Mice alternate between discrete strategies during perceptual decision-making

Zoe C. Ashwood[1], Nicholas A. Roy[2], Iris R. Stone[3], The International Brain Laboratory[4], Anne K. Churchland[5], Alexandre Pouget[6], and Jonathan W. Pillow[7]

[1]Dept. of Computer Science, Princeton University
[1,2,3,7]Princeton Neuroscience Institute
[4]University College London
[5]David Geffen School of Medicine, The University of California, Los Angeles
[6]Faculty of Medicine & Dept. of Basic Neurosciences, University of Geneva
[7]Dept. of Psychology, Princeton University

October 19, 2020

Abstract

Classical models of perceptual decision-making assume that animals use a single, consistent strategy to integrate sensory evidence and form decisions during an experiment. Here we provide analyses showing that this common view is incorrect. We use a latent variable modeling framework to show that decision-making behavior in mice reflects an interplay between different strategies that alternate on a timescale of tens to hundreds of trials. This model provides a powerful alternate explanation for "lapses" commonly observed during psychophysical experiments. Formally, our approach consists of a Hidden Markov Model (HMM) with states corresponding to different decision-making strategies, each parameterized by a distinct Bernoulli generalized linear model (GLM). We fit the resulting model (GLM-HMM) to choice data from two large cohorts of mice in different perceptual decision-making tasks. For both datasets, we found that mouse decision-making was far better described by a GLM-HMM with 3 or 4 states than by a traditional psychophysical model with lapses. The identified states were highly consistent across animals, consisting of a single "engaged" state, in which the strategy relied heavily on the sensory stimulus, and multiple biased or disengaged states in which accuracy was low. These states persisted for many trials, suggesting that lapses were not independent, but reflected state dynamics in which animals were relatively engaged or disengaged for extended periods of time. We found that for most animals, response times and violation rates were positively correlated with disengagement, providing independent correlates of the identified changes in strategy. The GLM-HMM framework thus provides a powerful lens for the analysis of decision-making, and suggests that standard measures of psychophysical performance mask the presence of slow but dramatic alternations in strategy across trials.

1 Introduction

In order to understand the computations performed in the brain, we need a comprehensive characterization of behavior [4, 21, 27]. This realization has fueled a recent surge in methods devoted to the

measurement, quantification, and modeling of natural behaviors [1, 10, 26, 43, 48]. However, studies of perceptual decision-making behavior have tended to rely on simplified models derived from signal detection theory [22]. In these models, decisions are assumed to rely on a single, static strategy that does not meaningfully change over trials or sessions.

One problem that has afflicted classic models of perceptual decision-making is the presence of lapses—errors committed on "easy" trials, or trials on which there is strong evidence for the correct choice. These lapses arise with surprisingly high frequency, particularly in rodent experiments: it is not uncommon for lapses to comprise 10-20% of all trials [24, 34, 35].

Why do lapses occur? Explanations for presence of lapse errors range from inattention, to motor error, to incomplete knowledge of the task, and "uncertainty-guided exploration" [11, 28, 36, 47]. However, a common thread to these explanations is that lapses arise independently, at a roughly constant rate across trials. The standard model for lapses involves augmenting the classic signal-detection theory model with a "lapse parameter", which characterizes the probability that the observer simply ignores the stimulus on any given trial and guesses randomly [11, 37, 47]. However, this stationarity assumption has not yet been directly tested.

Here we show that the classic model is incorrect, and that rodent perceptual decision-making behavior can be more accurately described in terms of discrete strategies, each associated with a distinct latent state, which alternate slowly over time. In this framework, lapses arise far more frequently in some states than others. This implies that: (i) lapses tend to be clustered together in time; and (ii) animals with high apparent lapse rates may nevertheless be capable of high-accuracy performance during extended blocks of trials.

Our modeling framework consists of a hidden Markov Model (HMM), with each state corresponding to a unique decision-making strategy, parametrized by a Bernoulli generalized linear model (GLM). The resulting GLM-HMM includes the classic lapse model as a particular parameter setting, but allows for more general and long-enduring strategies. We take advantage of the recent availability of high throughput behavioral training paradigms [7, 24, 34, 35], which can readily produce hundreds of thousands of trials of choice data; and we analyze the choice data of large cohorts of mice performing two perceptual decision-making tasks [24, 34]. Using the GLM-HMM, we show that lapses do not arise independently over time. Instead, lapses are the result of mice switching between relatively engaged and disengaged decision-making strategies *within* a session. Furthermore, these strategies persist for tens to hundreds of trials at a time. Compared to the classic lapse model, a test set of choices for an example animal is vastly more likely under the GLM-HMM with long-enduring engaged and disengaged strategies (2×10^{24} times). We study the statistics of data simulated from the best-fitting GLM-HMM, and find that the long choice runlengths that are observed in real choice data exist in the simulated data; however the classic lapse model cannot produce such long choice runlengths. Finally, we validate that the strategies that our model recovers are semantically meaningful when we compare the response times and violation rates associated with engaged and disengaged states. We find that the most extreme response times are typically associated with the disengaged states for the IBL task, which is consistent with previous findings linking accuracy and response times [23, 38, 44]. Altogether, our analysis suggests that a significant rethinking of rodent choice behavior in perceptual decision-making experiments is necessary, and that acknowledging state dependence will be critical when analyzing any neural activity measured while mice perform these tasks.

2 Results

2.1 Generalizing the classic lapse model of choice behavior

The traditional approach to modeling choice behavior in binary choice perceptual decision-making experiments involves fitting a psychometric curve [11, 37, 47] in order to predict, as a function of the stimulus intensity, the probability that the animal chose to go rightward (for example). The psychometric curve assumes that an animal's decision is only influenced by the stimulus and the animal's bias. This can be extended this to incorporate choice dependence on stimulus, choice and reward history [9, 19]. In what follows, we will refer to such an extension of the psychometric curve as the "classic lapse model of choice behavior":

$$p(y_t = 1|\mathbf{x}_t) = \big(1 - (\gamma_R + \gamma_L)\big)\frac{1}{1 + e^{-\mathbf{w}\cdot\mathbf{x_t}}} + \gamma_R \tag{1}$$

where $y_t \in \{0, 1\}$ represents the choice (left or right) that an animal makes at trial t, and $\mathbf{x}_t \in \mathbb{R}^M$ is the set of covariates that influence an animal's choice (such as the stimulus strength, but also choice and reward history). The vector $\mathbf{w}$ is the set of weights that reflects how each covariate influences the animal's choice; while γ_R and γ_L are the animal's lapse rates. A psychometric curve and its associated lapse rates are shown graphically in Fig. 1c.

Whereas the classic lapse model in eq. 1 takes on its particular mathematical form in order to fit empirical data, we observe that this model can also be interpreted as a mixture model [15, 30, 32]. It is as if, on each trial, the animal is flipping a biased coin so as to determine its internal state, z_t. With probability $(1 - (\gamma_R + \gamma_L))$, the animal is in its engaged decision-making state; and with probability, $(\gamma_R + \gamma_L)$, the animal is in its lapse state (Fig. 1 - panels a and b). In the engaged decision-making state, the animal makes its choice according to a sigmoid: $p(y_t = R|\mathbf{x}_t, z_t = \text{engaged}) = \frac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x_t}}}$; while in the disengaged state, the animal makes its choice according to a bias: $p(y_t = R|\mathbf{x}_t, z_t = \text{lapse}) = \frac{\gamma_R}{\gamma_R+\gamma_L}$.

Viewing the classic lapse model in this way highlights some of its limitations. First, the model assumes that mice switch between only two decision-making strategies. Second, it assumes that lapses occur independently in time, according to a Bernoulli random variable or biased 'coin flip' on each trial. (Fig. 1b). Finally, the model assumes that choices in the "lapse" state are fully independent of the stimulus, neglecting the possibility that they are still weakly stimulus dependent, or are influenced by other covariates such as reward or choice history.

Recognizing these limitations, we propose to instead examine perceptual decision-making using a framework based on Hidden Markov models (HMMs) with generalized linear model (GLM) observations [10, 17]. The resulting "GLM-HMM" framework, also known as an "input-output HMM" [3], includes the classic lapse model as a special case, but allows for arbitrary numbers of states, states that persist for an extended number of trials and states with different dependencies on external covariates.

2.2 The GLM-HMM allows for switching between multiple strategies

The GLM-HMM consists of a Hidden Markov Model (HMM), with each state's emission probabilities given by a Bernoulli generalized linear model (GLM), so as to include dependence on external inputs.

This model can also be referred to as an Input-Output HMM [3]. Concretely, we model an animal's choices, $\mathbf{y} \in \{0,1\}^T$, across T trials in a binary choice perceptual decision making task once the animal's choice behavior has reached stationarity (see Fig. S3). Choice at trial t depends on external inputs, $\mathbf{x_t} \in \mathbb{R}^M$ (the stimulus intensity, as well as history regressors and a bias); as well as the animal's latent state at that trial, $z_t \in \{1, ..., K\}$, according to:

$$p(y_t = 1|z_t = k, \mathbf{x}_t, \mathbf{w}_k) = \frac{1}{1 + e^{-\mathbf{w_k} \cdot \mathbf{x_t}}} \tag{2}$$

where $\mathbf{w}_k \in \mathbb{R}^M$ is the set of GLM weights associated with state k. Transitions between latent states are governed by a stationary transition matrix, $A \in \mathbb{R}^{K \times K}$:

$$p(z_{t+1} = k|z_t = j, A) = A_{jk} \tag{3}$$

If the animal's strategy at trial $t + 1$ is independent of its strategy at the previous trial t, A will be structured such that the entries across the rows are identical for a particular column of the matrix. That is, the best fitting transition matrix has a single value for $A_{:,k}$. By not restricting the structure of the transition matrix in this way, we allow for the possibility that the animal uses time-dependent strategies.

The joint probability for states and choices is given by:

$$p(\mathbf{y}, \mathbf{z}|\{\mathbf{x}\}_{t=1}^T, \Theta) = p(z_1) \prod_{t=2}^{T} p(z_{t+1}|z_t) p(y_t|z_t, \mathbf{x}_t) \tag{4}$$

where $\Theta \equiv \{\phi, A, \{\mathbf{w}_k\}_{k=1}^K\}$ is the set of GLM-HMM parameters and $\phi \in \mathbb{R}^K$ is the initial state distribution such that $\phi_k \equiv p(z_1 = k)$. This set of parameters is fit to empirical animal choice data via Maxmimum a Posteriori estimation via the Expectation Maximization algorithm [14, 31] as described later (see Inference of GLM-HMM parameters). We also verified that we were able to recover the generative parameters when we simulated choice data from GLM-HMMs in different parameter regimes (see Fig. S1 and Fig. S2).

2.3 Classic lapse model is 2 state GLM-HMM subject to constraints

We observe that the classic lapse model of choice behavior given in eq. 1 corresponds to a 2 state GLM-HMM where the GLM weights for state 1 are $\mathbf{w}$, while the GLM weights for state 2 are zero for all covariates except the bias. The weight for the bias is $w_{\text{bias}} = -\log \frac{\gamma_L}{\gamma_R}$. Furthermore, the transition matrix for the lapse model has $A_{:,1} = 1 - (\gamma_R + \gamma_L)$, and $A_{:,2} = \gamma_R + \gamma_L$. In Fig. 1a, we show example engaged and lapse states for the classic lapse model, as well as the transition probabilities between these states (for $\gamma_R = 0.13$ and $\gamma_L = 0.07$). Notice that, independent of the animal's state at the previous trial, with probability 0.8, the animal is in the engaged state at the next trial. In Fig. 1b, we show an example (simulated) sequence for the animal's internal state across a session when generated according to these transition probabilities. Observe that the lapse state tends to persist for only a single trial. Based on this sequence of states, the animal generates a sequence of choices using the state-specific GLMs of Fig. 1a; these choices can then be fit to give a single psychometric curve (Fig. 1c).

In Fig. 1d, we consider a different parameter regime for the GLM-HMM. We, instead, consider a 3 state model where the states represent engaged, disengaged and rightward biased behavior; and the states

are now long-enduring (the self-transitions are large and are 0.95, 0.86 and 0.75). The large values for the self-transition probabilities result in a sequence for the animal's internal state, where each state lasts for many trials in a row. When choice data is generated using the state-specific GLMs shown in Fig. 1d and for the internal state sequence of Fig. 1e, the resulting psychometric curve is that shown in Fig. 1f. Despite the types of behavior responsible for generating Fig. 1c and Fig. 1f being vastly different, the psychometric curve masks this, and they are almost identical in these two figures.

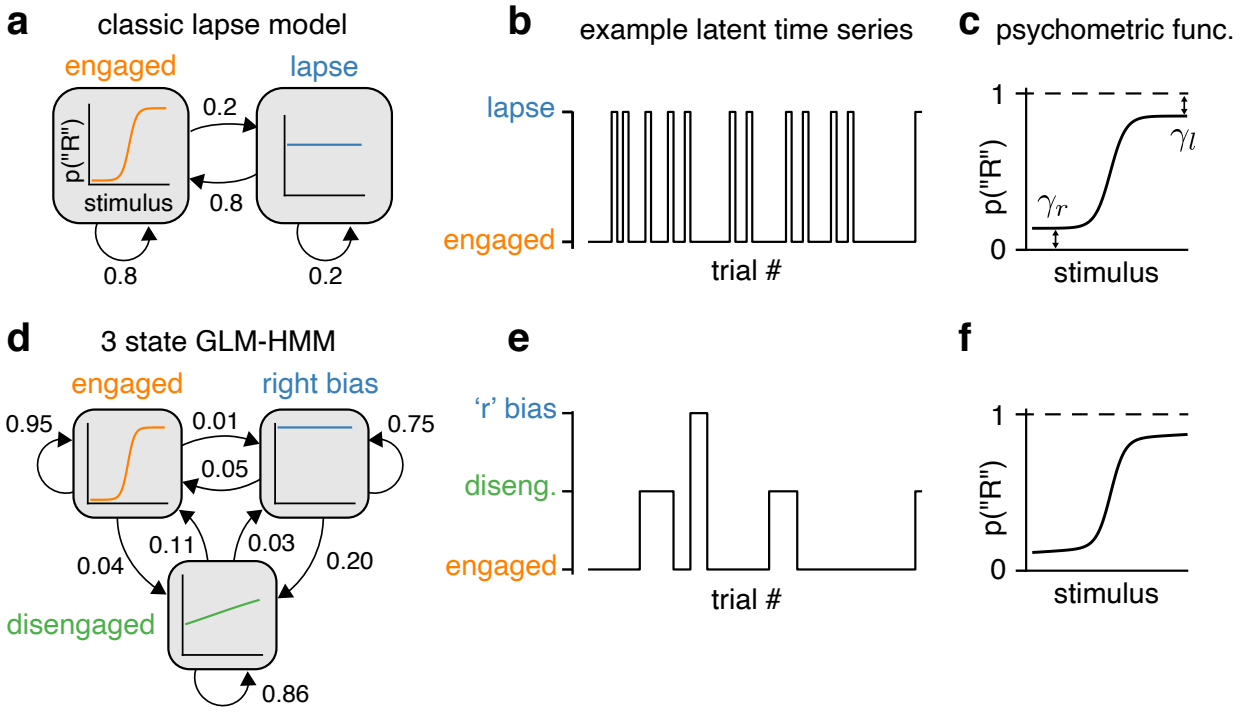

Figure 1: **GLM-HMM is generalization of classic lapse model. (a)** The classic lapse model of choice behavior is formulated as a 2 state GLM-HMM. For this schematic, the external inputs that influence choice behavior are only the stimulus intensity and the animal's bias. The curves inside each of the boxes indicate the GLM that the animal will use to make its decision when in each of the two states. Arrows between these boxes indicate the transition probabilities between states. Note that, regardless of the animal's state at trial t, there is a 0.8 probability that the animal will end up in the engaged state at trial $t + 1$. **(b)** An example (simulated) sequence for the animal's internal state when the transitions between states are governed by the probabilities in (a). Notice that the lapse state tends to persist for only a single trial. **(c)** When the animal's internal state fluctuates as in (b) and when it makes choices according to the state-specific psychometric curves in (a), a sequence of choices is generated. This is the overall psychometric curve (in terms of the probability of going rightward) that results from fitting this choice data. We also show the lower and upper lapse rates, γ_r and γ_l. **(d)** We now consider a different parameter regime for the GLM-HMM, where the model has three states (engaged, disengaged and rightward biased states) and where the corresponding states are long-enduring (the self transition probabilities are 0.95, 0.86 and 0.75). **(e)** An example sequence for the animal's internal state when generated by the GLM-HMM shown in (d). **(f)** The corresponding psychometric curve when choice data is generated with the GLM-HMM in (d) and when the animal's internal state fluctuates as in (e). Observe that the generating processes for the decision-making data of (a)-(c) and (d)-(f) are vastly different; yet the resulting psychometric curves look almost identical and cannot be used to distinguish these two modes of behavior.

2.4 Mice alternate strategies on a slow timescale during the IBL task

We fit the GLM-HMM to the choice data from 37 mice performing the IBL decision-making task [8, 24] after behavior has reached stationarity (accuracy across sessions is consistent: see Fig. S3). During the task (shown in Fig. 2a), a mouse faces a screen and a Gabor patch, of variable contrast, appears either on the left or right side of the screen. The mouse then has to indicate this side by turning a wheel so as to bring the patch into the center of the screen. If the mouse is correct, it receives a water reward; if it is incorrect, there is a noise burst and a 2 second time out. Each IBL animal that we study has at least 3000 trials of choice data from across multiple sessions (see Fig. S1 for a histogram of the number of trials per animal). In Fig. 2 we show the results of applying the GLM-HMM to the choice data of an example IBL animal. We fit GLM-HMMs with 2, 3, 4 and 5 states to the choice data from each mouse, as well as a GLM (equivalent to a 1 state GLM-HMM), and a lapse model (as defined in eq. 1; this is a constrained 2 state GLM-HMM). We show the results of 5-fold cross-validation in Fig. 2b where, in grey, we plot the test set loglikelihood in units of bits per trial (see Normalized Loglikelihood for an in-depth description of this quantity). This represents the increase in test set loglikelihood per trial compared to a baseline model which assumes that an animal is making its choices according to a Bernoulli coin flip (where the probability of choosing to go rightward is determined by the proportion of rightward choices in the training set). We see that the normalized loglikelihood increases as the number of states in the GLM-HMM is increased, but that performance starts to level off after 3 states. For the animal in question, there are 900 trials in its test set and the normalized loglikelihood for the 3 state model is 0.43, while that for the lapse model is 0.34. This corresponds to the test set data being 3×10^{116} times more likely under the 3 state GLM-HMM compared to the Bernoulli coin flip model described above, and 2×10^{24} times more likely under the 3 state GLM-HMM compared to the traditional lapse model. We also calculated each model's predictive accuracy on a held-out dataset. We find that, while the classic lapse model can predict the choices that the animal makes with 79.5% accuracy, the 3 state GLM-HMM can predict the choices that the animal makes with 83.0% accuracy (Fig. 2b in purple).

In Fig. 2c we show the retrieved GLM-HMM parameters for the 3 state model for this example animal. On the left is the best fitting transition matrix, while on the right, we show the best fitting GLM weights for each of the 3 states. We first notice that the transition matrix has large entries on the diagonal: this indicates that mice use the same decision-making strategy for multiple trials in a row. We then look at the weights that the animal places on four covariates in each of the 3 states; the covariates we found to be most predictive of the mouse choice behavior are the z-scored stimulus intensity, the animal's bias (set to 1), the animal's past choice (in {-1, 1} for left and right respectively) and a win-stay-lose-switch covariate defined as $r_{t-1} \times \left(\frac{2y_{t-1}-1}{2}\right)$ where $r_{t-1} \in \{-1, 1\}$ indicates if the animal was rewarded on the previous trial, and the choice on the previous trial is scaled to take on a value in $\{-1, 1\}$. In state 1, the animal places a large weight on the stimulus intensity when making its decision, while the weights on the bias, past choice and win-stay-lose-switch covariates are almost zero. In comparison, when the animal is in states 2 and 3, the animal places a much smaller weight on the stimulus intensity, and has a large bias for choosing to go rightward in state 3, and choosing to go left in state 2. In Fig. 2e, we show that the 3 state model allows us to decompose the psychometric curve for this animal into the 3 constituent state GLMs that generate the data. Specifically, the GLM-HMM allows us to assign trials to states, and we show the psychometric curves generated for trials with posterior state probabilities greater than 0.9. We see that the trials that are associated with state 1 result in a psychometric curve

with a sharp slope and perfect performance on the easy contrasts. For this reason, we interpret the retrieved state 1 as corresponding to 'engaged' decision-making behavior. In contrast, as is also clear from the GLM weights, the animal has a bias to choose to go to the left when in state 2 – even on easy contrasts – and to go rightward when in state 3. The animal has an accuracy of 90% in the engaged state, while it is only 60% and 58% accurate in the biased leftward and biased rightward states. Finally, as mentioned before, the GLM-HMM allows us to obtain the posterior state probabilities, $\gamma_{s,t,k} \equiv p(z_{s,t} = k|\mathbf{y}, \{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta)$, which give the probability that the animal is in state k at trial t in session s (see Expectation Step for details on how to calculate these). We plot these for a few example sessions in Fig. 2f, where we see that the mouse switches between multiple decision-making strategies *within* a session, and, contrary to the prediction made by the classic lapse model, these strategies persist for tens of trials at a time. We show the overall occupancies of each state in Fig. 2d, and find that this mouse spends 69% of its 5040 trials in the engaged state, while it spends 15% and 16% of trials in the biased leftward and rightward states.

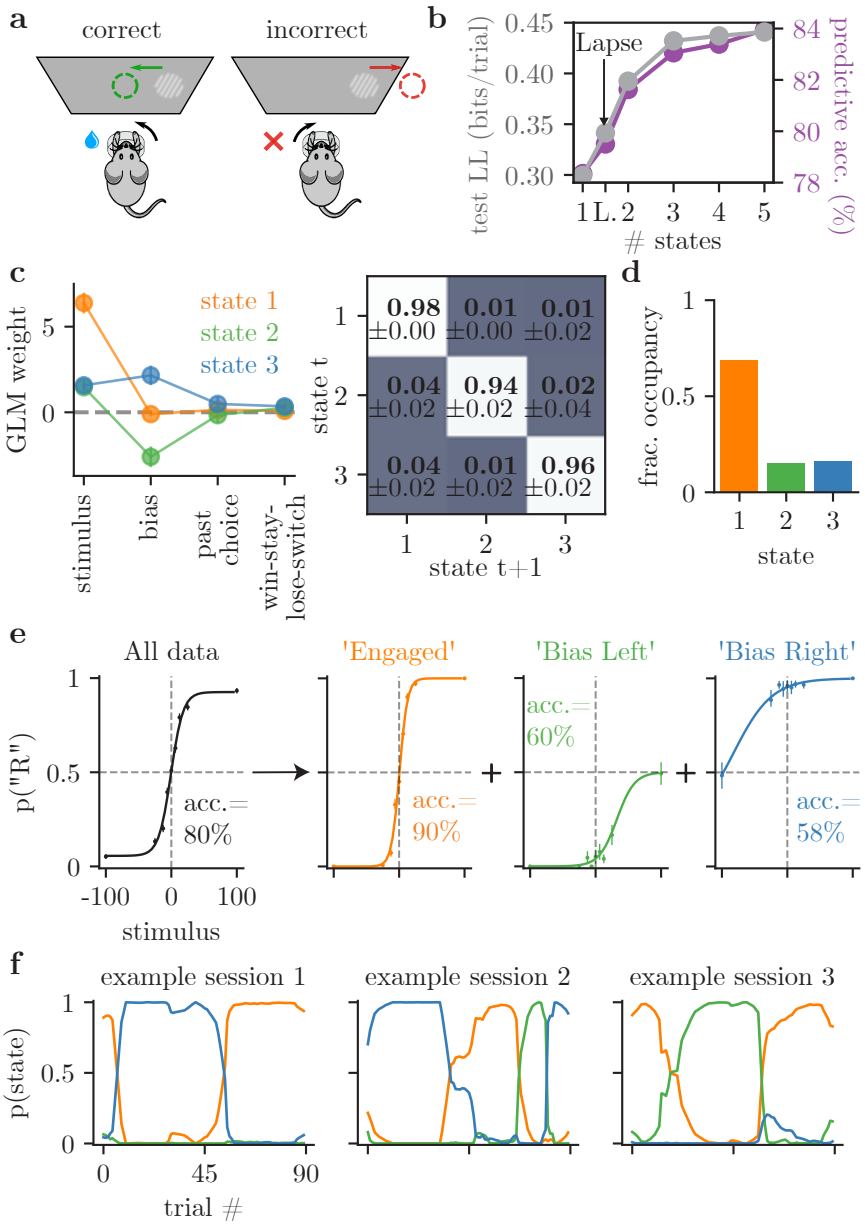

Figure 2: **GLM-HMM application to single IBL animal.** **(a)** IBL task schematic. A Gabor patch appears on the left or right side of a screen, and the mouse has to indicate this side by turning a wheel. If correct, the mouse receives a water reward; if incorrect, there is an auditory tone and a time out. **(b)** 5-fold cross-validation applied to choice data of a single mouse. Grey: test set loglikelihood for a single GLM (1 state GLM-HMM), the classic lapse model (labeled 'L') and for GLM-HMMs with 2-5 states. Purple: test set predictive accuracy for each model. This describes the fraction of held-out trials where we can successfully predict the choice that the mouse made. **(c)** Retrieved parameters of the 3 state GLM-HMM for this mouse. In state 1, the mouse places a large weight on the stimulus intensity when making its decision. In states 2 and 3, the animal's choice is strongly influenced by its leftward (state 2) and rightward biases (state 3). The best fitting transition matrix has large entries on the diagonal, indicating long-enduring states. **(d)** We use the posterior probabilities shown in (f) to assign trials to states and show the relative occupancies of each of the 3 states. **(e)** Black: the regular psychometric curve for this animal for all of its choice data. Orange, green, blue: state-specific psychometric curves. To obtain these, we use the posterior state probabilities to assign trials to states, and fit a psychometric curve to trials for which the posterior state probability for the relevant state is over 0.9 (we show the empirical choice probabilities and the 68% confidence intervals for each stimulus value on top of each curve). For each curve, we also report the animal's accuracy on the task when in the state. **(f)** Posterior state probabilities, $\gamma_{s,t,k}$, for three example sessions.

In Fig. 3, we expand our analysis and study the decision-making behavior of all 37 IBL animals through the lens of the GLM-HMM. We find that our conclusions for the example animal in Fig. 2 generalize to the larger cohort of animals. In Fig. 3a, we show the results of model comparison for all 37 animals: the test set loglikelihood is higher for *all* animals for the GLM-HMM compared to for the classic lapse model (to see each curve separately, see Fig. S6). In solid black, we show the mean curve across the population: while test set performance continues to increase as the number of states in the GLM-HMM is increased, the improvement with the addition of each state decreases after 3 states. Furthermore, as can be observed by looking at the 4 state model fits shown in Fig. S9, the interpretation of the retrieved states does not significantly change when moving from the 3 state to 4 state model (the engaged state of the 3 state model simply splits into two engaged states where there are small leftward and rightward biases for the 'difficult' trials); thus we discuss the 3 state model in what follows. Relative to a GLM model (1 state GLM-HMM), the 3 state GLM-HMM results in a 4.2% improvement in test set predictive accuracy (our ability to predict the animal's choice at each trial) on average across the cohort (black solid line in right panel of Fig. 3a). In Fig. 3d, we show the retrieved GLM weights for the 3 state model for each of the 37 animals, and we find that the retrieved states have a similar interpretation across the cohort as 'engaged', 'biased left' and 'biased right'. (We discuss our our method to align states across animals in Comparing states across animals and GLM-HMM parameter initialization). In Fig. 3e, we use the diagonal elements of the retrieved transition matrices to calculate each animal's expected dwell time in state k as $\frac{1}{1-A_{kk}}$ (state dwell times, in the Hidden Markov Model, are assumed to be geometrically distributed; we show this is indeed the case in Fig. S4 and Fig. S8). Since the length of each session is 90 trials, we see further evidence that mice switch between multiple states *within* a session, and that disengaged behavior persists for more than a single trial. In Fig. 3c, we obtain the posterior state probabilities as in Fig. 2e, and plot the fraction of each session in which the mouse is in both states 1 and 2. While there are some sessions (shown in grey) where mice spend the entire session in a single state, the vast majority of sessions (83% of 2017 sessions) involve switching between two or more states. Yet, for 31 out of 37 animals, the overall time spent in the engaged state (shown in red) is greater than 50% of trials.

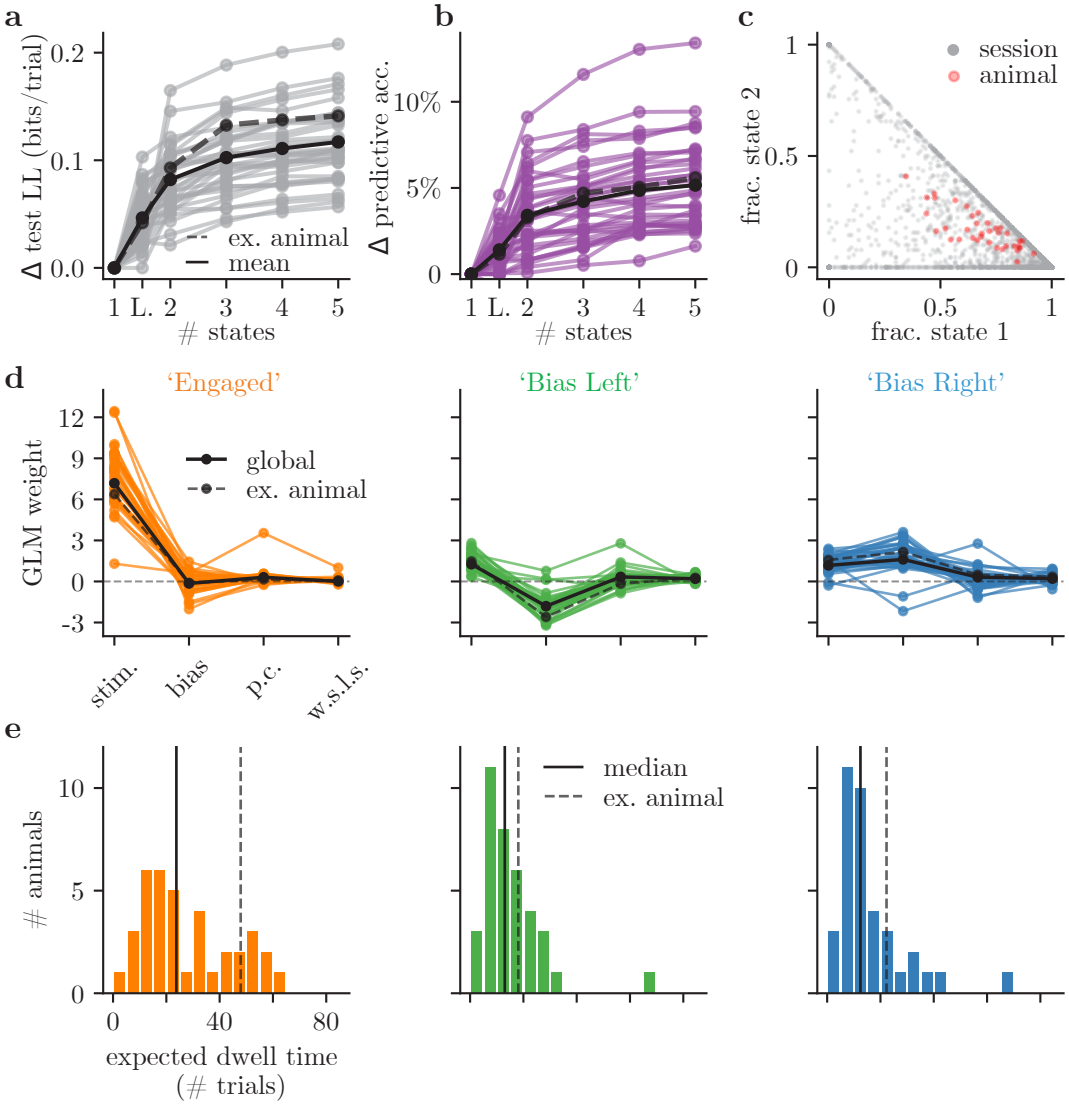

Figure 3: **GLM-HMM application to 37 IBL animals. (a)** We perform 5-fold cross-validation and show the change in test set loglikelihood relative to a GLM for all 37 animals in the IBL cohort for the classic lapse model (here labeled 'L') and for GLM-HMMs with 2-5 states. In solid black, we show the mean across all animals; in dashed black, we show the example animal from Fig. 2. **(b)** For each animal, we show the change in test set predictive accuracy relative to a GLM for each of the models studied. The solid black curve represents the population mean, and the dashed curve is the example animal from Fig. 2. **(c)** For each of the 2017 total sessions of choice data studied, we calculate posterior state probabilities and use these to calculate the fraction of each session spent in states 1 and 2. In red we also plot, for each of the 37 IBL animals, their overall fractional occupancies of states 1 and 2. **(d)** We show the retrieved GLM weights for each animal in the cohort when we fit a separate 3 state GLM-HMM to each animal's choice data. The solid black curve represents the global fit across all animals (see Algorithm 1), while the dashed line is the example animal from Fig. 2. **(e)** From each animal's retrieved transition matrix, we calculate the expected number of trials that the animal will spend in states 1 (orange), 2 (green) and 3 (blue), assuming a geometric distribution for state dwell times (we validate that posterior state dwell times are geometrically distributed in Fig. S8). In solid black, we show the median dwell time across the population for each state, while the dashed black line represents the example animal.

2.5 Mice have discrete – not continuous – decision-making states

While it is clear that the GLM-HMM with long-enduring decision-making states describes the choice data of IBL animals better than the classic lapse model, it is possible that mouse decision-making is governed by continuous, rather than discrete states. For example, one could imagine a mouse's attention gradually increasing or decreasing over time, rather than changing suddenly. In order to address this possibility, we performed model comparison with the PsyTrack model of [39, 40]. PsyTrack assumes that an animal makes its choice according to

$$p(y_t = 1|\mathbf{w}_t, \mathbf{x}_t) = \frac{1}{1 + e^{-\mathbf{w}_t \cdot \mathbf{x}_t}} \tag{5}$$

where $\mathbf{w}_t$ evolves according to

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \eta_t \tag{6}$$

where $\eta_t \sim \mathcal{N}(0, \text{diag}(\sigma_1^2, ..., \sigma_M^2))$ and $\mathbf{w}_0 \sim \mathcal{N}(0, \text{diag}(\sigma_{\text{init},1}^2, ..., \sigma_{\text{init},M}^2))$. Specifically, the animal is assumed to use a set of slowly changing weights to make its decision on each trial. This model is typically used to characterize the learning behavior of mice [39, 40], but we here apply it to model stationary decision-making. While we are able to successfully recover PsyTrack as the best fitting model if we generate choice data from PsyTrack (Fig. S11c), we find that the test set loglikelihood for all 37 IBL animals is considerably larger for a 3 state GLM-HMM compared to for PsyTrack (Fig. S11a). We, thus, believe that a model with discrete decision-making states is a better description of mouse choice behavior than one with continuous decision-making states.

2.6 GLM-HMM simulated data shares statistics of real data

In order to understand why the 3 state GLM-HMM with long-enduring strategies is equipped to describe the choice data of IBL mice, we set out to examine the statistics of GLM-HMM simulated data. Because the GLM-HMM compared to the classic lapse model allows for temporal dependencies in both strategy and choice, a natural statistic to study is choice runlength. In Fig. 4a, we define a choice runlength as the number of trials during which an animal repeats a decision, and we show examples of choice runlengths of 3, 4, and 9 trials. In panel Fig. 4b, we plot the distribution of choice runlengths observed in the 181530 trials of real mouse data (the data from across all 37 IBL animals), and we compare this to the choice runlengths that we would see if these animals performed the task perfectly. Whereas perfect performance results in 98.8% trials being in runlengths lasting 10 trials or less, real choice data results in only 83.7% of trials being in runlengths lasting 10 trials or less. In order to model the choice behavior of real mice, it is thus important to be able to capture this long-tail in the choice runlength distribution. In Fig. 4c, we simulate 100 datasets from three different models: the classic lapse model (with stimulus intensity and bias regressors), the classic lapse model with the addition of history regressors (past choice and win-stay-lose-switch regressors), and a 3 state GLM-HMM with the usual 4 regressors (stimulus intensity, bias, past choice and win-stay-lose-switch). We then subtract the mean choice runlength histogram from across the 100 simulations from the histogram generated from the real data (the red histogram of panel b). We show the subtracted histograms in panel c. Whereas, there are large differences between the histogram generated from the lapse model and the real data histogram, we see that the inclusion of history regressors allows the lapse model to generate longer choice runlengths.

However, better yet is the full GLM-HMM, which is able to generate the choice runlength histogram that is closest to the real data. A related statistic that we can calculate is the number of choice runlengths lasting more than 5 trials. The real data has 6111 runlengths lasting more than 5 trials (dashed line, Fig. 4d) and only the GLM-HMM can generate simulations with as many choice runlengths lasting more than 5 trials as are observed in the real data (for the GLM-HMM, 15 out of 100 simulations have as many choice runlengths lasting 5 or more trials). Thus, compared to the lapse model, the GLM-HMM is better equipped to capture the long choice runlengths observed in real animal choice data.

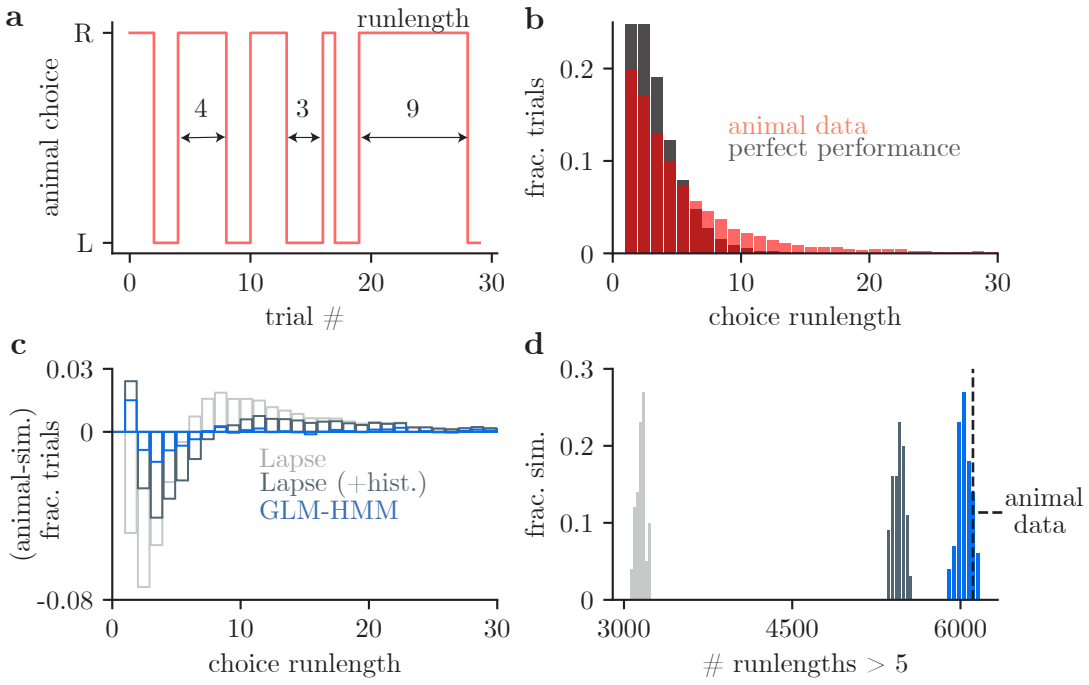

Figure 4: **Simulated data from GLM-HMM captures statistics of real choice data. (a)** We plot the choices that an example IBL animal makes over the course of 30 trials. Choice runlength is defined as the number of trials during which an animal repeats the same decision (we use arrows to highlight choice runlengths of 4, 3 and 9 trials in the above plot). **(b)** For each value of choice runlength, we plot the fraction of trials included in a runlength of that size. If IBL animals perfectly tracked the stimulus, of the 181530 trials that we study, 98.8% of these trials would be in a choice runlength of size 10 or less (black). Yet, real IBL animals often have choice runlengths lasting more than 10 trials (red), and only 83.7% of trials are in a choice runlength of size 10 or less. A model of mouse choice behavior should be able to capture this long tail in the choice runlength distribution. **(c)** We simulated data from each of three models: a lapse model with only stimulus intensity and bias regressors, a lapse model that also has history regressors (past choice and win-stay-lose-switch), and a GLM-HMM with all four regressors (stimulus intensity, bias, past choice and win-stay-lose-switch). We fit each of these models to all data from all animals and used the best fitting parameters to simulate 100 example choice sequences (also of length 181530 trials) for each model. To make the plot shown here, we take the mean histogram across the 100 simulations so as to generate a histogram analogous to the one shown in (b) and we subtract this mean histogram from the red histogram of real choice data shown in (b). If the statistics of the simulated data perfectly resembled real choice data, we would expect the difference in the histograms to be uniformly zero. **(d)** In the 181530 trials of real choice data, there are 6111 runlengths lasting more than 5 trials (as shown with the dashed line). When we simulate choice data according to each of the models shown in (c), we find that only the GLM-HMM can generate simulations with as many runlengths lasting more than 5 trials (15/100 simulations have 6111 or more runlengths lasting more than 5 trials for the GLM-HMM compared to 0/100 for both of the lapse models).

2.7 Mice performing other tasks also switch between long-enduring strategies

In order to check that state-dependent decision-making behavior was not merely a consequence of the structure of the IBL task or IBL training protocol, we investigated the choice behavior of mice performing a different task. In the task of [34], mice face an LED box, and have to indicate if the flash rate is above or below 12Hz by nose poking to the right or to the left (Fig. 5a). Once again, we find that the GLM-HMM compared to the lapse model better describes animal choice behavior (Fig. 5b), with the mean difference in the test set loglikelihood for the 4 state GLM-HMM relative to the lapse model being 0.025. This makes the held-out choice data for each of the 15 animals studied at least 1×10^{18} times more likely under the GLM-HMM compared to the lapse model (each animal has at least 12,000 trials from across 20 sessions, and we perform 5-fold cross-validation). In Fig. 5d, we show the retrieved weights for each animal for the 4 state model. We focus on the 4 state model here because cross-validated performance was better for 10 out of 15 animals for the 4 state model compared to for the 3 state model (see Fig. S14 for each individual curve shown in Fig. 5b), but we show the retrieved parameters for a 3 state model in Fig. S17. In Fig. 5e, we visualize the retrieved GLMs as a function of the stimulus intensity. Specifically, we plot the GLMs associated with the global weights (shown in black in panel d; the result of fitting the GLM-HMM to all data from all animals together. These global weights are then used to initialize the GLM-HMM for each individual animal so as to obtain separate fits for individual animals (see Comparing states across animals and GLM-HMM parameter initialization)). The four curves correspond to all possible combinations of values for the past choice and win-stay-lose-switch covariates. By plotting the retrieved state GLMs in this way, we see that the behavior of mice in each of these four states can be interpreted as 'engaged', 'bias left', 'bias right' and 'win-stay' (where the animal is likely to repeat its choice if the choice was rewarded, but will use the stimulus to make its choice if it was unsuccessful). We can also compare the retrieved GLMs to the psychometric curve that we would get if we concatenated the data from all animals together (Fig. 5c), and we see that when mice are in the engaged state, their accuracy is higher than the average accuracy across all trials (92% compared to 84%). Finally, we obtain the expected dwell times in each of the four states from the retrieved transition matrices. The average session lasts 650 trials, so we see that mice are again switching between strategies within a session. This is also made clear in Fig. S16, where we use the posterior state probabilities to plot the fraction of each session that an animal spends in state 1 compared to the fraction of the session that the animal spends in states 2 or 3. For the 300 sessions (across 15 animals) that we study, only 5 sessions result in an animal staying in the same state for the entire session.

That an animal's task accuracy varies so dramatically between states (from 67% to 92%) may suggest that there are different neural circuits responsible for the behavior observed in each state: both [45] and [16] found vastly different neural signatures for engaged and disengaged behavior, so we might expect as much here too. That animals switch between states within a session may also suggest that a new type of preprocessing of trials is necessary when examining any neural data collected as mice perform these perceptual decision making tasks, and that the GLM-HMM should be used as a principled method for separating out trials corresponding to different modes of decision-making behavior.

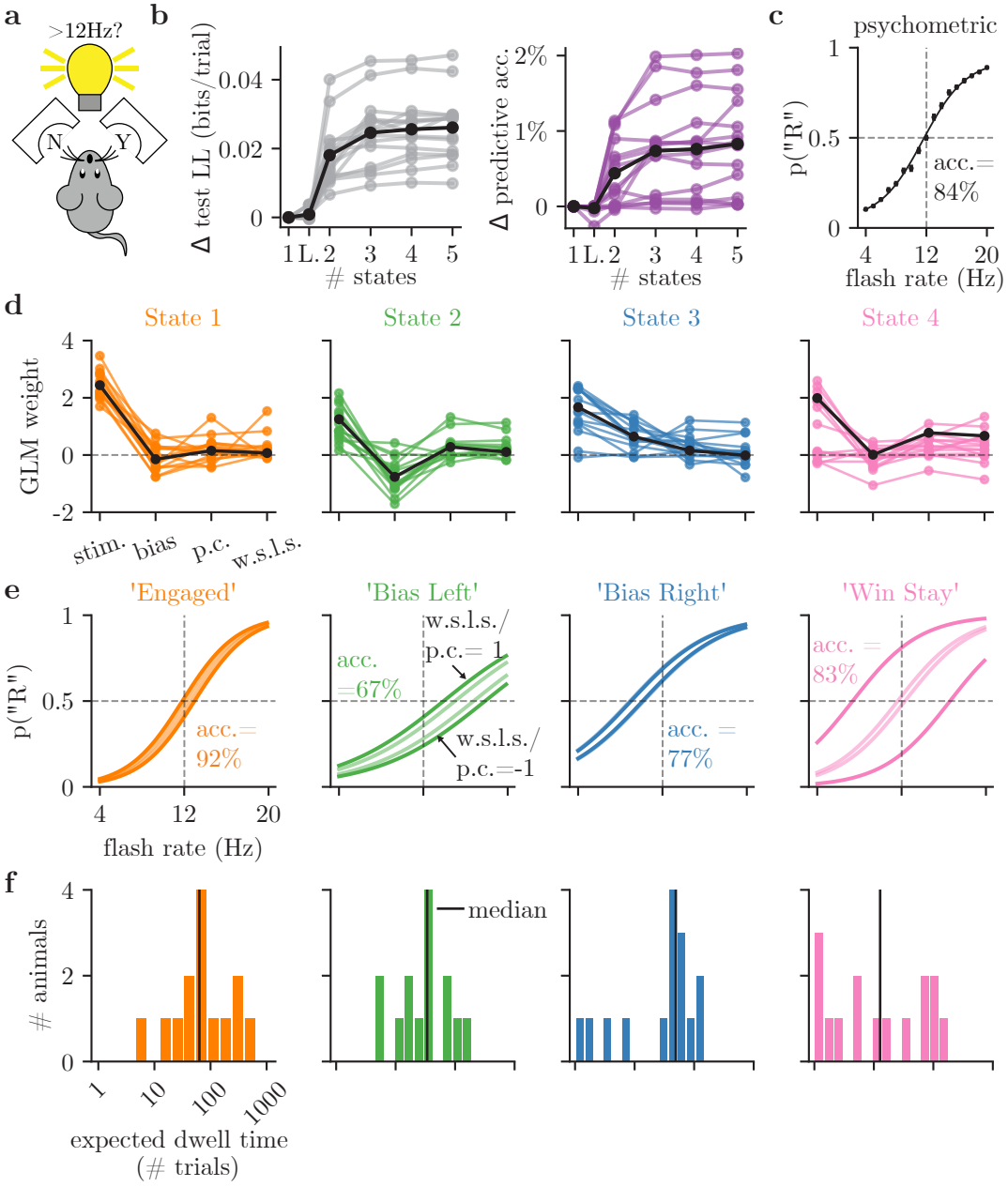

Figure 5: **GLM-HMM application to Odoemene et al. (2018) dataset. (a)** Mice in the Odoemene et al. task have to indicate if the flash rate of light pulses from an LED panel is above or below 12Hz by nose poking to the right or left. **(b)** Left: the test set loglikelihood for 15 animals in this dataset. Black is the mean curve across animals. Right: the change in test set predictive accuracy relative to a GLM for each animal in the dataset. Black is the mean curve across all animals. **(c)** The psychometric curve when all data from all animals are concatenated together. **(d)** The retrieved weights for a 4 state GLM-HMM for each animal. The covariates influencing choice in each of the four states are the stimulus intensity, the animal's bias, its choice on the previous trial ('p.c') and the product of reward and choice on the previous trial ('w.s.l.s.'). Black indicates the global fit across all animals. **(e)** The retrieved GLMs for each state as a function of the stimulus intensity. This is another way to visualize the weights shown in black in panel (d). The four lines correspond to each of the four possible combinations of past choice and win-stay-lose-switch values. Since the weights in all states on these covariates are non-negative, setting the past choice and win-stay-lose-switch regressors to 1 corresponds to the upper curve in the these plots, while setting the past choice and win-stay-lose-switch regressors to -1 results in the lower curves. **(f)** The expected dwell times for each animal in each state, as obtained from the best fitting transition matrices. Black indicates the median dwell time across animals.

2.8 Response times and violation rates are different in engaged and disengaged states

Finally, we wanted to examine how response times and violation rates differed, if at all, in the retrieved states. For humans and monkeys performing perceptual decision-making experiments it has been observed that, for the trials of the same stimulus difficulty, trials resulting in longer reaction times are more likely to be error trials [23, 38, 44]. We, thus, looked at the distributions of response times for the engaged state, compared to for the disengaged states (states 2 and 3), for animals performing the IBL task. Within the IBL task, response time is the time from when the stimulus appears on the screen to when the animal receives feedback on the outcome of its decision (when it receives a reward, or hears an auditory cue to indicate an error trial). The median response time across trials and across IBL animals is just 0.34s, but it is not uncommon for trials to last much longer (up to tens of seconds). We show response time Q-Q plots for each of the 37 IBL animals in Fig. 6a. The engaged and disengaged response time distributions are statistically different for all 37 animals (Komogorov-Smirnov tests reject the null hypothesis with $p < 0.05$ for all 37 animals). Examining the Q-Q plots, it is clear that the most extreme response times for each animal are typically associated with the lower accuracy disengaged states. In Fig. 6b, we show the difference in the 90th quantile response times for the disengaged compared to the engaged states for each animal, and find that, for the majority of animals, 90th quantile response times are longer or the same for the disengaged states compared to for the engaged state. The median difference in the 90th quantile response time, across all animals, is 0.95s (shown in blue) and this is statistically different from 0s (we calculate 95% bootstrap confidence intervals).

In Fig. 6c, we examine the difference in the violation rates for each animal in the Odoemene et al. data set for the disengaged (states 2, 3 and 4) compared to engaged states (state 1). The mean violation rate across all animals and all trials in the dataset is 21% (much higher than in the IBL dataset, where the violation rate was less than 1%), and we find that, across all animals, the violation rate is 3.2% higher in the disengaged states compared to in the engaged state (shown in blue).

Given that no information about response times or violation rates was used to obtain the retrieved GLM-HMM decision-making states, we think that these analyses are a useful external validation that the states that we obtain from the GLM-HMM are semantically meaningful.

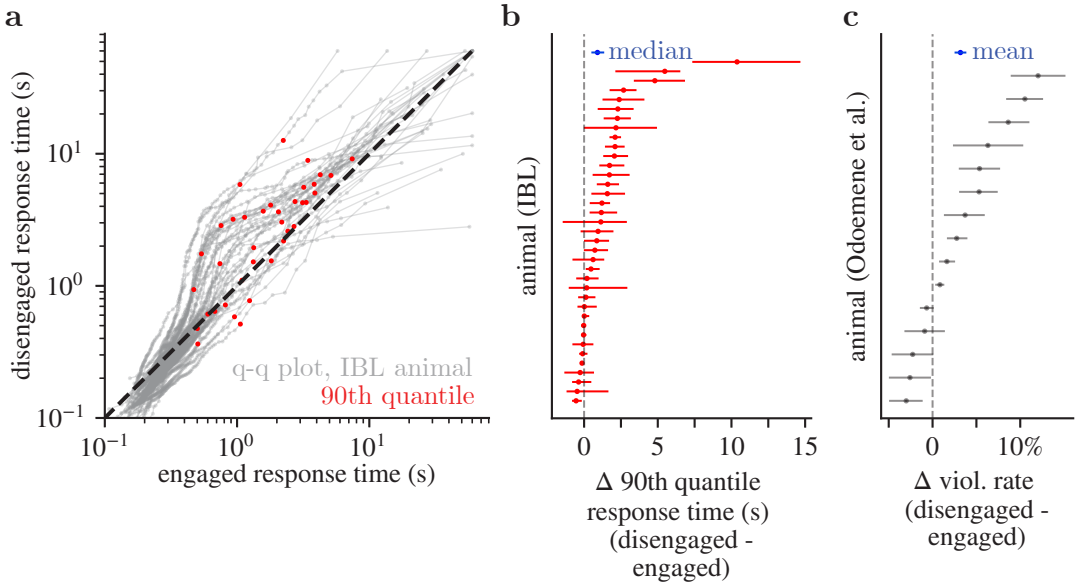

Figure 6: **Behavioral correlates for GLM-HMM states. (a)** We examine the response time distributions for IBL animals associated with the engaged and disengaged states. We assign trials to states using the posterior state probabilities, and examine trials that result in posterior state probabilities greater than 0.9. We show response time Q-Q plots for the engaged (state 1) and disengaged states (states 2 and 3); each curve is an individual animal, and the red dots indicate the 90th quantile response times. For all 37 animals, a KS-test indicates that the engaged and disengaged response time distributions are statistically different. Furthermore, as can readily be observed from the Q-Q plots, the longest response times for each animal are typically associated with the disengaged states. **(b)** We show the difference in the 90th quantile response time for the engaged and disengaged states for each IBL animal, as well as 95% bootstrap confidence intervals. We also show the median difference across animals in blue, as well as the 95% bootstrap confidence interval. Across all animals, the median difference in the 90th quantile response time is almost 1 second (0.95s), but for some animals this can be as large as 10 seconds. **(c)** We show the difference in violation rates for animals in the Odoemene et al. dataset when in the engaged (state 1) and disengaged states (states 2, 3, 4). Error bars are 95% bootstrap confidence intervals, and in blue is the mean difference in violation rate across all animals (3.2%), which is large relative to the mean violation rate across all animals in this dataset (21%).

3 Discussion

In this work, we have used the GLM-HMM framework to identify discrete strategies underlying mouse perceptual decision-making behavior. We showed the classic lapse model corresponds to a particular parameter setting of the GLM-HMM, in which state probability is independent of the state on the previous trial. However, we found that the mouse behavior in two different perceptual decision-making tasks [24, 34] was better described by a GLM-HMM with sustained engaged and disengaged states. Unlike the classic lapse model, these states alternated on the timescale of tens to hundreds of trials.

Additionally, we found that for most IBL animals, the most extreme response times were associated with the disengaged states, and animals performing the Odoemene et al. task had higher violation rates in the less engaged states.

That discrete states underpin mouse choice behavior calls for new normative models to explain why mice may develop these states to begin with. While the disengaged and engaged states may correspond to exploratory and exploitative behavior respectively [12, 13, 36]; the existence of this seemingly suboptimal behavior may instead be the result of an optimal learning strategy as in [33]. We leave the development and testing of normative theories for the behavior that we describe here to a future work. Another exciting future direction will be to parameterize the transition matrix so as to allow transitions between states to depend on external covariates - as in [10]. This will enable us to determine the factors that influence switches between engaged and disengaged behavior and, thus, to potentially control such transitions.

We feel that these results call for a significant reconsideration of rodent perceptual decision-making behavior, and open up a variety of avenues for future research. Crucially, the GLM-HMM can be used to automatically identify switches in strategy, allowing researchers to select trials when the animal is engaged or disengaged for further analysis. Identifying the neural correlates of different states, and of the switches between them, will provide new insights into a previously inaccessible dimension of perceptual behavior.

4 Methods

4.1 Inference of GLM-HMM parameters

4.1.1 GLM-HMM objective function

The parameters of the GLM-HMM, $\Theta \equiv \{\phi, A, \{\mathbf{w}_k\}_{k=1}^K\}$, are fit to empirical animal choice data using Maximum A Posteriori estimation via the Expectation Maximization Algorithm [14]. Specifically, the objective function that we seek to maximize is:

$$f(\Theta|\mathbf{y}, \{\mathbf{x}\}_{t=1}^T) \equiv \log \sum_s \sum_{\mathbf{z}_s} p(\mathbf{y}_s, \mathbf{z}_s|\{\mathbf{x}_{s,t}\}_{t=1}^T, \Theta) + \log p(\Theta) \tag{7}$$

where s indexes the session in which the data was collected (out of S sessions), and the sum over $\mathbf{z}_s$ is over all possible state allocations for trials in that session. $p(\Theta)$ is the prior distribution on the parameters given by:

$$p(\Theta) \equiv p(W)p(A) = \mathcal{N}(W|0, \Sigma)p(A|\alpha) \tag{8}$$

Here, $W \equiv \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_K\end{bmatrix}^\intercal$, $W \in \mathbb{R}^{K\times M}$ (where K is the number of states, and M is the number of covariates (including the bias) used in each state's GLM) is the matrix created by concatenating the weight vectors for each state together. $\Sigma \equiv \sigma \times \mathbb{1}$, $\Sigma \in \mathbb{R}^{KM\times KM}$ is the prior covariance matrix parametrized by a single σ value. The prior distribution that we consider for the transition matrix is the Dirichlet distribution parametrized by a single α value: $p(A|\alpha) = \prod_{j=1}^K \frac{1}{B(\alpha)} \prod_{k=1}^K A_{jk}^{\alpha-1}$. In order to select the prior hyperparameters for a given dataset, we perform a grid search for $\sigma \in \{0.5, 0.75, 1, 2, 3\}$

and $\alpha \in \{1, 2\}$ and select the set of hyperparameters that result in the best performance on a held-out validation set. For IBL animals, the prior hyperparameters selected were $\sigma = 2$ and $\alpha = 2$. While for mice in the Odoemene et al. dataset, the best fitting hyperparameters were $\sigma = 0.75$ and $\alpha = 2$.

4.1.2 Expectation Maximization Algorithm

We employ the Expectation Maximization iterative algorithm [14, 31] in order to maximize the objective function, $f(\Theta|\mathbf{y}, \{\mathbf{x}\}_{t=1}^{T})$, given in eq. 7 with respect to the GLM-HMM parameters. The Expectation Maximization algorithm involves forming the Expected Complete Loglikelihood (ECLL) during the E-step using the forward-backward algorithm [2]. The ECLL is a lower bound on $f(\Theta|\mathbf{y}, \{\mathbf{x}\}_{t=1}^{T})$ [5, 14, 31], which is then maximized with respect to Θ during the M-Step. Concretely, the Expected Complete Loglikelihood is:

$$\begin{aligned}
\text{ECLL}(\Theta) &\equiv \sum_{s} \sum_{\mathbf{z}_s} p\big(\mathbf{z}_s|\mathbf{y}_s, \{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta^{\text{old}}\big) \log p\big(\mathbf{z}_s, \mathbf{y}_s|\Theta, \{\mathbf{x}_{s,t}\}_{t=1}^{T}\big) + \log p(\Theta) \\
&= \sum_{s=1}^{S} \Bigg[\sum_{k=1}^{K} \gamma_{s,1,k} \log \phi_k + \sum_{t=1}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi_{s,t,j,k} \log A_{jk} + \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{s,t,k} \log p(y_{s,t}|z_{s,t} = k, \mathbf{x}_{s,t}, \mathbf{w}_k) \Bigg] \\
&+ \log p(\Theta)
\end{aligned} \tag{9}$$

where we substituted the definition of the joint distribution (eq. 4) in order to get to the second line. $\gamma_{s,t,k} \equiv p\big(z_{s,t} = k|\mathbf{y}_s, \{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta^{\text{old}}\big)$ and $\xi_{s,t,j,k} \equiv p\big(z_{s,t+1} = k|z_{s,t} = j, \mathbf{y}_s, \{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta^{\text{old}}\big)$ are the quantities estimated through the forward-backward algorithm, as we will describe below; $p(y_{s,t}|z_{s,t} = k, \mathbf{x}_{s,t}, \mathbf{w}_k)$ is obtained from eq. 2.

4.1.3 Expectation Step

During the E-Step of the EM algorithm, the quantities $\{\gamma_{s,t,k}\}$ and $\{\xi_{s,t,j,k}\}$ are estimated with the forward-backward algorithm [2] at the current setting of the GLM-HMM parameters, Θ^{old}. In particular, the forward-backward algorithm involves calculating $a_{s,t,k} \equiv p\big(\mathbf{y}_{s,[0:t]}, z_{s,t} = k|\{\mathbf{x}_{t'}\}_{t'=1}^{t}\big)$ and $b_{s,t,k} \equiv p\big(\mathbf{y}_{s,[t+1:T]}|z_{s,t} = k, \{\mathbf{x}_{t'}\}_{t'=t+1}^{T}\big)$ so as to be able to calculate $\gamma_{s,t,k}$ and $\xi_{s,t,j,k}$ as follows:

$$\begin{aligned}
\gamma_{s,t,k} &\equiv p\big(z_{s,t} = k|\mathbf{y}_s, \{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta^{\text{old}}\big) \\
&= \frac{p\big(\mathbf{y}_{s,[0:t]}, z_{s,t} = k|\{\mathbf{x}_{s,t'}\}_{t'=1}^{t}, \Theta^{\text{old}}\big) p\big(\mathbf{y}_{s,[t+1:T]}|z_{s,t} = k, \{\mathbf{x}_{s,t'}\}_{t'=t+1}^{T}, \Theta^{\text{old}}\big)}{p\big(\mathbf{y}_s|\{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta^{\text{old}}\big)} \\
&= \frac{a_{s,t,k} b_{s,t,k}}{\sum_{k=1}^{K} a_{s,T,k}}
\end{aligned} \tag{10}$$

In an analogous fashion,

$$\xi_{s,t,j,k} = \frac{a_{s,t,j} A_{jk} b_{s,t+1,k} p\big(y_{s,t+1}|z_{t+1} = k, \mathbf{x}_{t+1}, \mathbf{w}_k\big)}{\sum_{k=1}^{K} a_{s,T,k}} \tag{11}$$

where, once again, $p\big(y_{s,t+1}|z_{t+1}=k,\mathbf{x}_{t+1},\mathbf{w}_k\big)$ comes from eq. 2.

The forward-backward algorithm is a recursive algorithm and calculates the forward probabilities, $\{a_{s,t,k}\}$, as follows:

$$a_{s,1,k}=\phi_k p\big(y_{s,1}|z_{s,1}=k,\mathbf{x}_{s,1},\mathbf{w}_k\big) \tag{12}$$

and, for $1<t\leq T$:

$$a_{s,t,k}=\sum_{j=1}^{K}a_{s,t-1,j}A_{jk}p\big(y_{s,t}|z_{s,t}=k,\mathbf{x}_{s,t},\mathbf{w}_k\big) \tag{13}$$

Similarly, the backward probabilities are calculated recursively as follows:

$$b_{s,T,k}=1 \tag{14}$$

and, for $t\in\{T-1,...,1\}$:

$$b_{s,t,j}=\sum_{k=1}^{K}b_{s,t+1,k}A_{jk}p\big(y_{s,t+1}|z_{s,t+1}=k,\mathbf{x}_{s,t+1},\mathbf{w}_k\big) \tag{15}$$

The advantage of using the forward-backward algorithm is that it enables evaluation of the sum over all possible state allocations in eq. 7 and eq. 9 in linear time in the number of trials, rather than in exponential time.

4.1.4 Maximization Step

After running the forward-backward algorithm, we form the ECLL as in eq. 9 and then maximize it with respect to the GLM-HMM parameters, Θ. For the initial state distribution ϕ and the transition matrix A, this results in closed-form updates:

$$\phi_k^{\text{new}}=\frac{\sum_{s=1}^{S}\gamma_{s,1,k}}{\sum_{s=1}^{S}\sum_{j=1}^{K}\gamma_{s,1,j}} \tag{16}$$

$$A_{jk}^{\text{new}}=\frac{\alpha-1+\sum_{s=1}^{S}\sum_{t=2}^{T}\xi_{s,t,j,k}}{K(\alpha-1)+\sum_{s=1}^{S}\sum_{t=2}^{T}\sum_{k=1}^{K}\xi_{s,t,j,k}} \tag{17}$$

For the GLM weights, there is no such closed form update, but a Bernoulli GLM falls into the class of functions mapping external inputs to HMM emission probabilities considered in [17], so we know that the ECLL is concave in the GLM weights. As such, we are able to numerically find the GLM weights that maximize (not just locally but globally) the ECLL using the BFGS algorithm [6, 18, 20, 42] as implemented by the scipy optimize function in python [46].

4.1.5 Comparing states across animals and GLM-HMM parameter initialization

In Fig. 3 and Fig. 5, we show the results from fitting a single GLM-HMM to each animal; however it is nontrivial to map the retrieved states across animals to one another. As such, we employ a multistage

fitting procedure that allows us make this comparison, and we detail this procedure in Algorithm 1. In the first stage, we concatenate the data for all animals in a single dataset together (for example, in the case of the IBL dataset, this would be the data for all 37 animals). We then fit a GLM (a 1 state GLM-HMM) to the concatenated data using Maximum Likelihood estimation. We use the fit GLM weights to initialize the GLM weights of a K-state GLM-HMM that we again fit to the concatenated dataset from all animals together (to obtain a "global fit"). We add Gaussian noise with $\sigma_{\text{init}} = 0.2$ to the the GLM weights, so that the initialized states are distinct, and we initialize the transition matrix of the K-State GLM-HMM as $0.95 \times \mathbb{1} + \mathcal{N}(0, \Sigma_{\text{trans.}})$ where $\Sigma_{\text{trans.}} \in \mathbb{R}^{K^2 \times K^2}$ and $\Sigma_{\text{trans.}} = 0.05 \times \mathbb{1}$. We then normalize this so that that rows of the transition matrix add up to 1, and represent probabilities. While the EM algorithm is guaranteed to converge to a local optimum in the log probability landscape of eq. 7, there is no guarantee that it will converge to the global optimum [41]. Correspondingly, for each value of K, we fit the model 20 times using 20 different initializations.

In the next stage of the fitting procedure, we wish to obtain a separate GLM-HMM fit for each animal, so we initialize a model for each animal with the GLM-HMM global fit parameters from all animals together (out of the 20 initializations, we choose the model that resulted in the best training set loglikelihood). We then run the EM algorithm to convergence, and the recovered parameters are the ones that we plot in Fig. 3 and Fig. 5. By initializing each individual animal's model with the parameters from the fit to all animals together, we no longer need to permute the retrieved states from each animal so as to map semantically similar states to one another.

Algorithm 1 Multistage GLM-HMM fitting procedure

```
Fit GLM (1 state GLM-HMM) to all data from all animals
Fit global GLM-HMM to all data from all animals:
for K ∈ {2, ..., 5} do
    for init. ∈ {1, ..., 20} do
        Initialize K-state GLM-HMM using noisy GLM weights
        Run EM algorithm until convergence
    end for
end for
Fit separate GLM-HMM to each animal by initializing with global fit:
for each individual animal do
    for K ∈ {2, ..., 5} do
        Initialize K-state GLM-HMM using best global GLM-HMM parameters for this K
        Run EM algorithm until convergence
    end for
end for
```

We note that the initialization scheme detailed above is sufficiently robust so as to allow recovery of GLM-HMM parameters in various parameter regimes of interest. In particular, we simulate datasets of comparable sizes to those that inidividual animals have from a GLM-HMM with the global fit parameters from across all IBL animals, the global fit parameters from across all Odoemene et al. animals, and a global fit lapse model. We show the generative and recovered parameters in Fig. S1 and Fig. S2.

4.2 Assessing Model Performance

4.2.1 Cross Validation

There are two ways in which to perform cross-validation when working with Hidden Markov Models. Firstly, it is possible to hold out entire sessions of choice data for assessing test-set performance. That is, when fitting the model, the objective function in eq. 7 and the ECLL in eq. 9 are modified to only include 80% of sessions (since we use 5-fold cross-validation throughout this work); and the loglikelihood of the held-out 20% of sessions is calculated using the fit parameters, and a single run of the forward-pass on the held-out sessions:

$$\text{LL}_{\text{test}} \equiv \sum_{s \in S \setminus S'} \log p\big(\mathbf{y}_s | \{\mathbf{x}_{s,t}\}_{t=1}^{T}, \Theta'\big) = \log \sum_{s \in S'} \sum_{k=1}^{K} a_{s,T,k} \tag{18}$$

where $S \setminus S'$ is the set of held out sessions, and Θ' is the set of GLM-HMM parameters obtained by fitting the model using the trials from S'.

The second method of performing cross-validation involves holding out 20% of trials within a session. When fitting the model, the third term in the ECLL is modified so as to exclude these trials and is now $\sum_{t \in T'} \sum_{k=1}^{K} \gamma_{s,t,k} \log p\big(y_{s,t,k} | z_{s,t} = k, \mathbf{x}_{s,t}\mathbf{w}_k\big)$, where T' is the set of trials to be used to fit the model. Furthermore, the calculation of the posterior state probabilities, $\gamma_{s,t,k}$ and $\xi_{s,t,j,k}$, is also modified so as to exclude the test set choice data. In particular, $\gamma_{s,t,k}$ is now $p\big(z_{s,t} | \{y_{s,t'}\}_{t' \in T'}, \{\mathbf{x}_{s,t'}\}_{t' \in T'}, \Theta^{\text{old}}\big)$ and similarly $\xi_{s,t,j,k}$ is now $p\big(z_{s,t+1} = k | z_{s,t} = j, \{y_{s,t'}\}_{t' \in T'}, \{\mathbf{x}_{s,t'}\}_{t' \in T'}, \Theta^{\text{old}}\big)$. The method of calculating these modified posterior probabilities is as detailed in eq. 10 and eq. 11, but now the calculation of the forward and backward probabilities, $a_{s,t,k}$ and $b_{s,t,k}$ in eq. 12, eq. 13, eq. 14 and eq. 15 is modified so that, on trials that are identified as test trials, the $p\big(y_{s,t} | z_{s,t} = k, \mathbf{x}_{s,t}, \mathbf{w}_k\big)$ term in these equations is replaced with 1.

In Fig. 2, Fig. 3 and Fig. 5, we perform cross-validation by holding out entire sessions. Our intuition was that it would be harder to make good predictions on entire held out sessions, compared to single trials within a session, since we thought that mice would exhibit more variability in behavior across sessions compared to within sessions. When we compare the performance of the GLM-HMM against the PsyTrack model of [39] in Fig. S11, we use the second method of cross-validation so as to use the same train and test sets as PsyTrack (PsyTrack cannot make predictions on entire held-out sessions).

4.2.2 Normalized Loglikelihood

In Fig. 2, Fig. 3 and Fig. 5 we report the normalized loglikelihood of different models on held-out sessions. This is calculated as follows:

$$\text{NLL}(\Theta) = \frac{\text{LL}_{\text{test}} - \text{LL}_0}{n_{\text{test}} \log(2)} \tag{19}$$

where, for the GLM-HMM, LL_{test} is the test set loglikelihood as calculated in eq. 18, and LL_0 is the loglikelihood of the same test set under a Bernoulli model of animal choice behavior. Specifically, this baseline model assumes that animals flip a coin on each trial so as to decide to go Right, and the

probability of going Right is equal to the fraction of trials in the training set in which the animal chose to go to the Right. n_{test} is the number of trials in the test set, and is important to include since LL_{test} depends on the number of trials in the test set. Dividing by $\log(2)$ gives the Normalized Loglikelihood the units of bits per trial. Clearly, larger values of the Normalized Loglikelihood are better, with a value of 0 indicating that a model offers no improvement in prediction compared to the crude baseline model described above. However, even small values of normalized loglikelihood can indicate a large improvement in predictive power. For a test set size of $n_{\text{test}} = 500$, a normalized loglikelihood value of 0.01 indicates that the test data is 31.5 times more likely to have been generated with the GLM-HMM compared to the baseline model. For a test set of $n_{\text{test}} = 5000$, and the same value of NLL, the test set becomes 1×10^{15} times more likely under the GLM-HMM compared to the baseline model!

4.2.3 Predictive Accuracy

In Fig. 2, Fig. 3 and Fig. 5, we also report the predictive accuracy of the GLM-HMM. When calculating the predictive accuracy, we employ a method similar to the second method described above in the Cross Validation section. In particular, we hold out 20% of trials and then obtain the posterior state probabilities for these trials, $t'' \in \{T \setminus T'\}$, as $\gamma_{s,t'',k} = p\big(z_{s,t''}|\{y_{s,t'}\}_{t'\in T'}, \{\mathbf{x}_{s,t'}\}_{t'\in T'}, \Theta\big)$, using the other 80% of trials (this latter set of trials being labeled T'). We then calculate the probability of the held-out choices being to go Right as:

$$p_{R,s,t''} \equiv p\big(y_{s,t''} = 1|\{y_{s,t'}\}_{t'\in T'}, \{\mathbf{x}_{s,t'}\}_{t'\in T'}, \mathbf{x}_{s,t''}, \Theta\big) = \sum_{k=1}^{k} \big(\gamma_{s,t'',k}\big) p\big(y_{s,t''} = 1|z_{s,t''} = k, \mathbf{x}_{s,t''}, \mathbf{w}_k\big) \tag{20}$$

We then calculate the predictive accuracy as:

$$\text{predictive accuracy} = \frac{\sum_{t''\in T\setminus T'} \Big(\big(\mathbb{1}\big(y_{s,t''} = 1\big) \mathbb{1}\big(p_{R,s,t''} > 0.5\big) \big) + \big(\mathbb{1}\big(y_{s,t''} = 0\big) \mathbb{1}\big(p_{R,s,t''} \leq 0.5\big) \big) \Big)}{|\{T \setminus T'\}|} \tag{21}$$

4.3 Datasets studied

In this paper, we apply the GLM-HMM to two publicly available behavioral datasets associated with recent publications. Firstly, we study the data associated with [24] that is made available via figshare at https://doi.org/10.6084/m9.figshare.11636748. We use the framework developed in [25] to access the data. We model the choice data for the 37 animals in this dataset which have more than 30 sessions of data during the 'biased-block' regime. We focus on this regime because of the fact that mice, when they have reached this regime, understand the rules of the task and exhibit stationary behavior. For each session, we subset to the first 90 trials of data because, during these trials, the stimulus is equally likely to appear on the left or right of the screen. After the first 90 trials, the structure of the task changes and for a block of trials, the stimulus appears on the left with a probability of either

80% or 20%; the block probability then switches between 80% and 20% multiple times throughout the session. We subset to the animals with more than 30 sessions of data because we are able to confidently recover GLM-HMM and lapse model generative parameters when we simulate datasets with this number of trials: see Figures S1 and S2. As a sanity check to make sure that the recovered states and transitions are not a consequence of the animals we study having been exposed to biased blocks in earlier sessions, we obtained data for 4 animals that were never exposed to biased blocks (not included in the publicly released dataset) and fit the GLM-HMM to the choice data for these animals. In Fig. S10, we show that the retrieved states, dwell times and model comparison results for these animals look very similar to those shown in Fig. 3.

The second dataset that we study is that associated with [34], with the data being made available at `https://doi.org/10.14224/1.38944`. For this dataset, the retrieved states are less distinct compared to those for the IBL dataset, and as such, we require more trials to be able to recover the generative parameters in simulated data: see Fig. S1. We thus subset to the 15 animals with more than 20 sessions of data and 12,000 trials of data. Compared to the IBL dataset, where the violation rate across all animals' data is less than 1% of trials (where a violation is where the animal chooses not to respond), the violation rate across the 15 animals that we study from this second dataset is 21%. Thus, it's important to develop a principled method for dealing with violation trials. We treat violation trials as trials with missing choice data, and we handle these trials in a similar way to how we handle test data when performing the second type of cross-validation described in the Cross Validation section above. That is, we modify the third term of the ECLL given in eq. 9 to exclude violation trials, and we modify the definition of the posterior state probabilities for these trials to be $\gamma_{s,t,k} = p\big(z_{s,t}|\{y_{s,t'}\}_{t'\in T'}, \{\mathbf{x}_{s,t'}\}_{t'\in T'}, \Theta^{\text{old}}\big)$ and $\xi_{s,t,j,k} = p\big(z_{s,t+1} = k|z_{s,t} = j, \{y_{s,t'}\}_{t'\in T'}, \{\mathbf{x}_{s,t'}\}_{t'\in T'}, \Theta^{\text{old}}\big)$, where T', rather than representing the training set data, is now the set of non-violation trials. The calculation of the forward and backward probabilities, $a_{s,t,k}$ and $b_{s,t,k}$, is modified so that, on violation trials, in eq. 12, eq. 13, eq. 14 and eq. 15, the $p\big(y_{s,t}|z_{s,t} = k, \mathbf{x}_{s,t}, \mathbf{w}_k\big)$ term is replaced with 1.

5 Acknowledgments

We are grateful to Miles Wells, Rebecca Terry and the Cortexlab at University College London for providing us with the data for the 4 mice plotted in Fig. S10. We are grateful to Scott Linderman for developing the beautiful Bayesian State Space Modeling framework of [29]: we built our code on top of this framework. We thank members of the Pillow lab, the International Brain Laboratory (IBL), and specifically the Behavior Analysis Working Group within the IBL for helpful feedback throughout the project. We thank Peter Dayan, Sebastian Bruijns and Liam Paninski for acting as the IBL Review Board for this paper. We thank Anne Urai and Emily Dennis for their feedback at various points during the project. We thank Abigail Russo and Matthew Whiteway for providing feedback on drafts of this manuscript.

References

[1] Batty, E., Whiteway, M., Saxena, S., Biderman, D., Abe, T., Musall, S., Gillis, W., Markowitz, J., Churchland, A., Cunningham, J. P., Datta, S. R., Linderman, S., and Paninski, L. (2019). BehaveNet: nonlinear embedding and Bayesian neural decoding of behavioral videos. In Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F. d., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32*, pages 15706–15717. Curran Associates, Inc.

[2] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Annals of Mathematical Statistics*, 41(1):164–171.

[3] Bengio, Y. and Frasconi, P. (1995). An Input Output HMM Architecture. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems 7*, pages 427–434. MIT Press.

[4] Berman, G. J. (2018). Measuring behavior across scales. *BMC Biology*, 16(1):23.

[5] Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

[6] Broyden, C. G. (1970). The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90. Publisher: Oxford Academic.

[7] Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science (New York, N.Y.)*, 340(6128):95–98.

[8] Burgess, C. P., Lak, A., Steinmetz, N. A., Zatka-Haas, P., Bai Reddy, C., Jacobs, E. A. K., Linden, J. F., Paton, J. J., Ranson, A., Schröder, S., Soares, S., Wells, M. J., Wool, L. E., Harris, K. D., and Carandini, M. (2017). High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Reports*, 20(10):2513–2524.

[9] Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Schölvinck, M. L., Zaharia, A. D., and Carandini, M. (2011). The Detection of Visual Contrast in the Behaving Mouse. *Journal of Neuroscience*, 31(31):11351–11361.

[10] Calhoun, A. J., Pillow, J. W., and Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. *Nature Neuroscience*, 22(12):2040–2049. Number: 12 Publisher: Nature Publishing Group.

[11] Carandini, M. and Churchland, A. K. (2013). Probing perceptual decisions in rodents. *Nature Neuroscience*, 16(7):824–831.

[12] Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879.

[13] Dayan, P. and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4):429–453. Publisher: Springer.

[14] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

[15] Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern Classification*. John Wiley & Sons. Google-Books-ID: Br33IRC3PkQC.

[16] Engel, T. A., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., and Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144. Publisher: American Association for the Advancement of Science Section: Report.

[17] Escola, S., Fontanini, A., Katz, D., and Paninski, L. (2011). Hidden Markov models for the stimulus-response relationships of multistate neural systems. *Neural computation*, 23(5):1071–1132.

[18] Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322. Publisher: Oxford Academic.

[19] Fründ, I., Wichmann, F. A., and Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision*, 14(7):9–9. Publisher: The Association for Research in Vision and Ophthalmology.

[20] Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26.

[21] Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., and Mainen, Z. F. (2014). Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, 17(11):1455–1462. Number: 11 Publisher: Nature Publishing Group.

[22] Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics*, volume 1. Wiley New York.

[23] Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8. Publisher: Frontiers.

[24] International Brain Laboratory, T., Aguillon-Rodriguez, V., Angelaki, D. E., Bayer, H. M., Bonacchi, N., Carandini, M., Cazettes, F., Chapuis, G. A., Churchland, A. K., Dan, Y., Dewitt, E. E., Faulkner, M., Forrest, H., Haetzel, L. M., Hausser, M., Hofer, S. B., Hu, F., Khanal, A., Krasniak, C. S., Laranjeira, I., Mainen, Z. F., Meijer, G. T., Miska, N. J., Mrsic-Flogel, T. D., Murakami, M., Noel, J.-P., Pan-Vazquez, A., Sanders, J. I., Socha, K. Z., Terry, R., Urai, A. E., Vergara, H. M., Wells, M. J., Wilson, C. J., Witten, I. B., Wool, L. E., and Zador, A. (2020a). A standardized and reproducible method to measure decision-making in mice. *bioRxiv*, page 2020.01.17.909838.

[25] International Brain Laboratory, T., Bonacchi, N., Chapuis, G., Churchland, A., Harris, K. D., Hunter, M., Rossant, C., Sasaki, M., Shen, S., Steinmetz, N. A., Walker, E. Y., Winter, O., and Wells, M. (2020b). Data architecture for a large-scale neuroscience collaboration. *bioRxiv*. Publisher: Cold Spring Harbor Laboratory _eprint: https://www.biorxiv.org/content/early/2020/02/23/827873.full.pdf.

[26] Johnson, R. E., Linderman, S., Panier, T., Wee, C. L., Song, E., Herrera, K. J., Miller, A., and Engert, F. (2020). Probabilistic Models of Larval Zebrafish Behavior Reveal Structure on Many Scales. *Current Biology*, 30(1):70–82.e4.

[27] Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, 93(3):480–490.

[28] Law, C.-T. and Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neuroscience*, 12(5):655–663. Number: 5 Publisher: Nature Publishing Group.

[29] Linderman, S., Antin, B., and Zoltowski, D. (2020). SSM: Bayesian learning and inference for state space models.

[30] Lindsay, B. G. (1995). Mixture Models: Theory, Geometry and Applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 5:i–163. Publisher: Institute of Mathematical Statistics.

[31] McLachlan, G. J. and Krishnan, T. (2007). *The EM Algorithm and Extensions*. John Wiley & Sons. Google-Books-ID: NBawzaWoWa8C.

[32] McLachlan, G. J. and Peel, D. (2004). *Finite Mixture Models*. John Wiley & Sons. Google-Books-ID: c2_fAox0DQoC.

[33] Mendonça, A. G., Drugowitsch, J., Vicente, M. I., DeWitt, E. E. J., Pouget, A., and Mainen, Z. F. (2020). The impact of learning on perceptual decisions and its implication for speed-accuracy tradeoffs. *Nature Communications*, 11(1):2757. Number: 1 Publisher: Nature Publishing Group.

[34] Odoemene, O., Pisupati, S., Nguyen, H., and Churchland, A. K. (2018). Visual Evidence Accumulation Guides Decision-Making in Unrestrained Mice. *Journal of Neuroscience*, 38(47):10143–10155.

[35] Pinto, L., Koay, S. A., Engelhard, B., Yoon, A. M., Deverett, B., Thiberge, S. Y., Witten, I. B., Tank, D. W., and Brody, C. D. (2018). An Accumulation-of-Evidence Task Using Visual Pulses for Mice Navigating in Virtual Reality. *Frontiers in Behavioral Neuroscience*, 12. Publisher: Frontiers.

[36] Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., and Churchland, A. K. (2019). Lapses in perceptual decisions reflect exploration. *bioRxiv*, page 613828. Publisher: Cold Spring Harbor Laboratory.

[37] Prins, N. (2012). The psychometric function: The lapse rate revisited. *Journal of Vision*, 12(6):25–25.

[38] Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in cognitive sciences*, 20(4):260–281.

[39] Roy, N. A., Bak, J. H., Akrami, A., Brody, C., and Pillow, J. W. (2018). Efficient inference for time-varying behavior during learning. In *Advances in Neural Information Processing Systems*, pages 5695–5705.

[40] Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D., and Pillow, J. W. (2020). Extracting the Dynamics of Behavior in Decision-Making Experiments. *bioRxiv*, page 2020.05.21.109678.

[41] Salakhutdinov, R., Roweis, S. T., and Ghahramani, Z. (2003). Optimization with EM and expectation-conjugate-gradient. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 672–679.

[42] Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656.

[43] Sharma, A., Johnson, R., Engert, F., and Linderman, S. (2018). Point process latent variable models of larval zebrafish behavior. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 10919–10930. Curran Associates, Inc.

[44] Shevinsky, C. A. and Reinagel, P. (2019). The Interaction Between Elapsed Time and Decision Accuracy Differs Between Humans and Rats. *Frontiers in Neuroscience*, 13. Publisher: Frontiers.

[45] Steinmetz, N. A., Zatka-Haas, P., Carandini, M., and Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786):266–273. Number: 7786 Publisher: Nature Publishing Group.

[46] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

[47] Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8):1293–1313.

[48] Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abraira, V. E., Adams, R. P., and Datta, S. R. (2015). Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, 88(6):1121–1135. Publisher: Elsevier.

[49] Yu, S.-Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, 174(2):215–243.

Supplementary Information

"Mice alternate between discrete strategies during perceptual decision-making"

Ashwood *et al*, 2020

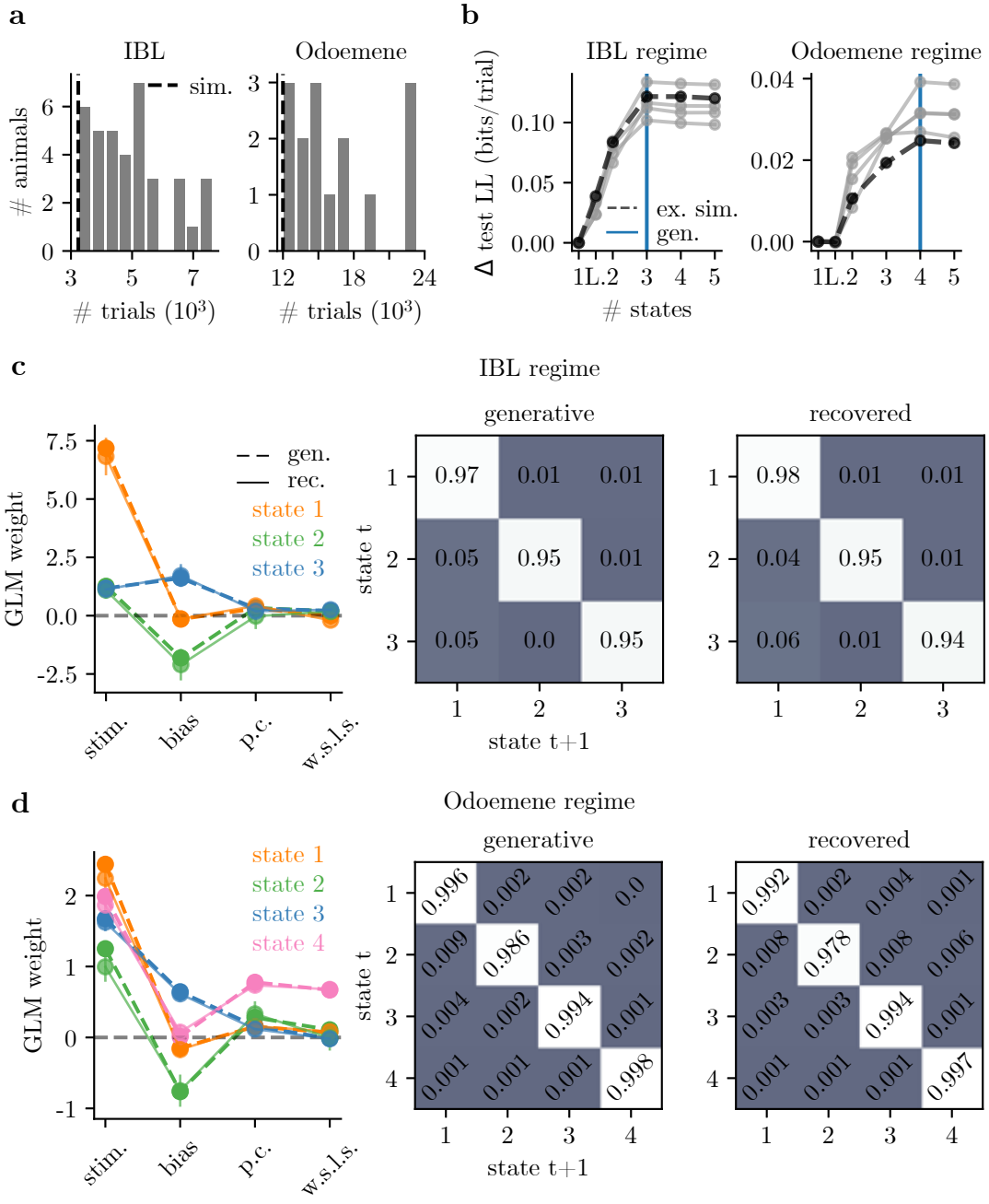

Figure S1: **GLM-HMM Recovery Analysis 1** For dataset sizes comparable to those of real animals, we can recover the IBL and Odoemene global parameters in simulated data. **(a)** Dataset sizes for each of the 37 IBL animals studied (left) and 15 mice from Odoemene et al. (right). The dashed vertical line indicates the number of trials that we used in simulation data in panels b, c and d (3240 for the IBL parameter regime and 12000 for the Odoemene regime simulation). **(b)** Normalized test loglikelihood for each of 5 simulations is maximized at 3 states (blue vertical line) after we simulate according to the (IBL regime) parameters shown in panel c. Similarly, in the right panel, normalized test loglikelihood is maximized at 4 states when we simulate choice data with the (Odoemene regime) parameters shown in panel d. **(c)** Left: the generative and recovered GLM weights (for the simulation shown with the dashed line in panel b) when we simulate choice data in the IBL parameter regime. Middle and right: the generative and recovered transition matrices. **(d)** The generative and recovered parameters in the Odoemene et al. parameter regime.

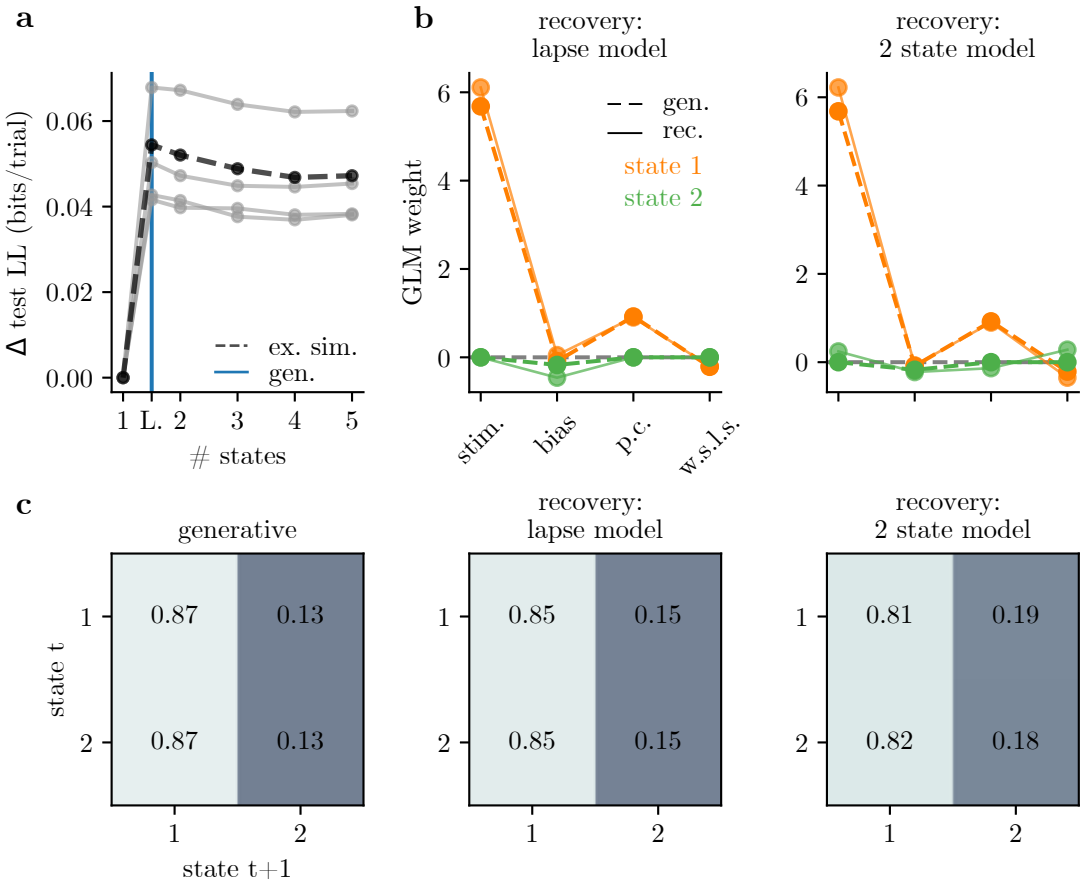

Figure S2: **GLM-HMM Recovery Analysis 2: We can recover lapse behavior. (a)** We simulate 5 datasets, each with 3240 trials, according to the best fitting lapse model for IBL animals. We then fit these simulated datasets with GLM-HMM models, as well as a lapse model (a constrained 2 state GLM-HMM). The normalized test loglikelihood is highest for the lapse model in all simulations, indicating that lapse behavior can be distinguished from the long-enduring multi-state behavior that best described the real data. **(b)** Left: the generative and recovered weights when recovery is with a lapse model. Right: the generative weights are the same as in the left panel, but we now recover with an unconstrained 2 state GLM-HMM (thus the stimulus, past choice and w.s.l.s. weights for the second state can be non-zero) **(c)** The generative (left) transition matrix and the recovered transition matrices when we recover with a lapse model (middle) and an unconstrained 2 state GLM-HMM (right). While the lapse model and 2 state GLM-HMM results don't perfectly agree, if mice were truly lapsing, the transition matrix would not have the large entries on the diagonals that we observe in the real data.

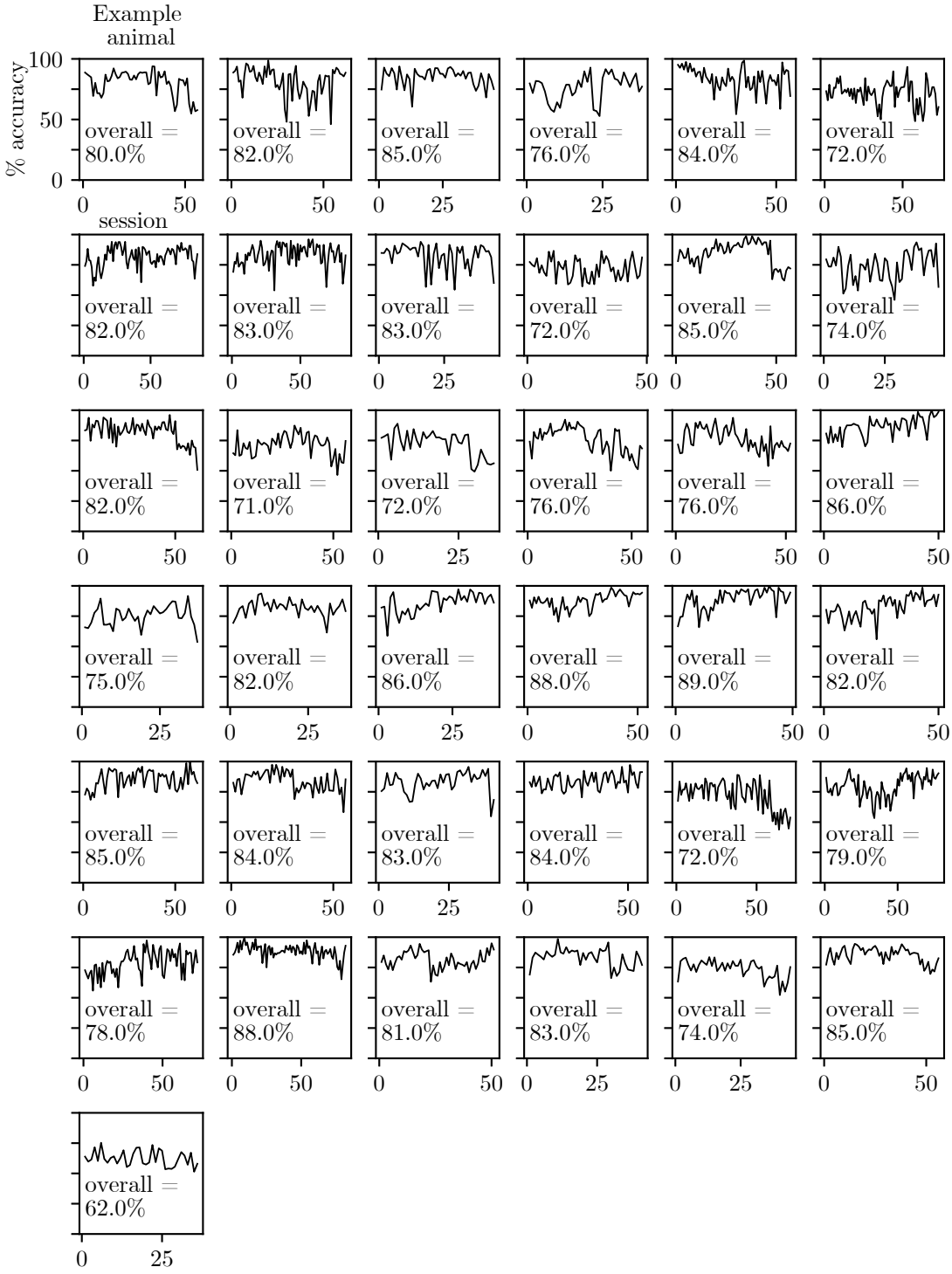

Figure S3: **Raw behavioral data for IBL animals (1): Accuracy across sessions** We plot the accuracy of IBL animals across sessions as evidence that mice have learned the task, and that their choice behavior has reached stationarity. We also report the overall accuracy of each animal when aggregated across sessions.

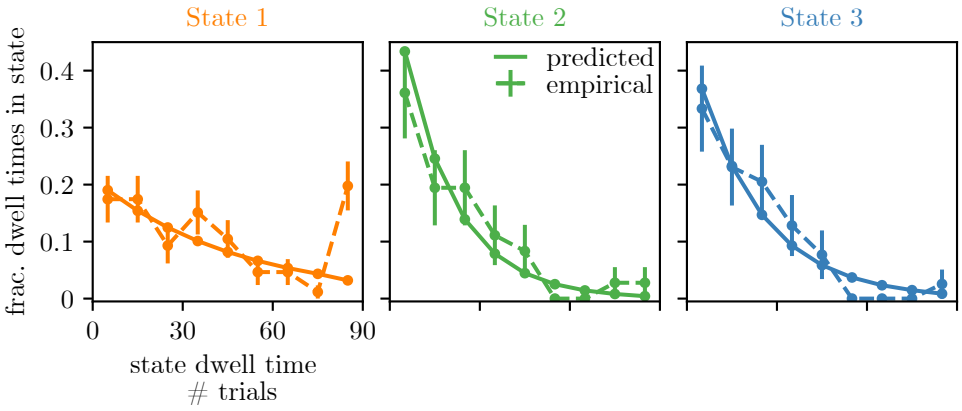

Figure S4: **Retrieved state dwell times are approximately geometrically distributed**. With the solid line, we show the predicted dwell times (according to the retrieved transition matrix) in each of the three states for the example animal of Fig. 2. Predicted dwell times can be obtained from the transition matrix as $p(\text{dwell time } = t) = (1 - A_{kk})A_{kk}^{t-1}$ because state dwell times in the Hidden Markov Model are geometrically distributed. We then use the posterior state probabilities to assign states to trials in order to calculate the dwell times that are actually observed in the real data (shown with the dashed line); we also show the 68% confidence intervals associated with these empirical probabilities. We find that the empirical dwell times for the biased leftward and rightward states seem to be geometrically distributed. For the engaged state, because there are some entire sessions (each session is 90 trials) during which the animal is engaged, we see that the empirical dwell times associated with this state are not as well described by a geometric distribution. A future direction may be to allow non-geometrically distributed state dwell times by replacing the Hidden Markov Model with e.g. the Hidden semi-Markov Model [49].

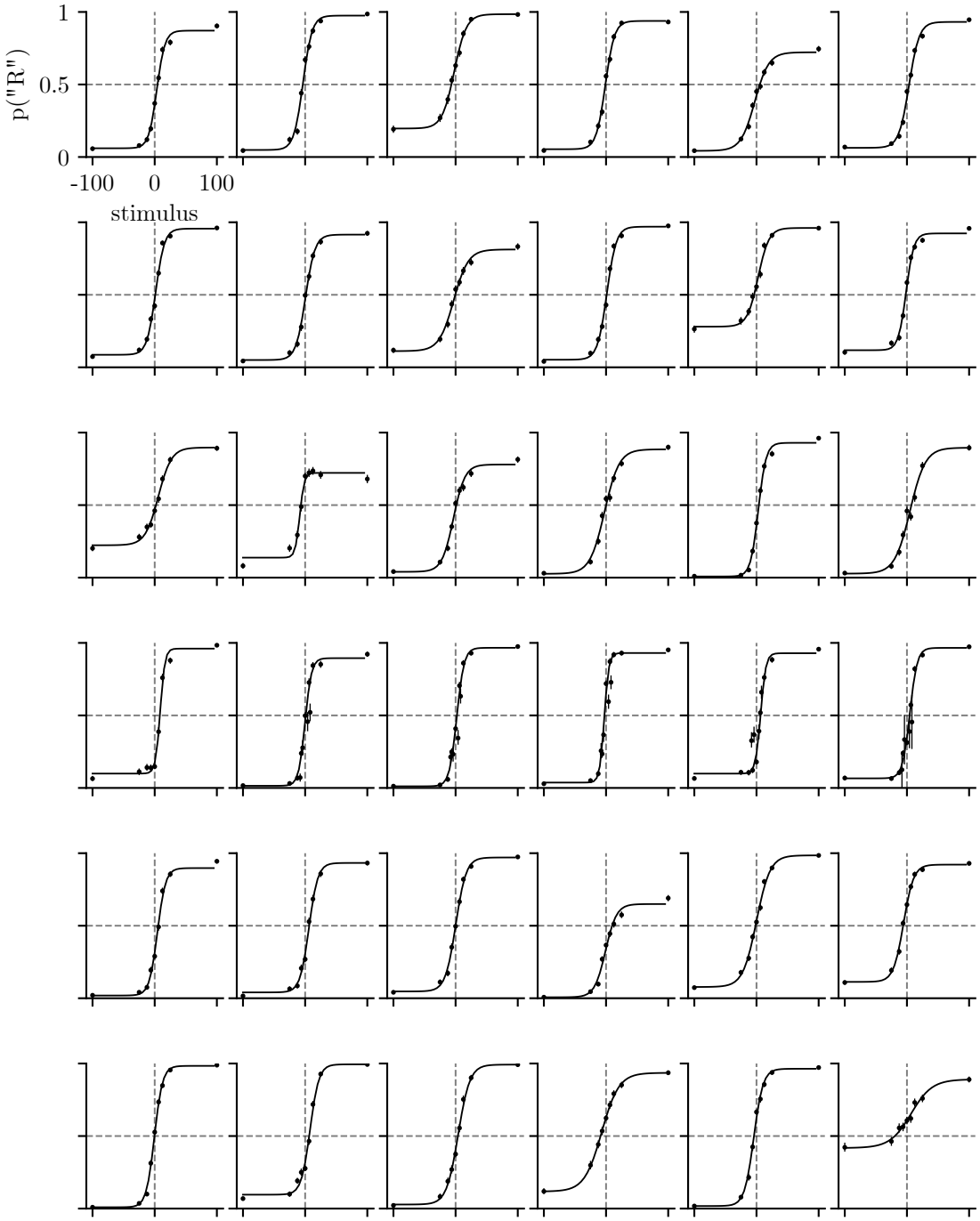

Figure S5: **Raw behavioral data for IBL animals (2): Psychometric curves** We plot the psychometric curves for 36 IBL animals whose choice data we study (we exclude the psychometric for the example animal shown in Fig. 2 since we already include it there); but animals are otherwise ordered in the same way that they are in Fig. S3 when considering row-major order, so plots can be compared across the two figures. In each plot, we also show each animal's empirical choice probabilities; error bars are 68% confidence intervals.

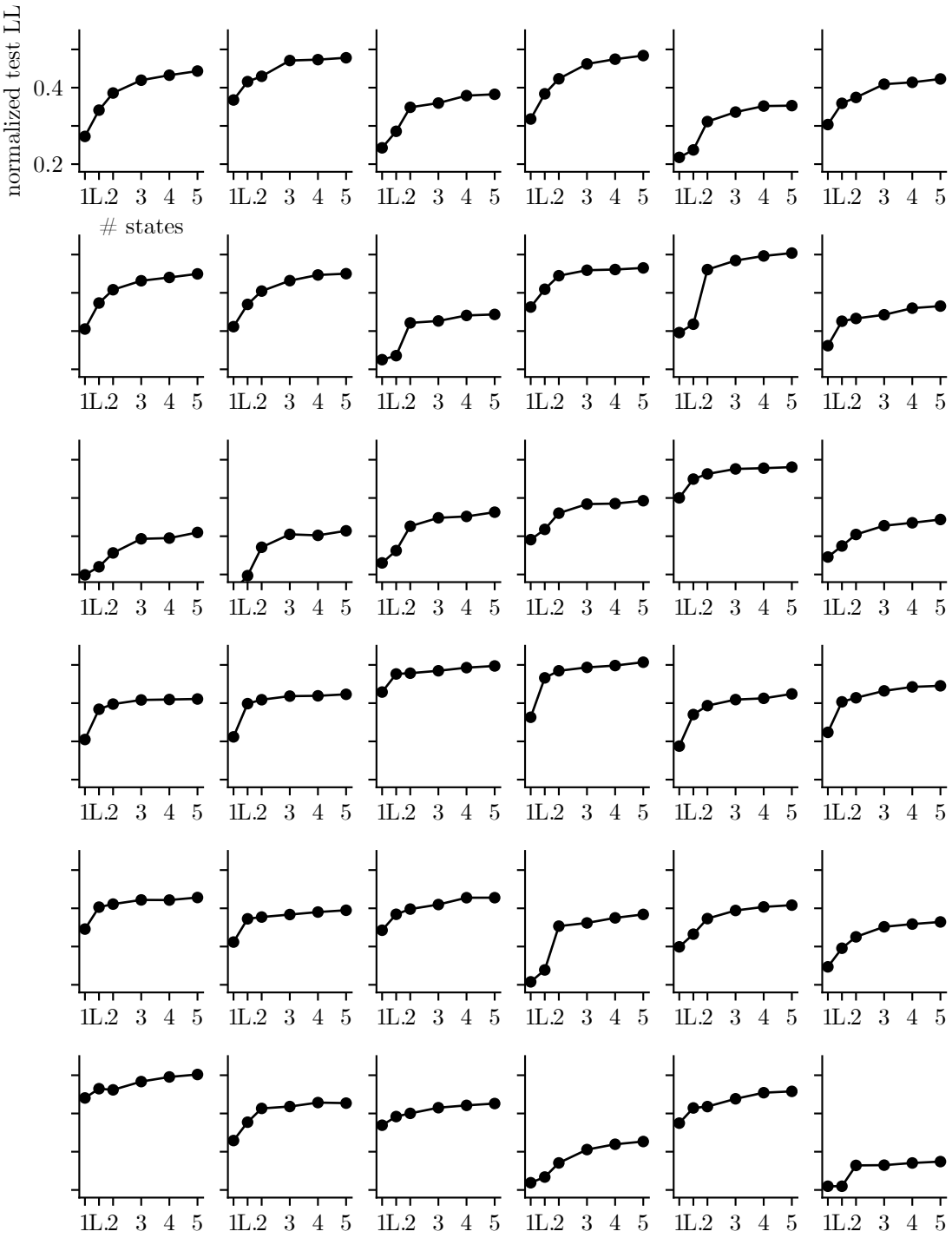

Figure S6: **Model comparison for all IBL animals** Rather than plotting curves on top of one another as in Fig. 3a, we plot each individual animal's normalized test set loglikelihood in a grid. We do not show model comparison results for the example animal studied in Fig. 2 since we show the curve for that particular animal there. Animals are ordered in the same way that they are in Fig. S5, so plots can be compared across the two figures.

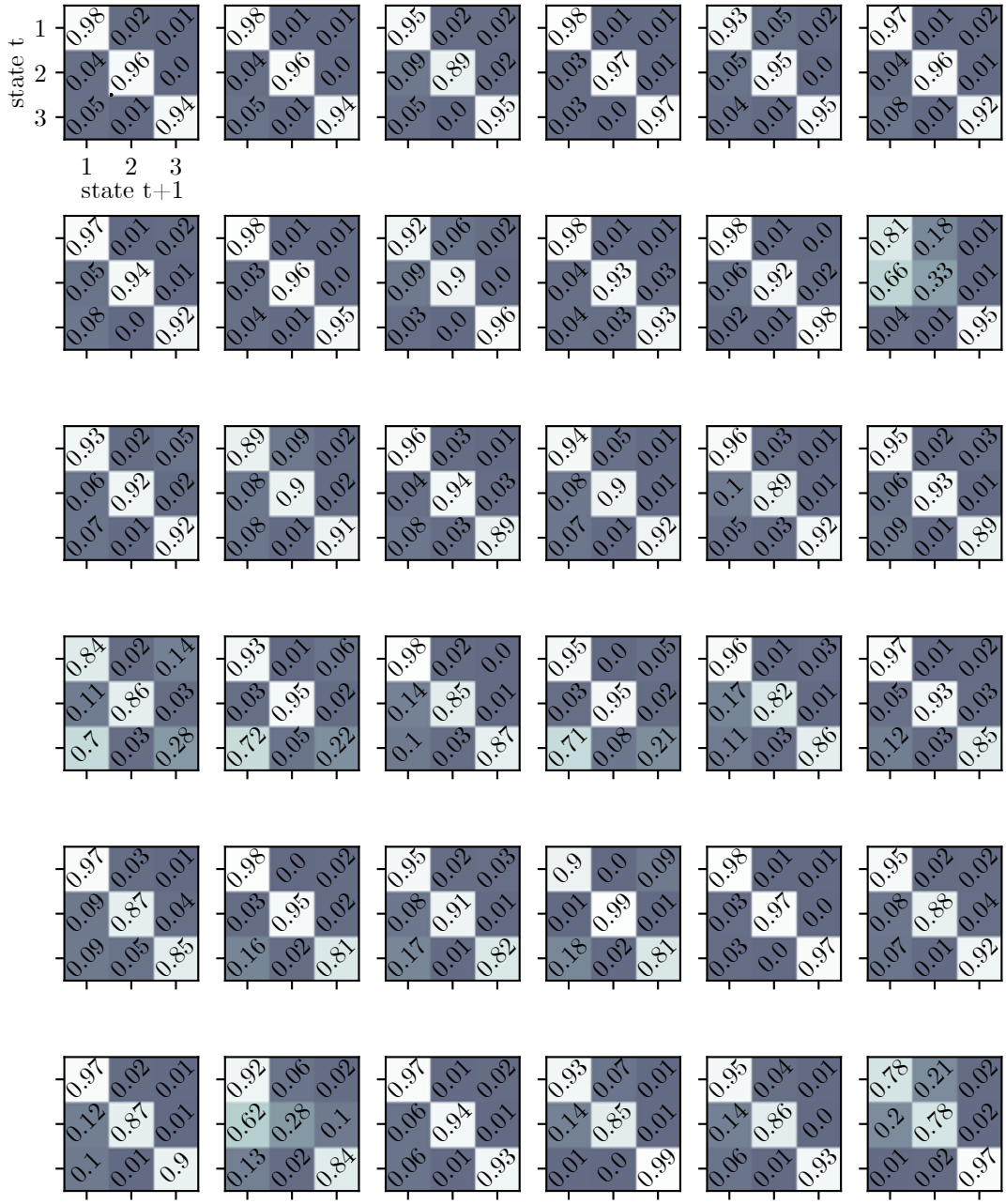

Figure S7: **Retrieved transition matrices for all IBL animals** We do not plot the transition matrix for the example animal studied in Fig. 2 since the retrieved transition matrix is shown there. Animals are ordered in the same way that they are in Fig. S5 and Fig. S6 so plots can be compared across figures.

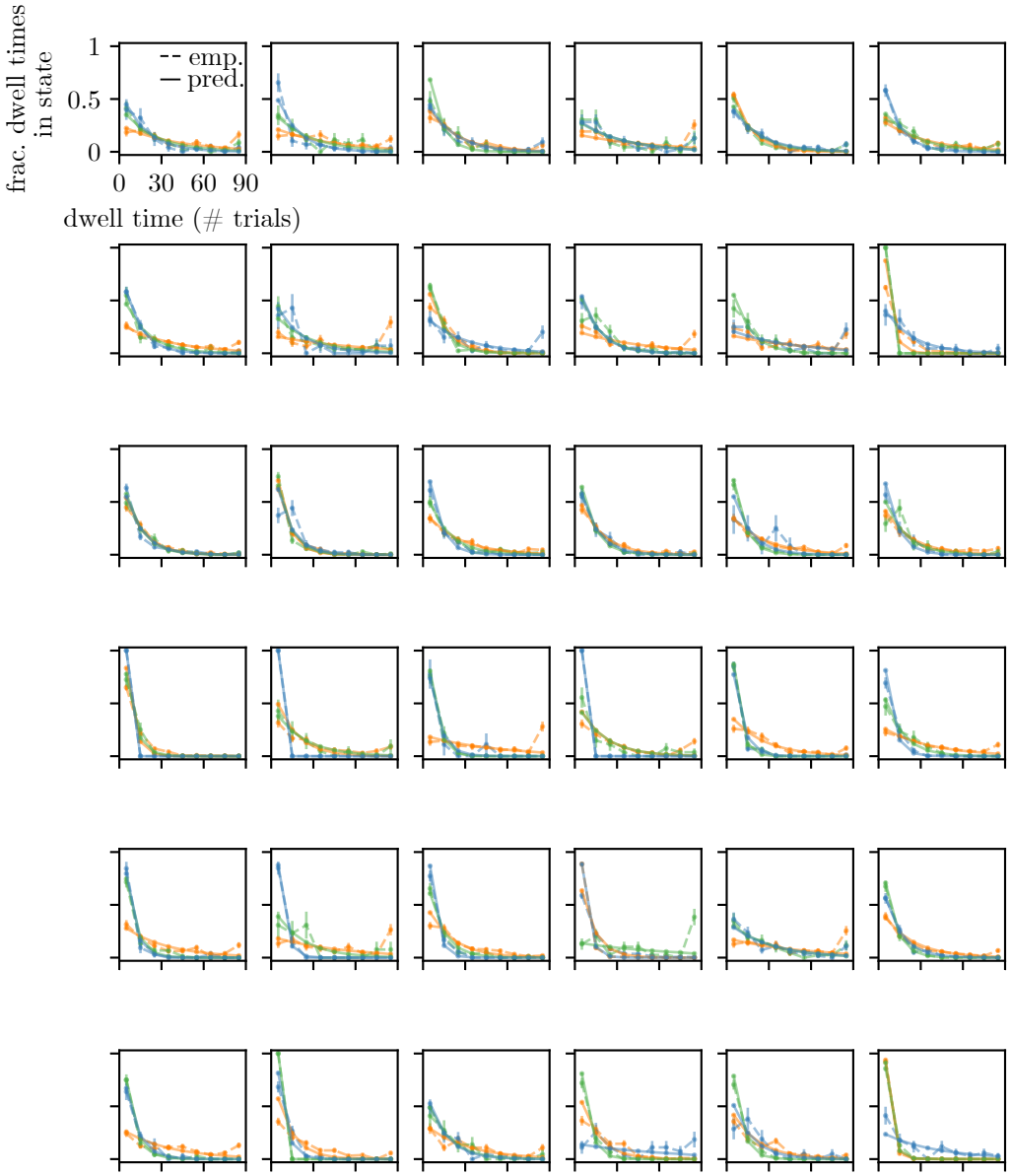

Figure S8: **State dwell times are approximately geometrically distributed for all IBL animals.** We plot the predicted ("pred."; shown with the solid line) and empirical ("emp."; shown with the dashed lines) state dwell time probabilities for each IBL animal and each state (excluding the example animal since the state dwell time probabilities for the example animal are shown in Fig. S4)). Colors map to states in the usual way (orange is state 1, green is state 2, blue is state 3). The empirical state dwell times are obtained by using the posterior state probabilities to assign state labels to trials, and the predicted state dwell time probabilities are obtained from the transition matrix according to $p(\text{dwell time} = t) = (1 - A_{kk})A_{kk}^{t-1}$.

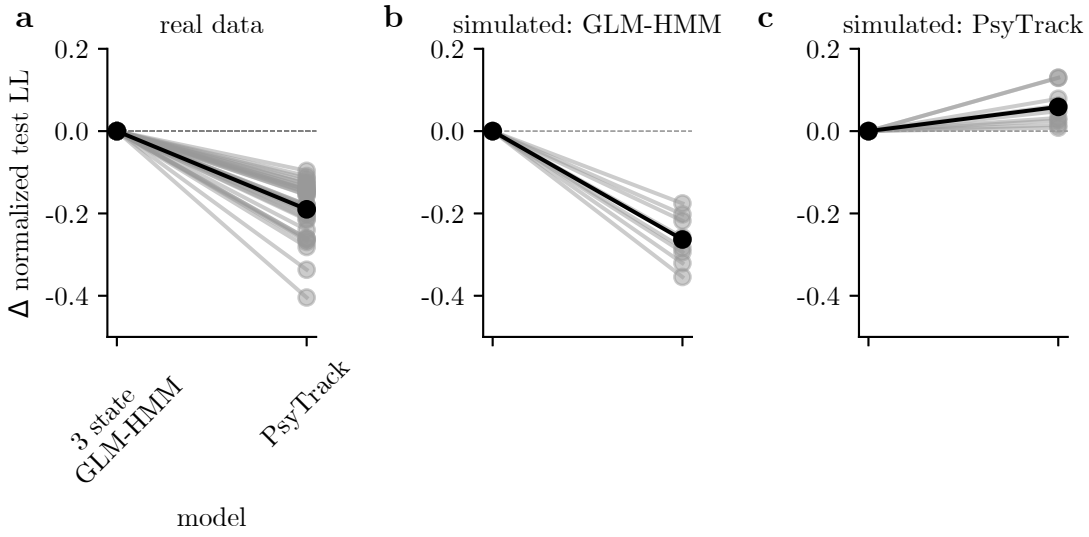

Figure S9: **GLM-HMM 4 state fits to IBL data.** When 4 state GLM-HMMs are fit to IBL animals, the engaged state splits into two engaged states where one state has a slight leftward bias, and the other has a slight rightward bias. This is a sister figure to Fig. 3 and Fig. 5, and each panel can be interpreted in the same way as in panels d, e and f of Fig. 5.

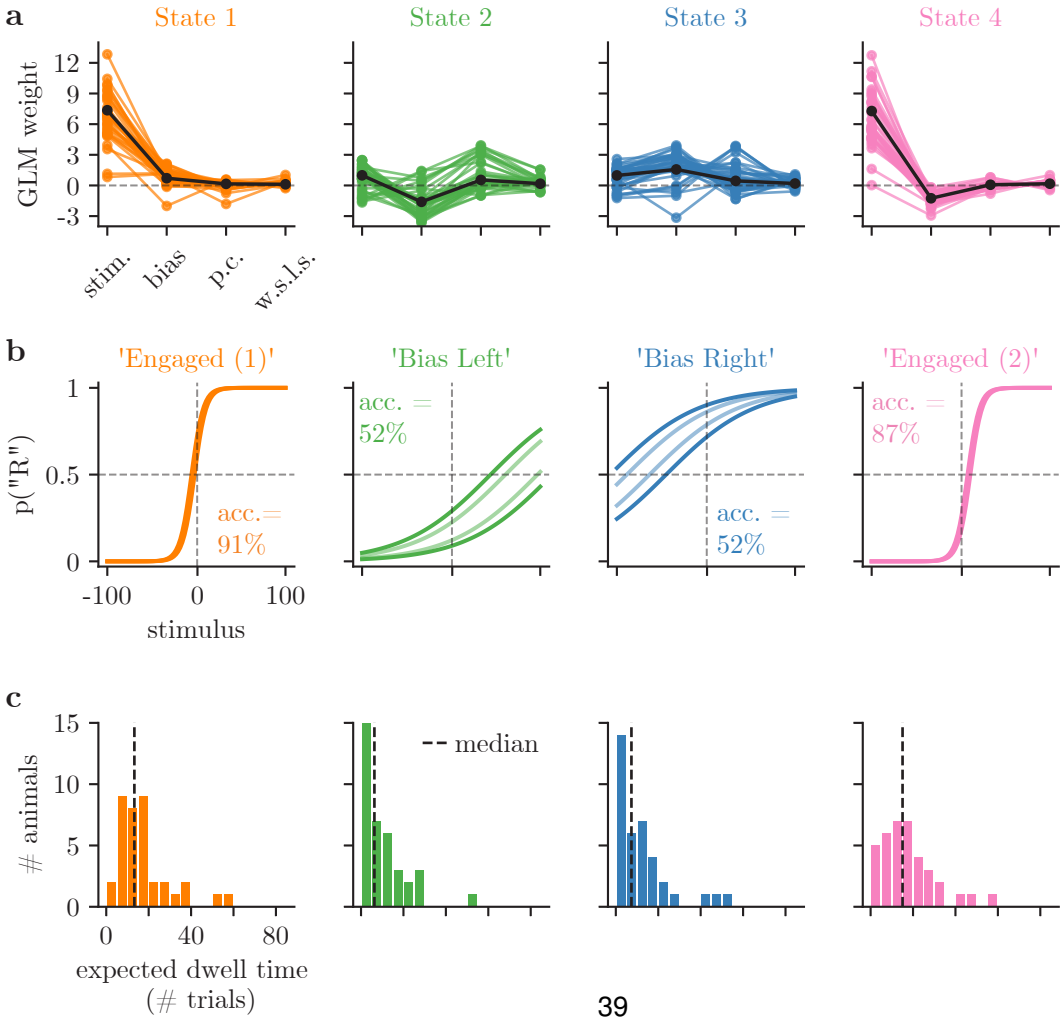

Figure S10: **GLM-HMM application to 4 mice not exposed to bias blocks in IBL task.** We confirm that mice that have never been exposed to bias blocks in the IBL task continue to show state-dependent decision-making. This is a sister figure to Fig. 3, and each panel can be interpreted in the same way as in Fig. 3.

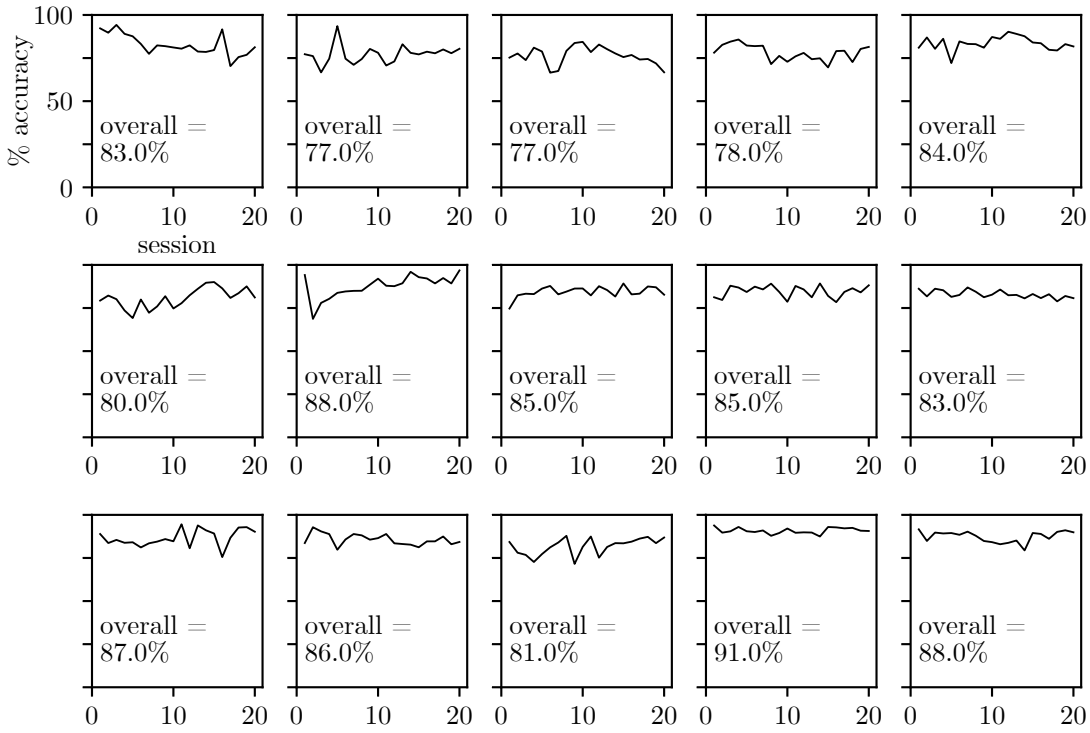

Figure S12: **Accuracy across sessions for Odoemene et al. animals** We plot the accuracy across sessions for Odoemene et al. animals as evidence that the animals' choice behavior has reached stationarity.

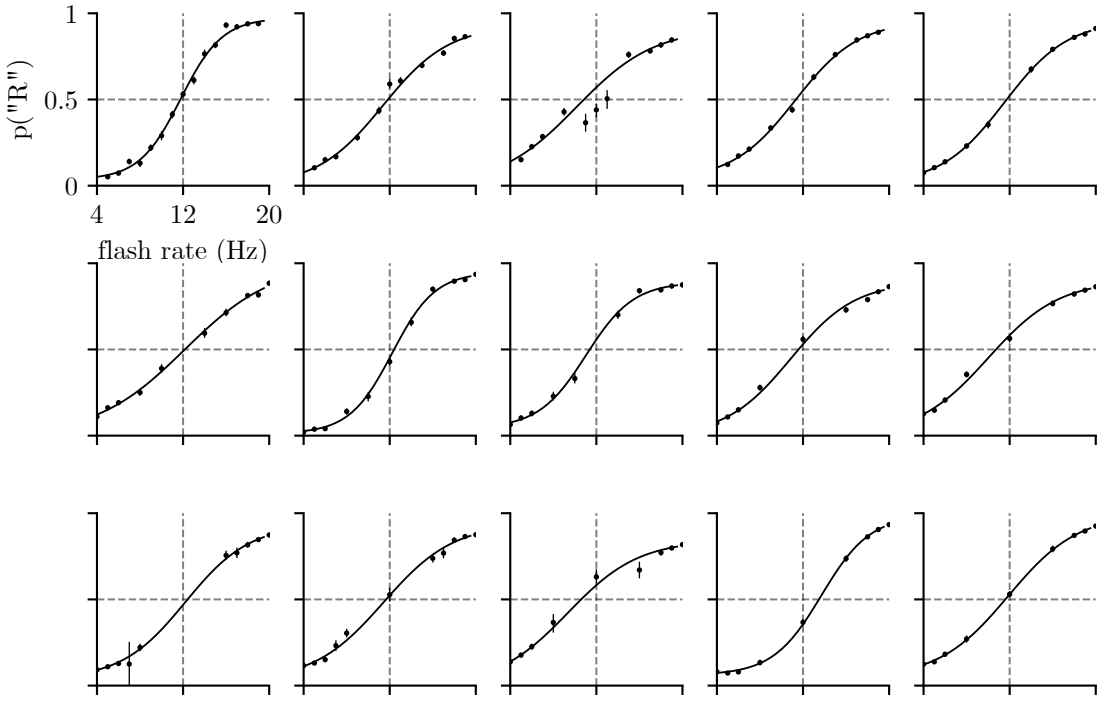

Figure S13: **Psychometric curves for Odoemene et al. animals** We plot the psychometric curves for each of the 15 animals whose choice data we study. We also show each animal's empirical choice probabilities; error bars are 68% confidence intervals. Animals are ordered as in Fig. S12.

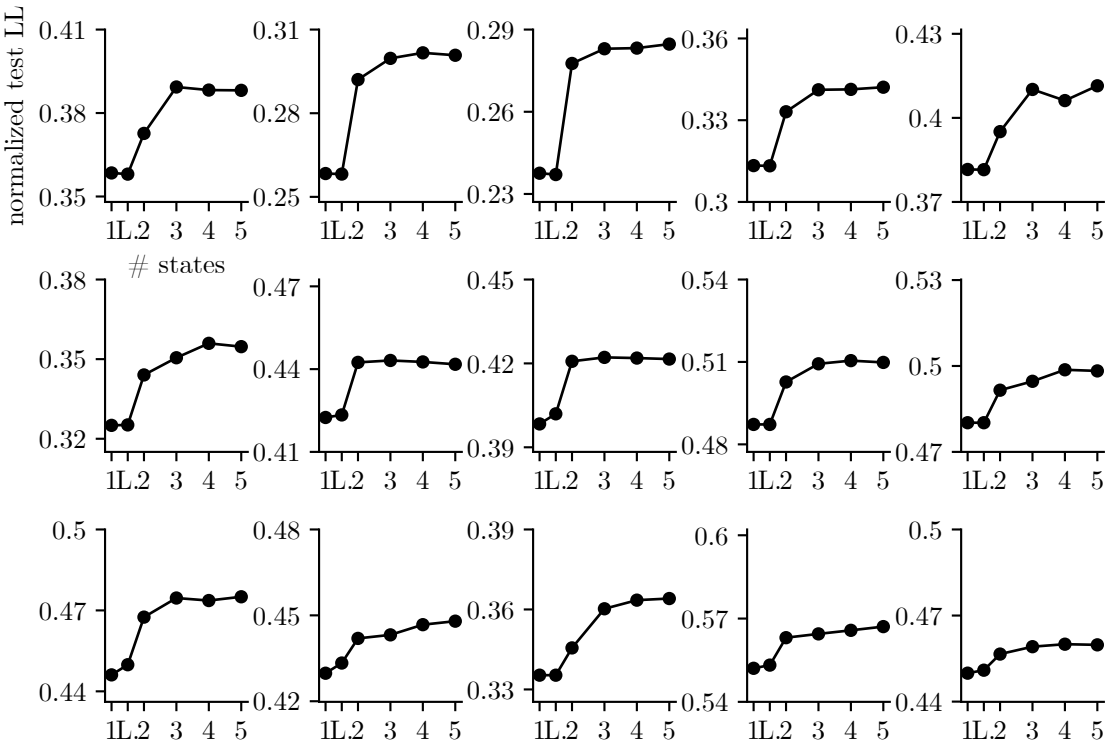

Figure S14: **Model Comparison for all Odoemene et al. animals.** Rather than plotting curves on top of one another as in Fig. 5b, we plot each individual animal's normalized test set loglikelihood in a grid. Animals are ordered in the same way that they are in Fig. S12 and Fig. S13, so plots can be compared across figures.

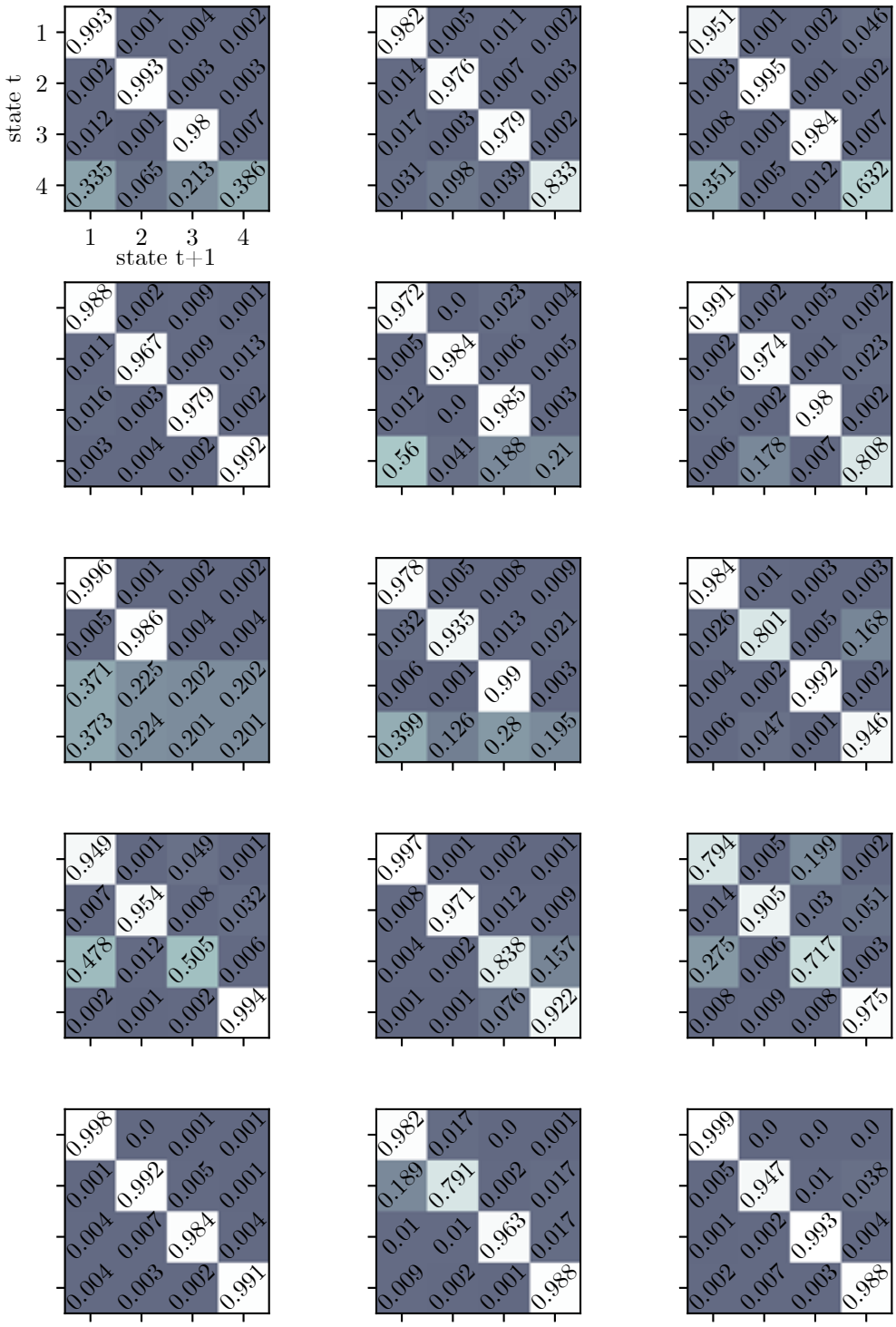

Figure S15: **Retrieved transition matrices for Odoemene et al. animals** Each individual transition matrix is the best fitting transition matrix for 1 of the 15 Odoemene et al. mice that we study. Row-major order has animals ordered in the same way that they are in Fig. S14.

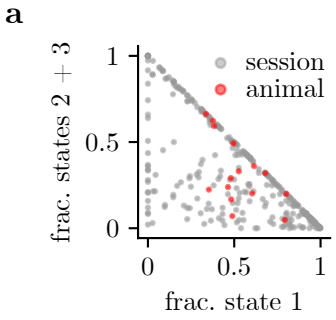

Figure S16: **State occupancies for Odoemene et al. animals** Analogous figure to Fig. 3d for Odoemene et al. animals. Each individual grey point is a single session of choice data, and we plot the fraction of the session that the animal spends in state 1 vs the fraction of the session that the animal spends in either states 2 or 3 (with the fraction of time spent in state 4 being calculable from these fractions). For the 300 sessions (across 15 animals) that we study, only 5 sessions result in an animal staying in the same state for the entire session. Each individual red dot represents an individual animal and shows the fractional occupancy of states.

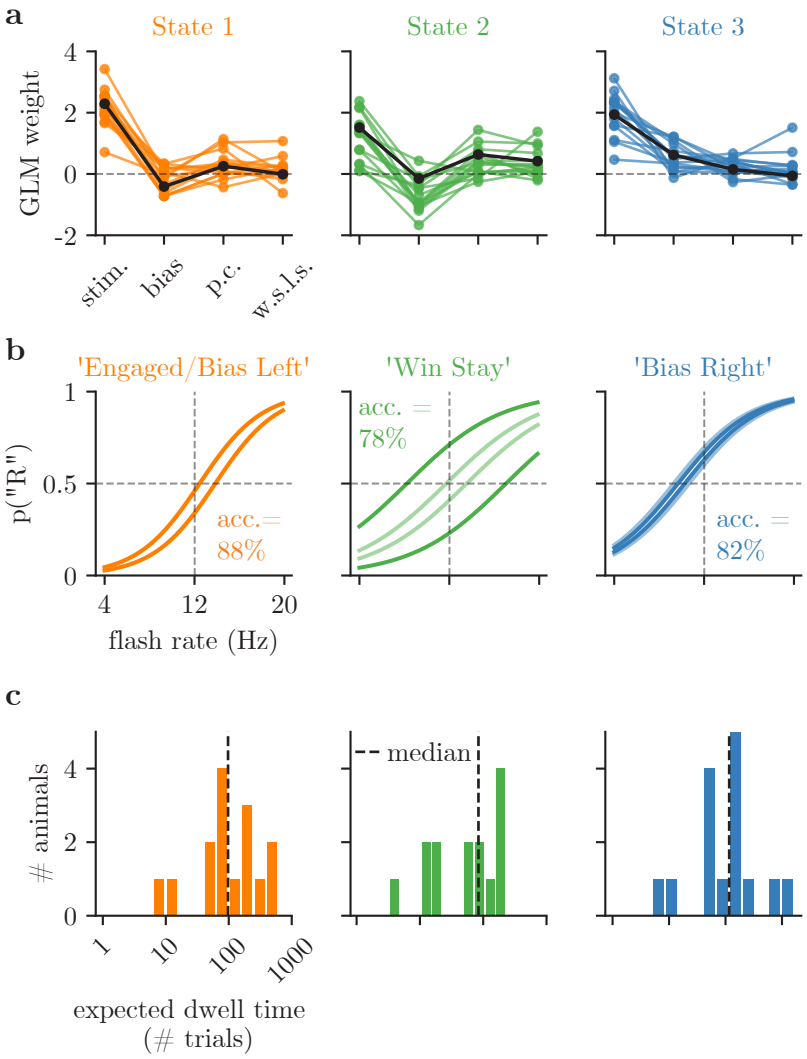

Figure S17: **GLM-HMM 3 state fits to Odoemene et al. data** When 3 state GLM-HMMs are fit to the Odoemene et al. data, the engaged and bias left states are merged to form a single (mostly engaged) state (accuracy remains high at 88%), while the win-stay and bias right states are largely unchanged. This is a sister figure to panels d, e, and f of Fig. 5, where the panels can be interpreted in the same way that they are there.