TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 10/38 pp1487–1500
DOI: 10.26599/TST.2025.9010086
Volume 31, Number 3, June 2026

Research Article

Revealing the Body Language of Social Interaction in Free-Moving Mice Using BL-BERT

Yaning Han, Zhiwei Jiang, Furong Ju, Liping Wang, Quanying Liu, and Pengfei Wei*

Abstract: Social behavior in mice is critical for understanding their natural interactions and underlying neural mechanisms. Traditional Markov models, however, face limitations in capturing the sequential dynamics of body language associated with social behaviors. To address these challenges, we developed the body language-bidirectional encoder representation from transformers (BL-BERT) framework, which surpasses the Markov model in extracting complex sequential behavioral patterns. BL-BERT effectively differentiates the body language of the mice within different social interaction paradigms and produces results consistent with manual annotations. Notably, BL-BERT achieves higher extraction accuracy than the Markov model by reducing the complexity of the recurrent state transitions in behavior sequences. These advantages enable BL-BERT to accurately quantify high-order sequential behavioral structures in mice, paving the way for more detailed insights into the brain's mechanisms controlling complex behavior.

Key words: computational neuroethology; social interaction; behavior sequence; transformer model

1 Introduction

Mice are widely utilized in neuroscience research due to the similarities between their brain structure and function and those of humans[1]. Unlike humans, mice are unable to report their subjective experiences during neural recordings, necessitating the interpretation of brain activity through correlations with specific behaviors[2]. Recent advancements in deep learning-based pose estimation techniques have significantly enhanced the precision of mouse behavior analysis[3–8]. These approaches enable the tracking of multiple body parts in three-dimensional (3D) space, providing detailed kinematic data. Subsequently, supervised or unsupervised classification methods are applied to decompose body trajectories, assigning behavioral significance within specific time windows[9–13]. Despite these advancements, existing methods are primarily sensitive to sub-second behavioral modules and struggle to capture long-term temporal dependencies. This limitation causes a challenge in studying behaviors with extended durations, such as those crucial for understanding social communication in mice[14].

Traditionally, the Markov model is employed to extract the long-term time dependencies on behavior modules[10, 13, 15]. It provides a state transition

• Yaning Han, Zhiwei Jiang, Furong Ju, Liping Wang, and Pengfei Wei are with the State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. E-mail: hanyn@sustech.edu.cn; zw.jiang2@siat.ac.cn; 635201692@qq.com; lp.wang@siat.ac.cn; 101014012@seu.edu.cn.
• Yaning Han is also with the Joint Laboratory of Guangdong-Hong Kong Universities for Vascular Homeostasis and Diseases, SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology, Shenzhen 518055, China.
• Quanying Liu is with Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen 518055, China. E-mail: liuqy@sustech.edu.cn.
• Pengfei Wei is also with School of Biological Science and Medical Engineering, State Key Laboratory of Digital Medicine, Southeast University, Nanjing 211189, China.
* To whom correspondence should be addressed.
Manuscript received: 2024-12-24; revised: 2025-02-28; accepted: 2025-04-22

description of a sequence of behavior modules. Nevertheless, emerging evidence suggests that behaviors may exhibit strongly non-Markovian characteristics[16]. Consequently, reliance on the Markov model could lead to diminished accuracy in behavior sequence identification. Moreover, disentangling the recurrent structure of transitions between behavioral states into distinct sequences proves challenging. These factors collectively contribute to inaccuracies in mapping behavior sequences to neural activities.

Recent advancements in Transformer-based large language models have yielded notable strides in language comprehension[17, 18]. The Transformer, an artificial intelligence (AI) model incorporating a self-attention mechanism, has proven adept at discerning semantic nuances within token sequences[17, 19]. Given that behavior sequences can be construed as akin to animal body language, there exists potential to leverage Transformer-based large language models for uncovering the inherent semantic structures within these sequences. The self-attention mechanism inherent in the Transformer architecture offers a solution to the challenge of modeling non-Markovian dynamics[20]. Additionally, unlike the Markov model, the Transformer learns semantics directly from sequences, thus obviating the need to disentangle recurrent structures within behavior sequences.

Here, inspired by these Transformer-based models, we developed a bidirectional encoder representation from Transformers (BERT[21])-based body language (BL) extraction computational framework termed BL-BERT. Through utilizing BL-BERT, we found that the meaningful behavior sequences of the neural-recorded mice are significantly different when they interact with trapped or free partners. Moreover, the social behavior sequences demonstrate consistency between BL-BERT and manually assigned labels. Comparative analysis with a Markov model reveals that BL-BERT achieves a more precise quantification of social interactions. Our findings underscore the potential of AI-driven BL-BERT in facilitating a deeper comprehension of animal behavior sequences and the underlying neural mechanisms governing such sequences.

2 Experimental

2.1 Animals

Ten male C57BL/6J mice (8−10 weeks old) were used in the experiment. The mice were housed at 5 mice per cage under a 12 h light-dark cycle at 22–25 °C with humidity of 40%–70%, and were allowed to access water and food ad libitum (Shenzhen Institutes of Advanced Technology, Shenzhen, China). Five of them accepted the surgery for the neural recording of two-photon microscopy. The mouse was anesthetized with isoflurane and placed in a stereotactic apparatus (RWD). The skull was exposed, and a craniotomy was performed for virus injection. To cover the entire somatosensory cortex, the mouse was microinjected at the following coordinates: antero-posterior (AP), −0.60 mm; medio-lateral (ML), −2.40 mm; dorso-ventral (DV), −0.80 mm. AAV9-CaMKII-GCaMP6s viruses were used. The procedure and experiences of imaging have been described in previous studies[22−24]. All husbandry and experimental procedures were approved by the Animal Care and Use Committee at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China.

2.2 Apparatus

The apparatus is based on the MouseVenue3D system[22, 25]. The MouseVenue3D system has three parts including a multi-view camera array, an automatic camera calibration board, and a neural recording device. The multi-view camera array includes four Intel RealSense D435 cameras. They are installed at the steel stand by camera forceps. The width of the steel stand is 90 cm. The distance between the cameras and the center of the bottom of the open field is 75 cm. The view angles of cameras need to be adjusted to cover the full bottom of the open field. The frame rate of each camera is 30. The resolution is 640 × 480. The frames have RGB three channels. The cameras are connected to a high-performance workstation (Intel i7-7900K CPU, 16 GB RAM, 1 TB SSD) through a (PCI-E) USB-3.0 data acquisition card. The control and synchronization of cameras are based on the pyrealsense2 Python application programming interface (API). At the same time, the workstation controls the automatic camera calibration board. The automatic camera calibration board is composed of one square screen, four servos, and customized steel supports. The neural recording device is a system of fast high-resolution, miniaturized two-photon microscope (FHIRM-TPM)[23, 24]. The synchronization between cameras and FHIRM-TPM is through the synchronization module integrated with FHIRM-TPM.

The workstation sends one transistor-transistor logic (TTL) marker to the synchronization module every second and the computer time is recorded at the same time. The open field is an acrylic transparent cylinder with 50 cm diameter and 50 cm height.

2.3 Behavior recordings

Before the behavior recording, four cameras are calibrated through the automatic camera calibration board. The workstation controls the screen in the automatic camera calibration board to show the checkboard with rotations and translations. At the same time, the servos will change the direction of the screen towards each camera. Four cameras will take frames of the checkboard until around 400 pairs of checkboard frames. The camera parameters are estimated using a MATLAB toolbox called pose3d[26].

Ten male mice were used in the behavior recording. Five of them accepted the surgery of FHIRM-TPM implantation to distinguish their identities. The trapped and free groups have different experimental procedures. The first step of the behavior recording of the trapped group is to put the trapped partner mouse in the open field. One metal pen container with mesh is the cage to restrain the locomotion of the trapped partner mouse. The height of the pen container is around 10 cm and the diameter is around 8 cm. The mouse accepting the surgery is then put inside the open field for 15 min of behavior recording. The first step of the behavior recording of the free group is to put the mouse without surgery inside the open field. Further, the mouse accepting the surgery is then put inside the open field for 15 min of behavior recording. The mice in the cage or free-moving in the open field are used repeatedly but the pairs between the mouse with surgery and the naïve one are always unique. For example, the pairs of surgery mouse 1 will be only used once with the naïve mouse 1. Therefore, each mouse with surgery or without surgery will be put into the open field twice with different partners. One time is in the trapped group and another time is in the free group. The sequences of the trapped group and free group are shuffled.

2.4 3D pose estimation

After the video recording, DeepLabCut[4, 5] is used to estimate the two dimensional (2D) poses of the free-moving mice in each frame. The version of DeepLabCut is for a single animal. Considering that the FHIRM-TPM can be the identity marker of one of the mice, we train two models separately for the mouse with neural recording and another untreated free mouse partner.

The model trained for the mouse with neural recording is on both trapped and free groups. The videos of trapped and free groups are mixed in one DeepLabCut project for image labeling. A total of 2615 frames are labeled to train the pose estimation model in DeepLabCut. The model trained for the free mouse partner is only on free group videos. Totally 1852 frames are labeled to train the pose estimation model in DeepLabCut. The backbones are both ResNet50 and the iteration numbers are both 1 030 000. Sixteen body points are labeled including “0: nose”, “1: left_ear”, “2: right_ear”, “3: neck”, “4: left_front_limb”, “5: right_front_limb”, “6: left_hind_limb”, “7: right_hind_limb”, “8: left_front_claw”, “9: right_front_claw”, “10: left_hind_claw”, “11: right_hind_claw”, “12: back”, “13: root_tail”, “14: mid_tail”, and “15: tip_tail”[27]. The accuracy of the pose estimation is evaluated manually. Combining with the parameters from four cameras, pose3d can reconstruct 2D poses of each frame to 3D. The pixel coordinates are transformed to the cm coordinates through a transform index of 8.8 pixels/cm calculated from the ratio between the maximum locomotion range of mice and the size of the open field.

2.5 Classical behavior analysis

The classical behavior analysis includes distance analysis and position preference analysis. The distance in the trapped group is calculated based on the minimum distance between the edge of the cage and the closest body parts to the edge. The center of the cage is the same as the center of the open field so that the edge coordinates can be estimated through the center position adding the radius of 4 cm. The distance in the free group is calculated based on the minimum distance between the two mice. Further, the temporal averaged distance of them is calculated in each video for comparison.

The peripheral area preference analysis is based on the back points of the mice. The open field is separated into the center part and peripheral part according to the half radius of 12.5 cm. The peripheral area preference index is calculated according to the ratio between the total time of the mouse back point in the peripheral and in the center area.

2.6 Locomotion heatmap

Before the calculation of the locomotion heatmap, the skeletons of the mouse in the same video are aligned according to their back points and tail base points. Firstly, the skeletons are aligned to the zero center by subtracting the position of the back points using all of the body points. Secondly, the skeletons are rotated in the same direction according to the vector [tail base, back]. The aligned skeletons are cumulative to the same 2D empty background, including [x, y] and [x, z] planes. The stacked skeletons are then normalized to show in the heatmap, which is the locomotion heatmap.

2.7 Process of behavior modules

3D behavior modules of the mice with neural recording are reconstructed, decomposed, and annotated following the procedure of behavior atlas (BeA)[10]. In total, 41 behavior modules are identified. They are "walking rising", "right step-ping", "body grooming type 1", "body grooming type 2", "body grooming type 3", "nose grooming", "face grooming type 1", "left movement", "right movement", "poking", "grooming type 1", "looking up", "creeping", "transition movement type 1", "stepping", "darting", "standing grooming", "left wiggling type 1", "transition movement type 2", "transition movement type 3", "rising front paw", "right wiggling grooming", "rearing", "grooming type 2", "left wiggling type 2", "walking sniffing type 1", "right darting", "wiggling", "rising", "left turning", "left wiggling grooming", "pausing type 1", "pausing type 2", "right turning", "approaching", "walking sniffing type 2", "pausing type 3", "face grooming type 2", "wiggling grooming", "left grooming", and "face grooming type 3". The behavior modules are shorted and merged in the visualization of the behavior sequence fingerprints for concision.

2.8 Training, validation, and prediction of BL-BERT

BL-BERT is trained on one GeForce RTX 3090 GPU. The layers of the Transformer encoder and decoder are set to 6. The batch size is set to 200. The number of epochs is set to 20. The number of accumulation steps is set to 10. The initial learning rate is set to 1.0. Other configurations of BL-BERT are the same as the Annotated Transformer[28] like the hyper-parameters including 0.1 dropouts of multi-head attention module, 512 input and output dimensions of linear layer, 2048 dimensions of the hidden linear layer, 0.1 label smoothing, and eight parallel attention layers. Totally 200 000 behavior sequences are randomly resampled from the raw behavior data. Each behavior sequence includes 60 behavior modules. And 80% and 20% of the 200 000 behavior sequences are split separately into training and validation datasets. The prediction of BL-BERT is based on greedy decoding[28]. The token with the highest probability of being selected as its next token is the output. The output length is set to 60.

2.9 Behavior sequence detection

The detection of different behavior sequences includes precision screening, continuity judgment, significant selection, and clustering. The precision screening uses hard precision as the criteria. The behavior sequences generated by the Monte Carlo method are screened with hard precision. The behavior sequences reaching 100% hard precision are kept for further processing.

The screened behavior sequences have random position masks. According to the continuities of the mask positions, the segments of behavior sequences with semantic dependencies are extracted. The maximum length of the continuous masks is 12 because the mask percentage is 20% and the length of sampled behavior sequences is 60.

After the continuity judgment, ten thousand of the behavior sequence segments are kept. In order to reduce their complexities and focus on the differences between the trapped and free groups, the t-test is used to detect the behavior sequences with significant differences. The reason to use a t-test instead of two-way analyses of variance (ANOVA) is the limitation of random access memory (RAM) for interaction terms. Hence, we simplified the conditions of interaction terms.

The comparison finally left hundreds of significantly different behavior sequences with semantic dependencies. Hierarchical clustering with standard Euclidean distance and ward linkage is used to sort and cluster these sequences. Principal component analysis is used in the sequence dimensions to decide the cluster number. The threshold of the explained variance is 0.9, so the minimum number of dimensions, 18, is set as the cluster number. Therefore, 18 different behavior sequences are detected for further visualization.

2.10 Visualization of behavior sequence fingerprints

The visualization of behavior sequence fingerprints is based on the word cloud[29]. In total, 18 different behavior sequences have inner temporal distributions of the probabilities of their sub-clusters. The probabilities of each time step are mapped to five levels for better visualization of the word cloud. The different colors of the word cloud are decided by the mean fraction values of the sequences in the trapped and free groups. The shape of the word cloud is set to the rectangle.

2.11 Distribution estimation of the Markov model

The Markov model for behavior transition quantification is based on the Markov chain[30]. The parameters of the transition matrix of the Markov model are calculated according to statistical estimation. Considering the variabilities of behavior transitions[31], the transition matrixes were not iterated to the steady-state distribution. Each behavior module is regarded as a state in the Markov model. The transition probability is estimated using

$$P(b_i \to b_j) = \frac{\sum_{t=1}^{T-1} 1(b_t = i,\, b_{t+1} = j)}{\sum_{k=1}^{N} \sum_{t=1}^{T-1} 1(b_t = i,\, b_{t+1} = k)},$$

where $P(b_i \to b_j)$ is the transition probability, b_i is the previous behavior state, b_j is the current behavior state, $1(\cdot)$ is the indicator function, N is the total number of behavior states, and T is the length of behavior sequences.

2.12 Performance of classification

Performance of classification uses a generalized linear model (GLM) as the base model[32, 33]. The behavior transitions with significant differences in Markov models are extracted to fit the GLM to improve the feature separation. The behavior sequences with significant differences are equivalent to be extracted for the fitting. The GLMs are fitted using cross-validation methods[6]. The datasets are randomly separated into 60% training dataset and 40% validation dataset for 100 times fitting of GLMs. The accuracy is calculated using (true positive + true negative) / (true positive + true negative + false positive + false negative).

2.13 Statistic

Before hypothesis testing, data were first tested for normality by the Shapiro–Wilk normality test and for homoscedasticity by the F test. For normally distributed data with homogeneous variances, parametric tests were used; otherwise, non-parametric tests were used. All the ANOVA have been corrected by the recommended options of Prism v.8.0.

3 Result and Discussion

3.1 Context-dependent social behaviors of mice toward trapped and free partners

We modified the three-chamber test to study social interactions within the MouseVenue3D system[22, 34] (Fig. 1a). Initially, a mouse is placed in a cage positioned at the center of an open field (left panel of Fig. 1b). Subsequently, another mouse equipped with a head-mounted neural recording device is put into the open field for a 15-min social interaction period, constituting the trapped group. The free group engages in a 15-min social interaction, consisting of two freely moving mice, one of which is also mounted with a neural recording device (right panel of Fig. 1b). The neural recording devices serve to distinguish the identities of the mice, and subsequent behavior analysis focuses exclusively on those with neural recordings. The 15-min interactions are divided into three 5-min segments for further analysis.

Initially, the mean body distances between free-moving mice and trapped/free partners are calculated to quantify their preferences (Fig. 1c). The free-moving mice are significantly closer to the trapped partners than the free partners. In addition, the peripheral area preference indexes are calculated to validate the trapped partner preferences (Fig. 1d). The peripheral area preference indexes in the trapped group are significantly lower than the free group. These results demonstrate that the free-moving mice tend to stay longer with the trapped partners and it also causes the reduced intrinsic peripheral preferences of mice. It suggests that the trapped partner and the cage have more attraction to the free-moving mice than the free partner and the peripheral area.

Further, BeA is employed to extract subtle behavior modules, and subsequent comparisons are made regarding the fractions of these modules[10] (Fig. 1c). Notably, the proportion of the “rising” behavior

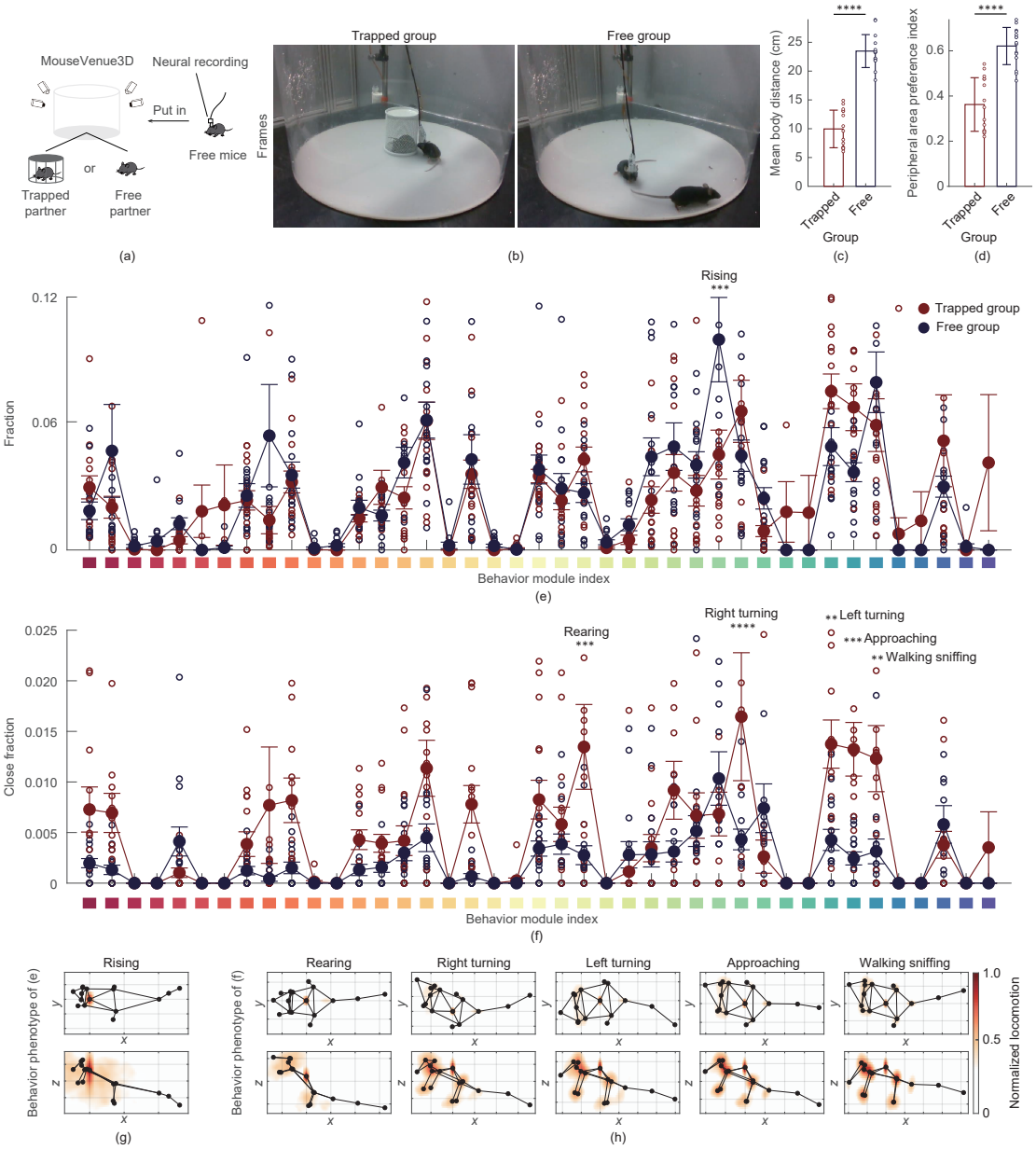

Fig. 1 Behavior phenotypes of mice interacting with trapped versus free partners. (a) Diagram of behavioral paradigm. (b) Video frames of the mouse interacting with trapped or free partners. Left, a free-moving mouse with a trapped partner named the trapped group. Right, a free-moving mouse with a free partner named free group. (c) Mean body distance between two mice in trapped and free groups (mean ± standard deviation (sd.), Mann Whitney test, $n = 15$). The T-shaped line segments indicate error bars. The dots represent the raw data corresponding to each bar. (d) Peripheral area preference index of the free-moving mice in trapped and free groups (mean ± sd., Mann Whitney test, $n = 15$). The threshold to distinguish peripheral and center is the half radius. (e) Comparison of behavior fractions using BeA. Mouse interacting with the free partner shows significantly more rising behavior than the trapped partner (mean ± standard error of the mean (s.e.m), two-way ANOVA followed by Sidak's multiple comparisons test, $n = 15$). (f) Comparison of fractions of behavior in close interactions. 5 cm is defined as the threshold of close interaction. The rearing, right turning, left turning, approaching, and walking sniffing behaviors are significantly higher in the trapped partner group (mean ± s.e.m, two-way ANOVA followed by Sidak's multiple comparisons test, $n = 15$). (g) 3D skeleton of rising behavior. The heatmap shows the locomotion of each body point. The skeletons (x, y, and z) are normalized for illustration. (h) 3D skeleton of rearing, right turning, left turning, approaching, and walking sniffing behaviors. The heatmap shows the locomotion of each body point. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. P represents the P-value, which indicates the probability of observing the obtained results under the null hypothesis. A smaller P-value suggests stronger evidence against the null hypothesis.

module exhibits a significant increase in the free group, indicative of heightened behavior flexibility compared to the trapped group.

Subsequently, a rigid threshold (set at 5 cm) is applied to differentiate between close-social and far-individual behavior modules, facilitating a comparison of interactions during close proximity (Fig .1d). The proportions of five behavior modules—rearing, right turning, left turning, approaching, and walking sniffing, show significant increases in the trapped group. These modules, recognized for their association with body language[14], are instrumental in conveying messages through behavioral sequences. Phenotypic skeletons of the behavior modules exhibiting significant differences are depicted in Figs. 1e and 1f.

Our findings underscore significant distinctions in the behavioral structures of mice when faced with trapped versus free partners. The social interactions of trapped partners may intermingle with body language indicative of prosocial tendencies, likely influenced by the restraint experienced by the partner mice.

3.2 Sequence-level representation of social interaction differences

To delve deeper into the variances in social interactions at the behavioral sequence level, the BL-BERT is developed, as depicted in Fig. 2a. Initially, the behavior modules decomposed by BeA (top left panel of Fig. 2a) are transformed into sequences specific to each mouse. These individual behavior sequences are subsequently resampled into segments (top right panel of Fig. 2a), facilitating their organization into a dataset for training the model. Concurrently, the behavior modules are tokenized into discrete behavior tokens analogous to words in natural language. Following the design of the BERT, the tokenized behavior sequences undergo random masking. The Transformer model within BERT is then tasked with processing the masked behavior sequences, with its output representing the unmasked behavior sequences (bottom panel of Fig. 2a). Upon completion of Transformer training, the Monte Carlo method is applied to predict the behavior sequences of each mouse. This involves the repeated random masking of behavior sequences one hundred times, followed by input into the Transformer. Subsequently, the congruent outputs between prediction and unmasked sequences are retained, signifying the presence of meaningful body language sequences.

The gradient for mask percentage ranges from 0% to 90%, aimed at identifying optimal mask parameters conducive to the extraction of nuanced body language cues (Fig. 2b). Initial findings indicate a direct correlation between mask percentage augmentation and heightened loss, coupled with a reduction in precision. Precision metrics exhibit a discernible inflection point at approximately 30% mask. To uphold the efficacy of body language extraction, a mask percentage of 20% is determined as the most optimal threshold.

Furthermore, two specific scenarios, namely shuffled sequences and wrong mask answers, are employed to validate the robustness of BL-BERT against noise inherent in behavior sequences (Fig. 2c). BL-BERT demonstrates an ability to converge to shuffled sequences but precision levels diminish compared to unshuffled counterparts. This observation suggests that certain behavior sequences may not inherently rely on sequential semantics to convey information. Despite this, BL-BERT exhibits insensitivity towards such non-semantic sequences, as inferred from the predicted precision metrics. Additionally, results of wrong mask answers highlight BL-BERT's incapacity to conform to noise, with approximately 80% precision affirming its fidelity solely to accurate representations.

Using BL-BERT, the behavior sequences of trapped and free groups are extracted, and their fractions are compared to reveal their differences (Fig. 2d). Total of 181 behavior sequences with significant differences are identified between the two groups. To depict the intricate behavioral sequences, we extend the analysis from the static behavioral module fingerprint[35], to encompass a dynamic behavioral sequence fingerprint (Fig. 2e). This visualization reveals that within the trapped group, behavioral modules such as approaching and turning are concatenated. The fractions of behavioral sequences formed by these modules are significantly higher compared to those in the free group. These findings align with prior research on prosocial behavior[36], indicating that mice exhibit more altruistic body language when confronted with their trapped counterparts. Conversely, the behavioral sequences of the free group demonstrate greater flexibility. The constituent modules of these sequences encompass a broader array of locomotion behaviors such as walking and rising. These outcomes underscore the considerable divergence in the behavioral structures underlying social interactions between the trapped and free groups.

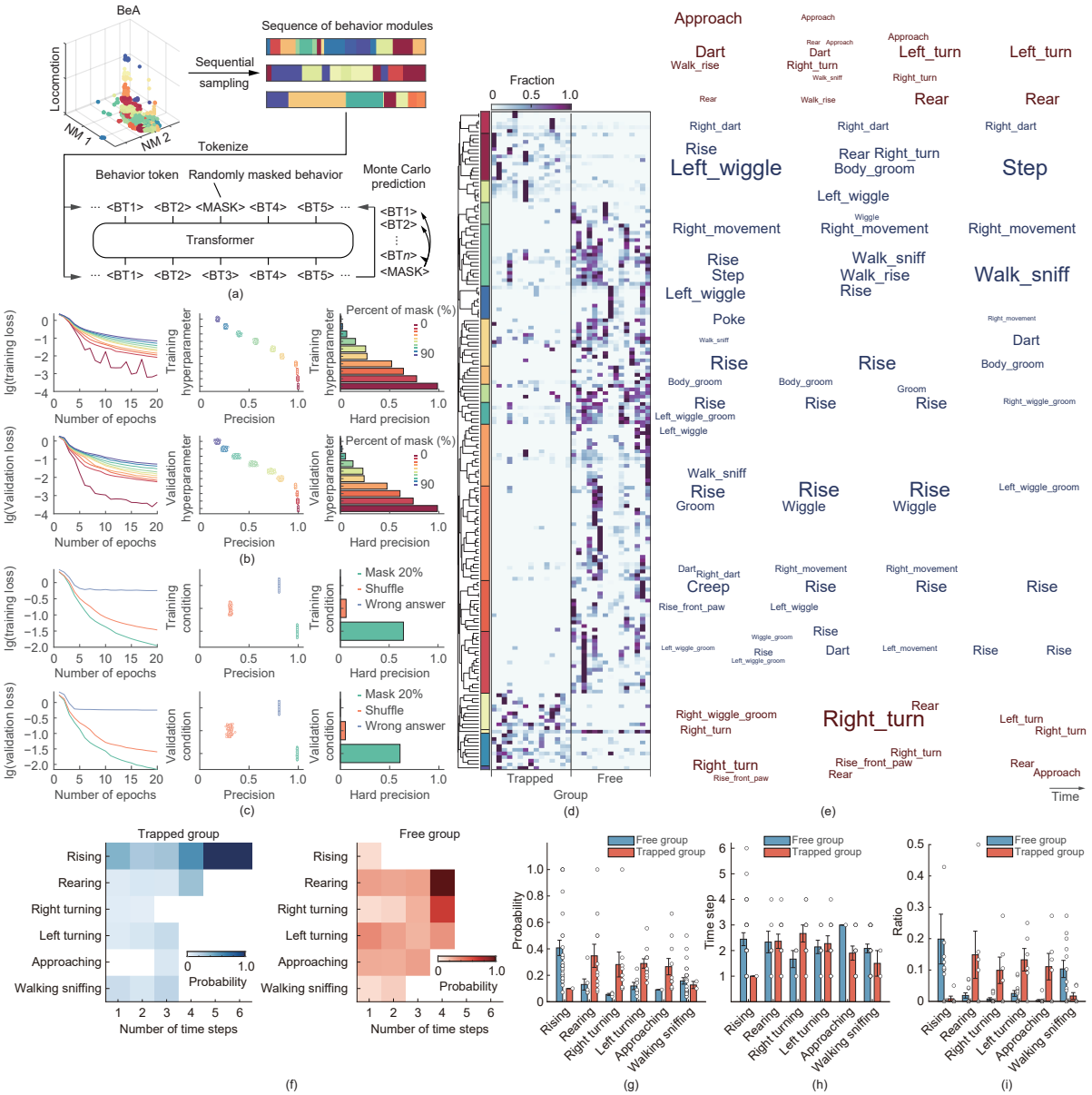

Fig. 2 BL-BERT-based detection of sequential behavior differences between trapped and free groups. (a) Diagram of BL-BERT. NM means non-locomotor movement. (b) Performance comparison of BL-BERT using different percentages of masks. Precision is the full-length comparison between model output and ground truth. Hard precision is calculated by only comparing the masked tokens. (c) Performance comparison of BL-BERT using shuffled sequences and wrong answers of masks. (d) Cluster gram of behavior sequences of trapped and free groups. (e) Behavior sequence fingerprint of two groups. Red words represent that the behavior sequences are significantly higher in the trapped group (t-test or Wilcoxon rank sum test, n = 15). Blue words represent that the behavior sequences are significantly higher in the free group (t-test or Wilcoxon rank sum test, n = 15). The size of the words represents the probability of the behavior modules in a behavior sequence. (f) Probabilities of the significantly different behavior modules of each time step in the trapped and free groups. (g) Comparison of the probabilities of these behavior modules in behavior sequences between trapped and free groups (mean ± s.e.m, two-way ANOVA followed by Sidak's multiple comparisons test, n = 27, 1, 6, 11, 3, 9, 7, 11, 1, 10, 24, and 2). The T-shaped line segments indicate error bars. The dots represent the raw data corresponding to each bar. (h) Comparison of the time steps of these behavior modules in behavior sequences between trapped and free groups (mean ± s.e.m, two-way ANOVA followed by Sidak's multiple comparisons test, n = 27, 1, 6, 11, 3, 9, 7, 11, 1, 10, 24, and 2). (i) Comparison of the ratios of these behavior modules in all of the different behavior sequences between trapped and free groups (mean ± s.e.m, two-way ANOVA followed by Sidak's multiple comparisons test, n = 12 for the free group and n = 6 for the trapped group).

The relationships between significantly different behavior modules (Figs. 1e–1h) and behavior sequences (Fig. 2e) are systematically analyzed. First, the probabilities of each behavior module at each time step within the corresponding sequences are quantified (Fig. 2f). The results indicate that the rising behavior module exhibits a higher probability in the trapped group compared to the free group, whereas the rearing behavior module shows a higher probability in the free group than in the trapped group. Furthermore, sequences containing rising or rearing modules in the trapped and free groups demonstrate peak probabilities of these two modules at the final stage. These findings suggest that the significant differences in rising and rearing modules are associated with variations in the probabilities of sequence-ending modules. In contrast, the probability distributions of other significantly different modules, including left turning, right turning, approaching, and walking sniffing, remain relatively consistent across time steps. These results imply that the rising and rearing modules may serve as key transition points in behavior sequences composed of diverse behavioral modules.

Although the probability differences in their mean values exhibit similar distributions in their fractions, no statistically significant differences are observed between trapped and free groups (Fig. 2g). Similarly, the temporal order of behavior modules within sequences do not show significant differences between the two groups (Fig. 2h). The overall ratio of different behavior modules across all behavior sequences follows a trend similar to that of module fractions but remains no statistically significant (Fig. 2i). The absence of significant differences in probabilities, time step orders, and ratios may be attributed to the high variability of behavior sequences. While different behavior modules serve as fundamental components of behavior sequences, their primary influence appears to be on the temporal dynamics of stepwise probabilities rather than on their global probabilities, orders, or ratios. These findings suggest that the contributions of significantly different behavior modules primarily manifest in their temporal probability distributions. Furthermore, the roles of behavior modules within sequences vary, for instance, the rising and rearing modules may function as punctuation points, whereas left and right turning modules are more likely involved in defining sequential semantics.

BL-BERT discernes variances in social interactions between the trapped and free groups at the sequence level. Findings from the behavior sequence fingerprint analysis imply that restrained mice may elicit prosocial-like behavioral phenotypes in their partners. This suggests potentially distinct neural mechanisms governing social interactions in mice when encountering trapped versus free partners.

3.3 Consistency between BL-BERT-extracted behavior sequences and manual annotation

To further assess whether BL-BERT can accurately capture the behavior sequences akin to those observed by humans, we conduct manual annotations on all behavior sequences related to social interaction within the trapped group[36] (Fig. 3a). These manual annotations are visually represented through behavior sequence fingerprints (Fig. 3b). The visual representation illustrates that the behavior sequences associated with social interaction entail modules of approach and turning, mirroring the findings of BL-BERT. Subsequently, we employ the Hamming distance metric to quantify the degree of similarity between the manually annotated sequences and those generated by BL-BERT[37] (Fig. 3c). The trapped group exhibit a significantly smaller Hamming distance compared to the free group, indicating a closer resemblance between the behavior sequences of the trapped group and the manually annotated ones. Furthermore, certain segments of the Hamming distances in the free group are zero, signifying an exact match with the manual annotations. This observation suggests that the social interactions observed in the trapped group may represent a subset of the social interactions observed in the free group.

The behavior sequences identify by BL-BERT demonstrate concordance with manually annotated data. Moreover, the spectrum of social sequences observed within unrestricted social interactions may encompass a broader array of social behavior phenotypes. Conversely, the social sequences exhibit by trapped groups may represent a distinct subset encapsulated within the broader spectrum of free social interactions.

3.4 Enhanced behavior sequence quantification by BL-BERT compared with the Markov model

Finally, we conduct a comparative analysis between the performance of the BL-BERT and the Markov model. The Markov models have been extensively

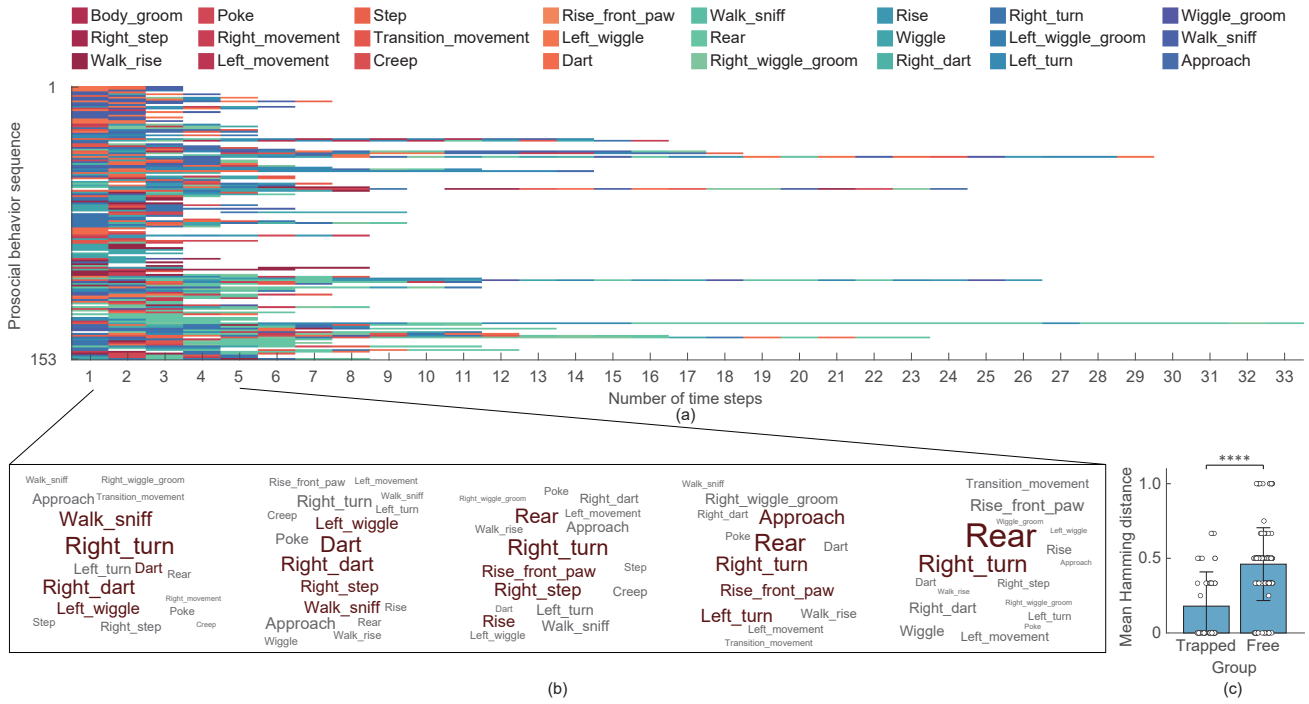

Fig. 3 Comparison between manual social behavior sequence annotations and BL-BERT. (a) Totally 153 manual annotated social behavior sequences are illustrated by the behavior modules. (b) Behavior sequence fingerprint of manually annotated social behaviors. The first five behavior modules are illustrated. (c) Mean hamming distances between manually annotated social behaviors and BL-BERT detected behavior sequence differences (mean ± s.e.m, Mann Whitney test, n = 45 and 136). The T-shaped line segments indicate error bars. The dots represent the raw data corresponding to each bar. ****$P < 0.0001$.

utilized for quantifying action selection[38], innate avoidance[13], and circadian behavior[10], rendering their application to our dataset a pivotal point of reference. The state transitions of behavioral modules within both the trapped and free groups are evaluated using the state transition matrix and depicted through an undirected graph (Fig. 4a). To enhance clarity regarding the distinctions in state transitions between the two groups, we further employ a directed graph with force-directed placement[39] (Fig. 4b). While tracing the paths delineated by the arrows in the directed graph, it becomes evident that the behavioral modules and sequence structures closely resemble those identified by BL-BERT. Nevertheless, discerning the initiation node of each sequence remains challenging even with the aid of the directed graph.

Moreover, we conduct a comparative assessment of the efficacy of the Markov model and BL-BERT in segregating the two groups (Fig. 4c). Enhanced performance may demonstrate a more precise extraction of behavioral sequence features[32]. Our findings indicate that BL-BERT outperforms the Markov model in classifying the trapped group, while the classification performance for the free group remains comparable between the two methods. The overall accuracy achieved with BL-BERT is superior to that of the Markov model (Fig. 4d). This heightened performance may be attributed to BL-BERT's capability to extract behavior sequences characterized by non-Markovian patterns.

Our findings show that BL-BERT is capable of generating behavior sequences comparable to those produced by the Markov model. Notably, BL-BERT facilitates the identification of initial points within behavior sequences more effectively than the Markov model. Moreover, BL-BERT successfully circumvents the challenge associated with possible non-Markovian sequence extraction encountered by the Markov model.

4 Conclusion

This study demonstrates the effectiveness of the BL-BERT model in decoding the body language associated with social interactions in free-moving mice. By leveraging the power of the Transformer model, BL-BERT surpasses traditional models, such as the Markov model, in accurately capturing non-Markovian patterns and complex social behaviors. Through a detailed analysis of behavior sequence fingerprints, we

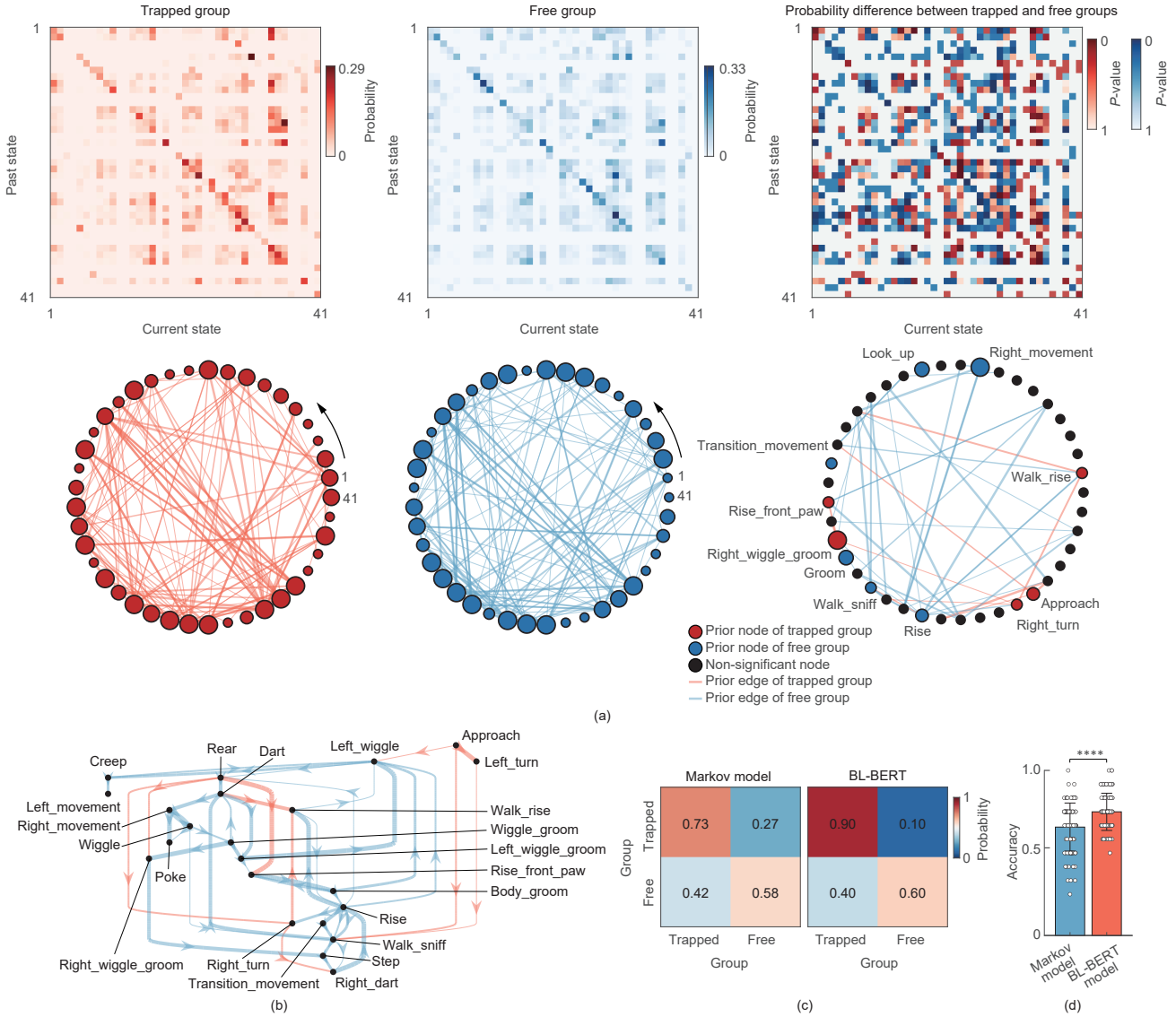

Fig. 4 Comparison between the Markov model and BL-BERT. (a) Markov model quantifies the temporal correlations of behavior modules. The top row is the transition matrix of behavior modules. The bottom row is the undirected graph of behavior state transition. (b) Directed graph of behavior state transitions with significant differences between the two groups. The red arrow line represents that the behavior transitions in the trapped group are significantly higher. The blue arrow line represents that the behavior transitions in the free group are significantly higher (t-test or Wilcoxon rank sum test, $n = 15$). (c) Accuracy of classifying trapped and free groups using Markov state transitions or BL-BERT behavior sequences. (d) Accuracy comparison between the Markov model and BL-BERT (mean ± s.e.m, Mann Whitney test, $n = 100$). The T-shaped line segments indicate error bars. The dots represent the raw data corresponding to each bar. ****$P < 0.0001$.

reveal distinct social sequence profiles in mice when interacting with trapped versus free counterparts. BL-BERT's ability to identify intricate behavior sequences not only aligns closely with manual annotations but also reflects the broader spectrum of social interactions in unrestricted settings, with trapped social scenarios representing a more prosocial subset of these interactions. The behavioral differences revealed by BL-BERT can improve the understanding of the high-order functions of neural circuits behind social interactions.

The social interaction test in this study is modified from the classical three-chamber test[40] to concentrate more on the differences in body language between trapped and free partners. Although Phase 1 of the classical three-chamber tests verify that the mice in a cage attract the free-moving mice more than the empty cage, the empty cage is also a novelty object for the

free-moving mice. Hence, a more comprehensive explanation of the body language in the trapped group will be the coexistence of both the prosocial tendencies of the trapped partner and the novelty of the cage. Further studies such as the preference test of free-moving mice toward the empty cage can be performed to isolate the novelty object-related body language. With BL-BERT, the mouse preference for new objects or partners can be quantified in a more detailed granularity to clarify the behavior components of social novelty and social attraction.

Brain intelligence arises from the dynamic interplay between the body and its environment[41]. Historically, neuroscientific research has relied on behaviorally restrained animals, as quantifying the intricate structures of natural animal behavior has challenges[42]. Nevertheless, such restraint represents a profound external stimulus to the brain, introducing notable artifacts in neural activity, as demonstrated in our trapped mouse paradigm. These artifacts not only affect the neural responses of the restrained animal but also influence the behaviors of its social partners. The BL-BERT framework offers a robust approach to uncovering the semantic structure of animal behavior, with potential parallels in neural mapping. This framework provides critical insights into how the brain processes sequential information. Advances in AI have further enriched brain research, with deep learning-based pose estimation methods enabling higher-dimensional behavioral analyses to better explain neural activity. By enhancing the abstraction of behavioral labels, BL-BERT bridges the gap between observed behaviors and the brain's high-level functional organization, offering a promising avenue for a deeper understanding of neural mechanisms.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 32222036 and T2394533), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB1010101), the Research Fund for International Senior Scientists (No. T2250710685), the Science and Technology Innovation 2030-Major Project (No. 2021ZD0203900), Shenzhen Key Basic Research Project (No. JCYJ20220818100805013), Shenzhen Institute of Advanced Technology Distinguished Young Scholars (No. E4G024), Shenzhen Science and Technology Program (No. KJZD20230923115114028), and the Key Area R&D Program of Guangdong Province (No. 2018B030338001).

References

[1] H. H. W. Wong, C. Y. C. Chou, A. J. Watt, and P. J. Sjöström, Comparing mouse and human brains, *eLife*, vol. 12, p. e90017, 2023.

[2] S. R. Datta, D. J. Anderson, K. Branson, P. Perona, and A. Leifer, Computational neuroethology: A call to action, *Neuron*, vol. 104, no. 1, pp. 11–24, 2019.

[3] J. Lauer, M. Zhou, S. Ye, W. Menegas, S. Schneider, T. Nath, M. M. Rahman, V. Di Santo, D. Soberanes, G. Feng, et al., Multi-animal pose estimation, identification and tracking with DeepLabCut, *Nat. Methods*, vol. 19, no. 4, pp. 496–504, 2022.

[4] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning, *Nat. Neurosci.*, vol. 21, no. 9, pp. 1281–1289, 2018.

[5] T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, and M. W. Mathis, Using DeepLabCut for 3D markerless pose estimation across species and behaviors, *Nat. Protoc.*, vol. 14, no. 7, pp. 2152–2176, 2019.

[6] M. Marks, Q. Jin, O. Sturman, L. von Ziegler, S. Kollmorgen, W. von der Behrens, V. Mante, J. Bohacek, and M. F. Yanik, Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments, *Nat. Mach. Intell.*, vol. 4, no. 4, pp. 331–340, 2022.

[7] Y. Han, K. Chen, Y. Wang, W. Liu, Z. Wang, X. Wang, C. Han, J. Liao, K. Huang, S. Cai, et al., Multi-animal 3D social pose estimation, identification and behaviour embedding with a few-shot learning framework, *Nat. Mach. Intell.*, vol. 6, no. 1, pp. 48–61, 2024.

[8] B. Liu, Y. Qian, and J. Wang, Rodent arena multi-view monitor (RAMM): A camera synchronized photographic control system for multi-view rodent monitoring, *Tsinghua Science and Technology*, 2025, vol. 30, no. 5, pp. 2195–2214.

[9] K. V. Gris, J. P. Coutu, and D. Gris, Supervised and unsupervised learning technology in the study of rodent behavior, *Front. Behav. Neurosci.*, vol. 11, p. 141, 2017.

[10] K. Huang, Y. Han, K. Chen, H. Pan, G. Zhao, W. Yi, X. Li, S. Liu, P. Wei, and L. Wang, A hierarchical 3D-motion learning framework for animal spontaneous behavior mapping, *Nat. Commun.*, vol. 12, no. 1, p. 2784, 2021.

[11] H. Jhuang, E. Garrote, X. Yu, V. Khilnani, T. Poggio, A. D. Steele, and T. Serre, Automated home-cage behavioural phenotyping of mice, *Nat. Commun.*, vol. 1, p. 68, 2010.

[12] C. Weinreb, J. E. Pearl, S. Lin, M. A. M. Osman, L. Zhang, S. Annapragada, E. Conlin, R. Hoffmann, S. Makowska, W. F. Gillis, et al., Keypoint-MoSeq: Parsing behavior by linking point tracking to pose dynamics, *Nat. Methods*, vol. 21, no. 7, pp. 1329–1339, 2024.

[13] A. B. Wiltschko, M. J. Johnson, G. Iurilli, R. E. Peterson,

J. M. Katon, S. L. Pashkovski, V. E. Abraira, R. P. Adams, and S. P. Datta, Mapping sub-second structure in mouse behavior, *Neuron*, vol. 88, no. 6, pp. 1121–1135, 2015.

[14] C. L. Ebbesen and R. C. Froemke, Body language signals for rodent social communication, *Curr. Opin. Neurobiol.*, vol. 68, pp. 91–106, 2021.

[15] Q. Liu, X. Yang, M. Luo, J. Su, J. Zhong, X. Li, R. H. M. Chan, and L. Wang, An iterative neural processing sequence orchestrates feeding, *Neuron*, vol. 111, no. 10, pp. 1651–1665.e5, 2023.

[16] V. Alba, G. J. Berman, W. Bialek, and J. W. Shaevitz, Exploring a strongly non-Markovian animal behavior, arXiv preprint arXiv: 2012.15681, 2020.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Proc. 31st Int. Conf. Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 6000–6010.

[18] L. Floridi and M. Chiriatti , GPT-3: Its nature, scope, limits, and consequences, *Minds Mach.*, vol. 30, no. 4, pp. 681–694, 2020.

[19] Z. Chen, D. Jia, Y. Sun, L. Yang, W. Jin, and R. Liu, Univariate time series anomaly detection based on hierarchical attention network. *Tsinghua Science and Technology*, vol. 29, no. 4, pp 1181–1193, 2024.

[20] B. Tang and D. S. Matteson, Probabilistic transformer for time series analysis, in *Proc. 35th Int. Conf. Neural Information Processing Systems*, Virtual Event, 2021, p. 1807.

[21] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proc. 2019 Conf. the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies, Volume 1* (*Long and Short Papers*), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[22] Y. Han, K. Huang, K. Chen, H. Pan, F. Ju, Y. Long, G. Gao, R. Wu, A. Wang, L. Wang, et al., MouseVenue3D: A markerless three-dimension behavioral tracking system for matching two-photon brain imaging in free-moving mice, *Neurosci. Bull.*, vol. 38, no. 3, pp. 303–317, 2022.

[23] W. Zong, R. Wu, M. Li, Y. Hu, Y. Li, J. Li, H. Rong, H. Wu, Y. Xu, Y. Lu, et al., Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice, *Nat. Methods*, vol. 14, no. 7, pp. 713–719, 2017.

[24] W. Zong, R. Wu, S. Chen, J. Wu, H. Wang, Z. Zhao, G. Chen, R. Tu, D. Wu, Y. Hu, et al., Miniature two-photon microscopy for enlarged field-of-view, multi-plane and long-term brain imaging, *Nat. Methods*, vol. 18, no. 1, pp. 46–49, 2021.

[25] Y. Han, K. Huang, K. Chen, L. Wang, and P. Wei, An automatic three dimensional markerless behavioral tracking system of free-moving mice, in *Proc. 2021 IEEE 11th Annu. Int. Conf. CYBER Technology in Automation, Control, and Intelligent Systems* (*CYBER*), Jiaxing, China, 2021, pp. 306–310.

[26] S. Sheshadri, B. Dann, T. Hueser, and H. Scherberger, 3D reconstruction toolbox for behavior tracked with multiple cameras, *J. Open Source Softw.*, vol. 5, no. 45, p. 1849, 2020.

[27] G. Tang, Y. Han, X. Sun, R. Zhang, M. Han, Q. Liu, and P. Wei, Anti-drift pose tracker (ADPT): A transformer-based network for robust animal pose estimation cross-species, *eLife*, vol. 13, p. RP95709, 2025.

[28] A. M. Rush, The annotated transformer, in *Proc. Workshop for NLP Open Source Software* (*NLP-OSS*), Melbourne, Australia, 2018, pp. 52–60.

[29] W. Cui, Y. Wu, S. Liu, F. Wei, M. X. Zhou, and H. Qu, Context-preserving, dynamic word cloud visualization, in *Proc. 2010 IEEE Pacific Visualization Symp.* (*PacificVis*), Taipei, China, 2010, pp. 121–128.

[30] D. Claudio, S. Moyce, T. Albano, E. Ibe, N. Miller, and M. O'Leary, A Markov chain model for mental health interventions, *Int. J. Environ. Res. Public Health*, vol. 20, no. 4, p. 3525, 2023.

[31] J. Ye, Y. Xu, K. Huang, X. Wang, L. Wang, and F. Wang, Hierarchical behavioral analysis framework as a platform for standardized quantitative identification of behaviors, *Cell Reports*, vol. 44, no. 2, p. 115239, 2025.

[32] A. B. Wiltschko, T. Tsukahara, A. Zeine, R. Anyoha, W. F. Gillis, J. E. Markowitz, R. E. Peterson, J. Katon, M. J. Johnson, and S. R. Datta, Revealing the structure of pharmacobehavioral space through motion sequencing, *Nat. Neurosci.*, vol. 23, no. 11, pp. 1433–1443, 2020.

[33] N. Liu, Y. Han, H. Ding, K. Huang, P. Wei, and L. Wang, Objective and comprehensive re-evaluation of anxiety-like behaviors in mice using the Behavior Atlas, *Biochem. Biophys. Res. Commun.*, vol. 559, pp. 1–7, 2021.

[34] S. S. Moy, J. J. Nadler, A. Perez, R. P. Barbaro, J. M. Johns, T. R. Magnuson, J. Piven, and J. N. Crawley, Sociability and preference for social novelty in five inbred strains: An approach to assess autistic-like behavior in mice, *Genes, Brain Behav.*, vol. 3, no. 5, pp. 287–302, 2004.

[35] T. Gschwind, A. Zeine, I. Raikov, J. E. Markowitz, W. F. Gillis, S. Felong, L. L. Isom, S. R. Datta, and I. Soltesz, Hidden behavioral fingerprints in epilepsy, *Neuron*, vol. 111, no. 9, pp. 1440–1452.e5, 2023.

[36] I. B. A. Bartal, J. Decety, and P. Mason, Empathy and pro-social behavior in rats, *Science*, vol. 334, no. 6061, pp. 1427–1430, 2011.

[37] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge University Press, 2003.

[38] J. E. Markowitz, W. F. Gillis, C. C. Beron, S. Q. Neufeld, K. Robertson, N. D. Bhagat, R. E. Peterson, E. Peterson, M. Hyun, S. W. Linderman, et al., The striatum organizes 3D behavior via moment-to-moment action selection, *Cell*, vol. 174, no. 1, pp. 44–58.E17, 2018.

[39] T. M. J. Fruchterman and E. M. Reingold, Graph drawing by force-directed placement, *Softw.*: *Pract. Exp.*, vol. 21, no. 11, pp. 1129–1164, 1991.

[40] Y. Wang, X. Song, X. Chen, Y. Zhou, J. Ma, F. Zhang, L.

Wei, G. Qi, N. Yadav, B. Miao, et al., Estrus-tracking cortical neurons integrate social cues and reproductive states to adaptively control sexually dimorphic sociosexual behaviors, bioRxiv preprint bioXiv, 2024, DOI: 10.1101/2024.08.30.610466.

[41] K. Friston, The free-energy principle: A unified brain theory, *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, 2010.

[42] M. W. Mathis and A. Mathis, Deep learning tools for the measurement of animal behavior in neuroscience, *Curr. Opin. Neurobiol.*, vol. 60, pp. 1–11, 2020.

Yaning Han received the PhD degree from University of Chinese Academy of Sciences, China, in 2025. He finished the PhD program at the State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China. He is a research assistant professor at the SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology, China. His research interest focuses on using artificial intelligence to decipher the brain functions behind complex behaviors in the natural environment.

Furong Ju received the PhD degree in zoology from Lanzhou University, China, in 2018. He was a postdoctoral researcher at the State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China. His research interest is developing the approaches of two-photon imaging in freely behaving mice.

Liping Wang received the PhD degree from Max-Delbruck Center for Molecular Medicine, Germany, in 2005. He is the founder and director of the State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China. His research interest focuses on (1) dissecting the mechanisms of subcortical neural circuits involved in innate behaviors, including fear, feeding, and sleeping; (2) elucidating the structural and functional interactions between central nervous and peripheral systems under physiological and pathological conditions, and their comorbidity with mental disorders; (3) developing new research tools and technologies for fine behavioral identification and functional dissection of neural circuits.

Zhiwei Jiang received the PhD degree from University of Chinese Academy of Sciences, China, in 2022. He is currently a postdoctoral researcher at the State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China. His research interest is the neural computation mechanisms of controlling natural behaviors.

Quanying Liu received the PhD degree from Eidgenössische Technische Hochschule Zürich, Switzerland, in 2017. She is currently an assistant professor at Department of Biomedical Engineering, Southern University of Science and Technology, China. Her research interest focuses on interactions among neuroscience, machine learning, and control theory, including multi-modal neural signal processing (EEG, sEEG, fMRI, and DTI), AI for neuroscience (explainable AI to interpret the structure and function of the brain), optimization for neuromodulation (tES, TMS, and electrical stimulation).

Pengfei Wei received the PhD degree from University of Chinese Academy of Sciences, China, in 2014. He is currently a professor at School of Biological Science and Medical Engineering, State Key Laboratory of Digital Medicine, Southeast University, China. He is also a visiting principal investigator at the State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China. His research interest focuses on the development of precise neural modulation methods and the artificial intelligence-based neuroethological research system.