Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Basolateral amygdala circuits in defensive behavior regulation	
Principal Investigator:	Aubrey C. Chan	
Prepared by (IIAI):	Kang Lee	
Other investigators:	John A. Wemmie	
Date:	December 29, 2020	
Were specific aims fulfilled:	Yes	
Readiness for extramural proposal?	No	
If yes … Planned submission date		
Funding agency		
Grant mechanism		
If no … Why not? What went wrong?	The team works on fine-tuning specific aims – based on which, some additional preliminary results will likely be needed.	

Brief summary of accomplished results:

This project consisted of two phases:

1) automatic determination of mouse behavior status
2) prediction of mouse behavior from temporal recordings of brain neuron activity

For the project, 2 manually labeled videos and 10 unlabeled videos of a caged mouse were available together with temporally corresponding recordings of neuronal activity.

In the first phase for automatic mouse behavior determination, we aimed to detect Exploring (E), Freezing (F), and Jumping (J) mouse behavior states in a video. We first used the DeepLabCut software for mouse center detection after training the DeepLabCut to perform this task using labeled videos. Once the center of mouse was identified for each frame in each video, jumping and freezing states were independently identified as we found that each problem had their own solution. All frames not belonging to J or F states were labeled as E = Exploring. In the 2 labeled videos for which manually determined independent standard was available, the jumping detection module employed image projection techniques and achieved 100% precision (9 out of 9) and 90% recall (9 out of 10), although the number of examples was very low due to the nature of experiments. The freezing detection module leveraged machine learning techniques and achieved 96% precision and 95% recall. When the two modules were applied to the unlabeled videos and manually evaluated, we were unable to verify the jump detection performance because there were actually no jumps in the 10 unlabeled videos. Freezing detection did not work well, caused primarily by a very small training dataset with only 2 video sequences from 2 mice labeled and thus available for training. The observed manual-labeling ambiguity in defining how a freeze should be identified was another factor contributing to insufficient performance.

In the second phase devoted to mouse behavior prediction, we aimed to predict mouse behavior (exploring/freezing/jumping) from 8 features including Genotype, Sex, GasConc, Delta, Theta, Alpha, Beta, and Gamma. With 2 videos manually labeled, the highest accuracy of mouse behavior prediction was 87.7% from the random forest model, although the accuracy for jumping is zero due to data sparsity. With the addition of 10 videos automatically labeled, the highest accuracy of mouse behavior prediction is 89.9% from the random forest model, although the accuracy for freezing is very low (29%), which indicates that freeze detection needs improvement. (Note that the prediction performance assessment used the auto-determined behavior state.)

Research report:

Aims (provided by PI):

The overall objective of this study is to determine how neurons in the amygdala govern defensive behavior. We hypothesize that amygdala neuronal activity regulates and predicts transitions in mouse behavior.

- Specific Aim 1: Determine the patterns of neuronal activity that predict transitions in mouse behavior.
- Specific Aim 2: Determine if amygdala neuronal activity can predict mouse genotype.
- Specific Aim 3: Use optogenetic techniques to manipulate the defensive response to CO2.
    - Note that this aim was originally proposed but after assessment of complexity removed from the pilot project and left for future extramurally funded research

Data:

There are total 12 videos. 2 videos manually labeled generated 11,994 examples (6,757 E’s, 5,231 F’s, and 6 J’s) for the first pass. 10 videos automatically labeled generated total 71,964 examples (63,181 E’s, 8,687 F’s, and 96 J’s) for the second pass.

AI/ML Approach:

Detection of Mouse Location in Video Frames

The software DeepLabCut (http://www.mousemotorlab.org/deeplabcut) was used to facilitate labeling videos and training deep learning models for mouse detection, as presented in Figures 1 and 2. We manually selected 30 frames from a video and label the center of mouse in each frame.

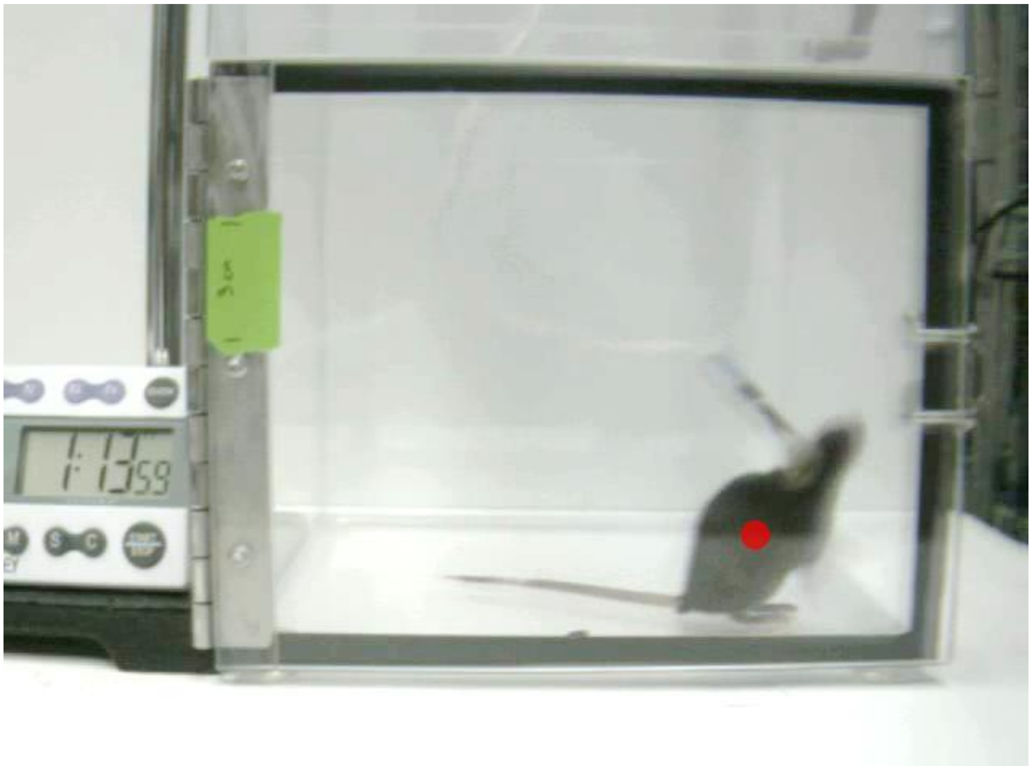
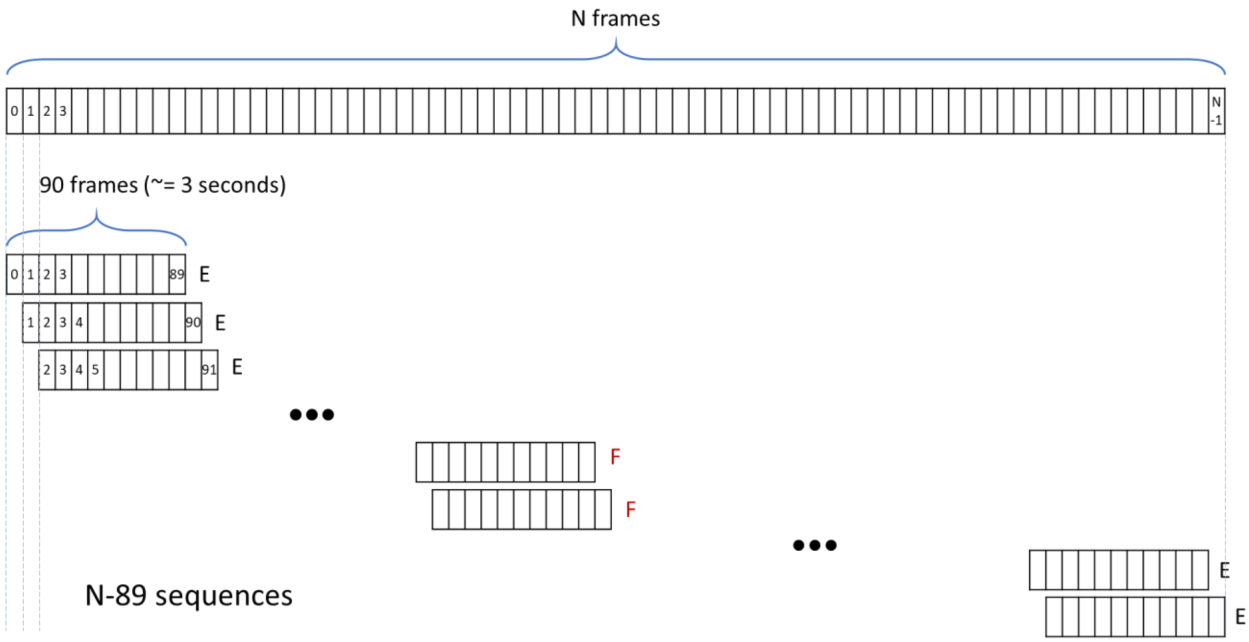


Figure 1. The center of mouse is detected by DeepLabCut. The behavior of this mouse is later then classified as Exploring by the mouse behavior detection module.

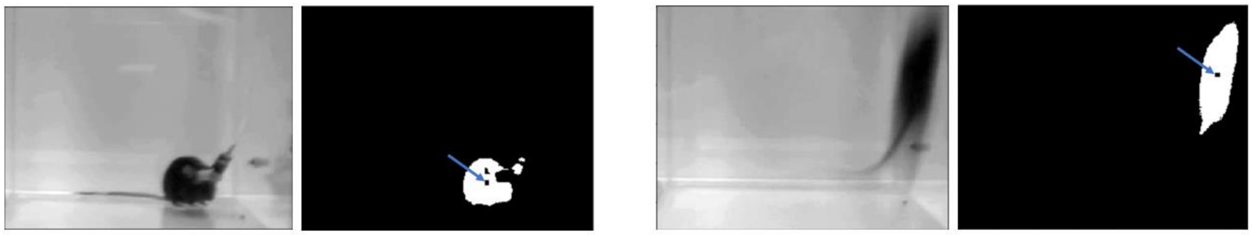
Figure 2. The behavior of this mouse is later classified as Jumping by the mouse behavior detection module.

Automated Labeling of Mouse Behavior in Video Frames

For freezing detection, 90 frames, which correspond to approximately 3 seconds, were considered a single sequence. Each sequence was labeled as Freezing only if all frames were labeled with Freezing and as Exploring otherwise. Different types of classical classification algorithms were employed, such as k-Nearest Neighbors (k-NNs), Logistic Regression, Decision Trees, Random Forest, Linear/Kernelized SVMs, and Neural Networks.



The jumping detection was facilitated by the estimation of the center of mass of the mouse in each video frame, using the SciPy package (scipy.org). The original 3-channel RGB video frames were converted into grayscale images, and were then thresholded to identify the foreground, which represented the mouse mass, and the background. Due to the interference of investigator's hand masses in various frames, which hampered the automated mouse mass detection algorithm, an optimal coverage threshold was applied to detect and eliminate those frames with human hands. Specifically, a true-positive mouse mass should not account for more than 5.8% of the total pixels in the frame. A jump is detected if the center of mass of the mouse lies above 65% of the cage height. This cut-off threshold was found to be effective in eliminating those incidents when the mouse was "standing" or "tiptoeing" instead of jumping. As a final step, a median filter with a window size of three was applied to the resulting frame series to identify those frames when the mouse was in a middle of the jump and its center of mass was outside the region of interest, which was the cage. Figure 3b shows an example of a frame, in which a jump was detected using the proposed automated method.



a) b)

Figure 3 An example of a frame without a jumping mouse (a) and a frame with a detected jumping mouse (b). The center of mass of the mouse is shown (arrow).

Prediction of Mouse Behavior in Response to Neuronal Stimulation

For mouse behavior prediction, there are three classes to predict: E (exploring), F (freezing), and J (jumping). The feature columns to use for prediction include "Behavior", "Genotype", "Sex", "GasConc", "Delta", "Theta", "Alpha", "Beta", and "Gamma". Then the 6 features of them such as "GasConc", "Delta", "Theta", "Alpha", "Beta", and "Gamma" are extended such that it further includes a set of those 6 feature values from the previous 1 second, another set of the 6 feature values from the previous 2 seconds, and another set of the 6 feature values from the previous 3 seconds. This was necessary because we believe it requires at most 3 seconds for a mouse to respond to a stimulus. For mouse behavior prediction, multiple types of classical classification algorithms were employed, as listed above.

Experimental methods, validation approach

Mouse Behavioral State Detection:

There are 2 videos with the mouse behavior column manually labeled (actually 3 videos, but one of them dropped as it has missing values) and 10 videos with the column automatically labeled. Labeling identified frame numbers associated with each behavioral state change.

For automatic mouse behavior detection, we separated the two modules for jumping and freezing, as we found that each problem had their own solution. The jumping detection module employed image projection techniques and achieved 100% precision (9 out of 9) and 90% recall (9 out of 10), although the number of examples was very low due to the nature of experiments. The freezing detection module leveraged machine learning techniques and achieved 96% precision (5214 out of 5417) and 95% recall (5214 out of 5461). When the two modules were applied to the unlabeled videos and manually evaluated, we were unable to verify the jump detection performance because there were actually no jumps in these 10 videos. Freezing detection did not seem to work well, possibly due to the ambiguity in defining what a freeze is.

Mouse Behavioral State Prediction:

For the first pass, only the 2 videos that are manually labeled were used. Again, there are 6 original features, 6 lag features from the previous 1 second, 6 lag features from the previous 2 seconds, and 6 lag features from the previous 3 seconds. There are 11,994 examples in total, of which 6,757 are E's, 5,231 are F's, and only 6 are J's. This is a biased dataset, but we did not address the bias issue due to the pilot character of the analyses. We randomly split the entire dataset into 80% for training and 20% for testing and then applied most of the commonly used classification algorithms as listed above.

For the second pass, we added the rest of the 10 automatically labeled videos to the model. Automated analysis of 71,964 video image frames in total identified 63,181 classified as E's, 8,687 classified as F's, and 96 classified as J's (J's are still scarce).

Results:

Performance summary for the first pass:

'k-NNs': 0.835
'Logistic Regression': 0.564
'Decision Trees': 0.564
'Random Forest': 0.877
'Linear SVMs': 0.592
'Kernelized SVMs': 0.770
'Neural Networks': 0.853

The random forest model yields the highest accuracy of 88%. The numbers appear to be good, but if we look at the classification report, we see that the accuracy for jumping is just zero due to the sparsity for the J class.

Performance of Random Forest classification:

Behavioral State	Precision	Recall	F1-score	Support
E	0.89	0.89	0.89	1,353
F	0.86	0.86	0.86	1,045
J	0	0	0	1

Performance summary for the second pass:

'k-NNs': 0.873
'Logistic Regression': 0.880
'Decision Trees': 0.880
'Random Forest': 0.899
'Linear SVMs': 0.880
'Kernelized SVMs': 0.880
'Neural Networks': 0.880

Again, the Random Forest model outperformed the other approaches. We see the low recall for the F class, which clearly indicates that our freeze detection module needs improvement, not to mention the low accuracy for the J class.

Performance of Random Forest classification:

Behavioral State	Precision	Recall	F1-score	Support
E	0.90	1.00	0.95	12,661
F	0.92	0.17	0.29	1,715
J	1.00	0.12	0.21	17

As the last step, based on the fact that a mouse would respond quite consistently to a certain setting of stimuli, we distinguished the mice for training from the mice for testing to check if the trained model still works well for different mice. To be specific, we selected four mice (X5493, X1582, X1323, and X5020) for training, the data for which constitutes approximately 75% of the whole data size, and three mice (X1578, X4972, and X4975) for testing, which constitutes approximately 25% of the whole data size. Below are the results, which are not much lower than the performance of the model not considering the mouse level.

'k-NNs':	0.819
'Logistic Regression':	**0.835**
'Decision Trees':	0.740
'Random Forest':	0.834
'Linear SVMs':	**0.835**
'Kernelized SVMs':	**0.835**
'Neural Networks':	**0.835**

This confirms that our models work well for different mice and in combination with the previous results suggest that additional improvements can be expected once more training data become available.

Ideas/aims for future extramural project:

The overall, neuroscientific goal is to determine what biological factors influence defensive behavioral responses to threat stimuli (in this case CO_2 exposure). Toward that goal, one intermediate goal of this project was to develop an automated behavioral classification system so that behavior videos could be scored efficiently and without subjective bias, as current manual scoring techniques are time-consuming and subject to intra- and inter-rater variability.

The results from this pilot study are promising. However, to apply for extramural funding, it is likely that the technical aspect of behavior detection will need to be functioning at an even higher performance as extramural funding would focus on the biological aspect of predicting animal behavior from measured features; the features represent neurophysiological activity and biological factors such as animal sex and genotype. To improve the behavior detection and classification, we have discussed the following plans:

1. Manual scoring of more behavioral videos to increase the size of the training set and subsequently improve the performance of the automatic behavior detection.
2. More explicit delineation of the subjective rules used by a human scorer to classify freezing behavior, so it can be translated into machine-compatible criteria.
3. Increase number of behavior videos that show jumping behavior to improve jump detection training and to verify jump detection.
4. Optimization of physical recording setup for improved consistency across samples; currently the chamber and camera both need to be moved for each recording which introduces variation in the field of view, distance, focus, etc.
5. Consideration of using multiple camera angles to improve visualization of mouse behavior (e.g., it is currently difficult to classify mouse behavior if only able to view it from behind).

Once behavioral detection has been improved, extramural funding for further technical improvements and biological aims would be feasible. Biological aims going forward will focus on behavior prediction from features and then artificially modifying those features (e.g., by electrical brain stimulation, optogenetic stimulation, or genetic manipulations) to change behavioral responses. In addition, explainability of how an AI uses features to predict behavior will shed light on the physiology and biology of the responses to different stimuli by highlighting the most relevant biological factors, and is of great interest to funding agencies. We have also considered addition of more features, such as firing rates of individual neurons, for prediction.

Further technical improvements would include adding more classes of behavior, such as grooming and rearing, both of which are currently classified as exploring; particularly rearing is of interest as it may be an intermediate behavior between exploring and jumping, and has already needed to be delineated from actual jumping in this pilot study.

Publications resulting from project:

N/A