A Computational Approach for Detailed Quantification of Mouse Parenting Behavior

A Computational Approach for Detailed Quantification of Mouse Parenting Behavior

A thesis presented by

Lynne Xie

to

the Faculty of the Committees on Degrees in Computer Science and Neuroscience

in partial fulfillment of the requirements

for the joint degree with honors

of Bachelor of Arts

Harvard University

Cambridge, Massachusetts

March 2022

The Harvard College Honor Code

Members of the Harvard College community commit themselves to producing academic work of integrity – that is, work that adheres to the scholarly and intellectual standards of accurate attribution of sources, appropriate collection and use of data, and transparent acknowledgement of the contribution of others to their ideas, discoveries, interpretations, and conclusions. Cheating on exams or problem sets, plagiarizing or misrepresenting the ideas or language of someone else as one's own, falsifying data, or any other instance of academic dishonesty violates the standards of our community, as well as the standards of the wider world of learning and affairs.

Signature: 

Acknowledgements

I would like to first and foremost thank the incredible Dr. Catherine Dulac and Dr. Mostafizur Rahman for their unending support throughout this thesis process. The evolution of this thesis undoubtedly stems from your hard work, perseverance, and guidance, and I am so honored to have had the opportunity to learn from and research with both of you. The Dulac Lab is also thanked for teaching me how to be a better researcher and how to courageously challenge revolutionary questions every week.

I also extend my deepest gratitude to Daniel Fu, Carissa Wu, and my roommates. From late night rants when my thesis was stalled to mini celebrations when my code finally worked, I could not have gotten through this thesis process without the love and support I have continuously received. I am so excited to share this beautiful piece of work that you all have helped make possible.

To the numerous adult and pup mice that made these results possible, I thank each one of you for your sacrifices.

My family and God – to you I extend the deepest appreciation and love of all. To my parents, your hope and wisdom sustain me. To my brother John, you inspire me to unceasingly challenge the norm. To my 姥姥 and 姥爷, thank you for your sacrifices in leaving your home country, raising me in difficult times, and instilling in me timeless traits for academic success. Finally, to God, thank you for your continuous blessings and unconditional love (Jeremiah 29:11).

List of Contributions

The overall design and conception of this project is credited to Dr. Catherine Dulac, Dr. Mostafizur Rahman, and Lynne Xie. Unless otherwise indicated, the python pipeline was coded by Lynne Xie. All behavioral assays and calcium recordings were completed by Dr. Mostafizur Rahman.

For Part 1's previously described parenting behaviors, Dr. Mostafizur Rahman and Lynne Xie jointly designed the algorithm. The extraction and correction of DeepLabCut points were also jointly done by Dr. Mostafizur Rahman and Lynne Xie. Manual annotations were completed by Stacey Sullivan.

For Part 2's hidden state parenting behaviors, Andrew Song adapted the Generalized Linear Model-Hidden Markov Model (GLM-HMM) from the Jan Clemens Lab. Lynne Xie further optimized and used the adapted model to run experiments.

For Part 3's newly described parenting behaviors, Dr. Mostafizur Rahman and Lynne Xie jointly conceptualized the design of the algorithm.

For Part 4's neuronal correlations, the analysis of the calcium recordings were completed by Dr. Mostafizur Rahman and Lynne Xie. The final interpretation and analysis of these results were completed by Lynne Xie, with assistance and guidance from both Dr. Catherine Dulac and Dr. Mostafizur Rahman.

Abstract

A major aim of neuroscience is understanding the brain circuitry underlying observable behaviors. Recently, an increasing number of tools have allowed researchers to investigate brain circuitry in unprecedented ways. However, behavior quantification methods have remained coarse and labor intensive, thereby limiting our understanding of underlying brain circuitry. Naturalistic behaviors like parenting are especially difficult to quantify, as they must often be reduced in a laboratory setting and usually lack trial structures. To address this problem, we have created a three-pronged approach to quantify previously described as well as newly defined behaviors and motor actions of parenting in mice. Our first approach is a user-defined system of geometric feature descriptions that were used to quantify 12 previously described parenting behaviors such as pup retrieval, pup investigation, etc. Our second and third approaches uses unsupervised methods like Generalized Linear Model-Hidden Markov Models (GLM-HMM) and Hidden Markov Models (HMM) to quantify newly described parenting behaviors. Firstly, we used a GLM-HMM to identify 3 hidden states underlying the interaction of the adult and pup mouse. Next, we used an unsupervised clustering approach combined with an HMM to identify 14 behavioral states such as forward locomotion, stretch and contract, idle with no head movement, etc. To demonstrate the utility of this behavior quantification framework, we used it to distinguish parenting behaviors in mothers, fathers, and virgin female mice. We also used it to identify behaviors encoded by parenting related neurons in the medial preoptic area (MPOA). This framework ultimately provides a robust method to quantify naturalistic behaviors.

Table of Contents

THE HARVARD COLLEGE HONOR CODE 2
ACKNOWLEDGEMENTS 3
LIST OF CONTRIBUTIONS 4
ABSTRACT 5
INTRODUCTION 7
MATERIALS AND METHODS 16
RESULTS 42
DISCUSSION 61
CONCLUSION 69

Introduction

One of the major objectives of neuroscience is to understand how underlying brain circuitry produces observable behavior, defined as "the internally coordinated response of whole living organisms to internal and/or external stimuli" (Levitis et al., 2009). In recent years, with the emergence of powerful new technologies (e.g. selective genetic targeting of neurons, imaging of neuronal activity in awake behaving animals, and optogenetic/chemogenetic circuit manipulations), researchers are now able to observe and manipulate brain circuitry with higher cellular and temporal precision than ever before and infer novel roles individual neurons play in producing observable behaviors. Despite these advances, technologies for quantifying behavior—defining the specific motor actions and motivational state sequences within a given behavior—have remained rather rudimentary. Without adequate behavior quantification, our ability to interpret experimental data and understand basic brain function principles is severely limited.

Previous work prioritizing the quantification of behavior has led to remarkable new findings in neuroscience (reviewed in Krakauer et al., 2017). For example, the circuitry for sound localization was found through an initially robust behavioral problem: how does the brain identify the location of a sound source based solely on auditory cues? Given this behavioral framework, Jeffress (1948) proposed a solution hypothesizing that avian brains could match temporal delays between the ears in receiving sound waves to identify unique sound source locations. Indeed, subsequent studies were able to identify specific nuclei in barn owls that detected such delay lines and contained coincidence detectors (Parks and Rubel, 1975; Carr and Konishi, 1990; Overholt et al., 1992; Joseph and Hyson, 1993; Wagner et al., 2002). Similarly, detailed behavioral quantifications have resulted in the discovery of other key neurons:

quantifying the initial loss of implicit motivation led to the discovery of neurons that reduce movement rigor in Bradykinesia (Mazzoni et al., 2007), quantifying the jamming avoidance response among electric fish led to the discovery of neurons that detect prey signal disturbances in electrolocation (Watanabe and Takeda, 1963), and quantifying motor behavior goals led to the discovery of neural circuits that compose motor learning behaviors (Taylor et al., 2014).

The history of research with the nematode *Caenorhabditis elegans* best illustrates the critical importance of rigorous behavior quantification and how it is difficult to infer the relationship between brain mechanisms and behaviors by solely looking at neuronal activity. Indeed, despite an in-depth knowledge of the nematode genome, cell types, and connectome—essentially every cell and its connections—researchers have been unable to fully understand how neural structure and activity can be translated into behavioral actions (Carandini, 2012; Gomez-Marin et al., 2014; Badre et al., 2015; Cooper and Peebles, 2015). Moreover, even if a connection is drawn between neuronal activity and behavior, this connection can be misinterpreted without proper behavior quantification. For example, in mammals, the activity of mirror neurons was initially misinterpreted as action understanding and interpretation, but improved behavior design and quantification has now revealed that mirror neurons are involved in action selection rather than action understanding (Hickok and Hauser, 2010). Such conclusions made based on the example of mirror neurons had already been phrased by Tinbergen (1963) in the 60s, noting that without first understanding the significance of a response in the context of an animal's natural life, there may be significant confusion in labeling a coordinated response to a stimulus as a "behavior". Finally, detailed behavior quantification ensures that the right neurons are being examined, especially given that the scale of neurons involved in producing behavior varies greatly between individual behaviors (Marder and

Goaillard, 2006). Previously, a small subset of neurons was first recorded during specific behavioral tasks. Then, these recorded neuronal activities would be correlated with the timing of behavior events within the tasks. Without a large quantity of both recorded neurons and detailed behavioral quantification, these correlations could not be very specific and would allow for misinterpretations of neuronal function. For example, it is hard to rigorously infer whether multiple circuits are responsible for a single behavior without enough recorded neurons or whether one circuit is responsible for multiple behaviors without detailed behavior quantification (Marom et al., 2009; Katz, 2016). However, researchers can now record from many neurons simultaneously during any given behavior (Musall et al., 2021). Therefore, a finer and more detailed description of behavior is needed accordingly to produce correlations with a high degree of reliability and specificity. To sum it up, to understand the neural basis of behavior more holistically and rigorously, researchers must develop techniques that can accurately quantify behavior with high detail.

Although more detailed behavior quantification is needed to better understand the neural basis of behavior, many current behavior quantification methods have remained rather coarse, especially while studying naturalistic behaviors. Naturalistic behaviors, such as foraging, parental care, predator-prey interaction, sexual selection, and social interaction (Dennis et al., 2021), are exhibited by animals in their natural, wild environments; they often differ from controlled behaviors, such as lever-pressing frequency (Winstanley and Floresco, 2016), head direction (Duan et al., 2015), or go/no-go success rates (Song et al., 2017), that researchers conceive and measure in laboratory settings. For years, researchers have faced three major options for naturalistic behavior quantification that all had limitations. As a first option, researchers could choose to quantify behavior by building it into an experimental paradigm (e.g.

forcing mice to turn their heads to the left for a reward) and measuring that behavior as the experiments progressed. While this quantification method produces high-throughput and consistent data, it captures behavior in a low-dimensional way, often reducing a naturalistic behavior down to a single observed measurement (e.g. the number of times a mouse turned its head to the right or left) (Gao and Ganguli, 2015). Alternatively, researchers could choose to measure a characteristic feature of a behavior that is not inherently built into the experiment (e.g. pupil diameter, locomotion velocity, face/whisker movement (Juavinett et al., 2018). Since these features do not constrain the design of the experimental paradigm, this option allows researchers to test more naturalistic behaviors. However, this option still limits researchers in analyzing naturalistic behaviors through a certain number of features, which may overlook various complexities within the naturalistic behaviors themselves. Finally, researchers can choose to develop their own classification system and quantify behaviors manually. Although this method provides numerous details and can capture the intended naturalistic behaviors, it is extremely labor-intensive, with hours spent devising the classification system and training others to use it. Most importantly, these methods suffer from user-specific variability and bias, making it hard to translate and use behavior quantifications derived from such methods at a larger scale (Berman, 2018). In fact, there have historically been many discrepancies even among expert behavior annotators on how to accurately quantify a behavior, especially the timing of a behavior's onset and offset (Levitis et al., 2009; Segalin et al., 2020). Furthermore, given the user-defined nature of these methods, personalized classification systems may also limit the number of behaviors that are quantified, since researchers often focus on a few specific behaviors and ignore more interesting or unique ones (Wiltschko et al., 2020).

In recent years, various machine learning algorithms have improved the three aspects mentioned above by allowing for more naturalistic, complex, and unbiased quantifications. First, recent advancements in machine learning tracking systems have enabled the development of new methods that can accurately track animal body postures without artificial markers by humans. In these tracking systems, animal behaviors can be abstracted and estimated by features related to the centroid of the animal's body, an ellipse of the animal's body, or select body points of the animal (Pereira et al., 2020). This automation has led to the quantification of more naturalistic and complex behaviors, mitigation of previous tracking obstacles (e.g. occluded, disappearing, or undetected objects), and tracking of animals in new environments (e.g. multiple animals, 3D pose estimation) (Pereira et al., 2020). For simple centroid and ellipse estimations, classical trackers such as Ctrax (Branson et al., 2009) and ToxTrac (Rodriguez et al., 2017) are practical solutions that do not require researchers to expend much time or effort. For more complex centroid or ellipse estimations and pose estimations, deep learning algorithms have become the state-of-the-art technique. Mature frameworks such as DeepLabCut (Mathis et al., 2018), SLEAP (Pereira et al., 2020), and DeepPoseKit (Graving et al., 2019) have allowed researchers to accurately quantify behavior solely from video frames by integrating custom and corrective labeling, active learning, and transfer learning features. Second, using these improved tracking methods, supervised and unsupervised machine learning algorithms have rapidly improved to quantify behavior more effectively and objectively. Supervised algorithms allow researchers to define a fixed set of criteria for classifying behavior and work well for quantifying previously described behaviors. Unsupervised algorithms define behaviors based on repeated patterns found in the data without researcher input and work better for quantifying previously undescribed behaviors. The general process for these algorithms has been to first extract postural estimations

from a video, then translate these estimations into a dynamic representation, and finally create a behavioral representation based on stereotyped actions (Berman, 2018). In recent years, supervised methods such as SimBA (Nilsson et al., 2020) and MARS (Segalin et al., 2020) have allowed for unprecedented findings in mouse social behavior with higher accuracy and more unbiased quantification. Unsupervised methods such as state-space models have also recently exploded in popularity due to their abilities to quantify previously described and undescribed complex behavior. Specifically, these unsupervised machine learning models have quantified complex mouse behavior as specific behavioral components with defined transition probabilities (Wiltschko et al., 2015), fly stereotyped behaviors as hierarchical states (Berman et al., 2016), fly mating behavior as a series of unobservable, hidden states (Calhoun et al., 2019), and drug-induced mouse behavior as specific behavior syllables (Wiltschko et al., 2020).

Despite these machine learning advances, behavior quantification remains rather coarse and needs further refinement. Indeed, researchers have still been unable to quantify many naturalistic behaviors. Yet, as established above, being able to quantify naturalistic behaviors offers significant opportunities for understanding neurons' functions in producing behavior. Social behaviors, specifically parenting behavior, is uniquely challenging to quantify as it involves several animals and multiple behavioral steps. Multi-animal tracking has long posed several challenges, namely that the correct body points are assigned to each animal and that body point continuity is maintained between frames for each animal (Pereira et al., 2020). In the context of parenting behavior, the body points of the pup are especially difficult to assign and maintain across frames, given that the pup often is occluded by other objects or blends in with the background nesting material.

Parenting behavior itself is a rather fascinating social behavior and a prime candidate for developing behavior quantification methods. Despite caregivers receiving no benefits from sacrificing time and effort into raising young, parenting behavior is robustly displayed by many animal species and consistently reinforced across generations without prior training (Dulac et al., 2014). This suggests that parenting behavior relies at least in part on evolutionary, hard-wired neural circuit function. Moreover, parenting behavior is a highly complex behavior that is made up of specific motor patterns, enhanced motivation for interactions with offspring, distinct hormonal states, and suppression of other social behaviors (Kohl et al., 2018). Thus, analyzing parenting behavior provides immense potential for understanding how a variety of brain circuits and mechanisms combine to produce complex behavior.

To thoroughly analyze parenting behavior, researchers have primarily relied on experiments involving varying assays of an adult mouse and pup. These adult mice could be virgin males, virgin females, mothers, or fathers, and the pup could be anywhere from a few days to weeks old. Prior work has noted how virgin females, mothers, and fathers naturally display parental behaviors when presented with a pup, whereas virgin males will attack and kill the pup (Vom Saal, 1985; Tachikawa et al., 2013; Wu et al., 2014).

More recently, researchers have attempted to illuminate the neural circuitry underlying differences in parenting behavior according to the animal sex and physiological state. In doing so, galanin-expressing neurons in the medial preoptic area of the hypothalamus (MPOA) were identified as having a key role in controlling parenting behavior in mice, since genetic ablation and optogenetic activation of these MPOA neurons respectively inhibited and facilitated parenting behaviors (Wu et al., 2014). MPOA neurons coordinate the motor, motivational, hormonal, and social aspects of mouse parenting behavior by integrating inputs from over 20

nearby brain areas and activating these inputs based on the animal's sex (e.g. female/male) and reproductive state (e.g. virgin female/male, mother, father) (Kohl et al., 2018). Specifically, the inputs from the pheromone processing pathways of the medial amygdala (MeA) and bed nucleus of the stria terminalis were all activated differently between parenting groups of virgin females, mothers, and fathers (Kohl et al., 2018). These MPOA neurons also exhibit heterogeneity, as they are grouped into functionally distinct neuronal pools that each elicit unique aspects of parenting behavior (Kohl et al., 2018). In particular, the pools of MPOA neurons that project to the periaqueductal grey, ventral tegmental area, and MeA were confirmed respectively to increase pup grooming, increase motivation to interact with pups, and decrease competing adult social interaction behavior (Kohl et al., 2018). Another study using single cell sequencing and spatial transcriptomics identified around 70 neuronal populations with unique molecular signatures and spatial organizations in the hypothalamic preoptic region. Within these neuronal populations, six were consistently activated during parenting behavior: I-27 and E-1 solely for mothers, I-2 and I-16 solely for fathers, I-10 solely for mothers and fathers, and I-14 for mothers, fathers, and virgin females (Moffitt et al., 2018).

These molecular and functional findings have rapidly expanded researchers' understanding of mouse parenting behavior, but such findings are inevitably limited by the lack of detailed quantification of mouse parenting behavior. For example, in the studies mentioned above, researchers were not able to definitively pinpoint the specific functions of the MPOA neurons due to the lack of detailed quantification of parenting behavior. While some motor actions can be easily classified as "parenting" or "not parenting", such as grooming or attacking the pup respectively (Wu et al., 2014), other motor actions' classifications are less clear. Without a detailed quantification of mouse parenting behavior, researchers then are unable to exactly

match a MPOA neuron's activity with its specific parenting behavioral output, limiting the insights we can gain about a given MPOA neuron's function. In addition, given the complexity of mouse parenting behavior, this ambiguity in quantifying mouse parenting behavior often prompts many researchers to rely on previous studies or expert annotators to define what motor actions or observable postures constitute mouse parenting behavior (Wu et al., 2014; Kohl et al., 2018). However, this reliance limits the findings of these researchers as well, since researchers are only looking at parenting behaviors that have been previously defined. Instead, there may be unknown behaviors related to parenting that are controlled by the MPOA neurons that researchers are overlooking. Furthermore, there could also be hidden factors such as motivational or emotional states that influence parenting behavior, much like more known factors such as sex and reproductive state (Kohl et al., 2018). Without defining mouse parenting behavior properly with such hidden factors, researchers could also be limiting themselves in the number of MPOA neurons that they can manipulate and observe changes from, thereby also limiting their holistic and full understanding of mouse parenting behavior itself.

This research will address these quantification needs of mouse parenting behavior through a three-pronged approach. First, I created a user-defined system of previously described parenting behaviors (e.g. grooming the pup, investigating the pup, retrieving the pup, etc.) by correcting body points of the adult mouse and pup and defining specific geometric features of previously described behaviors (e.g. area, angles, distances, etc.). With this user-defined system, I hope that researchers studying mouse parenting behavior can accurately and quickly obtain high-throughput data on which previously described parenting behaviors are happening at any given moment. Second, I created a data-defined system by using specific clustering, classifying, and inference methods to identify and quantify newly described parenting behaviors. Since this

method relies solely on the computer's abilities to identify patterns in the data, there is potential for the computer to identify patterns previously unknown to researchers and for the researchers to define newly described parenting behavior. In addition, the inference methods also offer the ability to observe hidden motivational and emotional states of mice that may affect parenting behavior. In fact, such methods were able to uncover three hidden states in fly courtship mating behavior and allow researchers to illustrate the neurons responsible for switching between states (Calhoun et al., 2019). This data-defined system will comprise of two subparts to ensure that all aspects of mouse parenting behavior, whether the adult is alone or with the pup, will be captured sufficiently: therefore, one subpart will be solely focused on adult-pup interactions, while the other subpart will be solely focused on individual adult mouse postures.

Thus, with this three-pronged approach, we will now be able to quantify previously described and newly described mouse parenting behaviors and incorporate any effects of hidden states. Beyond mouse parenting behavior, this approach can be applied to other naturalistic behaviors. The ultimate hope of this approach is that it can elucidate a framework to understand any naturalistic behavior in the future and thereby increase the ability of researchers to precisely identify the unique functions neurons have in causing naturalistic behavior.

Materials and Methods

Parenting Assay Setup

Behavior Assays. Three different types of adult mice were used: virgin females, mothers, and fathers. For each behavior assay, the adult was allowed to roam freely for around 30 minutes in its homecage (dimensions: 15" x 8" x 6") containing a nest. Around the 5-minute mark, the pup (1-4 day(s) old) was placed into the cage. Around the 20-minute mark, the pup was removed from the cage. All recordings were done with an Infrared (IR) camera (Model: Blackfly S,

manufacturer: FLIR) and recorded from an overhead perspective. The experiments were done in the dark with IR lighting. 12 behavior assays were used (8 mothers, 2 virgin females, 2 fathers).

Animal Tracking. All points for the adult, pup, and nest were tracked using DeepLabCut (Mathis et al., 2018) (Fig. 1). Each point was represented by its x-coordinate, y-coordinate, and confidence probability that DeepLabCut had detected it properly. All results were stored and outputted in H5 files.

Neuronal Recordings. Activity of galanin-expressing medial preoptic area (MPOA) neurons were recorded during parenting behavior. Briefly, calcium sensor GCAMP was expressed in MPOA neurons using a viral strategy. Next GRIN lens was implanted targeting MPOA neurons. After recovery, the animals were attached to a head mounted microendoscope (INSCOPIX) (Li et al., 2017). Calcium imaging was performed during parenting behavior. The images were processed using a standard CNMFE algorithm (Zhou et al., 2018). The calcium activity traces of MPOA neurons were obtained, and these signals were further used for analysis in this thesis.

Part 1 – Quantifying Previously Described Parenting Behaviors

The overall workflow of Part 1 is displayed in Fig. 2. We first extracted points for the adult mouse, pup, and nest from DeepLabCut (DLC). We then applied extrapolation and median filtering corrections to the DLC points. Finally, with the corrected DLC points, we quantified previously described parenting behaviors using various geometric features. This was done in all 12 behavior assays. We also visualized the corrected points and quantified behaviors to ensure that the correction and quantification steps were completed accurately.

Extracting Tracked Points from DeepLabCut. The outputted H5 files from DLC were used to extract the position of the tracked body points of the adult mouse and the pup and subsequently

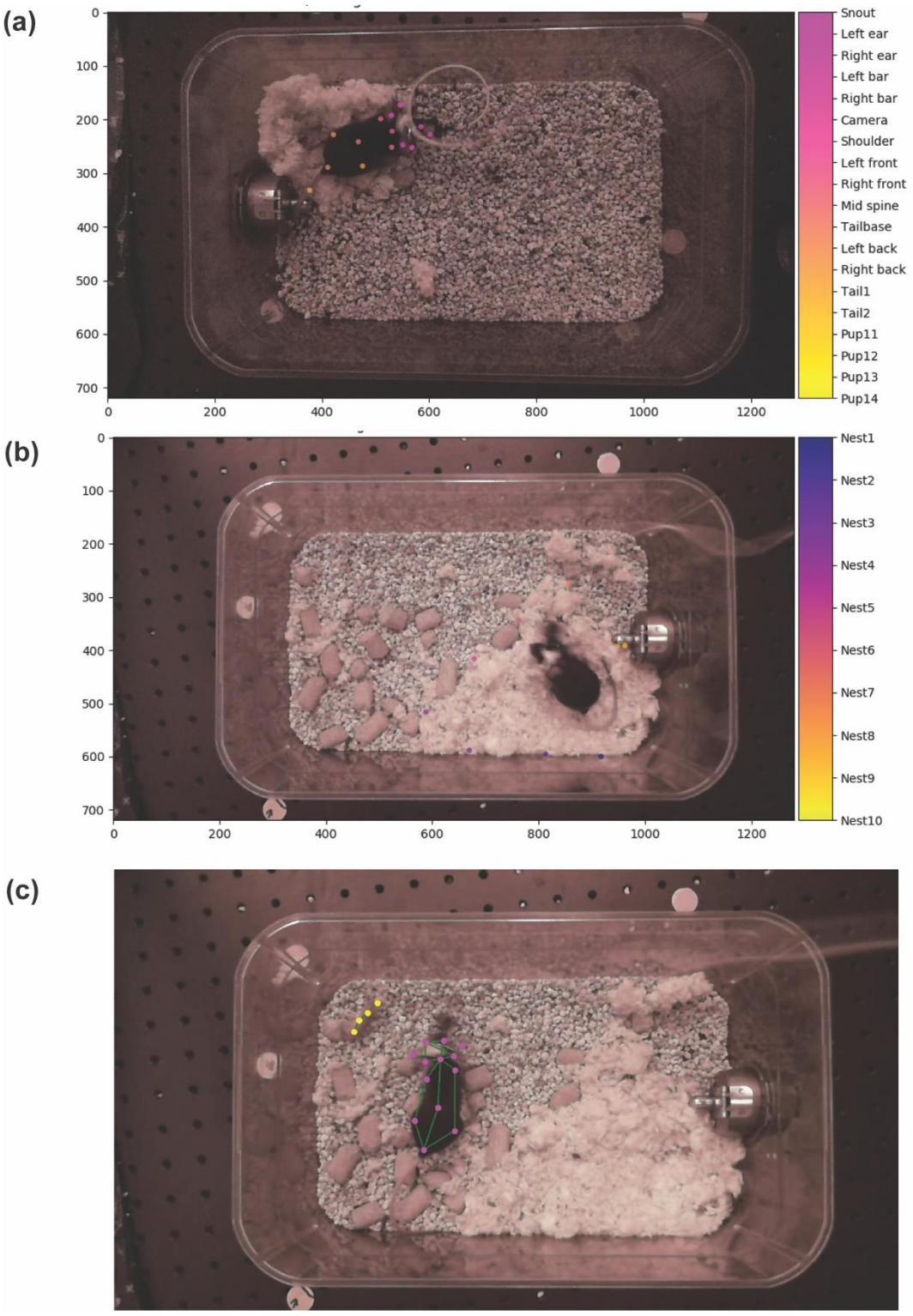


Fig. 1. Tracked Points from DeepLabCut. Examples of a tracked adult **(a)**, nest **(b)**, and pup **(c)**. There were 15 body points for the adult mouse, 10 points for the nest, and 4 body points for the pup. The 4 pup body points are highlighted in yellow in **(c)**.

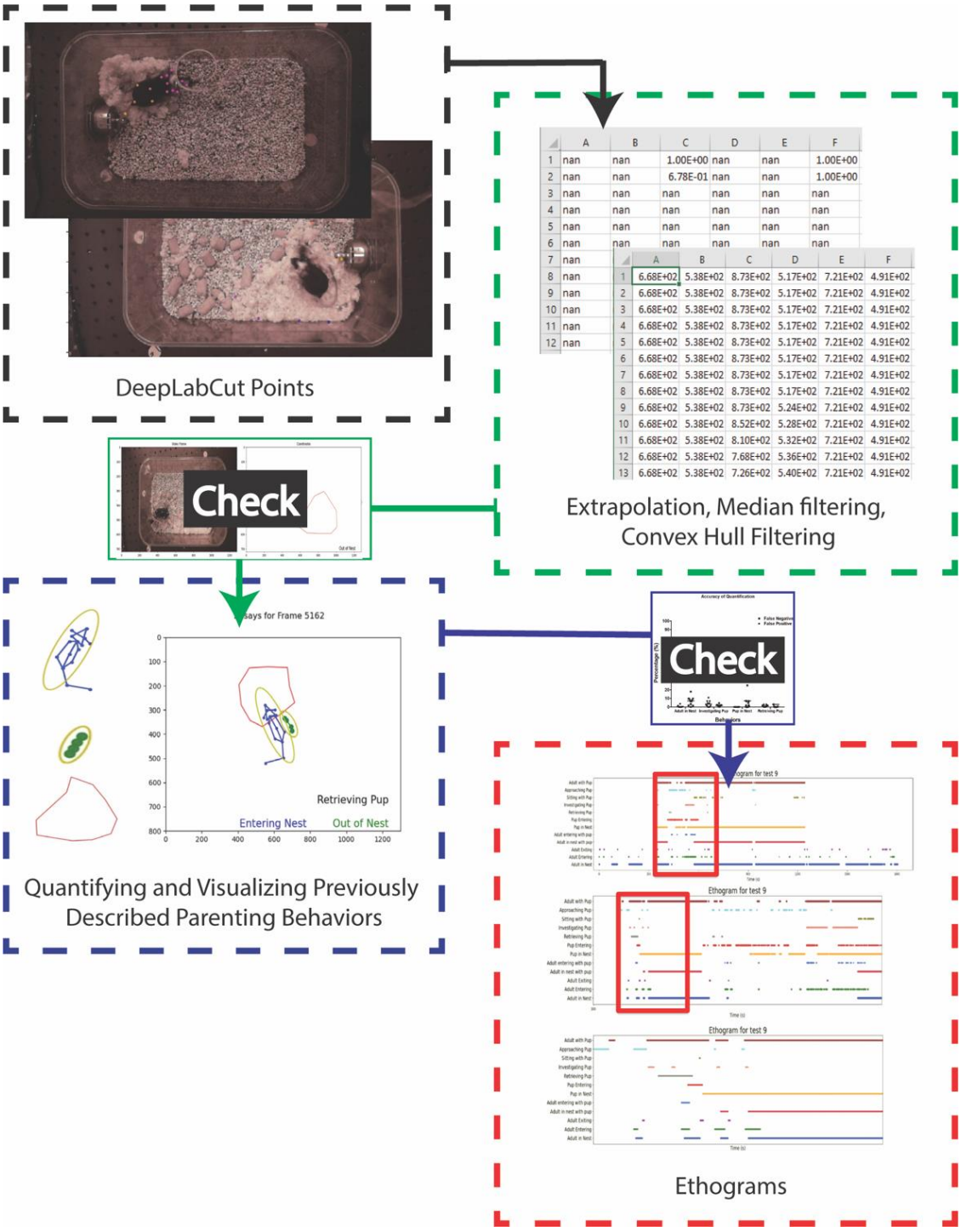


Fig. 2. Workflow for Part 1. Points from DeepLabCut were corrected, checked, and used to quantify previously described behaviors. After checking these behaviors, ethograms were outputted for each video.

converted into a python array using the *numpy* library. The array was of size (number of frames, (3 x number of points)), given each row was a video frame and contained points represented by their x-coordinates, y-coordinates, and confidence probabilities. The body points of the adult, pup, and nest were contained in different H5 files, The tracked points were organized into different numpy arrays for adult, pup and nest.

Correcting Irregular Points with Extrapolation. The imported points from DLC were preprocessed. A tracked point's x-coordinate or y-coordinate was considered invalid if DLC could not detect it in a certain frame (e.g. occluded points), indicated by a NaN value in the array, or had too low of a confidence probability (< 0.15 for nest and pup, <0.1 for adult mouse). If the invalid coordinate was during the first or last frame of the video, it was replaced by the closest frame's valid coordinate. If the invalid coordinate was during any other frame, we found the closest valid coordinates before and after it and then linearly extrapolated between these two valid coordinates to replace the invalid coordinates' values. Results are shown in Fig. 3.

Correcting Irregular Points with Median Filter. Some incorrectly detected tracked points with DLC had high confidence probabilities (e.g. detected wrong objects, points jumped randomly between frames), so they were not corrected by the extrapolation step above. To correct this possible issue with these incorrect points, we conducted median filtering with a sliding window of 1 second duration over every coordinate in the data. We set the current coordinate's value equal to the median value of the coordinates in the window. For most videos, we considered a set of 15 coordinates (7 frames before, 7 frames after, and current frame), given that 15 frames was the frame rate per second in behavior assay videos. For some behavior assay videos that were recorded at 7.5 frames per second (fps), we considered a set of 7 frames (3 before, 3 after, and

current frame). If the coordinate was in the first or last frame, we added the necessary number of frames before or after with the value 0. Results are shown in Fig. 3.

Correcting Irregular Nest Points with Convex Hull Filter. An additional filter was applied to the tracked nest points, given that the nest had a relatively fixed location and is a closed object. After applying extrapolation and the median filter to the nest points, a convex hull was formed around these points using the *scipy.spatial* library. If the area of this hull was larger than 1500 or less than 500, then the original nest points were replaced with the vertices of the convex hull. The numbers 1500 and 500 were chosen after visually inspecting multiple run-throughs for the average size of the nest, but future behavior assays may have different area threshold numbers. Occasionally, the number of convex hull vertices were less than the original tracked nest points. If this was the case, to maintain consistency across frames, the first convex hull's vertex was duplicated until 10 nest points were achieved. Results are shown in Fig. 3.

Visualization of Original Tracked Frames and Preprocessed Points. All plotting for visualization purposes was done using the *matplotlib* library. First, using the *cv2* library, each frame from the video was saved as a PNG file. This ensured that the frame could be uploaded to a *matplotlib* subplot while still maintaining the original tracked frame's dimensions for ease of comparability. Second, two subplots were created using *matplotlib* to combine the original tracked video frame with the plot of the corrected points. The corrected adult and pup points were plotted using the *matplotlib* library, and the corrected nest points were plotted using the *matplotlib.patches* library. Using the interactive feature of *matplotlib*, the plot values were dynamically replaced and redrawn each frame. Results are shown in Fig. 3.

Quantifying Previously Described Behaviors. For the previously described parenting behaviors (adult-pup, adult-nest, pup-nest, adult-pup-nest), we quantified them through a

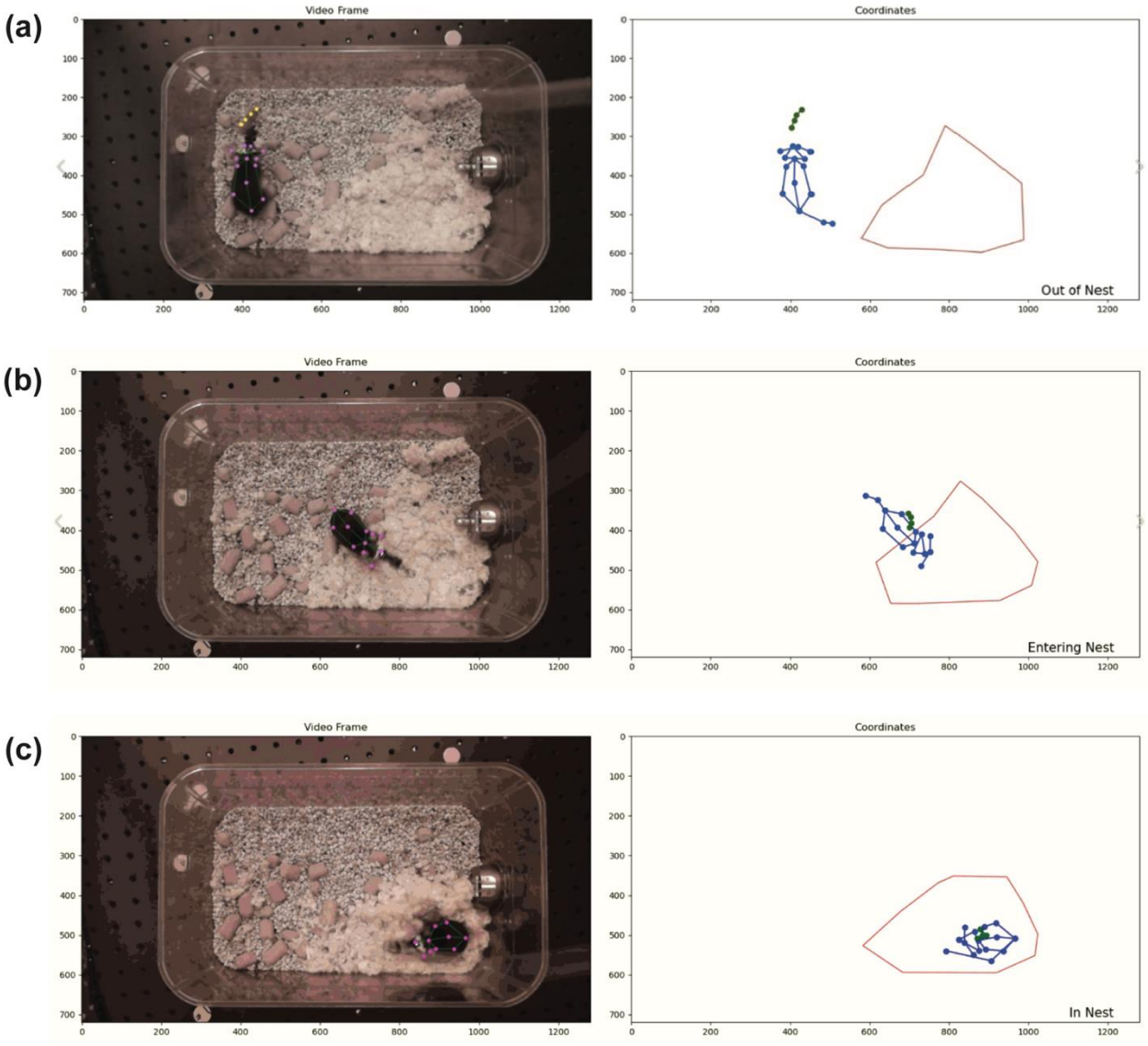


Fig. 3. Visualization of Corrected Points. Selected frames taken from a behavior assay, where the left image shows the original tracked frame and the right shows the preprocessed points. In **(a)**, the pup (green), adult (blue), and nest (red) are all shown to visually match the original tracked frame on the right. In **(b)**, it is shown that the extrapolation and filter methods are effective at correcting frames where points (e.g., pup and nest) were not originally detected. In **(c)**, it is shown that the extrapolation and filter methods allow occluded points (e.g., the pup) to still be detected.

combination of geometrical features of the nest, adult, and pup. Using the *shapely.geometry* library, we created two polygons with the corrected points to represent the adult mouse and nest and drew a line object with the corrected points to represent the pup. In addition, two ellipses were created using the *shapely.affinity* library, one centered around the adult polygon and another centered around the pup line. These ellipses defined the interaction areas around the adult and the pup. Based on previously described parenting behaviors, we used 9 parameters and various geometric features to quantify 12 parenting behaviors. The descriptions of the parameters and parenting behaviors can be found in Table 1. To save the behaviors present in each behavior assay, a numpy array of size (number of frames, number of behaviors) was initialized with all 0's. If a given frame contained a behavior, that behavior's cell value would be replaced with a 1. This file was ultimately saved and exported in a csv file.

Correcting Pup Retrieval Quantification. Although most behaviors were quantified accurately through the 9 parameters and various geometric features, we noticed that the pup retrieval behavior had some errors and sometimes single frames got labelled as pup retrieval. Since pup retrieval is a continuous behavior, there should not have been any isolated single frames with this behavior. To correct for this error, if the pup retrieval behavior was detected, we checked if this behavior happened during the first frame, last frame, or an isolated frame. If so, the cell value in the array was changed to 0. These corrections happened before the array was exported in the "Quantifying Previously Described Behaviors" method.

Visualizing Quantified Previously Described Behaviors. Using the *shapely.geometry* library, polygon shapes were plotted for the adult and nest, and a line shape was plotted for the pup. For the adult polygon, the head and body were first plotted as separate polygons before joining them together, since *shapely* does not allow for any point overlaps in one polygon. For the nest

(a)

Parameter	Description
1	Nest polygon area
2	Adult polygon area
3	Intersection area of the adult and nest polygons
4	Percent overlap of adult polygon area with nest polygon area $(3 \div 2)$
5	Percent overlap of pup with nest $(intersection\ between\ nest\ polygon\ and\ pup\ shape \div length\ of\ pup\ shape)$
6	Distance between adult and pup shapes
7	Distance between pup and nest shapes
8	Angle of adult polygon $(relative\ to\ horizon)$
9	Angle of pup polygon $(relative\ to\ horizon)$

(b)

Behavior/Description	Criteria
1. Adult in nest	Percent overlap of adult with nest > 60%
2. Adult entering nest	Percent overlap of adult/nest is larger than 1 second ago (15 frames) and not 100%, snout is in the nest
3. Adult exiting nest	Adult not in the nest, percent overlap of adult/nest is smaller than 1 second ago (15 frames) and not 0%, snout is not in nest
4. Pup in nest	Nest polygon contains pup polygon
5. Pup entering nest	Percent overlap of pup/nest is larger than 1 second ago (15 frames) and not 100%
6. Adult with pup	Adult ellipse intersects with pup ellipse
7. Investigating pup	Adult not in nest, and either adult snout intersects with pup polygon or adult snout is within 15 units of distance from pup
8. Retrieving pup	Distance between pup and nest has decreased by more than 15% over 1 second
9. Sitting with pup	Adult polygon contains pup polygon
10. Approaching pup	Distance between pup and adult has decreased by more than 15% over 1 second, and both behaviors 6 and 7 are not true
11. Adult in nest with pup	Both behaviors 1 and 4 are true
12. Adult entering nest with pup	Behavior 2 is true, and either behavior 5 or 8 is true

Table 1. The descriptions for the 9 parameters are listed in **(a)**. The 12 previously described behaviors are listed in **(b)**, along with the criteria we used to quantify them.

polygon, the points were first sorted in a counterclockwise order to ensure proper plotting. The ellipse shapes denoting interaction areas were created using the *shapely.affinity* library, and they were centered and rotated around the adult polygon or pup line.

In addition, three text objects were inserted: one showed adult only behaviors (Entering/Exiting Nest, In/Out Nest), another showed pup only behaviors (Entering Nest, In/Out Nest), and finally one showed parenting behaviors between the adult and pup (Investigating Pup, Retrieving pup, Sitting with Pup, Entering Nest with Pup, Approaching Pup, With the Pup, Not with the Pup). These texts were shown based on the saved csv file ("Quantifying Previously Described Behaviors" method) that contained which behaviors were present in each frame (indicated by a 1 in the numpy array in the csv file). If multiple behaviors were present, the code was arranged in a hierarchal fashion that displayed the first behavior according to the order presented in the parentheses above.

Each shape and text object were redrawn for each frame using the interactive feature of *matplotlib.* After all the frames had been drawn, using the *matplotlib.animation* library and the ffmpeg setting, the frames were combined and outputted in one mp4 file. Snapshots of the videos can be found in Fig. 4a-f.

Checking Quantified Previously Described Behaviors.

Behaviors were manually quantified by a trained annotator Stacey Sullivan. These annotations were exported as vectors containing 1 for a frame when a given behavior was occurring and 0 for frame when the given behavior was not occurring. These vectors were then compared to those predicted by our algorithm for four behaviors. False positive was quantified as the percentage of frames when the algorithm generated vector was 1 but the manual annotated vector was 0. False

negative was quantified as percentage of frames when the algorithm generated vector was 0 but the manual annotated vector was 1. The results are displayed in Fig. 4g.

We also plotted ethograms for each behavior assay using the *matplotlib* library and saved csv file ("Quantifying Previously Described Behaviors" method). Each behavior had its own value, where a dot was plotted at that value if the behavior was present, and a dot was plotted on the 0 value if it was not present. Ultimately, to make the graph cleaner, the y-axis for the ethogram started slightly above the 0. The x-axis was converted from frames to seconds for ease of interpretability. A sample ethogram is shown in Fig. 5.

Part 2 – Quantifying Hidden Parenting Behavioral States

The overall workflow for Part 2 is displayed in Fig. 6. First, we modified a python-based Generalized Linear Model-Hidden Markov Model (GLM-HMM), modeled off of the one originally used in Calhoun et al., 2019. Second, we corrected overlapping quantified behaviors from Part 1, since the GLM-HMM could only take in one quantified behavior per frame. Finally, we ran the GLM-HMM with the most optimal parameters to find the hidden parenting behavioral states for each video. 5 behavioral assays were used (3 mothers, 1 virgin female, 1 father).

Setting Up the Generalized Linear Model-Hidden Markov Model (GLM-HMM). We adopted the algorithm from (https://github.com/janclemenslab/pyGLMHMM) published by the Jan Clemens Lab, who rewrote the MATLAB GLM-HMM model originally developed in Calhoun et al., 2019 in Python. We optimized the model for our purpose. We varied and tested the model parameters to see which were the optimal values for representing our data. We found that 3 states, 30 bins, 0.7 lambda, and a training set including all mothers was most optimal. Further details on other parameter variations and results can be found in Table 2.

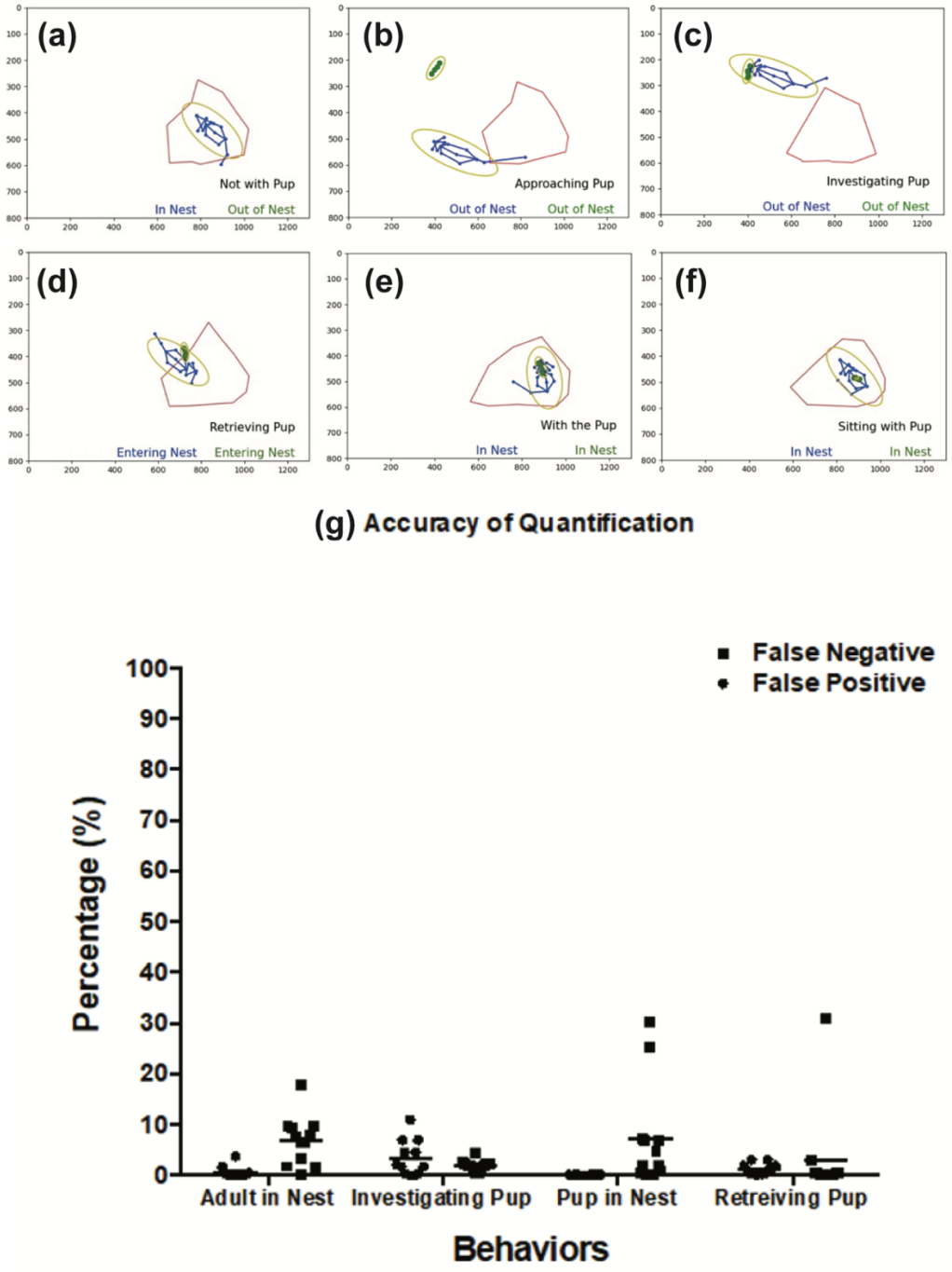


Fig. 4. Visualizations and Accuracies of Quantified Previously Described Behaviors. (a-f) Selected screenshots of the plots of a given behavior assay. Six different quantified previously described parenting behaviors are plotted, with the pup (green), adult (blue), nest (red), and ellipse (yellow) shapes. The text objects correspond with the parenting behavior (black), adult-nest interaction (blue), and pup-nest interaction (green). **(g)** The accuracy of four quantified previously described behaviors were cross-checked with manual annotations. The average false negative and positive rate was less than 10 percent, meaning that across these four behaviors, an average accuracy greater than 90 percent was observed for our Part 1 pipeline.

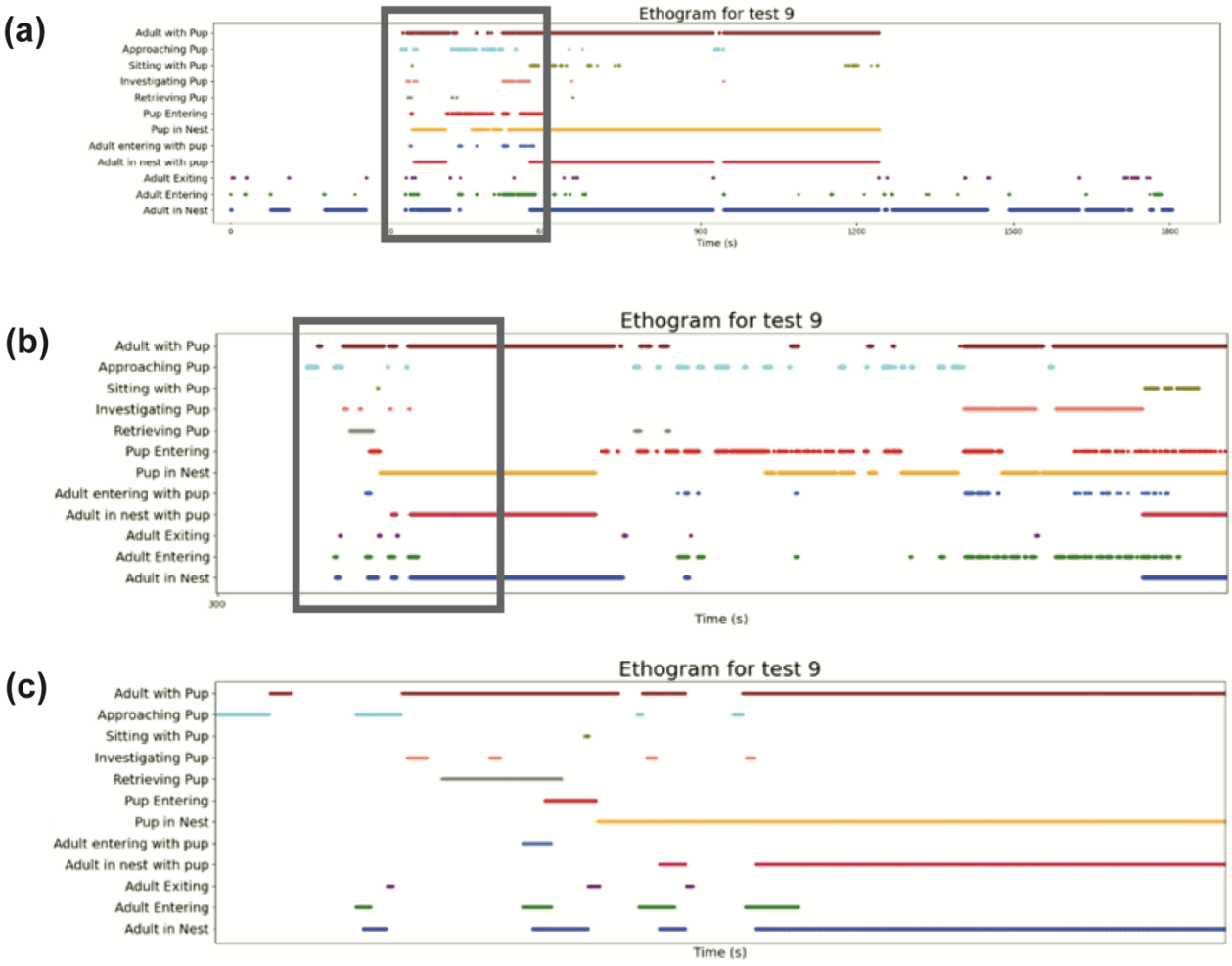


Fig. 5. An ethogram example for behavior assay #9. On the y-axis, each quantified previously described behavior is given a specific value. On the x-axis, the time in seconds is shown. **(a)** represents the entire behavior assay, whereas **(b)** is a zoomed in portion highlighted by the grey rectangle in **(a)**. **(c)** follows a similar pattern, being the zoomed in portion highlighted by the grey rectangle in **(b)**.

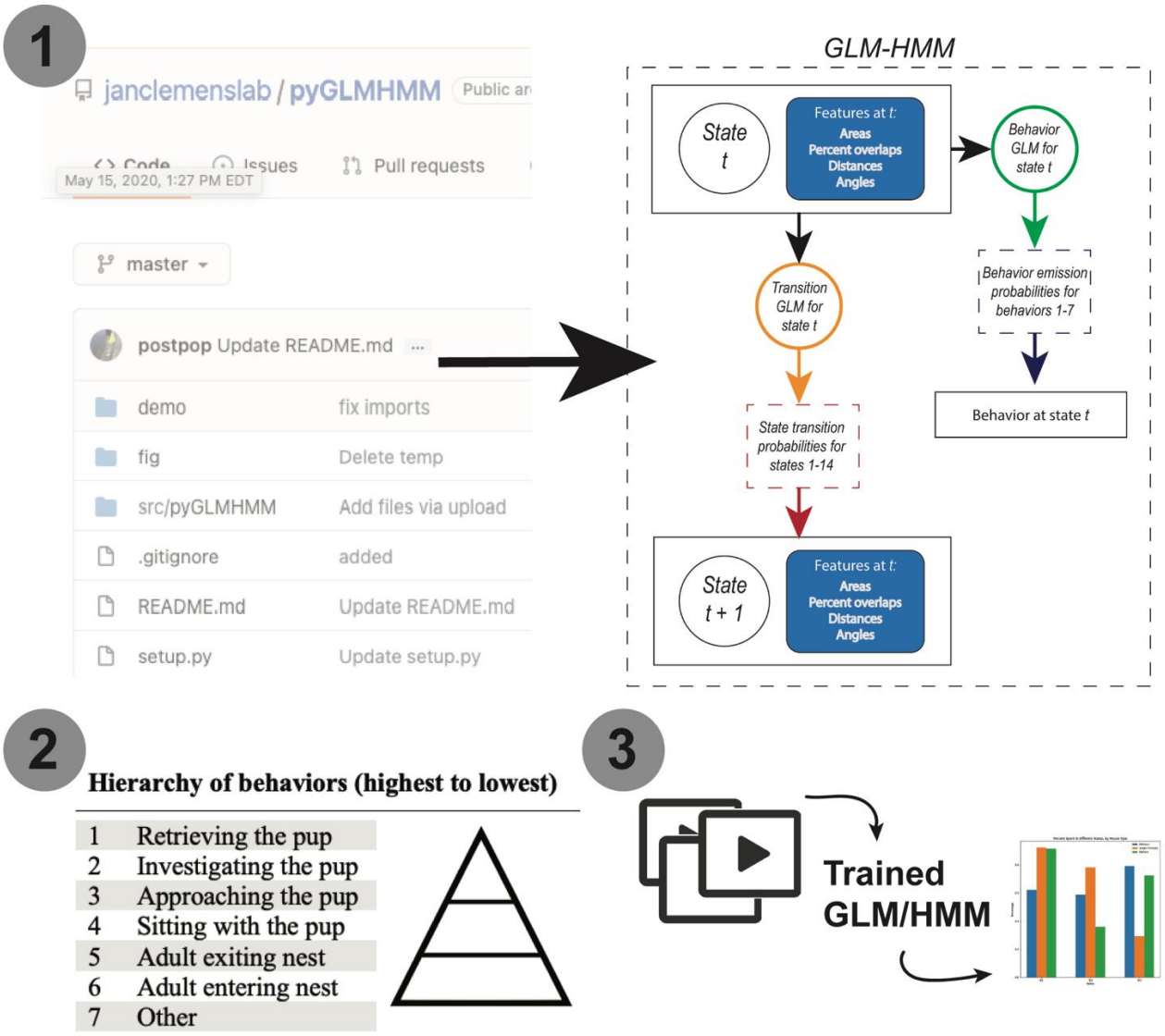


Fig. 6. The workflow for Part 2. We first modified another python-based GLM-HMM for our purposes in parenting. Then, we made sure to filter all the behaviors according to a hierarchy so that at each video frame, only one behavior was fed into the GLM-HMM. Finally, we trained and tested the GLM-HMM on all test set videos to output state transitions for each video.

(a)

Parameter	Variations
Number of States	2, 3, 4
Number of Bins	10, 30, 40, 50
Smooth Lambda	0.05, 0.2, 0.5, 0.7
Training Set	Two (F42/43) or three mothers (F42/43/45)
Test Set	Mother, virgin female, father (F45, FV4, MV1) or virgin female/father (FV4, MV1)

(b)

	Hierarchy of behaviors (highest to lowest)
1	Retrieving the pup
2	Investigating the pup
3	Approaching the pup
4	Sitting with the pup
5	Adult exiting nest
6	Adult entering nest
7	Other

Table 2. Parameter variations and behavior hierarchy for the GLM-HMM. In **(a)**, the different parameters are listed, as well as the different variations we tried. The default parameters were 2 states, 30 bins, 0.05 lambda, training set of two mothers, and test set of mother, virgin female, and father. In **(b)**, the hierarchy of behaviors is listed, where the higher behavior was assigned to the given frame when there was more than one behavior initially quantified.

Behavior Corrections. Our final model also took in two additional data inputs for training and testing it. First, we inputted 7 of the previously described quantified behaviors from Part 1 (Six behaviors 2, 3, 6, 7, 9, 10 from Table 1; all other behaviors were grouped under the seventh "Other" behavior). For each video, we only focused on the time periods when the pup was in the frame, since the hidden parenting behavioral states we were attempting to find in this part were based on adult-pup interactions. In addition, given that there were some overlapping behaviors (e.g. adult could be entering nest while also retrieving the pup), we filtered the data based on a hierarchal order (Table 2b) so that only one behavior was assigned to each frame at a given time. Second, we also inputted 8 features listed in Table 1a, where "Intersection Area" was eliminated since it contained overlapping information with the other parameters.

Training and Testing the GLM-HMM. To train the GLM-HMM, we needed to initialize the emission and transition probabilities, which represent the chance that any given behavior is outputted or state is chosen. For these, we initialized all emission probabilities at 1/(number of possible outputted behaviors) and transition probabilities at 1/(number of hidden states). In our case, the initial emission and transition probabilities were set at 1/7 and 1/3 respectively. After training the GLM-HMM, we predicted the state outputs for all videos in the test set.

Part 3 – Quantifying Newly described Parenting Behaviors (Individual Adult Postures)

The overall workflow for Part 3 is found in Fig. 7. First, we calculated and checked 14 features based on the adult body points found in Part 1. Then, we found the optimal number of clusters to conduct k-means clustering. Given the lengthy time it takes to cluster frames, we also trained two classifiers based on the k-means clustering output to use on future video frames. Finally, we found the optimal number of states for the HMM and trained it. After applying the trained HMM

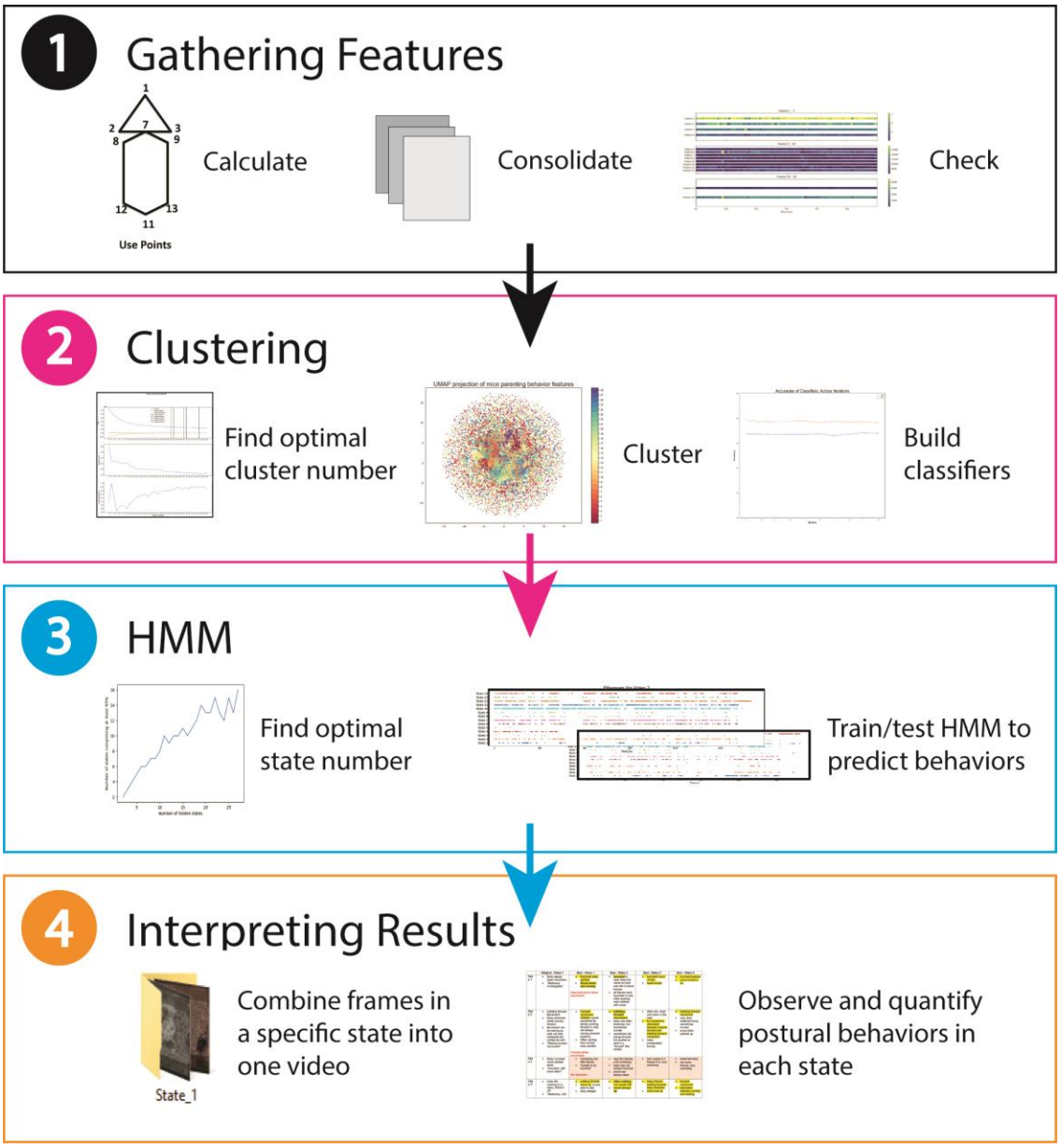


Fig. 7. The workflow for Part 3. First, features are calculated, consolidated, and checked. These features are then assigned into clusters through k-means clustering. To save time on clustering for future datasets, classifiers are also trained with the clustering output. Next, the clustering output is used to train a Hidden Markov Model (HMM), which is then used to predict behaviors for all frames. Finally, all frames in each state are combined to be analyzed.

to all videos, we newly described parenting behaviors that were consistently appearing in all frames within a given state. 12 behavioral assays were clustered, trained, and tested on.

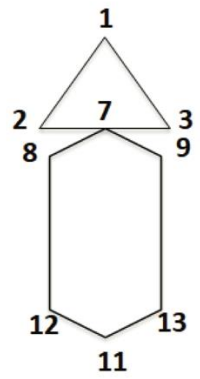
Calculating Features. Based on the corrected adult body points from Part 1, 14 features were calculated. Each feature's description is outlined in Fig. 8a. Each frame's features were calculated, and then the x-coordinates, y-coordinates, and calculated features were appended to separate lists. After each behavior assay was run, the x-coordinates, y-coordinates, and calculated features were combined and outputted into one numpy array. Additionally, a list of the total frames for each behavior assay was also created.

Checking Features. To check the features, we decided to use the ListedColormap and BoundaryNorm methods from the *matplotlib.colors* library to create a colormap. The colormap plotted a color based on the value in each frame. Since there were different types of features with widely ranging values, we split up the colormaps into three categories: angles, areas, and distances. Furthermore, to ensure that we could visually inspect the features on a similar scale, we normalized all points. Users were prompted for the test and time frame they wanted to check. If the resulting colormaps maintained a smooth gradient, then we knew that there were no massive jumps in the data, and we assumed that the features had been calculated and consolidated correctly. As an example, results for behavior assay #1 are shown in Fig. 8b.

Clustering. Using the k-means method from the *sklearn.cluster* library, we clustered the frames based on the calculated features. Since k-means requires an input of the optimal number of clusters, using the *sklearn.metrics* library, we first found the optimal number of clusters based on the Elbow, Silhouette and Davies-Bouldin methods (Fig. 9).

The Elbow method uses the sum of squared distance between each point and the centroid in a cluster as a metric (Within-Cluster Sum of Squares or WCSS) and plots WCSS as the number of

(a)



Use Points

Feature	Description
1	Angle 1, 7, 11
2	Angle 2, 7, 11
3	Angle 1, 2, 3
4	Angle 1, 3, 2
5	Distance 1, 7
6	Distance 7, 11
7	Distance 8, 9
8	Distance 12, 13
9	Distance 8, 12
10	Distance 9, 13
11	Distance 8, 13
12	Distance 9, 12
13	Area of Head
14	Area of Body

(b)

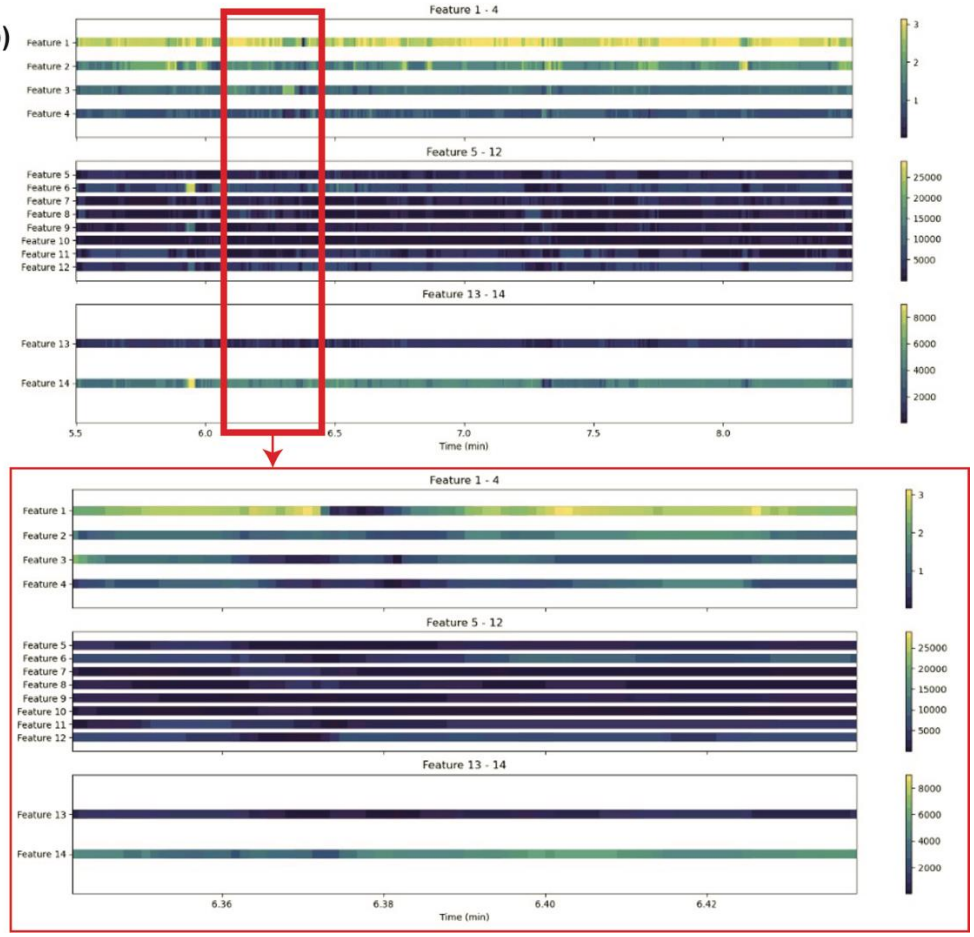


Fig. 8. Features from Part 3. (a) Descriptions of the 14 features collected of the adult mice. **Left**: the diagram of the tracked points on the adult are shown, labeled with numbers. **Right**: the descriptions of each feature are given with the numbers on the **left**. Notice that on the **right**, there are 4 angle features, 8 distance features, and 2 area features. **(b)** Checking feature consolidation colormaps for behavior assay #1. **Top:** the selected time frame that the user chose to examine. **Bottom:** a closer view of the **top** diagram to show the smooth gradient across all features, indicating that the features were calculated and consolidated correctly. Note that the feature numbers correspond to the those in **(a)**.

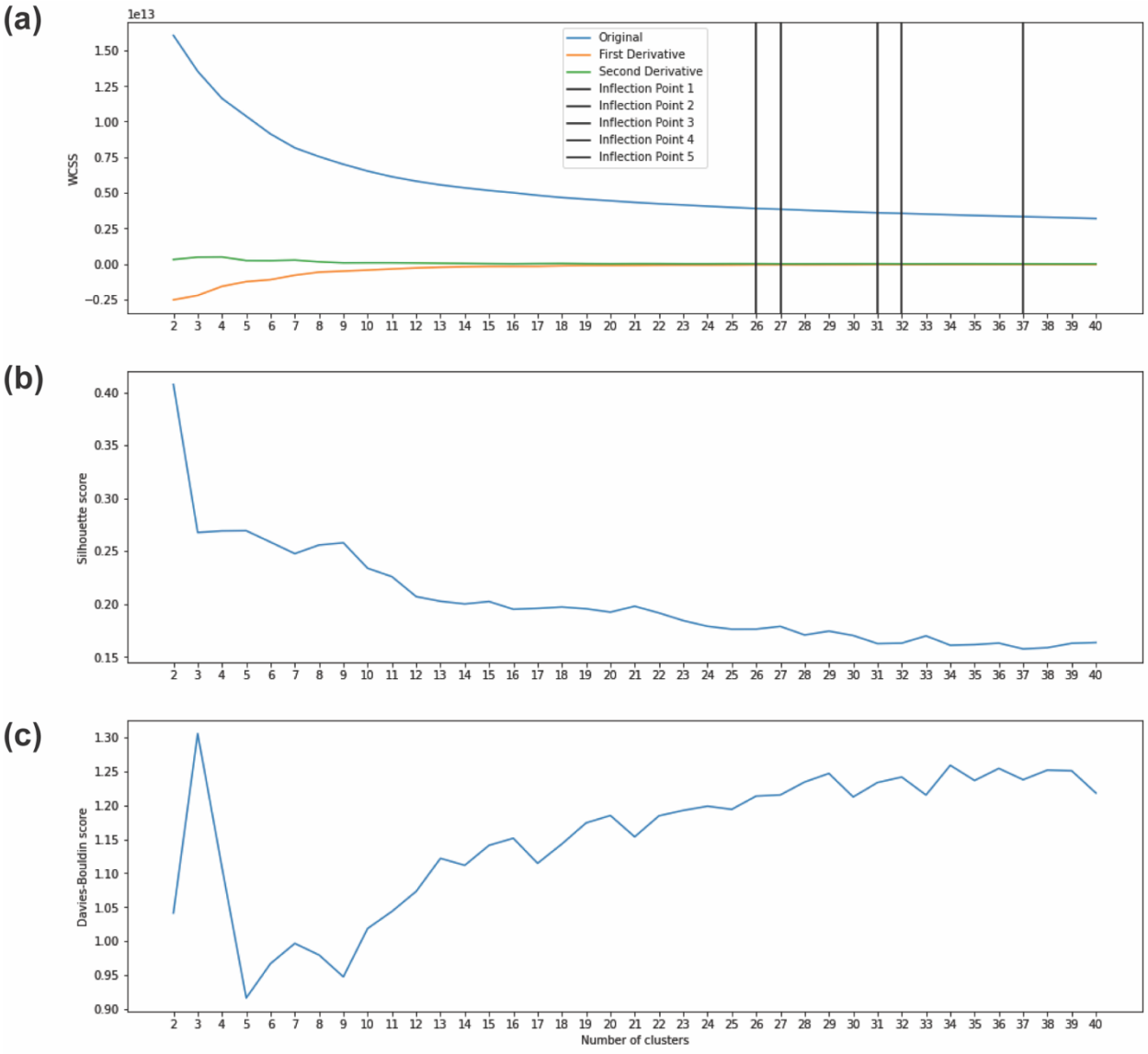


Fig. 9. Optimal Cluster Evaluations. The three optimal cluster evaluation methods: Elbow **(a)**, Silhouette **(b)**, and Davies-Bouldin **(c)** methods. In **(a)**, we can see that the second inflection point that often indicates the optimal number of clusters is around 27. In **(b)**, we can see that the slight peaks that often indicate the optimal number of clusters emerge at 9, 21, 27, and 33. Finally, in **(c)**, we see that the largest dips that indicate the optimal number of clusters emerge at 5 and 9. From this, we can conclude that cluster numbers of 9 or 27 are worth further pursuing.

clusters increase. As the number of clusters increase, the WCSS decreases. We took the optimal number of clusters to be the point at which the WCSS no longer decreases significantly to the visual eye. This result was confirmed by finding the second derivative of the WCSS curve, where around 27 clusters was considered optimal.

The Silhouette method is based on plotting the silhouette coefficient, which depends on the mean intra-cluster and nearest-cluster distance for each frame. As the number of clusters increase, the silhouette coefficient is expected to peak at the optimal number of clusters. Although there was not a prominent peak, local peaks were found at 9, 21, 27, and 33 clusters.

The Davies-Bouldin method plots the Davies-Bouldin score, which measures the ratio of within-cluster distances to between-cluster distances. As the number of clusters increase, the Davies-Bouldin score is expected to dip at the optimal number of clusters. Prominent dips were found at clusters 5 and 9.

From these results, it appeared that 9 or 27 clusters were promising. However, we ultimately chose 27 clusters based on two reasons: the Elbow method is still the most widely accepted method, and 9 clusters may overlook intricacies in the data. After k-means was trained on the 27 clusters, the clustering predictions for each frame was saved as a numpy array of size (1, number of frames). In addition, using the *umap.umap* library, I plotted and saved the clustered features on a 2-dimnesional space all together and separately (Fig. 10).

Classifiers. After clustering the frames, we wanted to form a template, so that we can cluster frames from new videos faster and more easily. Therefore, we trained a classifier to assign frames to clusters from new videos based on the same set of features. Two classifiers, SVC and RandomForest Classifier, were created using the *sklearn.svc, sklearn.ensemble, sklearn.pipeline,*

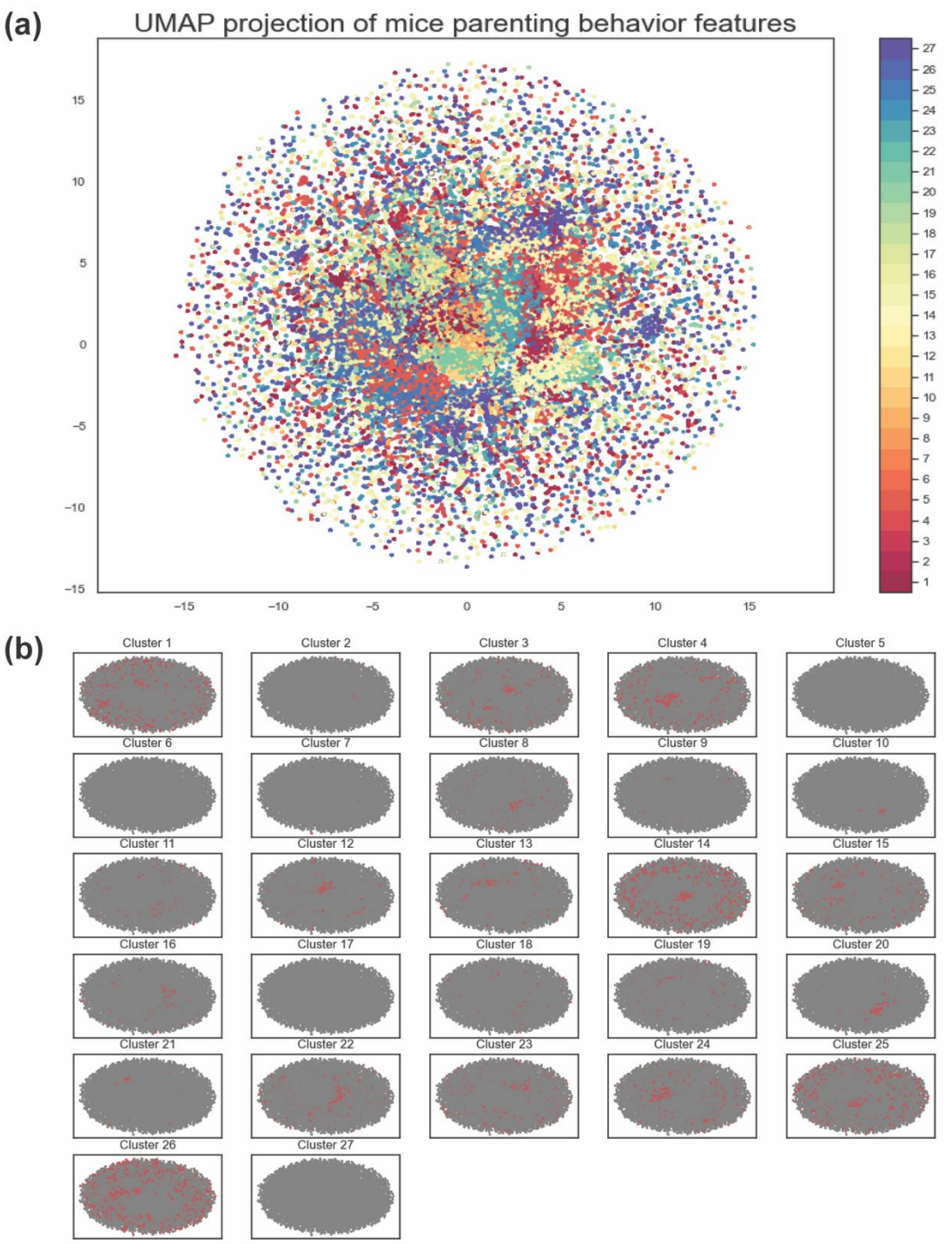


Fig. 10. The resulting plots of the k-means clustering for 27 clusters. Since there are 14 features, there are 14 dimensions needed to represent the data. However, for ease of comprehension and readability, the features have been reduced to 2-dimensions in **(a)** and **(b)**. **(a)** shows the conglomerated clusters, with the colormap on the right indicating which color corresponds to each feature. **(b)** shows the individual clusters in **(a)**, where red represents the selected cluster and gray represents all other feature clusters.

and *sklearn.preprocessing* libraries. Using the test_train_split method from the *sklearn.model_selection* library, frames were split into training and testing sets using a division of 80 percent train and 20 percent test. The accuracy for each training iteration is shown in Fig. 11. After being trained, they were saved so that users can choose to use them to predict clusters for frames next time instead of re-running k-means.

Hidden Markov Models (HMM). Using the *hmmlearn* library and given the number of clusters and states, an HMM was created. The start probability of being in any state was initially set to 1/(number of states). In addition, the transition probability between states was initially set to 0.5 that the behavior assay stays in the current state and 0.5/(number of states – 1) that it goes to another state. Finally, the emission probability of a state predicting any cluster was initially set to 1/(number of clusters).

The timeseries containing the cluster identity of each frame were used to train the model. Given the unsupervised nature of our approach, we selected all videos to be used in the training set for the HMM. This ensured that the HMM could be trained on the most information available. However, users have the flexibility to choose how many and which behavior assays they would like to include for their own training sets. The trained HMM was then returned to the user and saved using the *pickle* library.

To test the HMM, we used the same inputs for training the HMM and predict method from the *hmmlearn* library to predict states for the test set. In our case, since we wanted to see what the HMM would output for each video and interpret the results with the most data available, we chose to include every video in our test set. However, future users can select how many and which videos to include in their test sets. After the states were predicted, they are saved along with the videos tested and lengths of each video in a *pickle* dictionary.

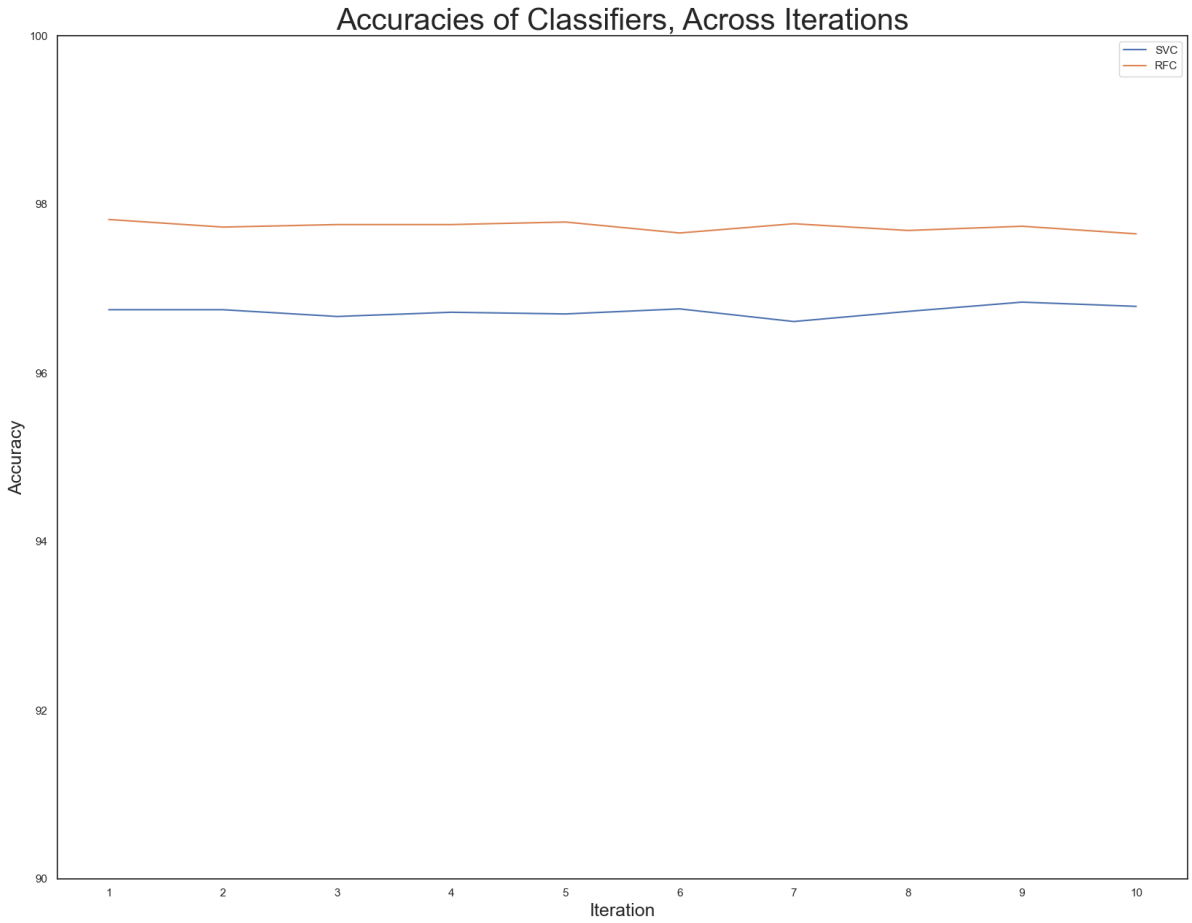


Fig. 11. The Accuracy of the Classifiers. Two classifiers were trained: SVC and Random Forest Classifier. Each classifier was trained for 10 iterations, and the accuracy of each iteration is displayed here. Each iteration had the data randomly ordered, of which 80 percent was train and 20 percent was test. For each classifier, the iteration with the highest accuracy was saved.

Finding Optimal Number of States for HMM. Using the outputted clusters ("Clustering" method), different HMMs were created with the number of states looping from 2 to the optimal number of clusters (found in "Clustering" method). For each loop, we found how many states contained 95 percent of the inputted clusters, given that most scientific research allows for around 5 percent of error. The number of states encompassing 95 percent of the cluster data was then plotted as number of states increased, ultimately plateauing around 14 states. Results are shown in Fig. 12.

Checking HMM Results. Using the predicted states for the assay and the *cv2* library, we combined frames from the original assay video into state specific videos (e.g. all frames predicted to be state 1 would be combined into one video). We then visually examined each state specific video across multiple behavior assays to determine common adult postural features found within the state (e.g. position of snout/head/body, curvature, direction). Using these postural features, we determined the newly described parenting behaviors centered around individual adult postures. Finally, we plotted ethograms and saved csv files for the specific states as we did for behaviors in Part 1 ("Checking Quantified Previously Described Behaviors" method).

Part 4 – Neuronal Correlations

In this part, we wanted to test the potential of our framework for understanding neuronal correlates of parenting behaviors. We achieved this through two main methods: (1) finding the difference between the activities of MPOA neurons when a select behavior/state was happening and when it was not and (2) correlating the activities of MPOA neurons with outputted parenting behaviors/states.

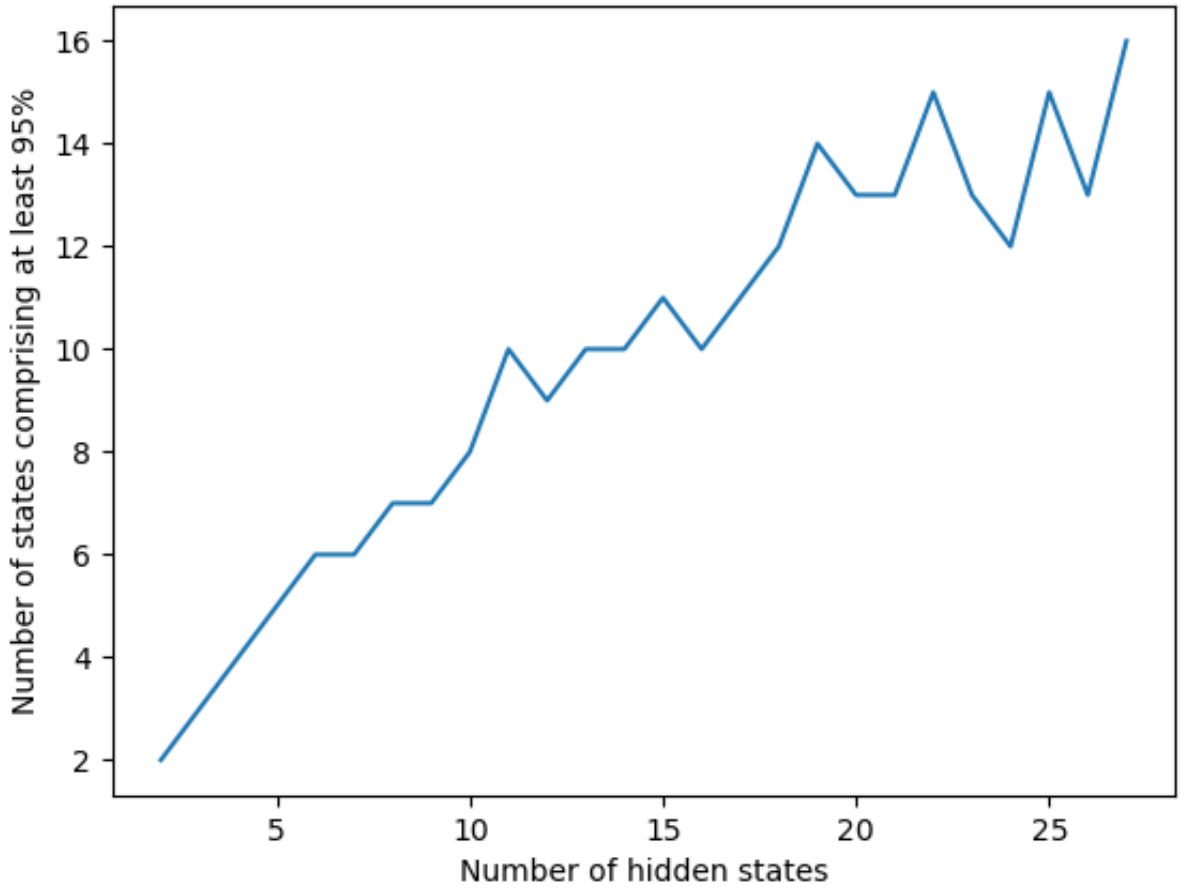


Fig. 12. Finding HMM Optimal Hidden States. The graph used to find the optimal number of hidden states across all behavior assays. Each iteration, the number of hidden states was increased in an HMM and trained using all behavior assays. Ultimately, as the number of states increase, more states are expected to be empty, as there is an optimal number of hidden states that can represent the data. Any more states than this optimal will not help, as the data has already been grouped into its distinct states. On the graph, this is shown by where the data appears to plateau. Accounting for noise, for 27 clusters, the optimal number of hidden states seems to be around 14 states.

Average Neuron Activity Difference. For each video in each part, we found the average difference of MPOA neuronal activities between when an outputted behavior or state was happening and when it was not. We then averaged the average differences for each behavior or state across all videos. ANOVA single factor and post hoc Tukey tests were then conducted to determine the significance of our conglomerated averages. This analysis was done only in mothers.

Correlations. For each video in each part, we also ran correlations between the activities of the MPOA neurons with the outputted behaviors/states. Since the outputted states contained a binary output that was not helpful against continuous correlation measurements, we decided to find the average of behavior and neuronal activity within a given time bin of 1 second. These time bins ensured that both the outputted behaviors/states and neuronal activities were not binary values and could be correlated. Next Pearson's correlation was calculated for each neuron against each behavior. The Pearson's r correlation coefficient was averaged across neurons for each behavior. We then further averaged the correlation coefficient across all videos and used ANOVA single factor and post hoc Tukey tests to determine the significance of the conglomerated averages.

Results

Part 1 Quantified Previously Described Parenting Behaviors

From the criteria listed in Table 2b, we were able to develop a user-defined system that quantified 12 previously described parenting behaviors. However, due to consistent overlaps between behaviors, we simplified this and focused on 6 previously described parenting behaviors, with an additional seventh category of "Other" that included all frames with the remaining 6 previously described parenting behaviors ("Adult in nest", "Pup in Nest", "Pup Entering Nest", "Adult with Pup", "Adult in nest with pup", and "Adult entering nest with pup")

or had no quantified behaviors. The 6 chosen previously described parenting behaviors are "Retrieving the pup", "Investigating the pup", "Approaching the pup", "Sitting with the pup", "Adult exiting nest", and "Adult entering nest". An example ethogram for these behaviors can be found in Fig. 13a. Furthermore, selected behaviors are also visualized in Fig. 13b-c.

Previously Described Parenting Behaviors Encompass a Minority of Parenting Assays

From the ethogram in Fig. 5, we observed that many of the animals spent most of their time in the "Other" category rather than the 6 previously described parenting behaviors. We confirmed this finding by plotting the distribution of time spent in the different behaviors (Fig. 14), where we found that across mothers, virgin females, and fathers, the time spent in the "Other" category was approximately 80 percent. This high percentage indicates that there may be a need for additional parenting behaviors to be quantified that are not currently described in research, since the "Other" category included many frames with behaviors that could not be quantified by the 6 previously described parenting behaviors. Furthermore, the difference between the bars were significant (Z-scores: Mothers (m) to virgin females (vf): -12.0, vf to fathers (f): 14.9, f to m: -6.62; all p-values < 0.05), where virgin females spent the most time in "Other" and fathers the least. Although it is consistent with prior research that virgin females spend less time performing the previously described parenting behaviors, these results indicate that there may be more mother-specific behaviors that are not currently quantified by existing research. In fact, current behavior quantifications might optimize for more father specific parenting behaviors. To come to a more solid conclusion the analysis should include more replicates of the assays. However, this may further reinforce the need to describe and quantify newly described parenting behaviors in mice.

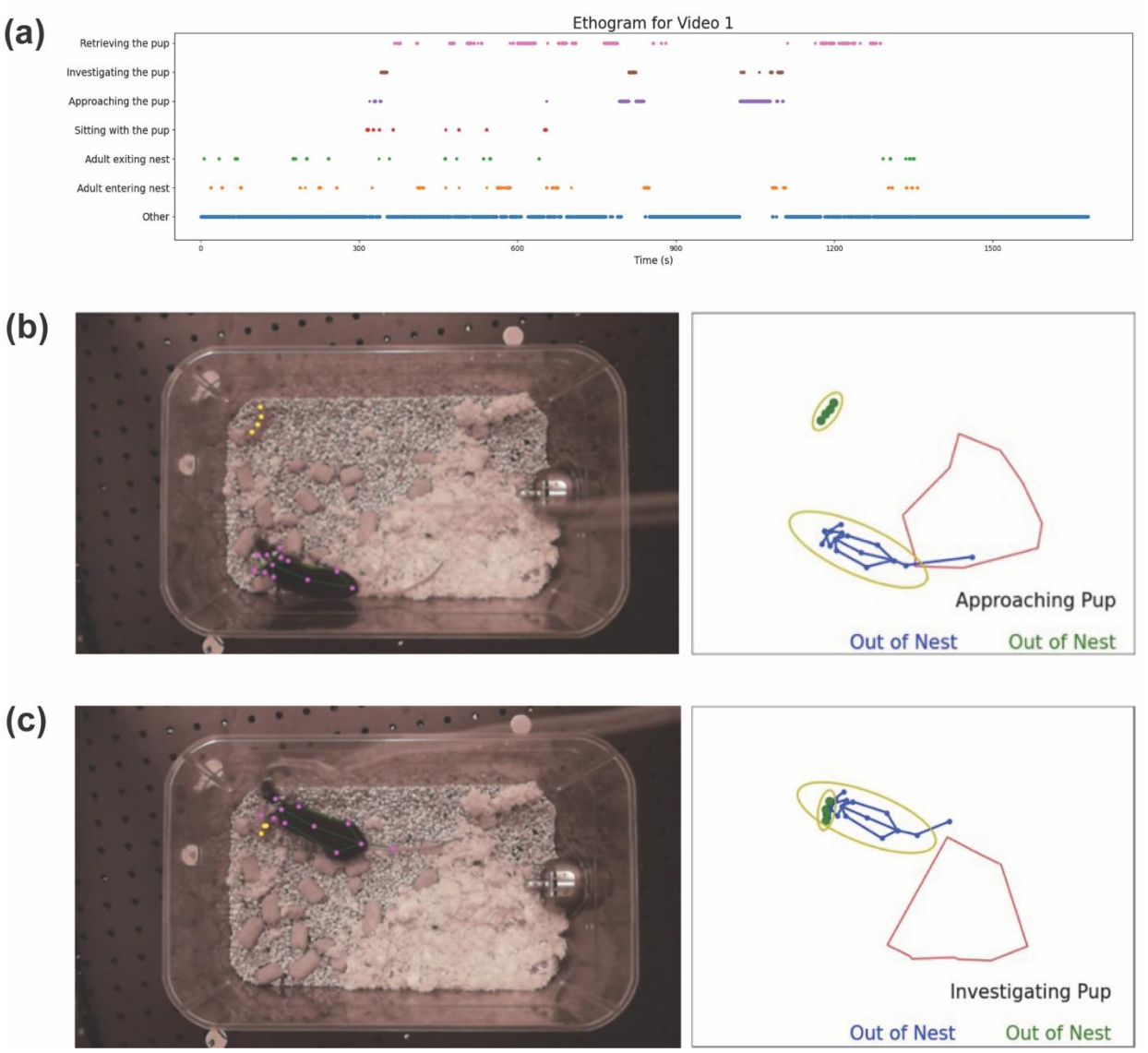


Fig. 13. Ethograms and Visualization for 6 Previously Described Parenting Behaviors. In **(a)**, a mother's parenting assay ethogram is shown. From the ethogram, it is visually clear that the behaviors logically follow each other, where animals enter before exiting the nest and approach before retrieving the pup. Ethograms were also made for virgin females and fathers that followed the same format (not displayed). **(b-c)** visualize two of the six previously described parenting behaviors, where the **left** displays the video frame and the **right** shows the corresponding behavior.

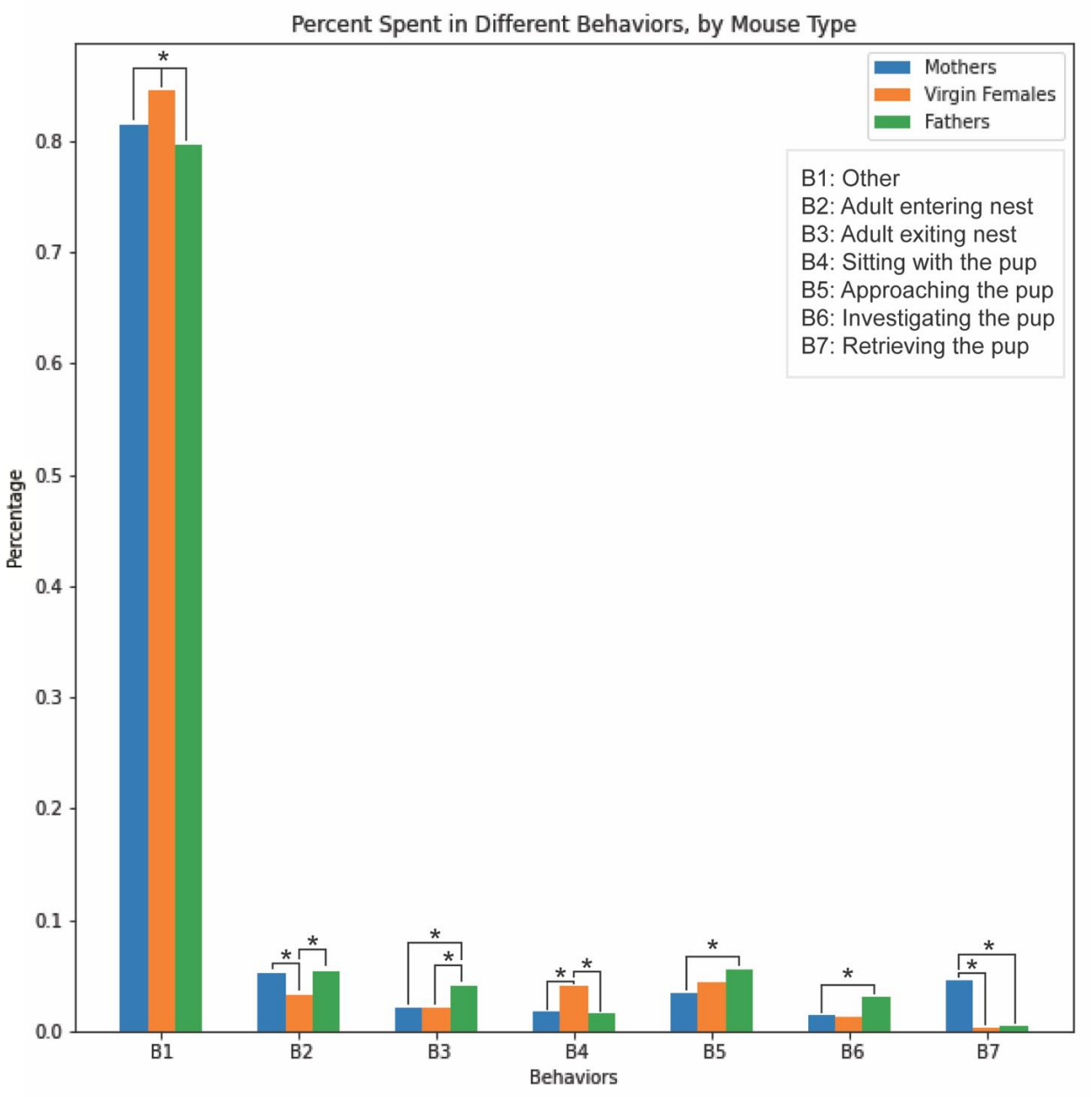


Fig. 14. Distribution of Time Spent Doing Previously Described Parenting Behaviors Across Mouse Types. We plotted the percentage of frames that our user-defined system denoted doing a previously described parenting behavior or "Other" for each mouse type. The key for the x-axis behaviors is shown in the graph. All * denote $p < 0.05$.

Father and Virgin Female Behaviors Consistent with Behavior Assay Observations

Based on Fig. 14, we also saw that the 6 previously described quantified behaviors consistently depicted differences across mouse types that corresponded with behavior assay observations. For example, we saw virgin females significantly entered the nest (B2) less than mothers and fathers (Z-scores: m to vf: 2.76, vf to f: -2.75; all p-values < 0.05). This was consistent with observations where virgin females tended to parent less and explore more outside the nest. In addition, for fathers, we saw that they significantly approached (B5) (Z-scores: f to m: 4.13; p-value < 0.05) and investigated (B6) (Z-scores: f to m: 3.06; p-value < 0.05) for longer time periods compared to mothers. This was consistent with observations during the behavior assays, which may suggest that fathers have a different sampling process before retrieving the pup when compared to mothers. Finally, we noticed that fathers significantly exited the nest (B3) more than other mouse types (Z-scores: vf to f: -2.41, f to m: 3.82; all p-values < 0.05). A possible explanation for this is that fathers tend to sit on the edge of the nest more. Given that our user-defined system used a criteria of overlapping percentage change between the animal and nest to determine B3, we hypothesize that the tendency of fathers often sitting near the edge of the nest could have resulted in the higher occurrence of B3 in fathers. This provides impetus to incorporate where the animal spends its time in the nest into description of mouse parenting behaviors.

Importance of Continuous Retrieval Processes

An interesting difference was observed in retrieving the pup (B7), where mothers spent significantly more time retrieving the pup compared to fathers and virgin females (Z-scores: m to vf: 2.20, f to m: -2.41; all p-values < 0.05) (Fig. 14). We hypothesize two possible causes for this

result. First, it is possible that mothers are gentler with their retrieval, which would make the overall process slower. Second, we also noted that our user-defined system relied on a continuous percentage change occurring in the distance between the pup and the nest to determine if B7 was happening. Given this result, we observed that in most behavior assays fathers and virgin females had many gaps in their retrieval process whereas mothers often had fewer but longer continuous retrieval processes. This suggests an importance of the length of retrieval that could be incorporated into current and future mouse parenting behaviors, since the last retrieval bout whereby an animal finally brings the pup into the nest could be processed differently across mouse types.

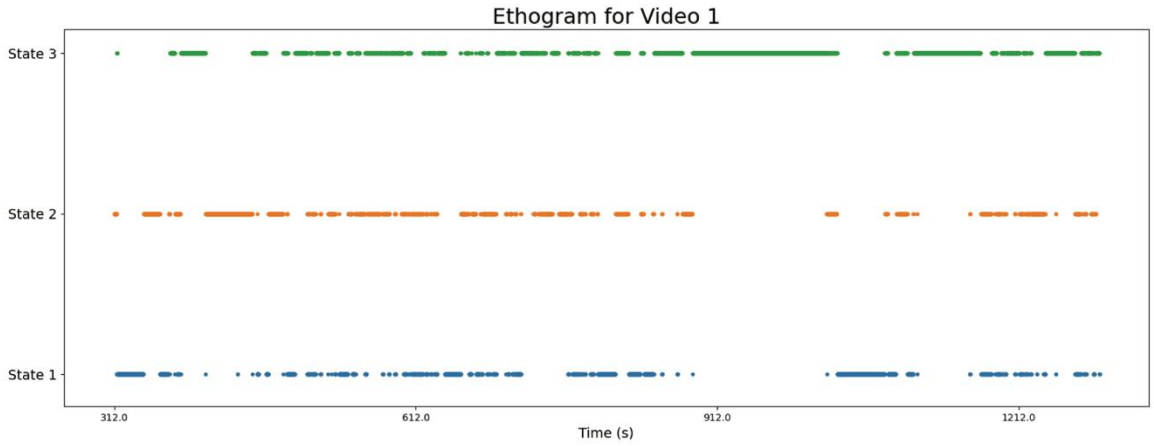
Part 2 Quantified 3 Hidden Adult-Pup Parenting Behavioral States

Using the 8 parameters from Table 2a (excluded parameter 3 due to redundancy with parameter 4) and the 7 previously described behaviors in Fig. 14, the GLM-HMM outputted three hidden states. Since the inputted data only encompassed the time when the pup was in the nest, we believe these three hidden states represent three adult-pup parenting behavioral states. A sample ethogram for these three behaviors is displayed in Fig. 15a. We have also averaged the features in each state to provide a more quantified description of each state (Fig. 15b).

Hidden State 1 Likely Represents a Non-Parenting State, and Hidden State 3 Likely Represents a Parenting State

Similar to the results of Part 1, we plotted the distribution of time spent in each state across mouse types (Fig. 16). We found that mothers spent significantly less time in hidden state 1 compared to virgin females and fathers (Z-scores: m to vf: -26.1, f to m: 25.0; all p-values <

(a)



(b)

	Feature	Hidden State 1	Hidden State 2	Hidden State 3
1	Nest polygon area *(std: +/- 0.0355)*	0.89	1.00	1.04
2	Adult polygon area *(std: +/- 0.0317)*	1.08	0.96	0.98
3	Percent overlap of adult and nest areas *(std: +/- 15.690)*	33.56	85.94	95.32
4	Percent overlap of pup with nest *(std: +/- 11.844)*	46.23	65.02	95.99
5	Distance between adult and pup *(std: +/- 30.435)*	139.72	91.47	11.87
6	Distance between pup and nest shapes *(std: +/- 24.670)*	109.66	67.50	5.51
7	Angle of adult polygon *(std: +/- 11.984)*	10.58	-39.76	8.43
8	Angle of pup polygon *(std: +/- 15.474)*	-24.85	27.57	-32.87

Fig. 15. Ethograms and Descriptions of Hidden States from Part 2. (a) shows the ethogram of the 3 hidden states plotted for a mother (virgin female/father ethograms not displayed). **(b)** is a table showing the quantified average features across all frames in each hidden state. Features 1 and 2 were normalized to their means. Features 7 and 8 were taken relative to the horizontal. These averages can be thought of as non-descriptive quantifications of the hidden states.

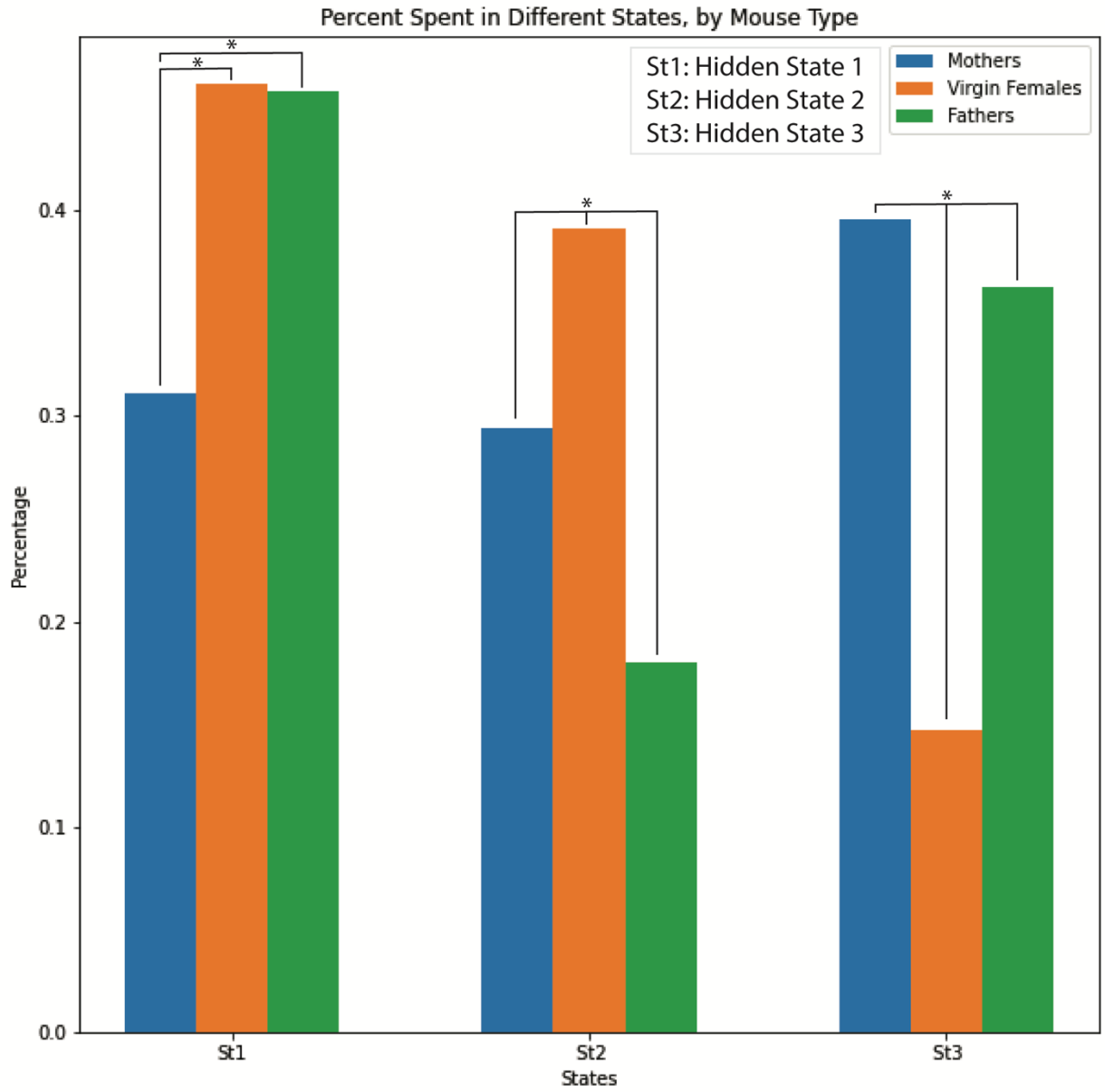


Fig. 16. Distribution of Time Spent in Hidden States Across Mouse Types. We plotted the percentage of frames that our GLM-HMM denoted an animal being in a certain hidden state. The key for the x-axis hidden states is shown in the graph. All * denote $p < 0.05$.

0.05), which suggested that hidden state 1 represented some form of a non-parenting state. Additionally, we observed the opposite for hidden state 3, where mothers and fathers spent significantly more time than virgin females in this state (Z-scores: m to vf: 27.2, vf to f: -21.9, f to m: -5.14; all p-values < 0.05). This indicates that hidden state 3 could be a parenting state. For hidden state 2, we observed that virgin females significantly spent more time in this state than mothers (Z-scores: m to vf: 27.2, vf to f: -21.9; all p-values < 0.05), who spent significantly more time in this state than fathers (Z-scores: f to m: -5.14; p-value < 0.05). Although these results hint that there could be sex specific differences in parenting behavior displayed in hidden state 2, further research should be conducted to draw more conclusive results.

Part 3 Quantified 11 Newly Described Adult Posture Parenting Behaviors

Based on the 14 features in Table 3a, our HMM outputted 14 hidden states that could each represent a potential newly described adult posture related to parenting behavior. In fact, the 14 features were averaged across all frames in each state to provide a more quantified description of each state (Table 3b). After observing the videos, we formed descriptions for each hidden state and ended up with 11 newly described behaviors (Fig. 17a). Specifically, these behaviors were named after consistent postures found in each state: “Hunched, with snout down movement”, “forward jerky movement”, “forward locomotion”, “alternating hunched/stretched”, “curved movement”, “stretched investigation”, “forward movement ending in snout up”, “stationary, with enhanced snout movement”, “stretch and contract”, “curved stationary”, and “stationary, with idle head”. 3 of the hidden states were deemed irrelevant for having too few frames. An example ethogram for these 14 hidden states is shown in Fig. 17b.

(a)

Feature	Measurement
F1	Angle between snout, head base, and tail base
F2	Angle between left head point, head base, and tail base
F3	Angle between snout, left head point, and right head point
F4	Angle between snout, right head point, and left head point
F5	Head length (distance between snout and head base)
F6	Body length (distance between head base and tail base)
F7	Distance between shoulders
F8	Distance between hind legs
F9	Left side body length (distance between left shoulder and left hind leg)
F10	Right side body length (distance between right shoulder and right hind leg)
F11	Left side cross-body length (distance between left shoulder and right hind leg)
F12	Right side cross-body length (distance between right shoulder left hind leg)
F13	Area of head
F14	Area of body

(b)

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
F1 *(+/- 0.03)*	2.65	2.58	2.18	2.45	2.49	2.65	2.62	2.46	2.56	2.51	2.61	2.64	2.69	2.59
F2 *(+/- 0.03)*	1.77	1.96	1.96	1.66	2.00	1.84	1.71	1.67	1.67	1.62	1.88	1.74	1.82	1.65
F3 *(+/- 0.01)*	1.16	1.07	1.05	1.19	1.05	1.05	1.15	1.15	1.13	1.14	1.12	1.15	1.09	1.18
F4 (+/- 0.02)	0.81	0.97	1.05	0.92	0.98	1.03	0.96	1.00	0.94	0.95	1.00	0.92	0.87	0.91
F5 *(+/- 175.63)*	913	1450	1080	3320	1270	888	1790	1590	1040	1730	1750	1850	670	2260
F6 *(+/- 1183.17)*	1220	2030	5020	15700	9350	848	2560	3740	2620	10300	7600	8280	827	10700
F7 *(+/- 288.048)*	4580	1770	2630	581	1290	3270	2540	1850	2300	1020	1740	1210	2730	439
F8 *(+/- 431.76)*	6370	2700	3160	686	996	3920	3420	2040	3110	721	1620	1300	4080	445
F9 *(+/- 353.21)*	179	1220	1650	4720	2930	258	619	1340	911	2760	2570	2280	345	3540
F10 *(+/- 596.97)*	1460	1450	2480	881	521	1430	1850	1710	9210	730	1050	1020	4680	398
F11 *(+/- 616.19)*	1110	5670	1580	5040	1070	2310	861	4200	9410	3840	2530	1570	5270	4080
F12 *(+/- 624.17)*	2610	892	6862	2170	5670	1300	6040	1160	8150	3600	4020	7620	4680	5130
F13 *(+/- 23.95)*	1010	979	762	978	790	948	1040	919	850	979	1050	974	856	1040
F14 *(+/- 215.49)*	4640	3870	4980	5850	4740	3600	4710	3540	6570	4980	4520	5300	5540	4940

Table 3. Part 3's Hidden States Descriptions. In **(a)**, the 14 features and their corresponding measurements are listed. In **(b)**, the hidden states (S1, S2, etc.) are quantitatively described by the 14 features in **(a)**, averaged across all frames in each hidden state.

(a)

Hidden State	Description
1	Hunched, with snout down movement
2	Forward jerky movement
3	Irrelevant
4	Forward Locomotion
5	Irrelevant
6	Alternating hunched/stretched
7	Curved movement
8	Irrelevant
9	Stretched Investigation
10	Forward movement ending in snout up
11	Stationary, with enhanced snout movement
12	Stretch and contract
13	Curved stationary
14	Stationary, with idle head

(b)

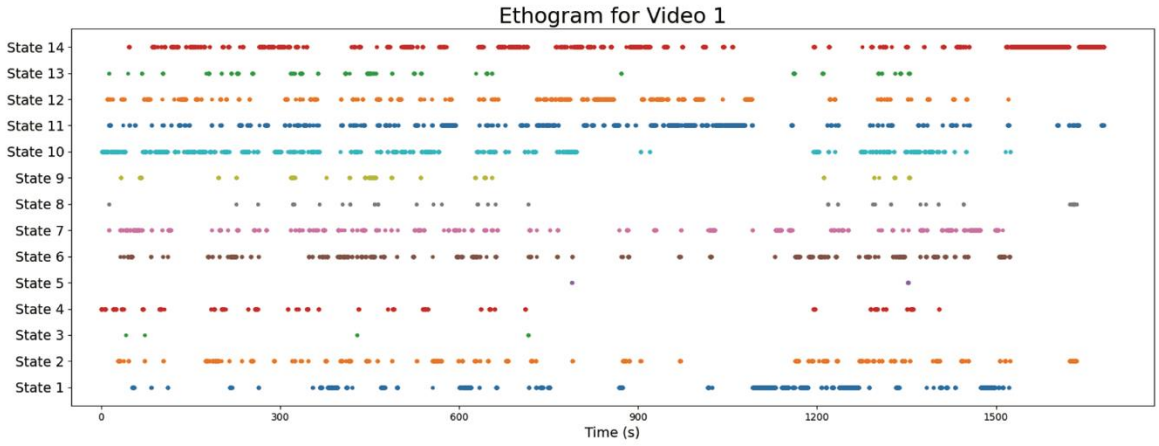


Fig. 17. Descriptions and Ethograms of Hidden States from Part 3. In **(a)**, the 14 hidden states are given qualitative names, based on the 14 features from Table 3a observed over our behavior assays. In **(b)**, the hidden state ethogram for a mother is shown (virgin female/father ethograms not displayed). The states on the y-axis follow the order described in **(a)**.

Hunched, Stationary, and Head Idle Postures Indicative of Parenting Behavior

Following similar results to Part 1 and 2, we decided to plot a distribution of the time spent in each hidden state across mouse types (Fig. 18). We noticed that mothers and fathers spent significantly more time in both "Hunched, with snout down movement" (St1) (Z-scores: m to vf: 2.58, vf to f: -10.6, f to m: 17.6; all p-values < 0.05) and "Stationary, with head idle" (St14) (Z-scores: m to vf: 10.4, vf to f: -6.88, f to m: -4.06; all p-values < 0.05) states. In checking the video frames where these states occurred, we hypothesize that these postures could indicate parenting behavior in mice. When the animal is hunched, it is often a mother or father in the nest nudging, grooming, or sitting on top of the pup. Moreover, we noticed that the head is more idle in mothers and fathers when they sit with the pup, which allows them to spend more time in the Stationary, with head idle" states. This was heavily contrasted with a result from our previously described behaviors, where we found that virgin females spent more time "Sitting with the Pup" (B4) than mothers and fathers (Z-scores: m to vf: -4.29, vf to f: 2.73; all p-values < 0.05). Since our user-defined system criteria for "Sitting with the Pup" was rather strict, where the behavior only occurred if the adult polygon completely encompassed the pup, we must acknowledge it is possible that the system simply missed out on instances where the mothers and fathers were supposedly sitting with the pup but did not exactly fulfill the criteria. However, an alternative explanation is that although virgin females are sitting with the pup, they are not actively focused on the pup. This hypothesis is reinforced by the observation that virgin females spend significantly more time in the "Forward movement ending in snout up" (St10) state (Z-scores: m to vf: -16.5, vf to f: 13.1; all p-values < 0.05) (Fig. 14), which often indicates exploratory movement. In fact, we observed that virgin females tended to rear and explore outside the nest more in the videos, resulting in them often having stretched and enhanced head movement

postures. Given these hypotheses, further research should be done to confirm the importance of hunched or steady head postures during newly described mouse parenting behaviors.

Contracting Movement Associated with Nest Building Parenting Behavior

When reviewing our videos, we noticed that there were two hidden states that often occurred during nest building phases: "Forward jerky movement" (St2) and "Stretch and contract" (St12). Fathers and mothers spent significantly more time in the "Forward jerky movement" state than virgin females (Z-scores: m to vf: 6.32, vf to f: -7.44, f to m: 2.99; all p-values < 0.05) (Fig. 18). In contrast, fathers and virgin females spent more time in the "Stretch and contract" state than mothers (Z-scores: m to vf: -18.3, vf to f: 8.4, f to m: 5.98; all p-values < 0.05) (Fig. 18). We hypothesized that these observations could be due to two types of nest building during parenting behavior. First, in the "Forward jerky movement" state, mothers and fathers are often building a nest for the pup, resulting in forward lunges to push nest material into a shape that is suitable for the pup. However, in the second "Stretch and contract" state, fathers and virgin females were observed both pushing and pulling nest material, which was interpreted as reshaping the nest material around the pup in the nest. This interpretation is consistent with previous behavior assay observations that mothers often build the best initial nests for pups, prior to the pup entering the nest. This results in mothers adjusting the nest much less than fathers and virgin females. Given these hypotheses, further research should be conducted to fully understand the different contracting movements in mice and how they could be integrated into quantifying newly described or better defining previously described mouse parenting behaviors.

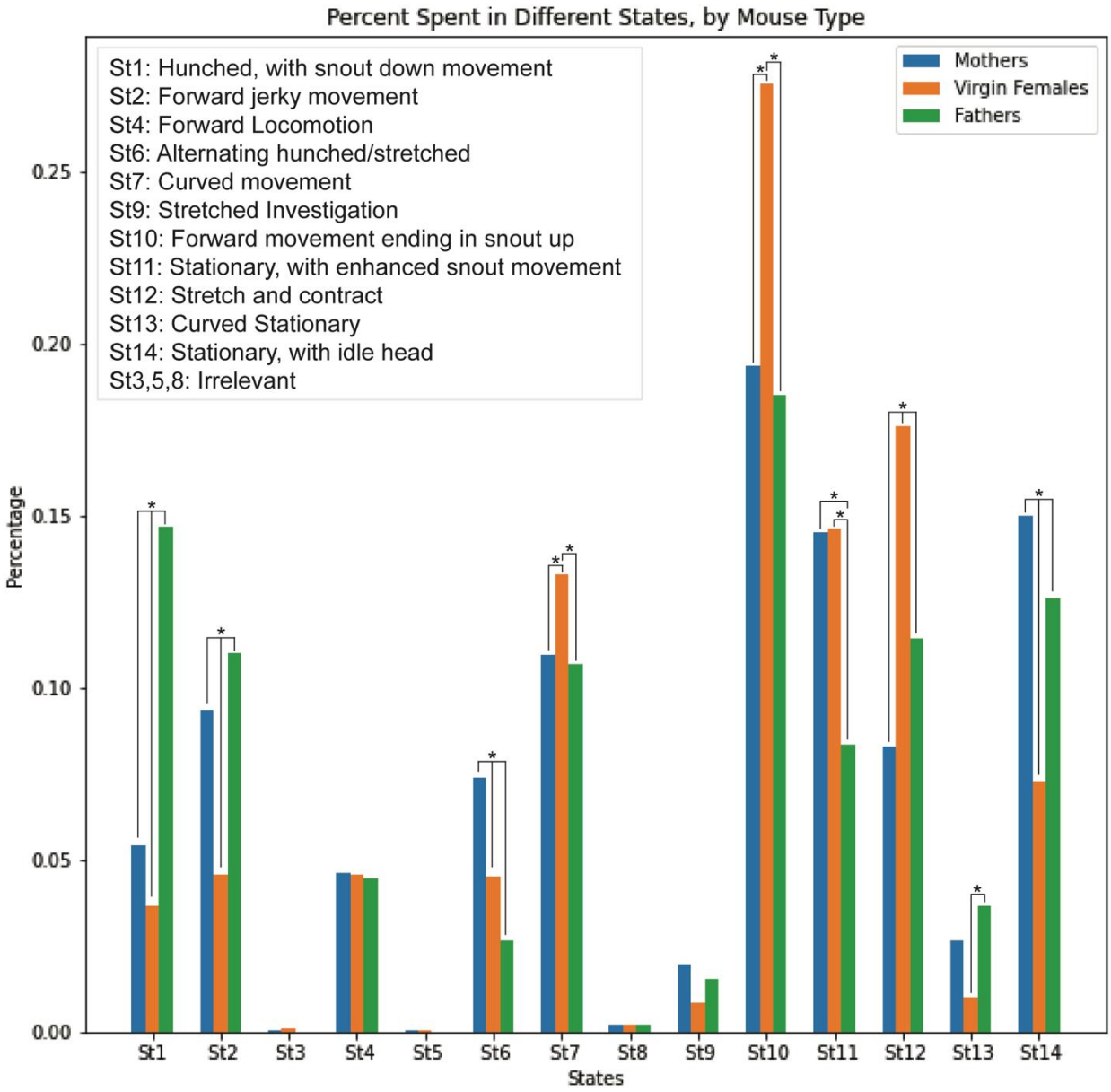


Fig. 18. Distribution of Time Spent in Hidden States Across Mouse Types. We plotted the percentage of frames that our HMM denoted an animal being in a certain hidden state. The key for the x-axis hidden states is shown in the graph. All * denote significance, where $p < 0.05$.

Body Position and Stationary Head Movement Reveal Potential Sex Specific Parenting Differences

There were two states that were significantly longer in duration for mothers and virgin females than fathers: "Alternating hunched/stretched" (St6) (Z-scores: m to vf: 4.11, vf to f: 2.38, f to m: -5.43; all p-values < 0.05) and "Stationary, with enhanced snout movement" (St11) (Z-scores: vf to f: 8.19, f to m: 5.98; all p-values < 0.05) (Fig. 18). After checking with our videos, the "Alternating hunched/stretched" state still seemed too ambiguous to concretely define its connection to female parenting patterns. However, it may be interesting to examine further the sex differences between how fast and often animals change body postures during parenting to determine if these differences play a role in newly described mouse parenting behavior. The "Stationary, with enhanced snout movement" state was more clearly associated with grooming and investigating, two activities that females tend to naturally do more frequently while parenting. Specifically, it has been observed that females tend to move their head more often than males and will often investigate at much closer ranges than males. Although grooming and investigation are previously described parenting behaviors, further research should be conducted to properly understand the role that head movement plays in parenting. If there is a sex difference in how females tend to care for pups, head movement may be a key posture to observe in future parenting assays to better understand the sex-specific circuits underlying newly described mouse parenting behavior.

MPOA Neurons Likely Encode "Approaching/Investigating the Pup" Behaviors

We found that the ANOVA single factor analysis of our data for Part 1 was significant (p-value: 0.00013). In running the post hoc Tukey tests, we found that "Approaching the Pup" (B5) and "Investigating the Pup" (B6) were significantly higher than the "Other" (B1) behavior (B5 vs. B1

p-value < 0.05, B6 vs. B1 p-value < 0.01) (Fig. 19a). Given that our "Other" behavior encompasses mainly non-parenting behaviors, we can take it as a form of control. Then, the significantly higher activity levels for B5 and B6 are indicative that MPOA neurons strongly encode these behaviors. Additional support is also provided for "Investigating the Pup", given that the activity of MPOA neurons was significantly higher for B6 when compared to any other behavior excluding B5 (B2 vs. B6 p-value < 0.06, B3 vs. B6 p-value < 0.01, B4 vs. B6 p-value < 0.05, B6 vs. B7 p-value < 0.01) (Fig. 19a).

MPOA Neurons do not Encode the Hidden Parenting Behavioral States

From our ANOVA test, we found that there was no significance in the activity differences of the MPOA neurons for any of the hidden parenting behavioral states (p-value: 0.219) (Fig. 19b). Thus, we believe that the MPOA neurons do not encode these specific hidden parenting behavioral states, and perhaps future research should explore other parenting related neurons that could govern such hidden parenting behavioral states.

No Significant Relationship between MPOA Neurons and Postural Parenting Behaviors

Given the irrelevant natures of state 3, 5, and 8, we removed them for our ANOVA and post hoc tests. Although the ANOVA test was significant (p: 0.413), there were no significant differences between the states found in our post hoc Tukey analysis (Fig. 19c). A result that was borderline significant was that state 9 "Stretched Investigation" had a higher difference in MPOA neuronal activity than state 1 "Hunched, with snout down movement" (p-value < 0.1) (Fig. 19c). This was intriguing given that "Stretched Investigation" is consistent with our previously described behavior results, where the "Investigating the Pup" behavior had a higher difference in MPOA

neuronal activity than all other previously described behaviors. The fact that multiple investigation-based behaviors resulted in similar activity differences in the MPOA neurons suggests that there may be a strong encoding of this parenting behavior in the MPOA neurons.

MPOA Neurons were Not Highly Correlated with Previously Described Behaviors

We saw that there were very low correlations across all previously described behaviors (Fig. 20). This could be because although these neurons encode parenting behaviors, the timing at the resolution we investigated (~1 second) might not be optimal. However, from our ANOVA and post hoc Tukey tests, we observed that there was significantly less correlation for the "Other" (B1) behaviors when compared to all other previously described behaviors in Part 1 (all p-values < 0.01) (Fig. 20a). Given that the MPOA neurons were associated with previously described parenting behaviors (Kohl et al., 2018), this result aligns well with previous observations. In addition, we observed that there was a higher correlation for "Investigating the Pup" (B6) when compared to "Adult exiting the nest" (B3) (p-value < 0.05), which again highlights the link between the MPOA neurons and investigation-related parenting behaviors.

Hidden States and Postural Behaviors had Unclear Correlations with MPOA Neurons

Overall, the correlations for Part 2's hidden parenting behavioral states and Part 3's postural parenting behaviors were rather low (Fig. 20b-c) again pointing to different time scales to look at. There was significant difference between the positive correlation for hidden state 1 and negative correlation for hidden state 3 (p-value < 0.01), which reinforces the idea that these hidden states are distinctly different from each other. For Part 3's postural states, we did remove states 3, 5, and 8 in Part 3 for their irrelevancy before conducting any ANOVA or post hoc tests.

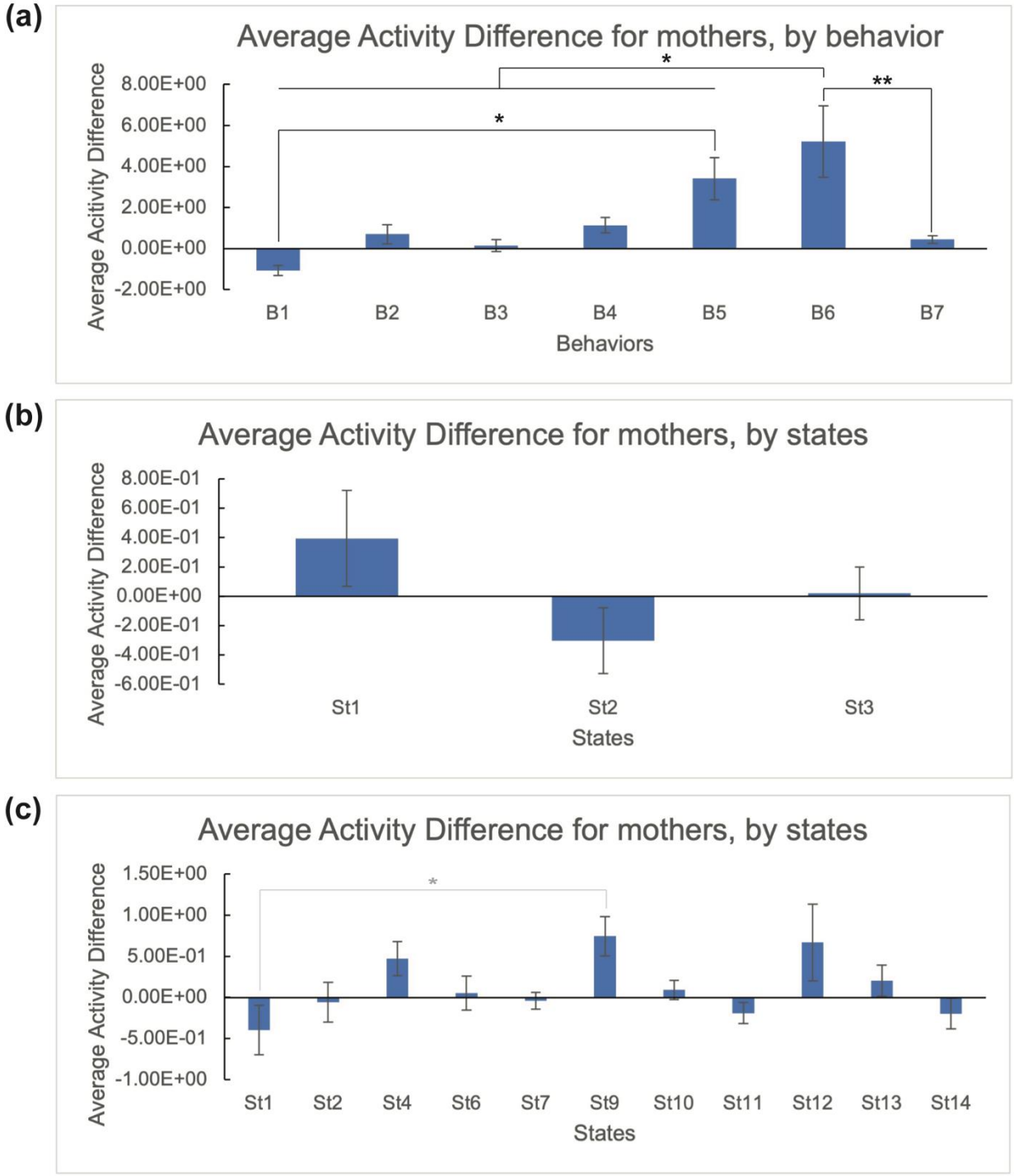
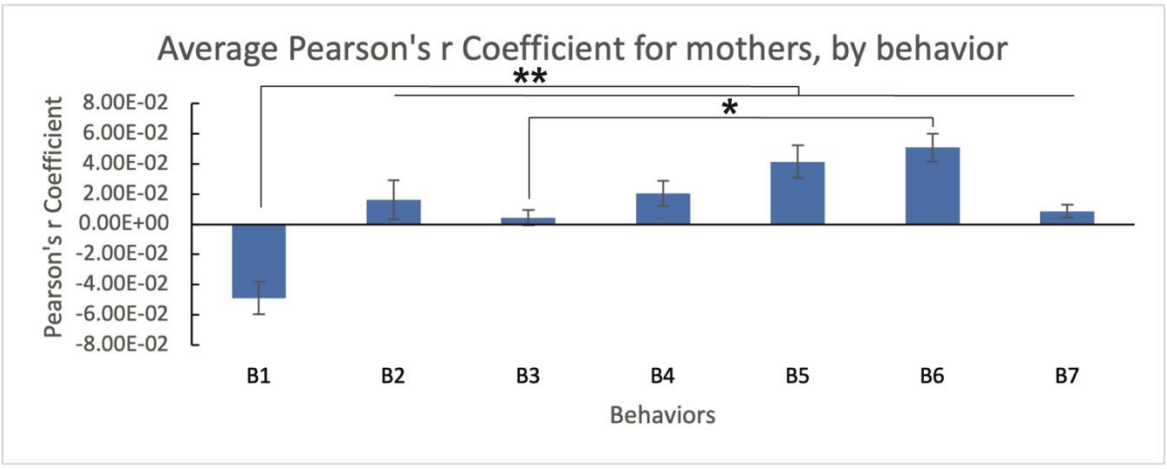
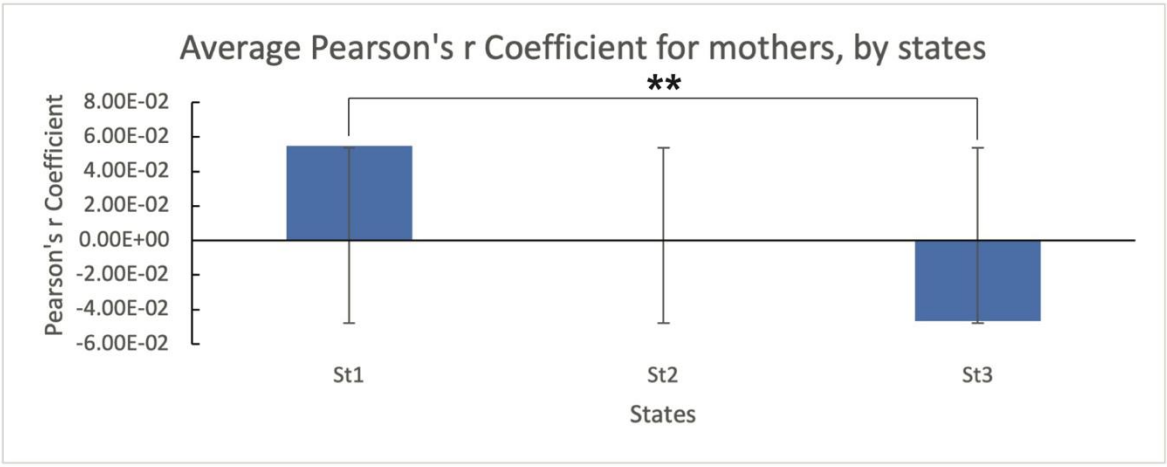


Fig. 19. Average MPOA neuronal activity differences for different behaviors and states. In each part, we took the average difference of MPOA neuronal activity between when the behavior or state was present in the frame and when it was not. We decided to soley focus on mothers due to smaller sample sizes in virgin females and males. **(a)** shows the differences for Part 1, **(b)** for Part 2, and **(c)** for Part 3. All ** denote $p < 0.01$, * denote $p < 0.05$, and grey * denote $p < 0.1$.

(a)



(b)



(c)

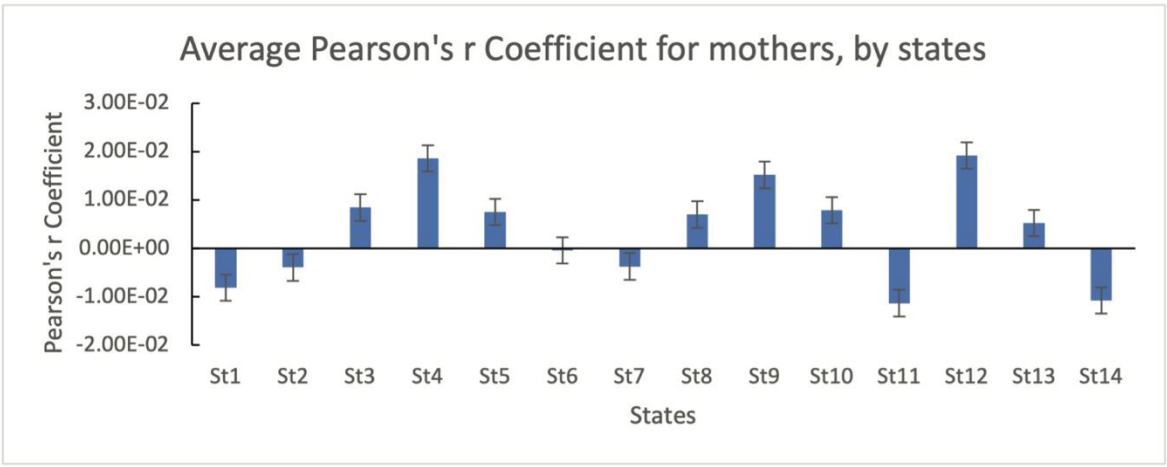


Fig. 20. Average Pearson's r Coefficient of MPOA neuronal activity and different behaviors or states. Since the outputted behaviors or states formed a step-wise function, we found the average activity or behavior/state over a 1 second duration, and ran correlations with these continuous values. The results for Part 1 are shown in **(a)**, Part 2 in **(b)**, and Part 3 in **(c)**. All ** denote $p < 0.01$ and * denote $p < 0.05$.

However, there was still no significance in the one-way ANOVA test. Further research should be conducted to confirm these results, especially in elucidating what hidden states 1 and 3 could imply for newly described parenting behaviors. In the future, different time scales should be investigated. Also, as we looked at activity across all the neurons, future analysis should focus on looking at specific subgroups of neurons instead of looking at all of them.

Discussion

Despite numerous advances in machine learning, a robust framework for quantifying naturalistic behavior has been difficult to develop. Within parenting alone, challenges like accurately tracking multiple animals, identifying hidden states related to parenting, and describing new parenting motor actions or postures have long hindered the development of such a framework. However, our three-pronged approach comprising of a user-defined system and unsupervised methods has demonstrated a novel framework that can accurately quantify previously described parenting behaviors, detect hidden behavioral states that affect parenting behavior, and find postures not previously associated with parenting behavior.

Novel Framework to Quantify Parenting Behavior

For our framework, we had three main components: a user-defined system that was more hardcoded and two models that were more unsupervised. Both approaches were chosen with different goals in mind. For the user-defined system, we wanted a way to gather high throughput data using a stable pipeline that would be robust under different conditions, especially pertaining to scale and modification of features across different videos. For the two unsupervised models, we needed a method to identify hidden states or body postures researchers had not previously incorporated into parenting behavior. Therefore, we went for an unsupervised approach. Within

our framework, we believe that the hardcoded user-defined system allows for faster and efficient processing of the previously described behaviors, whereas our unsupervised models have found interesting new hidden states and postures to explore further in parenting behavior.

However, there are improvements that we suggest to our framework, particularly for the unsupervised model approach. We were originally motivated to use a GLM-HMM because we needed to combine multiple variables (e.g. various features) and produce multiple outputs (e.g. behaviors/states probabilities) for a single frame. The GLM aspect of the GLM-HMM allows for both the combination of multiple inputs and outputs. In addition, the GLM-HMM approach has previously found hidden states that have revealed hidden motivations in fly mating behavior (Calhoun et al., 2019) that we believed could be helpful in the context of parenting behavior. Yet, given the somewhat ambiguous results of our GLM-HMM, further modifications on the intrinsic model itself may be warranted. We found that throughout our parameter variations, the most significant improvement occurred with increasing the lambda, where the time spent in each hidden state was longer and more stable. Given a higher lambda increasingly smooths our data, further work could be done in improving the intrinsic ability of the model to parse noise, which could include researching different optimization functions or regularizers. Furthermore, given that different variations of HMM algorithms can include the ability to provide multiple inputs, we believe other HMM variations could also be considered in the future.

The other unsupervised approach we used included a clustering step to identify similar frames and then an HMM model to combine these clusters into describable behaviors. We originally chose to use HMM for finding postural parenting behaviors due to its ability to incorporate history. Since behaviors are usually completed over multiple frames, we liked that the HMM transition and emission matrices stored historical information across multiple frames. However,

there are two areas to potentially research further. First, we had relied on k-means clustering to transform the features in each frame into one input for the HMM. Given k-means is a well-known clustering approach that is less time-intensive, we wanted to see if as a baseline, it could produce meaningful results. However, one clustering method we would explore further if given more time is spectral clustering. We initially examined spectral clustering, but it proved to be too time-intensive, since the runtime scales exponentially with each additional frame added. In the future, we would likely re-run spectral clustering on a smaller set of frames first to see if any results could be drawn from a smaller sample size. Additionally, dimensionality reduction methods like Principal Component Analysis (PCA) could be used to reduce the inputted feature space (Berman, 2018). Second, we would have liked to compare the HMM results with other models that also incorporate history, such as Bayesian inference models.

Ability to Collect High Throughput Behavior Quantification Data

We ultimately were able to quantify several previously described parenting behaviors in mice (Table 2b), where our initial cross-references between the user-defined system and human behavior annotations averaged at least 90 percent similarity for four of these behaviors. In addition, many of the findings from the user-defined system matched our visual observations during the behavior assays, whether that was seeing how virgin females explored more outside the nest, fathers investigated longer before retrieving the pup, and mothers had longer retrieval processes. These results indicate that this user-defined system potentially provides a high throughput method that would allow researchers to accurately analyze numerous videos in shorter time periods. A major strength of this user-defined system is that many of its behavior quantifications are largely based on ratios and percentages rather than environment specific

details. Thus, even if the environments of a behavior assay change, the user-defined system can still accurately and quickly quantify various behaviors. This is in contrast to previous supervised methods that are trained to be used in very specific contexts (Nilsson et al., 2020; Segalin et al., 2020). Given the time consuming and biased nature of human annotation, this user-defined system could provide unprecedented change for conducting more efficient analysis and larger sample experiments.

Yet, there were a few limitations for this user-defined system that should be noted. First, many of these behaviors were quantified based on unique features of a given parenting behavior. For instance, entering the nest could clearly be measured by the percent overlap change between nest and mouse; retrieving the pup could be succinctly measured by the distance change between the pup and nest (Table 2b). These features were very context dependent for parenting behavior, and other naturalistic behaviors (e.g. foraging, aggression, mating, social encounters, fly courtship behaviors, etc.) may not be as easily measured with simple geometric features (Hong et al., 2015; Berman et al., 2016; Calhoun et al., 2019). In those cases, this user-defined system could be limited in which behaviors it can accurately or efficiently quantify. In addition, our user-defined system is still affected by human bias. From the features that we chose to look at to the previously described behaviors we highlighted, there were significant human choices made throughout the design of this system. Thus, although this user-defined system did highlight that previously described parenting behaviors take up very little time in the behavior assays and that more parenting behaviors should be defined, this result could change if there were different features or behaviors were examined in the experiment. Further research should be done to confirm our results on whether these chosen features and behaviors are the most optimal.

Ability to Detect Hidden Factors in Parenting

We used our GLM-HMM to quantify 3 hidden states that are related to parenting behavior. Similar to previous work done in fly mating behavior (Calhoun et al., 2019), we were able to find hidden states that could indicate motivational or emotional factors related to parenting. Specifically, we were able to differentiate that hidden state 1 and hidden state 3 represented a non-parenting and parenting state respectively, whereas hidden state 2 may be a sex-specific parenting state. The GLM-HMM results here indicate an exciting and novel method for analyzing naturalistic behaviors like parenting, since previous research has not had the tools to examine similarly hidden factors with such detail. Given the motivation-related function of the hypothalamus where the MPOA neurons are located, these hidden states could provide new discoveries of parenting behavior and unprecedented findings for neuronal function.

However, these results are still preliminary, especially given that many of the hidden states here were not intuitive. While we were able to observe that these three hidden states visually best represented the behavior assays since they minimized constant state switches, we were unable to discern what exact patterns the model was seeing to separate the frames into the different hidden states. Our model still has room for improvement though. These included aspects like expanding to a larger training set, exploring different types of regularizers, increasing the number of inputted features, and even varying the types of inputted features. For instance, whether it was defining the "Entering the Nest" behavior in Part 1 with the snout body point or noticing numerous states differentiated solely by head movement in Part 3, our research indicates that an important feature to input in future experiments would be one that concerns the head of the mouse. If further research incorporates such improvements into future GLM-HMM models, we believe the hidden states could become more intuitive. With more precise descriptions of the

motivational or emotional factors that exist in these hidden states, new parenting behaviors could be described and existing ones could be examined in a new light.

Ability to Quantify Newly Described Parenting Behaviors

Our HMM quantified 14 postural behaviors not described earlier, 11 of which were occurring frequently during the parenting assays. These 11 postural parenting behaviors revealed the importance of the head and snout movement of mice in parenting. For instance, the hunched posture with an idle head appeared closely associated with parenting behavior. Furthermore, the stationary body with enhanced snout movement results appeared to indicate a difference between female and male parenting behavior. These postural parenting behaviors were not documented before, but the HMM was able to identify and highlight these new parenting behaviors. In addition, other behaviors such as the forward jerky and contracting movement, which ended up coinciding with aspects of nest building, were also new parenting behaviors previously undocumented. These preliminary results indicate the potential of this method to discover, quantify, and identify previously undescribed parenting behaviors. This is especially exciting, given that many researchers would not likely know to look for the behaviors identified in this part, and they represent new behaviors that can be incorporated into future experiments examining mouse parenting behavior.

The most surprising finding from this part was the number of similar states that were ultimately separated by head movement differences (e.g. Forward locomotion vs. Forward movement ending in snout up, Stationary with enhanced snout movement vs. Stationary with idle head, etc.). However, an underlying cause of these similar states could also be due to the limitation of human verbal descriptions. As researchers, we must describe new behaviors in the context of

existing behaviors or observable features. Yet, the wide range of behaviors likely spans beyond what humans can verbally describe. Thus, these results highlight a need for incorporating newly described behaviors into future experiments. Given the unconstrained nature of parenting and other naturalistic behaviors, there are likely new behaviors that cannot be objectively describe without further research and understanding. Therefore, by using the averages of the quantified features (Table 3b) in this method, researchers could perhaps go beyond the existing descriptive behaviors, discover new non-descriptive behaviors, and uncover new neuronal roles for parenting behavior.

MPOA Neurons May Relate to Initiation of Parenting

From our results in Part 1's previously described behaviors and Part 3's postural parenting behaviors, we saw that MPOA neurons had greater activity differences and stronger correlations with parenting behaviors related to investigation and approaching the pup. These results confirm that parenting behaviors are significantly encoded in the MPOA, consistent with previous research (Wu et al., 2014; Kohl et al., 2018). However, it was interesting to note that the behaviors that the MPOA neurons responded strongly to, investigation and approaching, all happen at the start of most parenting processes. Often, after a pup is presented, these are two common behaviors completed when the adult begins to realize the pup is in its presence and attempts to care for it. This may suggest that the MPOA neurons specifically play a role in the initiation of parenting, and further research should be conducted to confirm these findings.

Potential for Further Subdivision of MPOA Neurons

In general, our analyses of MPOA neuron activity differences were cleaner than those of MPOA neuron correlations. We hypothesize that this could be due to further specialization in a subset of MPOA neurons. Specifically, we were initially puzzled by how we saw there was a negative correlation between hidden state 3 and MPOA neuron activity but a higher distribution of time spent in hidden state 3 by mothers and fathers. However, we realized that this could be due to the fact there are subsets of MPOA neurons that may encode for separate and even opposing processes. Since we averaged our correlation and activity difference values across all MPOA neurons, the correlations, which assign equal weight to each neuron, will be more negatively affected than activity differences, which assign unequal weights to each neuron. Thus, this could explain why we saw almost no activity difference in Fig. 19b but much larger negative correlation in Fig. 20b for hidden state 3. This suggests that a potential subdivision of MPOA neurons should be examined in future parenting assays, where a single cell or smaller subset may be needed to get a fuller picture of the function of MPOA neurons in parenting behavior. It is also possible that the correlation results stem from too small of a time bin size. Since behaviors may take longer than one second to complete, the correlations may be inaccurate if the time bin is too short and does not take in the activity of the entire behavior. Additional research examining different time bin sizes can help elucidate these results further, and if the MPOA neurons are subdivided even more into specialized functions.

Impetus to Explore Other Parenting Neurons

We observed that the quantified mouse parenting behaviors yielded meaningful information about brain function. We were able to use activity differences and correlations to hypothesize

potential relationships between new hidden states and postures and MPOA neurons. However, there was still a large majority of "Other" behaviors from Part 1 and postural parenting behaviors from Part 3 that we could not find relationships for. This was particularly interesting, given that although MPOA seem to be a key neuron in driving parenting behavior, they do not control all aspects of it. This could provide motivation to therefore search other brain regions and neurons for parenting-related functions.

Conclusion

Using mouse parenting behavior as a motivating case study, we were able to develop a three-pronged behavior quantification framework (the code will be uploaded to a Github repository) that analyzes previously described parenting behaviors with high throughput, identifies hidden states associated with parenting behaviors, and illuminates newly described parenting behaviors. The framework's outputted quantifications were also proven to be tangibly useful for future experiments, whether that be in using the quantifications to analyze detailed differences in parenting behavior across mothers, fathers and virgin females or discovering how the quantifications reveal new findings for MPOA neuronal function. This framework holds immense potential for creating more detailed behavior quantification that not only builds upon existing behaviors but also expands the definitions for mouse parenting behavior. Ultimately, mouse parenting behavior is one case study of many naturalistic behaviors, and we hope that this framework can be applied to drive similarly unprecedented discoveries and analysis for other naturalistic behaviors.

References

Badre D, Frank MJ, Moore CI (2015) Interactionist neuroscience. *Neuron* **88**:855–860.

Berman GJ (2018) Measuring behavior across scales. *BMC Biol* **16**:1–11.

Berman GJ, Bialek W, Shaevitz JW (2016) Predictability and hierarchy in Drosophila behavior. *Proc Natl Acad Sci* **113**:11943–11948.

Branson K, Robie AA, Bender J, Perona P, Dickinson MH (2009) High-throughput ethomics in large groups of Drosophila. *Nat Methods* **6**:451–457.

Calhoun AJ, Pillow JW, Murthy M (2019) Unsupervised identification of the internal states that shape natural behavior. *Nat Neurosci* **22**:2040–2049.

Carandini M (2012) From circuits to behavior: a bridge too far? *Nat Neurosci* **15**:507–509.

Carr CE, Konishi M (1990) A circuit for detection of interaural time differences in the brain stem of the barn owl. *J Neurosci* **10**:3227–3246.

Cooper RP, Peebles D (2015) Beyond single-level accounts: The role of cognitive architectures in cognitive scientific explanation. *Top Cogn Sci* **7**:243–258.

Dennis EJ, El Hady A, Michaiel A, Clemens A, Tervo DRG, Voigts J, Datta SR (2021) Systems Neuroscience of Natural Behaviors in Rodents. *J Neurosci* **41**:911–919.

Duan CA, Erlich JC, Brody CD (2015) Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. *Neuron* **86**:1491–1503.

Dulac C, O'Connell LA, Wu Z (2014) Neural control of maternal and paternal behaviors. *Science (80-)* **345**:765–770.

Gao P, Ganguli S (2015) On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr Opin Neurobiol* **32**:148–155.

Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF (2014) Big behavioral data:

psychology, ethology and the foundations of neuroscience. *Nat Neurosci* **17**:1455–1462.

Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID (2019) DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* **8**:e47994.

Hickok G, Hauser M (2010) (Mis) understanding mirror neurons. *Curr Biol* **20**:R593–R594.

Hong W, Kennedy A, Burgos-Artizzu XP, Zelikowsky M, Navonne SG, Perona P, Anderson DJ (2015) Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc Natl Acad Sci* **112**:E5351–E5360.

Jeffress LA (1948) A place theory of sound localization. *J Comp Physiol Psychol* **41**:35.

Joseph AW, Hyson RL (1993) Coincidence detection by binaural neurons in the chick brain stem. *J Neurophysiol* **69**:1197–1211.

Juavinett AL, Erlich JC, Churchland AK (2018) Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility. *Curr Opin Neurobiol* **49**:42–50.

Katz PS (2016) Evolution of central pattern generators and rhythmic behaviours. *Philos Trans R Soc B Biol Sci* **371**:20150057 Available at: https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0057.

Kohl J, Babayan BM, Rubinstein ND, Autry AE, Marin-Rodriguez B, Kapoor V, Miyamishi K, Zweifel LS, Luo L, Uchida N (2018) Functional circuit architecture underlying parental behaviour. *Nature* **556**:326.

Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D (2017) Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**:480–490.

Levitis DA, Lidicker Jr WZ, Freund G (2009) Behavioural biologists do not agree on what constitutes behaviour. *Anim Behav* **78**:103–110.

Li Y, Mathis A, Grewe BF, Osterhout JA, Ahanonu B, Schnitzer MJ, Murthy VN, Dulac C (2017) Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. *Cell* **171**:1176-1190.e17 Available at: https://pubmed.ncbi.nlm.nih.gov/29107332.

Marder E, Goaillard J-M (2006) Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci* **7**:563–574.

Marom S, Meir R, Braun E, Gal A, Kermany E, Eytan D (2009) On the precarious path of reverse neuro-engineering. *Front Comput Neurosci* **3**:5.

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 Available at: https://doi.org/10.1038/s41593-018-0209-y.

Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *J Neurosci* **27**:7105–7116.

Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, Rubinstein ND, Hao J, Regev A, Dulac C (2018) Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science (80-)* **362**.

Musall S, Sun XR, Mohan H, An X, Gluf S, Churchland AK (2021) Pyramidal cell types drive functionally distinct cortical activity patterns during decision-making. *bioRxiv*.

Nilsson SRO, Goodwin NL, Choong JJ, Hwang S, Wright HR, Norville ZC, Tong X, Lin D, Bentzley BS, Eshel N (2020) Simple Behavioral Analysis (SimBA)–an open source toolkit for computer classification of complex social behaviors in experimental animals. *BioRxiv*.

Overholt EM, Rubel EW, Hyson RL (1992) A circuit for coding interaural time differences in the chick brainstem. *J Neurosci* **12**:1698–1708.

Parks TN, Rubel EW (1975) Organization and development of brain stem auditory nuclei of the chicken: organization of projections from n. magnocellularis to n. laminaris. *J Comp Neurol* **164**:435–448.

Pereira TD, Shaevitz JW, Murthy M (2020) Quantifying behavior to understand the brain. *Nat Neurosci* **23**:1537–1549.

Rodriguez A, Zhang H, Klaminder J, Brodin T, Andersson PL, Andersson M (2017) ToxTrac: A fast and robust software for tracking organisms. Methods in ecology and evolution. *Br Ecol Soc doi* **10**.

Segalin C, Williams J, Karigo T, Hui M, Zelikowsky M, Sun JJ, Perona P, Anderson DJ, Kennedy A (2020) The Mouse Action Recognition System (MARS): a software pipeline for automated analysis of social behaviors in mice. *BioRxiv*.

Song Y-H, Kim J-H, Jeong H-W, Choi I, Jeong D, Kim K, Lee S-H (2017) A neural circuit for auditory dominance over visual perception. *Neuron* **93**:940–954.

Tachikawa KS, Yoshihara Y, Kuroda KO (2013) Behavioral transition from attack to parenting in male mice: a crucial role of the vomeronasal system. *J Neurosci* **33**:5120–5126.

Taylor JA, Krakauer JW, Ivry RB (2014) Explicit and implicit contributions to learning in a sensorimotor adaptation task. *J Neurosci* **34**:3023–3032.

Tinbergen N (1963) On aims and methods of ethology. *Z Tierpsychol* **20**:410–433.

Vom Saal FS (1985) Time-contingent change in infanticide and parental behavior induced by ejaculation in male mice. *Physiol Behav* **34**:7–15.

Wagner H, Mazer JA, Von Campenhausen M (2002) Response properties of neurons in the core of the central nucleus of the inferior colliculus of the barn owl. *Eur J Neurosci* **15**:1343–1352.

Watanabe A, Takeda K (1963) The change of discharge frequency by AC stimulus in a weak electric fish. *J Exp Biol* **40**:57–66.

Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abraira VE, Adams RP, Datta SR (2015) Mapping sub-second structure in mouse behavior. *Neuron* **88**:1121–1135.

Wiltschko AB, Tsukahara T, Zeine A, Anyoha R, Gillis WF, Markowitz JE, Peterson RE, Katon J, Johnson MJ, Datta SR (2020) Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat Neurosci* **23**:1433–1443.

Winstanley CA, Floresco SB (2016) Deciphering decision making: variation in animal models of effort-and uncertainty-based choice reveals distinct neural circuitries underlying core cognitive processes. *J Neurosci* **36**:12069–12079.

Wu Z, Autry AE, Bergan JF, Watabe-Uchida M, Dulac CG (2014) Galanin neurons in the medial preoptic area govern parental behavior. *Nature* **509**:325–330 Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4105201/.

Zhou P, Resendez SL, Rodriguez-Romaguera J, Jimenez JC, Neufeld SQ, Giovannucci A, Friedrich J, Pnevmatikakis EA, Stuber GD, Hen R, Kheirbek MA, Sabatini BL, Kass RE, Paninski L (2018) Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *Elife* **7**:e28728 Available at: https://pubmed.ncbi.nlm.nih.gov/29469809.