 400–806 frames in all three experimental conditions (see an example in Movie 1). Given our resolution of 25 fps (that is, each frame captures 40 ms of action), we wondered whether there was a lower threshold in the duration of BMs to be considered ethologically unreliable. To answer this question, we quantified the cumulative percentage of time spent in BM bouts shorter than three frames using the following formula: $[(\text{number of frames in 1-frame long bouts} + \text{number of frames in 2-frames long bouts}) \times 100] / (\text{total number of frames})$ (Fig. 3d). To note, and in keeping with what shown in Figure 3c, manual quantification revealed that it is common to record behavioral bouts resolving within 100 and 200 ms, leading us to hypothesize that three frames might be the lowest reasonable threshold for reliable detection of ethologically relevant behaviors.

To our surprise, we found that sequences lower than 100 ms account for ~11% of the total time in the WN (irrespectively to the habituation condition), while only ~8% in the OF (Fig. 3d). More broadly, we found that BMs sequences are longer in the OF compared with both WN conditions (Fig. 3c; Kruskal–Wallis's test: mild vs heavy habituation $p = 0.49$; mild habituation vs OF $p < 0.0001$; heavy habituation vs OF $p < 0.0001$). Moreover, this effect

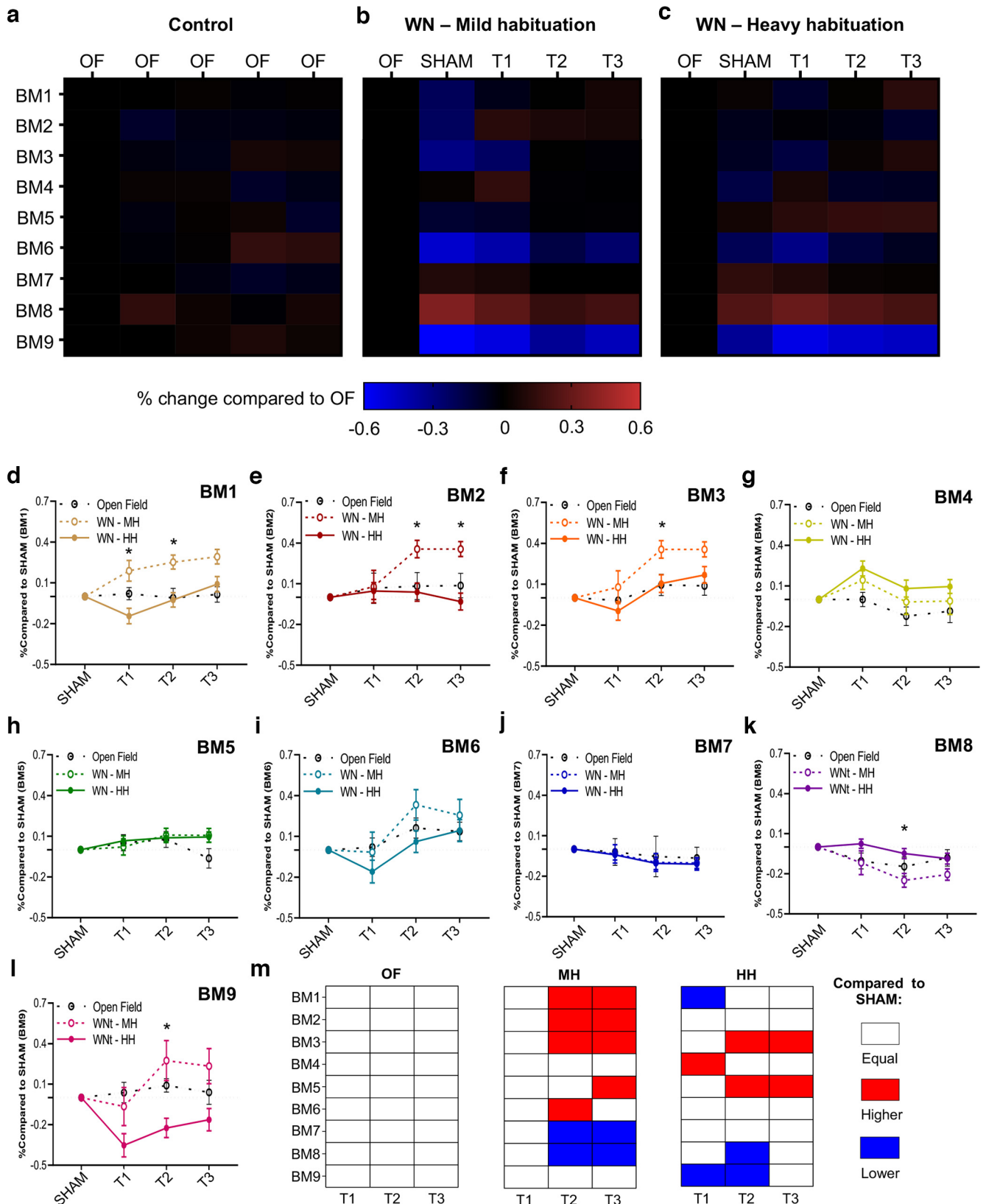


Figure 4. SEB3R discriminates stimulus-driven behavioral changes. **a**, Control condition (open field, OF). **b**, Whisker nuisance (WN) with heavy habituation (HH). **c**, WN with mild habituation (MH). The percentage change in time spent in BMs varies depending on the test session and experimental conditions. The heatmap scale indicates the percentage change in the time spent in each BM normalized over the habituation session (baseline value = 0%); where the habituation session is the first of the five sessions in each of

continued

the three experimental conditions (control, WN-HH, and WN-MH). (Statistics are summarized in Extended Data Fig. 4-1.) **d–l**, Each plot shows the percentage changes of BMs with respect to the SHAM session, for all three experimental conditions. Asterisks (*) indicate $p < 0.05$ between HH and MH. **m**, Diagram recapitulating significant differences between the SHAM session and each stimulation trial. White = no difference. Red = significantly higher than SHAM. Blue = significantly lower than SHAM. (Statistics are summarized in Extended Data Figs. 4-2, 4-3, 4-4, 4-5, 4-6, 4-7, 4-8, 4-9, 4-10 for repeated measure and Extended Data Figs. 4-11, 4-12, 4-13 for between groups multiple comparisons). Error bars refer to SEM.

is abolished in curiosity-driven BMs 3–6–9 (Extended Data Fig. 3-1). While our data do not allow to conclude that bouts lower than 100-ms capture complete behavioral sequences, the discrepancy observed between OF and WN suggests that the velocity of behavioral action increases in response to the invasive stimulation. This indicates that even the lowest detectable BMs sequences retain substantial biological valence. Speculatively, we interpret these short bouts as transitioning frames from one BM to another, rather than full behavioral sequences. However, since we do not have instruments to definitively establish this claim, we have included them as part of the total time of BMs expression. We believe this kind of analysis could be reliably used as an additional metric to investigate qualitative behavioral differences; hence, we have included a script to SEB3R toolbox to automatically extrapolate BMs sequence duration (SequenceDuration.m).

BMs identified using SEB3R discriminate stimulus-driven changes in mouse behavior

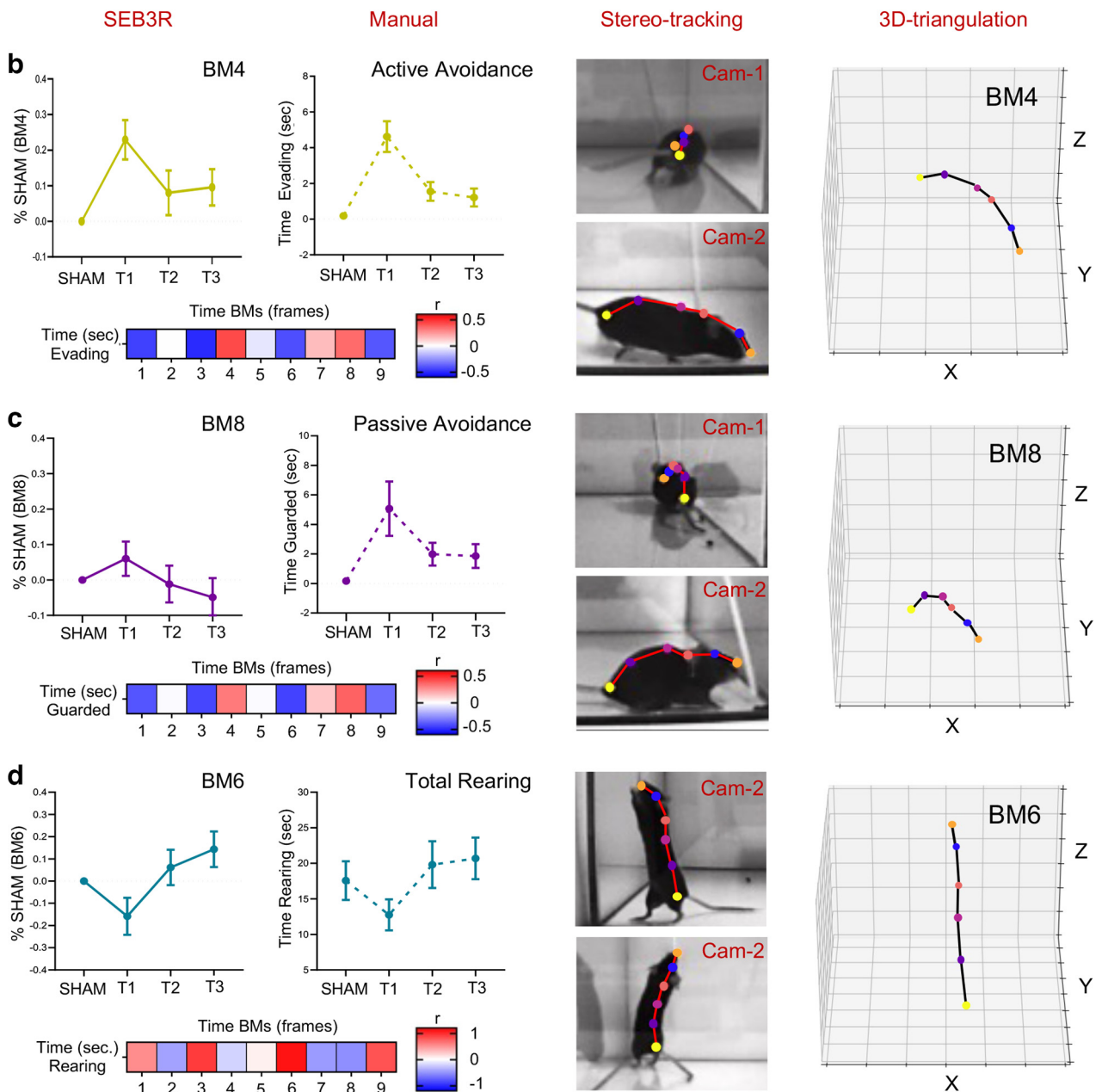
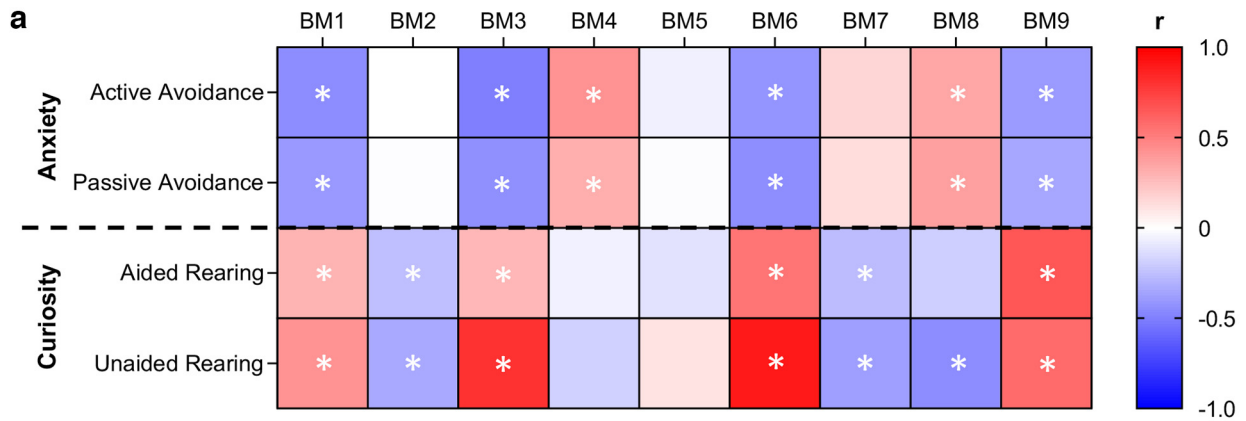
To assess SEB3R ability to discriminate between experimental conditions, we compared the percentage change in time spent in each BM during the habituation phase with the four consecutive sessions of the test. Predictably, no significant differences were observed at any timepoint in the open field (OF) condition (Fig. 4a). Conversely, we identified significant changes in the heavy habituation (BM1, $\text{prob} > \chi^2 > 0.02$; BM3, $\text{prob} > \chi^2 > 0.005$; BM4, $\text{prob} > \chi^2 > 0.0108$; BM5, $\text{prob} > \chi^2 > 0.011$; BM6, $\text{prob} > \chi^2 > 0.023$; BM8, $\text{prob} > \chi^2 > 0.0001$; BM9, $\text{prob} > \chi^2 > 0.0001$; Fig. 4b) and mild habituation (BM1, $\text{prob} > \chi^2 > 0.0013$; BM3, $\text{prob} > \chi^2 > 0.0022$; BM6, $\text{prob} > \chi^2 > 0.0112$; BM8, $\text{prob} > \chi^2 > 0.0001$; BM9, $\text{prob} > \chi^2 > 0.0001$; Fig. 4c) groups at several different timepoints (descriptive statistics and p -values of the Friedman's test for nonparametric repeated measure relative to this analysis are summarized in Extended Data Fig. 4-1). Notably, the majority of the BMs (seven out of nine) showed changes starting from the SHAM session, when the stimulus (stick) was introduced in the arena, but before the beginning of the whisker stimulation. This data demonstrates that the exposure to a novel stimulus, although not tactile, is sufficient to trigger a meaningful change in mice behavioral state. Then, to specifically assess the whisker-dependent changes in behavior, we normalized BMs intensity as percentage variation from the SHAM and compared the three consecutive sessions of whisker stimulation (T1–T2–T3) with the baseline value (SHAM = 0, after normalization), for all three experimental conditions. Data resulting from the control condition (open field, OF) were normalized in analogous way, using the second session of acquisition in replacement for the

SHAM. As a reference, we established that there were no changes in the open field, in none of the timepoints compared with the baseline (T1–T2–T3; Fig. 4d–l). In the mild habituation group, instead, significant increase in the expression of BM1 (T2 and T3, $p < 0.0001$), BM2 (T2, $p = 0.0006$, T3, $p < 0.0001$), BM3 (T2, $p < 0.0001$, T3, $p = 0.0002$), BM5 (T3, $p < 0.0034$), BM6 (T2, $p = 0.0077$), and BM9 (T2, $p < 0.0077$) were found, while a significant decrease was observed for BM7 (T2 and T3, $p < 0.0034$) and BM8 (T2, $p = 0.0006$, T3, $p < 0.0001$). Finally, in heavy habituation, a significant increase was found in BM3 (T2, $p = 0.0084$, T3, $p = 0.0016$), BM4 (T1, $p = 0.0016$) and BM5 (T2 and T3, $p = 0.006$), while a significant decrease was found in BM1 (T1, $p = 0.0084$), BM8 (T2, $p = 0.0036$) and BM9 (T1, $p < 0.0001$, T2, $p = 0.0005$). Descriptive statistics and the results of Friedman test with Wilcoxon test for *post hoc* repeated measure multiple comparisons are summarized in Extended Data Figures 4-2, 4-3, 4-4, 4-5, 4-6, 4-7, 4-8, 4-9, 4-10; significant differences with SHAM are graphically recapitulated in Figure 4m. Moreover, we found that the mild habituation condition showed significant increase in the expression of BM1 (T1, $p = 0.01$, T2, $p = 0.0027$), BM2 (T2, $p = 0.008$, T3, $p = 0.0004$), BM3 (T2, $p = 0.008$), BM8 (T2, $p = 0.008$), and BM9 (T2, $p = 0.005$) compared with heavy habituation (asterisks in Fig. 4d,f,k,l; descriptive statistics and results of Mann–Whitney test with Steel–Dwass nonparametric multiple comparisons are summarized in Extended Data Figs. 4-11, 4-12, 4-13). Extended Data Figure 4-14 shows the absolute time spent in BMs before normalization.

Altogether, these findings confirm that SEB3R classification successfully identifies time-driven and stimulus-driven changes in the expression of specific behaviors, discriminating between different experimental conditions.

BMs are informative of the mouse emotional response to whisker stimulation

To evaluate SEB3R reliability in identifying dynamic changes in the mouse emotional state, we assessed the linear correlation between the time spent in each BM and the time spent into discrete behavioral categories quantified by a manual user during the WN. We observed that behaviors indicative of fear and anxiety had a strong positive correlation with BMs 4 and 8, while being negatively correlated with BMs 1–3–6–9 (Fig. 5a). More specifically, among fear-related BMs, BM4 showed stronger correlation with evasive behaviors ($r = 0.42$, $p = 0.0002$; Fig. 5b; Movie 2), while BM8 showed stronger association with passive avoidance response ($r = 0.36$, $p = 0.001$; Fig. 5c; Movie 3). A mirrored trend was observed for explorative rearing, where strong positive correlations were found for



continued

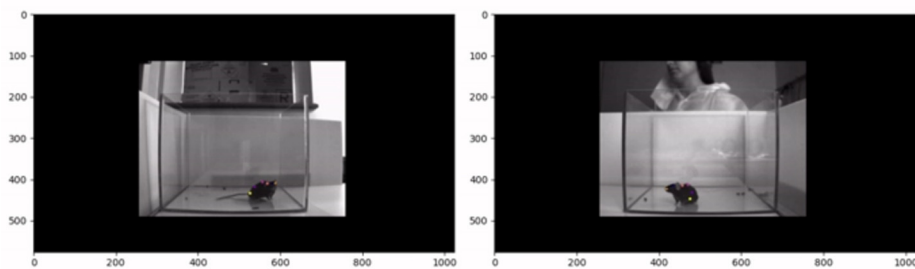
Figure 5. BMs correlate with discrete behavioral categories quantified by a trained user. **a**, Heatmap showing correlation between BMs and behavioral categories. Color indicates strength and direction of Pearson's r (summarized in Extended Data Fig. 5-1). Asterisk indicates significant correlation (summarized in Extended Data Fig. 5-2). **b-d**, Left panels, Top charts, Correlated BMs and behavioral categories show a similar temporal progression. Left panels, Bottom charts, Heatmaps showing the strength and direction of the correlations between BMs and behavioral categories. Center panels, Representative frames corresponding to each BMs displayed in the left panel. Right panels, 3D triangulation of corresponding BMs. Error bars refer to SEM.

BMs 1-3-6-9, while negative correlations was observed for BMs 2-7-8 (Fig. 5a). As an example, Figure 5d shows correlation between BM6 and total rearing time ($r = 0.92$, $p < 0.0001$; Movie 4). No correlation was found between BM5 and the categories assessed (Fig. 5a), likely indicating that this specific BM relates to an emotionally neutral behavior such as spontaneous spatial navigation (Pearson's correlation coefficients are summarized in Extended Data Fig. 5-1; p -values of the correlations are summarized in Extended Data Fig. 5-2). These data suggest that BMs identified by SEB3R can be reliably used as a proxy to isolate multiple emotionally charged behaviors.

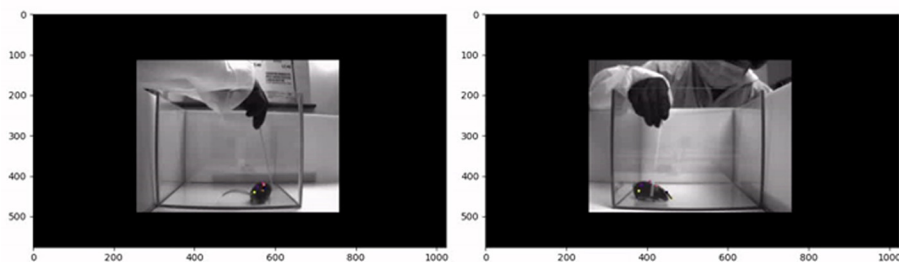
Fear-associated BM8 finds direct neural correspondence in the basolateral amygdala

To further provide proof-of-principle evidence of the biological relevance of our behavioral classification, we asked whether BMs could be mapped within specific brain areas.

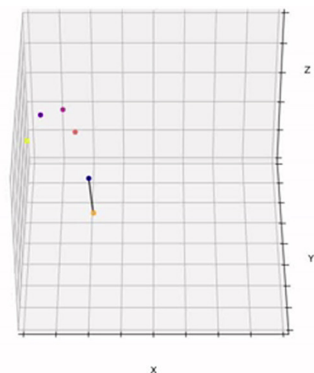
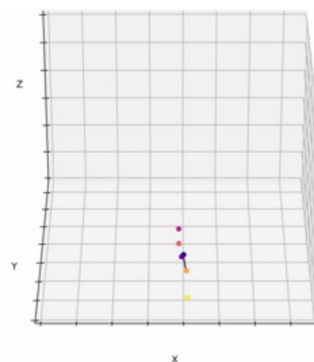
We therefore assessed the association between the activation of the amygdaloid complex with freezing-associated BM8 (as shown in Fig. 5a,c) in the mild habituation group, aiming to find correspondence with anxiety-related behaviors. To probe for neural activation, we used immunofluorescent labeling of ARC, a protein coded by an immediate early gene whose expression in the amygdala is activated by stressful stimuli (Dirven et al., 2022). We quantified the fluorescent intensity within the lateral (LA) and basolateral (BLA) nuclei. Strikingly, in BLA, we found that ARC expression was positively correlated with BM8 selectively ($r = 0.81$, $p = 0.001$; Fig. 6a,b). To the contrary, BLA ARC expression was negatively correlated with BM6 ($r = -0.61$, $p = 0.05$; Fig. 6a), which was identified as indicative of explorative behavior. No significant correlation was found between LA ARC expression and all BMs (Extended Data Fig. 6-1). These findings suggest that blindly identified BMs correspond to biologically relevant instances represented within specific neural circuits.

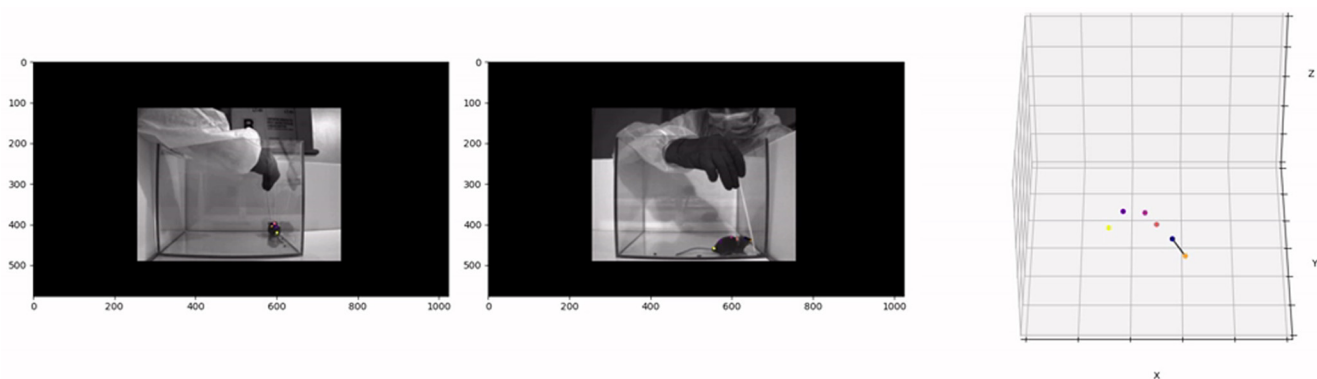


Movie 1. Example of the longest behavioral bout identified by SEB3R in the validation study; total length of the bout was 806 frames corresponding to ~ 32 s. [View online]



Movie 2. Example of a representative cluster of frames identified as BM4 taken from one of the original video files. BM4 was shown to be mostly correlated with active avoidance response. [View online]





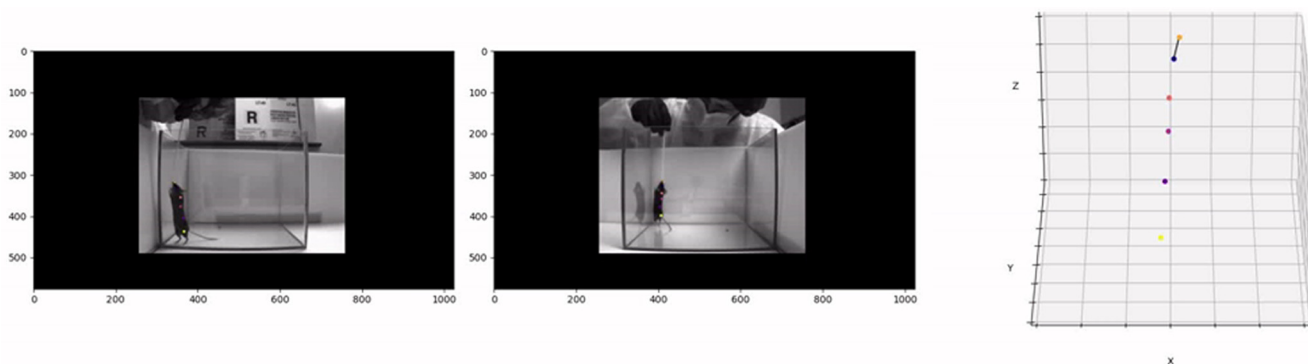
Movie 3. Example of a representative cluster of frames identified as BM8 taken from one of the original video files. BM8 was shown to be mostly correlated with passive avoidance response. [View online]

Discussion

In this work, we used a new-generation software for mouse body tracking combined with a strategy for automated posture detection to achieve successful and biologically relevant deconstruction of the mouse body language during freely moving behavior. This approach eliminates biases intrinsic of subjective rating strategies while significantly reducing the time required for quantification. Moreover, thanks to the frame-by-frame temporal resolution (Fig. 3), it offers the flexibility to study behavioral changes in selected time-windows as well as for extended periods. This feature renders our approach well-suited to ask questions beyond the simple categorical quantification looking at presence or absence of a behavioral outcome, allowing investigation of qualitative features such as complexity, persistency, and flexibility of the response to a stimulus. In addition, it detects the fluctuations of ethologically relevant behaviors in real time; as an example, SEB3R could be adapted to quantify freezing in fear-conditioning paradigms or to track pursuit during hunting. Finally, our approach would allow to investigate the neural correlates of specific behavioral hallmarks by combining behavioral parcellation with *in vivo* electrophysiological recordings or optical imaging. To this point, we showed that the frequency of BM8, putatively identified as associated with passive avoidance response (Fig. 6), was strongly correlated with activity-dependent ARC

immunolabeling in the BLA, thus suggesting that BMs are not only valuable proxies for behavioral interpretation, but have a well-defined neuroanatomical location to be exploited for *in vivo* studies.

Our validation experiments focused on a whisker-guided response (WN), a task that triggers complex and ambivalent reaction in wild-type mice (McNamara et al., 2010; Fontes-Dutra et al., 2018; Chelini et al., 2019; Balasco et al., 2022a). One major limitation of previous versions of this approach consists in the fact that most manually quantified behaviors show virtually the same temporal progression depending on their biological valence; anxiety-related behaviors dramatically increase during the first trial and are followed by a sharp decline in the second and third trials because of adaptation; an opposite tendency is observed for explorative categories (Chelini et al., 2019; Balasco et al., 2022a). Thanks to SEB3R and its unbiased extraction of BMs, we described independent behavioral categories characterized by their own temporal fluctuation as well as specific intrinsic interindividual variability, demonstrating a more scrupulous selection of behavioral states. At the same time, the automated quantification retained the capacity of discriminating between groups characterized by different affective states, as testified by the differences observed between heavy habituation and mild habituation (Fig. 4). Given the parallels and discrepancies observed with the manual quantification,



Movie 4. Example of a representative cluster of frames identified as BM6 taken from one of the original video files. BM6 was shown to be mostly correlated with unaided rearing. [View online]

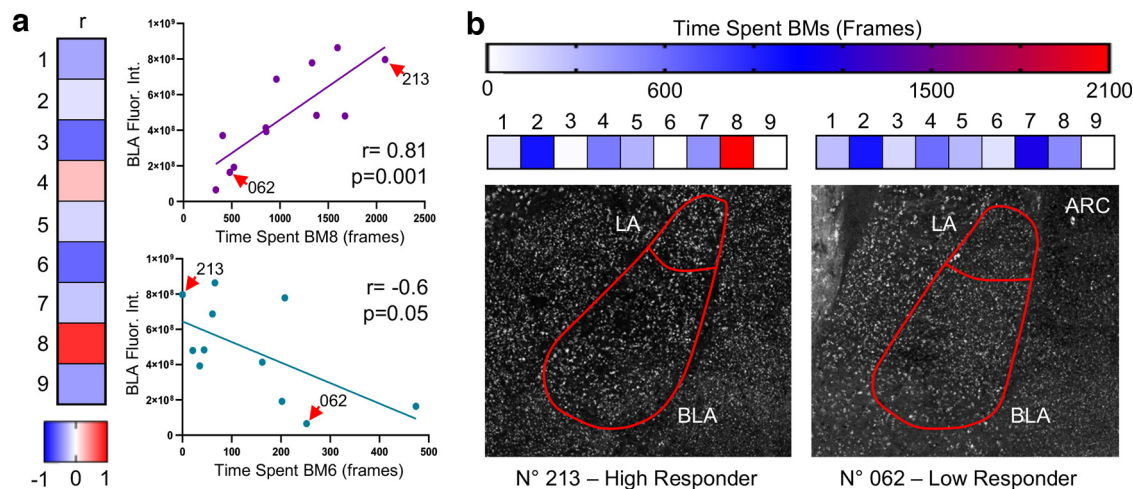


Figure 6. Stimulus-induced Arc protein expression in the amygdala. **a**, Arc protein expression in BLA correlates positively with BM8 and negatively with BM6. Red arrows point out the subjects shown in **b**. **b**, Representative diagrams show one animal displaying elevated fearful response and high Arc staining (left, No. 213) compared with a second subject showing low fearful response and low Arc staining in BLA (right, No. 062). BLA, basolateral amygdala; LA, lateral amygdala. (Lack of correlation between BMs and LA is shown in Extended Data Fig. 6-1.)

it is important to clarify the divergent nature of BMs with respect to traditional behavioral characterization. Especially for anxiety-related behaviors we have found significant correlation with behavioral categories, but the strength of the correlation was relatively weak (<0.5 ; Fig. 5a-c). This is not only because of the unavoidable noise included in both measurements, but also to the fact that the two measures are not directly overlapping. BMs are, by definition, indirect measures that can be used as a proxy to infer emotional states, they are not the behavioral state itself. By the same principle, correlation between manually and automatically quantified rearing is virtually perfect, as in that case the visual observer is directly quantifying the posture rather than a more complex behavioral construct. However, the remarkable similarities between automated and manual quantification, as in the case of the temporal dynamic of BM4 and its correspondence with evasive behaviors (Fig. 5a,b), confirm the efficacy of this computational approach in detecting emotion-driven behaviors, even in the context of a mildly emotionally loaded task, such as whisker stimulation. Thus, SEB3R promises to be an effective tool to study affective responses in tasks with high emotional salience such as fear conditioning.

This work aligns with recent advancement in system neuroscience and a variety of novel computational methods to study animal behavior (Wiltschko et al., 2015; Mathis et al., 2018; Nath et al., 2019; Dolensek et al., 2020; Bohoslav et al., 2021; Hsu and Yttri, 2021; Karashchuk et al., 2021; Luxem et al., 2022). Given this context, it is worth mentioning the advantages and limitations of this method. The main difference between SEB3R and other methods resides in its low computational complexity. *k-mean* cluster analysis is a relatively simple approach that does not compare in efficacy with emerging mathematical models used in other approaches. As a result, behavioral segmentation achieved with SEB3R is limited compared with previous studies (Wiltschko et al.,

2015; Dolensek et al., 2020). By contrast, SEB3R's reductionist approach degrades the behavioral complexity by forcibly compressing some datapoints in an array of interpretable behavioral modules, allowing the user to achieve a meaningful decoding of behavioral flow, to be used for translational inference. Another advantage of SEB3R stands in its technical feasibility. After obtaining behavioral tracking with DLC-3D, which has minimal technical needs per se (Mathis et al., 2018; Nath et al., 2019), the pipeline runs on any personal or professional device, without the need of a strong background in programming or computational skills. Thanks to minimal technical requirements of SEB3R we aim to provide the scientific community with a tool that unifies behavioral segmentation across research laboratories, allowing to investigate the affective content of behavior in freely moving mice in multiple experimental settings. Nonetheless, SEB3R's simplistic approach leaves room for future improvements. For instance, by discretizing the current classification it would be theoretically possible to include additional principal components to the cluster analysis, improving precision and specificity of the behavioral detection. In conclusion, the method proposed here will improve the translational validity of several behavioral tests by reducing quantification biases, ensuring replicability (within and beyond different laboratories), and expanding the number of behavioral categories to be analyzed. Finally, by applying novel system neuroscience approaches to a traditional ethology-based behavioral observation, we prove how the convergence of these two disciplines can contribute the study of the brain and its disorders.

References

Bailey KR, Crawley JN (2009) Anxiety-related behaviors in mice. In: Methods of behavior analysis in neuroscience (Buccafusco JJ, ed), Ed 2. Boca Raton: CRC/Taylor and Francis.

- Balasco L, Pagani M, Pangrazzi L, Chelini G, Ciancone Chama AG, Shlosman E, Mattioni L, Galbusera A, Iurilli G, Provenzano G, Gozzi A, Bozzi Y (2022a) Abnormal whisker-dependent behaviors and altered cortico-hippocampal connectivity in Shank3b^{-/-} mice. *Cereb Cortex* 32:3042–3056.
- Balasco L, Pagani M, Pangrazzi L, Chelini G, Viscido F, Chama AGC, Galbusera A, Provenzano G, Gozzi A, Bozzi Y (2022b) Somatosensory cortex hyperconnectivity and impaired whisker-dependent responses in Cntnap2^{-/-} mice. *Neurobiol Dis* 169:105742.
- Bohnslav JP, Wimalasena NK, Clausing KJ, Dai YY, Yarmolinsky DA, Cruz T, Kashlan AD, Chiappe ME, Orefice LL, Woolf CJ, Harvey CD (2021) DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. *Elife* 10:e63377.
- Campos AC, Fogaça MV, Aguiar DC, Guimarães FS (2013) Animal models of anxiety disorders and stress. *Rev Bras Psiquiatr* 35: S101–S111.
- Chelini G, Zerbi V, Cimino L, Grigoli A, Markicevic M, Libera F, Robbiati S, Gadler M, Bronzoni S, Miorelli S, Galbusera A, Gozzi A, Casarosa S, Provenzano G, Bozzi Y (2019) Aberrant somatosensory processing and connectivity in mice lacking Engrailed-2. *J Neurosci* 39:1525–1538.
- De Landtsheer S (2019) kmeans_opt. MATLAB central file exchange. Available at https://it.mathworks.com/matlabcentral/fileexchange/65823-kmeans_opt.
- Dirven BCJ, Botan A, van der Geugten D, Kraakman B, van Melis L, Merjenburgh S, van Rijn R, Waajen L, Homberg JR, Kozicz T, Henckens MJAG (2022) Longitudinal assessment of amygdala activity in mice susceptible to trauma. *Psychoneuroendocrinology* 145:105912.
- Dolensek N, Gehrlach DA, Klein AS, Gogolla N (2020) Facial expressions of emotion states and their neuronal correlates in mice. *Science* 368:89–94.
- Fontes-Dutra M, Santos-Terra J, Deckmann I, Brum Schwingel G, Della-Flora Nunes G, Hirsch MM, Bauer-Negrini G, Riesgo RS, Bambini-Junior V, Hedin-Pereira C, Gottfried C (2018) Resveratrol prevents cellular and behavioral sensory alterations in the animal model of autism Induced by valproic acid. *Front Synaptic Neurosci* 10:9.
- Gafford GM, Ressler KJ (2016) Mouse models of fear-related disorders: cell-type-specific manipulations in amygdala. *Neuroscience* 321:108–120.
- Hsu AI, Yttri EA (2021) B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nat Commun* 12:5188.
- Karashchuk P, Rupp KL, Dickinson ES, Walling-Bell S, Sanders E, Azim E, Brunton BW, Tuthill JC (2021) Anipose: a toolkit for robust markerless 3D pose estimation. *Cell Rep* 36:109730.
- Komada M, Takao K, Miyakawa T (2008) Elevated plus maze for mice. *J Vis Exp*. Advance online publication. Retrieved Jan 22, 2008. <https://doi.org/10.3791/1088>.
- Kraeuter AK, Guest PC, Sarnyai Z (2019) The open field test for measuring locomotor activity and anxiety-like behavior. *Methods Mol Biol* 1916:99–103.
- Luxem K, Mocellin P, Fuhrmann F, Kürsch J, Miller SR, Palop JJ, Remy S, Bauer P (2022) Identifying behavioral structure from deep variational embeddings of animal motion. *Commun Biol* 5:1267.
- Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 21:1281–1289.
- McNamara KC, Lisembee AM, Lifshitz J (2010) The whisker nuisance task identifies a late-onset, persistent sensory sensitivity in diffuse brain-injured rats. *J Neurotrauma* 27:695–706.
- Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW (2019) Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc* 14:2152–2176.
- Panlilio LV, Goldberg SR (2007) Self-administration of drugs in animals and humans as a model and an investigative tool. *Addiction* 102:1863–1870.
- Takao K, Miyakawa T (2006) Light/dark transition test for mice. *J Vis Exp*. Advance online publication. Retrieved Nov 13, 2006. <https://doi.org/10.3791/104>.
- Thorndike RL (1953) Who belongs in the family? *Psychometrika* 18:267–276.
- Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abaira VE, Adams RP, Datta SR (2015) Mapping sub-second structure in mouse behavior. *Neuron* 88:1121–1135.