nature neuroscience

Article

https://doi.org/10.1038/s41593-025-02130-x

Infinite hidden Markov models can dissect the complexities of learning

Received: 18 December 2023

Accepted: 15 October 2025

Published online: 30 December 2025

Check for updates

Sebastian A. Bruijns[1,2], International Brain Laboratory*, Kcénia Bougrova[3], Inês C. Laranjeira[3], Petrina Y. P. Lau[4,5], Guido T. Meijer[3], Nathaniel J. Miska[6], Jean-Paul Noel[7], Alejandro Pan-Vazquez[8], Noam Roth[9], Karolina Z. Socha[4,10], Anne E. Urai[11] & Peter Dayan[1,2]

Learning the contingencies of a task is difficult. Individuals learn in an idiosyncratic manner, revising their approach multiple times as they explore and adapt. Quantitative characterization of these learning curves requires a model that can capture both new behaviors and slow changes in existing ones. Here we suggest a dynamic infinite hidden semi-Markov model, whose latent states are associated with specific components of behavior. This model can describe new behaviors by introducing new states and capture more modest adaptations through dynamics in existing states. We tested the model by fitting it to behavioral data of >100 mice learning a contrast-detection task. Although animals showed large interindividual differences while learning this task, most mice progressed through three stages of task understanding, new behavior often arose at session onset, and early response biases did not predict later ones. We thus provide a new tool for comprehensively capturing behavior during learning.

Engaging with a new environment or task raises a multitude of problems—which sensory signals are pertinent to the task, and which are just noise? What actions are relevant to performance? How should observations inform actions? Particularly if the experimenter suddenly changes an aspect of the task (to manipulate or shape behavior), but also in stable environments, animals solve these problems through a mixture of apparent leaps in performance and slow accumulation of improvements[1–9]. This process of learning is marked by substantial variability across individuals, who progress at different speeds and over distinct intermediate stages[10]. Interindividual differences during learning are a known phenomenon[11], although relatively little studied (although, for instance, see refs. 12,13). Even if the resulting behavior is highly similar across animals, variability during learning can make comparisons across groups in this period challenging[11]. This is because behavior during learning is a complex mixture of differently competent decision-making modes, prone to sudden shifts in performance (for better or worse), all of which occur on widely different timescales across individuals. More generally, the idiosyncrasies of the learning path may leave a trace in performance even after learning has finished and behavior has stabilized[14].

Despite the richness of these dynamics, much of the work on the modeling of learning has ignored acquisition in its full breadth and generally considered only how animals adapt to ongoing changes in facets of tasks, such as reversing reward schedules. By this point in the task, the animals have learned the basics of the problem, and those that failed to learn it have been excluded. One of the reasons for this neglect of initial learning is that each animal provides only one sample of a learning curve, whereas for fully acquired behavior, every trial can typically be viewed as another sample from the learned behavior. This means that learning curve data are generally sparse, further aggravating the problem of large variability. Here we make use of the large-scale approach to data collection embodied by the International

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany. [2]University of Tübingen, Tübingen, Germany. [3]Champalimaud Foundation, Lisbon, Portugal. [4]University College London, London, United Kingdom. [5]The Chinese University of Hong Kong, Hong Kong, China. [6]Sainsbury Wellcome Centre, University College London, London, United Kingdom. [7]University of Minnesota, Minneapolis, MN, USA. [8]Princeton University, Princeton, NJ, USA. [9]University of Washington, Seattle, WA, USA. [10]University of California, Los Angeles, Los Angeles, CA, USA. [11]Leiden University, Leiden, the Netherlands. *A full list of authors and their affiliations appears at the end of the paper. e-mail: sabruijns@gmail.com

Brain Laboratory (IBL; ref. 11) to build a rigorous descriptive model of the multisession learning curves of more than 100 mice solving a perceptual decision-making task. While we used a large and varied dataset for development and testing, the final method does not require a large number of individuals and should therefore be broadly applicable to multisession learning data.

Previous work on task acquisition has sought to find the point in time at which an animal can be said to have 'learned' a task, often defined as reliably above chance performance[15]. Methods for solving this kind of change-point detection (for example, refs. 16,17) typically make a binary distinction between uninformed and learned behavior, rather than describing used strategies in detail, or finding possible intermediate stages. Other previous work addresses strategy inference more specifically and does consider learning[18]. This involves inference on a trial-by-trial basis over a set of simple, preselected strategies, decaying evidence exponentially over time to track the arrival and departure of various strategies.

We sought to accommodate the complexities of learning curves using a descriptive modeling framework that satisfies a number of desiderata. First, at any point along the curve, the model should capture an individual's current repertoire of behaviors, characterizing its performance. Second, it needs to track this repertoire as behavior evolves, introducing new components (which we identify as behavioral 'states') when change is abrupt (for example, refs. 9,19), detecting the reuse of a past state if it re-emerges and allowing for slow, gradual shifts in a component, with the steady development of skilled performance (for example, ref. 20). Third, the collection of components should be potentially unbounded because we cannot know ahead of time how many distinct behaviors any individual might exhibit.

We therefore built a model that combines and extends two recent approaches. One is from ref. 21 (additional related work in ref. 22), which describes decision-making performance after learning with a hidden Markov model (HMM). Each latent state of the model captures a single component of behavior as a map from task-relevant variables to a distribution over choices, via logistic regression. In the case of perceptual decision-making, this generalizes a psychometric function (PMF) to include other factors (for example, perseveration). The overall description of behavior is in terms of a mixture of different policies that can switch rapidly. However, the HMM approach assumes stationarity of behavior across time and is constrained to a fixed level of complexity by specifying the number of states a priori. This weakens its ability to characterize the dynamic and idiosyncratic progression through training. To address these issues, we adopted the HMM framework to capture abrupt changes, except that (1) the states come from a Bayesian nonparametric structure, allowing for a degree of behavioral complexity that is only constrained by an inbuilt Occam's razor and enabling the introduction of new states for suddenly appearing new behaviors[23–27]; and (2) we used a semi-Markov model so that latent states can persist for nongeometrically distributed numbers of trials.

The second approach is that of ref. 28, which effectively considers only a single state but allows the logistic regression weights implemented by that state to be dynamic, tracking changes in behavior through appropriate updates to the weights. We used this so that the characteristics of our hidden states can evolve slowly, capturing the other prevalent form of acquisition of skilled performance.

Showcasing our model on behavioral data from the IBL task, described in ref. 11, we reveal that learning progresses over a small number of distinct stages that are present in almost all animals. These stages apparently correspond to the sequential acquisition of elements of the task—in our case, particularly associated with taking into account different aspects of the sensory environment inherent to the task. Although this pattern was shared across the mice, the duration and diversity of the stages differed greatly between individuals.

We first describe the IBL task and our way of characterizing the behavior mice exhibit; then we discuss the details of the model by studying a representative fit to one animal in detail; and finally, we conclude by summarizing the fits of our model to 134 subjects, highlighting similarities and differences across the population.

Results

We analyzed the choices of 134 mice learning a perceptual decision-making task, each of which underwent, on average, 24.4 sessions (total, >3,200) and ~14,800 trials (total, >1.9 million)[11]. In this task, head-fixed mice were shown a sinusoidal grating of a controlled contrast, which had equal probability of being on either the right or left side of a screen (Fig. 1a). They then had to center it (within 60 s) by turning a steering wheel in the appropriate direction. Successful trials led to water reward, whereas unsuccessful trials resulted in a noise burst and a 1-s timeout. Trials were self-paced, with mice signaling their readiness by keeping the wheel still for a period.

Mice learned the task according to a shaping protocol that gradually introduced more difficult stimuli and actively removed action biases (Fig. 1b). Accordingly, shaping began with strong contrasts—100% and 50%. At the initial stage, there was no perceptual difficulty; the animals only had to learn the basic contingencies of the task. Once they had reached sufficient performance on these contrasts (≥80% correct for each contrast type on the last 50 trials), 25% contrasts were introduced. After performance was also good on this extended set (same criterion), the remaining contrasts were introduced in a staggered manner—12.5%, 6.125% and 0%, whereas the 50% contrast was dropped. For the 0% contrast, one side was randomly rewarded (50% probability per trial). A debiasing protocol increased the probability of repeating the stimulus that was just shown when the mouse made a mistake on an easy (100% or 50%) contrast. This deterred perseverative or biased strategies, but could lead to reward rates <50%.

To characterize the course of learning across trials, we developed a flexible model that segments an animal's behavior into discrete states that last for variable numbers of trials within a session and can recur across multiple sessions. As this is a descriptive model, we equate a behavior with its corresponding state and, generally, will not distinguish between the two in the text. We first describe how a single state generates choice probabilities on a trial for which it was responsible (Fig. 1c, within circles) and then how we treat multiple states (Fig. 1c, arrows).

As in previous work[21,28], we formalize the response probabilities for the binary choices of mice through logistic regression (omitting the rare trials in which the animal timed out by not responding within 60 s). Trial t of session n is described by features $\mathbf{f}_{n,t}$ comprising (1) the stimulus, that is, the contrast on the left and right of the screen, separated to allow for different sensitivities to leftwards and rightwards stimuli, as mice were frequently differently sensitive to the screen sides in this task; (2) task history, in the shape of an exponentially decaying average over the last actions—interestingly, mice only used a perseverative bias, but did not use reward information to implement a win-stay lose-shift strategy (as was also observed in refs. 29,30 and, in the same task at a later stage, in ref. 31); and (3) a bias term to allow for side preferences regardless of other features. Labeling the state that is active on this trial as $x_{n,t}$, the response $y_{n,t} \in \{\mathrm{L}, \mathrm{R}\}$ (for left and right) is modeled by the distribution

$$P(y_{n,t} = R) = \mathrm{sig}\left(\boldsymbol{f}_{n,t} \times \boldsymbol{w}_{x_{n,t},n}\right), \tag{1}$$

where the weights of the states $\mathbf{w}_{x,n}$, $\forall x$ are also indexed by session n, as they can drift across sessions. Here sig(·) is the standard logistic sigmoid function.

The model generalizes a standard HMM in the following three ways that make it especially suited to describe the phases of learning: (1) it is nonparametric about the number of states; that is, the number of states describing the behavior of each individual is separately determined, accommodating interindividual differences. This characteristic also

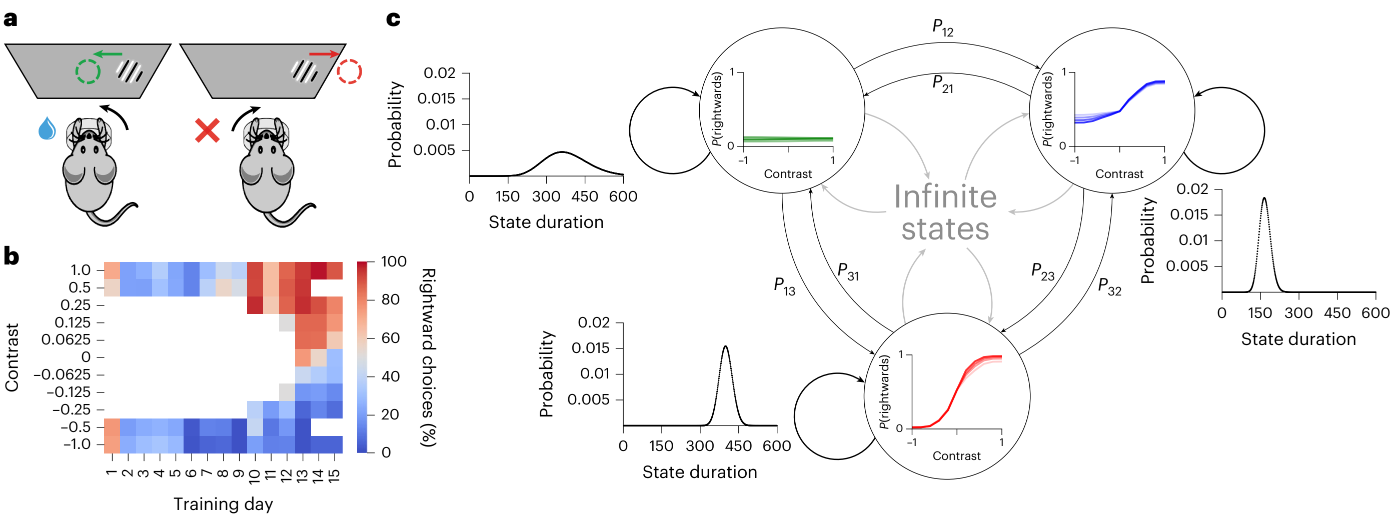


Fig. 1 | Behavior and modeling overview. a, Sensory decision-making paradigm. Mice indicated whether a contrast grating was on the left or right of a screen using a wheel. **b**, Representative behavior of mouse KS014 (also used later). This shows improvements in behavior and the concomitant extension of the contrast set. **c**, Visual representation of the main components of the model. Each state, represented by a circle, has an associated observation distribution, shown inside its circle. This is implemented via logistic regression, which considers the contrast on the current trial and a weighted history of previous choices (the latter is not shown here). The weights underlying these regressions can change from session to session, resulting in shifts of the PMFs they represent; we depict this evolution here with varying shades of color. States are connected to other existing states via transition probabilities *P*. In addition to that, states also have the option to transition into a new state, to describe a type of behavior that is not well captured by any of the existing states. Finally, staying in the same state for more than one trial is not modeled via a self-transition probability; instead, each state has its own duration distribution. Panel **a** reproduced from ref. 11 under a Creative Commons licence CC BY 4.0.

allows the model to capture sudden changes in behavior, as it is able to introduce a new state when behavior changes notably (we call this the 'fast process'; 'Infinite hidden semi-Markov model' section). (2) States are dynamic over sessions, allowing the behavior implied by a state to change gradually across session boundaries[28] (the 'slow process'; 'Dynamic logistic regression prior and sampling' section). (3) While for HMMs the numbers of trials for which a single state remains active always follow a geometric distribution, we adopt a semi-Markovian approach, allowing for more general distributions. Taking all these additions together, we end up with a dynamic infinite input–output hidden semi-Markov model (diHMM).

The transition matrix over a flexible number of states and the evolution of the psychometric weights are defined by priors, and the Bernoulli observation model provides a likelihood for each trial, allowing for approximate Bayesian inference (Methods). We performed this using a Markov chain Monte Carlo (MCMC) algorithm, namely Gibbs sampling. For a single animal, the entire response and feature data across all training sessions were fitted together. Individuals were fitted separately, meaning a large number of subjects is not necessary for the application of our model. Integrating across a number of Gibbs samples from multiple Markov chains led to a set of behavioral states defined by their session-varying weights $\mathbf{w}_{x,n}$ and duration distributions, as well as a hard assignment of every trial onto one of these states. While all other relevant random variables are specified hierarchically or ruled by vague priors, the variance for slow changes within states is set, as inference over this variable proved problematic; we revisit this parameter in the discussion.

Single animal fit

We visually summarize the model fit for mouse KS014 at the resolution of entire sessions in Fig. 2. This animal exemplified many of the interesting properties found across the population. The inferred model contains eight states, but these states were generally active for only a small number of sessions before being replaced by others. We number them in order of appearance. In a typical session, the majority of trials were explained by a single one; at most, a few were active. Later states generally represented more adept behavior, although not exclusively. The mouse started with state 1, which exhibited a flat PMF (far right of the plot), indicating that the animal did not take into account the side of the sensory input. This state was promptly replaced by state 2 in the next session, which also had a flat PMF, although shifted. This change in bias was strong enough to warrant a new state (rather than the slow process of changing the existing state), but there was no evidence that the animal advanced in its understanding of the underlying task.

State 2 lasted four sessions, indicating that behavior remained relatively consistent during this time. It was then predominantly replaced by state 3, which started with a mostly flat and strongly biased PMF (leading to a lower reward rate due to the bias correction) but improved considerably over the next few sessions, as can be seen in the evolving PMF (with darker colors showing later sessions). It seemingly considered only sensory information from the left side when making its choice, becoming increasingly random when that side was uninformative. The random behavior was doubly beneficial, as the animal would sometimes have been correct and also got foiled less by the bias-correction protocol. State 3 was accompanied by state 4, which described the behavior at the ends of the next few sessions (and later also at the ends of sessions 14 and 15). Puzzlingly, this state had a good PMF on both sides and a higher reward rate than state 3, but although this better state was available, the animal seemed incapable or unwilling to use it for the majority of a session.

The last major step in learning appeared abruptly as state 6, with good performance on both sides (albeit differently from state 4). Along with state 6, we observed the introduction of state 7, which captured a strong but transient decline in behavioral quality. Finally, state 8 represented another notable change in behavior, as performance on 100% contrasts increased abruptly enough to warrant a new state, allowing the mouse to conclude this part of training.

Various aspects of our model cannot be reproduced by existing treatments. The Psytrack model of the study discussed in ref. 28 can fit incremental changes in behavioral characteristics; however, because it lacks a concept of state, it does not natively support the identification of recurring behavioral patterns. We find that many states occur,

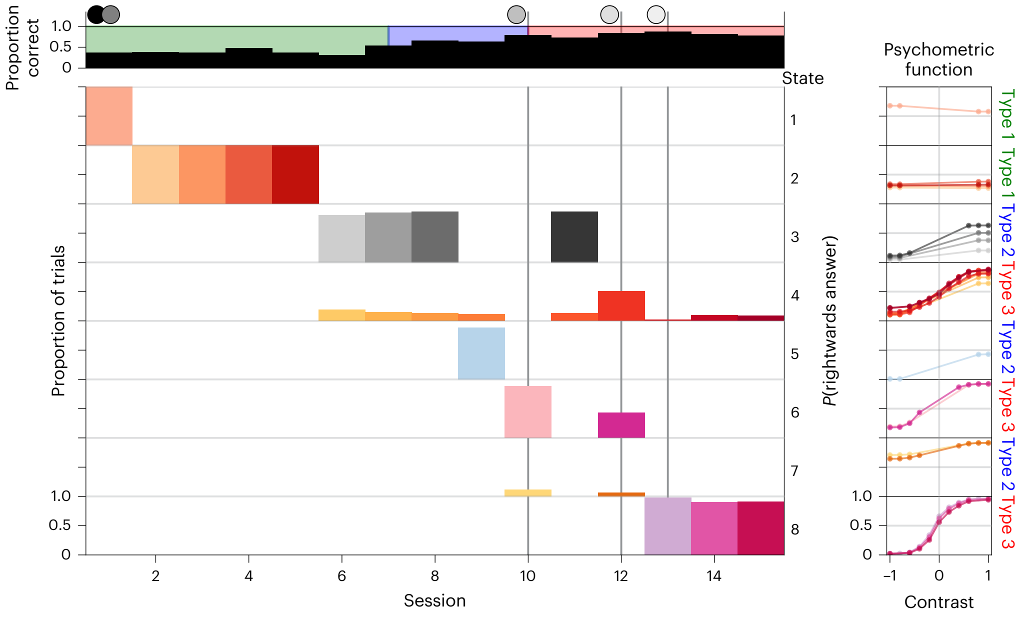


Fig. 2 | diHMM fit to mouse KS014. The topmost row shows the overall performance during the session, as proportion correct, and the current stage of learning as the background color (we elaborate on learning stages later in the text). Vertical lines with shaded circles at the top indicate the sessions during which new contrasts were introduced. The remaining rows show the prevalent behavioral states (label to the right) ordered by appearance, indicating which proportion of trials they explained during each session. To the far right of every state, we show its PMFs across time, ignoring the contribution from the history of previous choices. The saturation of the colors of the states indicates successive appearance and matches the PMF plots.

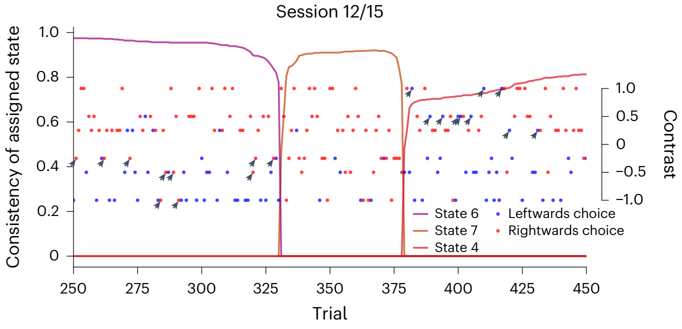


Fig. 3 | Excerpt of state assignments in session 12 from mouse KS014, also shown in Fig. 2. The left *y* axis serves as a scale for how connected a trial is to the other trials of that state (see Methods for details). The right *y* axis shows the contrast. The dot color indicates the animal's response. One can see how the drastic and sudden change in the response patterns, rightwards (red) answers for leftwards (negative) contrasts, from trial ~330 to ~380 was detected by the model with a transition to state 7. The PMFs of states 4 and 6 looked similar but did, in fact, represent significantly higher error rates on the right and left sides, respectively. These mistakes are highlighted with arrows.

then disappear, before reoccurring in a later session, such as states 3 and 6 in the animal shown in Fig. 2. This re-emergence of previously used strategies is an important feature of learning. Similarly, the static generalized linear model (GLM)–HMM described in ref. 21, which is aimed at asymptotic behavior, does not determine the number of states automatically. This implies that model selection is required for each individual animal, which the relatively small number of data points can make challenging. Furthermore, in the GLM–HMM, states cannot adapt their PMFs, which is a second important feature of learning. Without this, the GLM–HMM would tend to split states when behavior changes gradually but sufficiently as to elude a single set of weights.

Our model also provides a fine-grained view of the use of behavioral states within a session. Although the diHMM provides a full posterior over the states for each trial, this is not directly useful due to the technicalities of the sampling procedure. We therefore processed the sample chains to estimate how much a trial belonged to a state (Methods). We show an excerpt of this, for session 12 of mouse KS014, in Fig. 3. This shows two clear transitions between states. The reasons for the animal to have made such a transition are probably multifaceted and may have been both internal (for example, insights or motivational fluctuations[32]) and external (for example, a number of low contrast, perchance unrewarded trials demotivating the animal). We do not model these reasons and, instead, only describe observed changes.

The within-session fit shows that the model can detect temporary yet strong deviations in behavior. State 7 only explained a couple of dozen trials in two sessions, but represented extremely biased behavior (comparable, but flipped relative to the earlier state 2, albeit lasting for many fewer trials). We speculate that state 7 arose from a form of inattention because the animal had previously shown itself capable of performing appropriately. This change in behavior is directly evident in the response patterns of the animal.

We can also capture subtler differences in behavior. The model used different states to explain behavior before and after state 7, although performance appeared equally good. However, the model identified different error rates on easy contrasts for the two states, and this can be found in the choices—state 6 was associated with more incorrect responses to contrasts on the left side, whereas in state 4, performance on leftwards contrasts was good, but there were frequent lapses on rightwards contrasts (comparing two logistic regression models—one using contrast and state 4 (assigned to $n = 391$ trials) or 6 information (assigned to $n = 331$ trials) to predict responses, and the other a nested model that only uses contrast (for the total of $n = 722$ trials)—the state-split model is significantly better, as determined by a likelihood-ratio test with $P < 0.0006$ ($D = 14.96$, $df = 2$), with an effect size, as measured by the McFadden pseudo-R^2, of 0.025; Fig. 3, responses marked by arrows).

Fits across the population

The threefold progression we observed throughout learning in Fig. 2—from flat PMFs, to 'one-sided' behavior, to generally good performance—is

https://doi.org/10.1038/s41593-025-02130-x

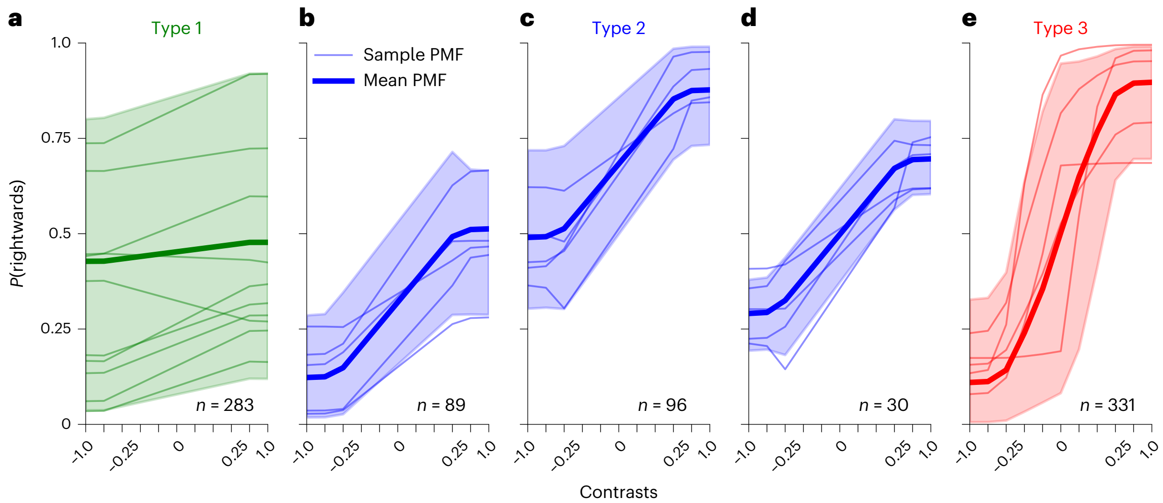


Fig. 4 | Summary of the PMFs associated with the different types. a–e, The first PMF of each state in each animal (representing response characteristics after a notable discontinuity in behavior) was collected. Each subplot shows a specific type—type 1 in green (**a**); type 2 in blue, further split by whether the PMF is left-biased (**b**), right-biased (**c**) or symmetric (**d**); and type 3 in red (**e**). The thick lines indicate the overall mean over PMFs of the type, which shows representative behavior of that type. The shaded regions show the range in which 95% of the PMFs fell (computed separately for each contrast level). The thin lines show samples of individual PMFs of these types. '*n*' indicates how many PMFs of each type were present across the entire population.

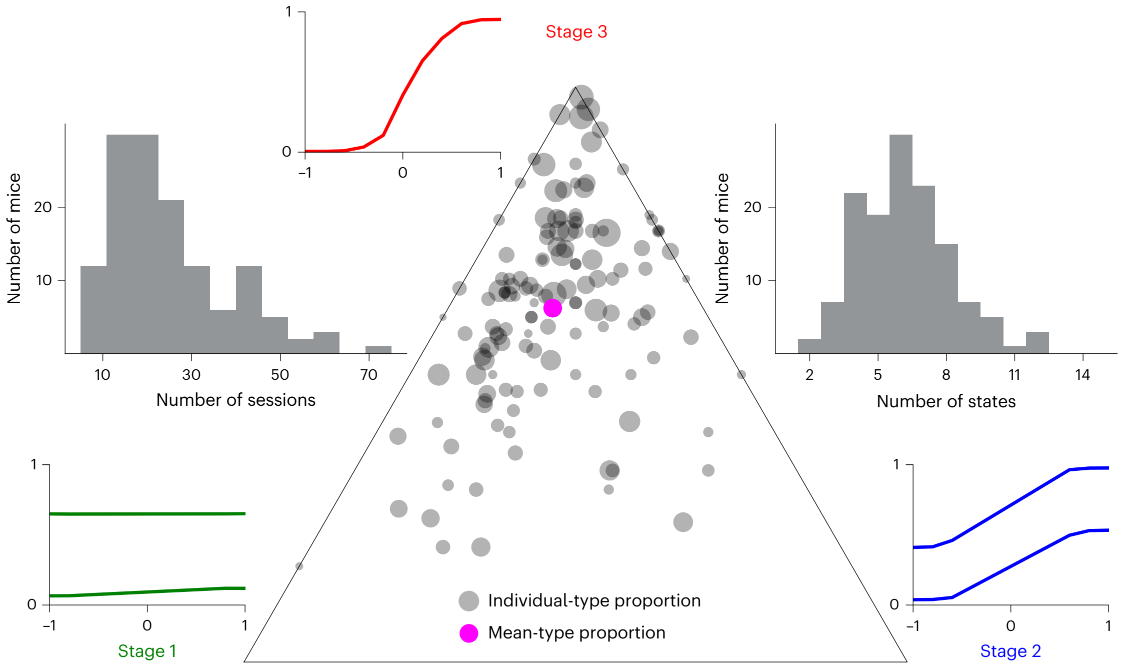


Fig. 5 | Proportions of sessions it took each mouse to reach the next major step in training, as defined by the three stages. Each individual is represented as a circle on the simplex (the larger the proportion of sessions within a specific stage, the closer the dot for that animal is to that corner of the simplex). Simplex corners are identified by example PMFs of the stage type. The marker area indicates the total number of sessions (min, 5; max, 75). The magenta circle marks the average proportion, and its size indicates the mean number of sessions (which was 24.4). See Supplementary Fig. 1 for a linear version of this plot. The histogram on the right shows the distribution over the number of states used by the model per mouse. The histogram on the left shows the distribution over the number of training sessions.

typical for the population of mice we fitted. To define this more objectively, we clustered the states into these three types based on their reward rate on easy trials (see 'Psychometric type classification' section for details). The boundary between types 1 and 2 is at a 60% reward rate, and the boundary between types 2 and 3 is at a 78% reward rate. We show an overview and examples of the different types in Fig. 4.

In addition to the state types, we define the stage at which an animal is on any given session as the highest type it has so far used for the majority of trials of any session up to this point. For instance, if up to session $n-1$, an animal only used type 1 states or type 2 states for fewer than 50% of trials, then it would be in stage 1 for those sessions. If, on session n, it then used type 2 states for more than 50% of trials, it would switch to stage 2 on that session. Because the state types delineate different aspects of task understanding, the stages allow us to determine how many sessions the animals stayed at a certain level of understanding. While the progression through state types was not monotonic (for example, session 11 of Fig. 2), the stage classification is, by definition, monotonically increasing.

https://doi.org/10.1038/s41593-025-02130-x

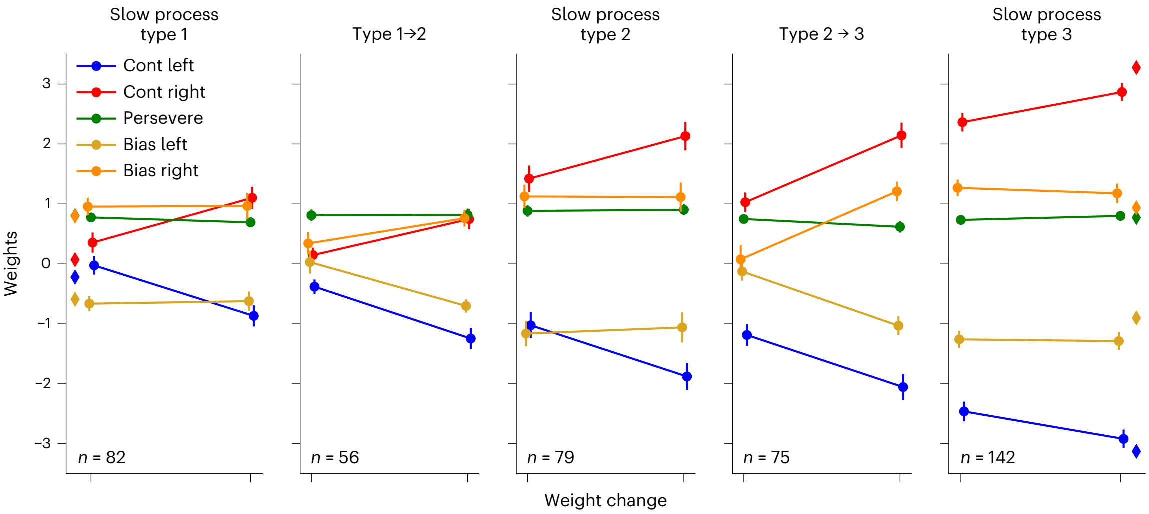


Fig. 6 | Evolution of the weights of states on average, through slow and sudden changes. Error bars indicate ±1 s.e.m. (lines are slightly offset along the *x* axis for visibility). Subplots titled by a type represent the weight changes from the first appearance of a state of this type to its last, so only showing state-internal slow changes (and only including states that were present between 5 and 15 sessions, as extremes would skew these averages). Subplots with a title indicating a transition from one type to another show how much each weight of the new state differed from the weights of the closest previously existing state and are based exclusively on the states that first brought the mouse into a new stage. That is, for 'type 1→2', we only took into account the first type 2 state exhibited by the mouse and only when that state was type 2 from its inception. For instance, for mouse KS014, this was state 4, which started as type 2 before using the slow process to become type 3. Colored diamond markers on the leftmost and rightmost plots indicate the average value of the weights of the very first state of each mouse and of the dominant state on the last session, respectively. To prevent biases from canceling out across the population, we split the bias weights into the following two groups: starting out below 0 (bias left) or starting out above 0 (bias right). While contrast sensitivities increased both through fast and slow changes, it is noticeable that biases stayed almost constant throughout the lifetime of a state on average, but changed more noticeably through sudden transitions.

Stage 1 consisted of states with flat PMFs of various biases, generally ignoring the contrast location. Stage 2 almost always involved asymmetric states, responding well to one side of the screen, but close to uniform guessing for the other (PMFs assigned to Fig. 4b,c account for 86% of those in stage 2; see 'Psychometric type classification' section for details). Only rarely were intermediate PMFs nearly equally good on both sides (the 14% in Fig. 4d). These rare cases did behave like the other type 2 states in terms of their time of appearance during training as well, rather than type 3. Finally, in stage 3, the animals started apparently paying attention to both sides. Generally, it took some further refinement of initial type 3 states, through the reduction of errors on easy trials on either side, to master this stage of training and progress to the next phase of shaping.

The three stages segment the learning process. We can analyze the proportion of training time the animals spent in the different stages by showing these proportions on a simplex (Fig. 5 and Supplementary Fig. 1, linear representation). The large majority of animals spent some time in each of the stages (that is, only a few mice are assigned to the edges of the simplex). Most animals spent the longest time in stage 3—going from moderately competent performance to passing the stringent training criteria. No fundamental change in understanding was necessary for this, unlike the changes from stage 1 to 2 or 2 to 3 (where the animal had presumably to learn to pay attention to the Gabor patch on one or both sides of the screen). However, reaching the required accuracy seemed difficult, even once the principles of the task were understood (possibly due to the small increase in reward rate afforded by the extra accuracy). Some of the longest trajectories (the largest circles) were associated with especially many sessions in stage 3, but overall the average fractional occupation was remarkably consistent across training lengths (the mean relative occupancy for stage 1, 2 and 3, respectively, were (0.24, 0.17, 0.59) and (0.21, 0.14, 0.65) for the shorter and longer halves of a median split on the total number of training sessions). Stage 2 consistently lasted for the fewest sessions, implying that the mice managed to pay attention to both sides not too long after starting to pay attention to one side.

Connected to this is the question of how slow and fast changes characterize behavior. We analyzed gradual changes within a state by comparing its PMF weights on its first and last appearances. We analyzed new state introductions by comparing their PMF weights to those of the closest previous state, as determined by the Wasserstein metric on their resulting PMFs (ignoring the perseverative weight). To highlight the changes, we focused on states that brought the animal into a new stage. These weight evolutions, split by type, are shown in Fig. 6. As the main driver of performance, contrast sensitivities reliably increased both over the lifetime of a state and when new states were introduced. Surprisingly, however, both the bias and perseverative weights were stable within a state. This was markedly different for the fast process—the changes through this were significantly larger (Supplementary Figs. 12 and 13; one-sided Mann–Whitney *U* test on absolute weight changes, two biases, two fast change points, three slow change processes; the fast process had significantly larger changes for all 12 comparisons at a 0.05 significance level, after applying the Benjamini–Hochberg procedure to control the false discovery rate, with effect sizes ranging from 0.43 to 0.94, quantified as standardized mean difference, using the fast change s.d.). We also see that the perseveration weight had a small but consistent role throughout learning (although its relative influence waned as the sensitivities grew).

The introduction of new states signifies notable changes in behavior, so by studying the patterns of their occurrences, we gain insight into when behavior was volatile or when substantial progress was made. The histograms in Fig. 7 show when new states first appeared across normalized training and session times. In later sessions, gradually fewer states were introduced, indicating that behavior saw fewer drastic changes as training progressed. We noted earlier that animals spent most of their time in stage 3, that is, perfecting their behavior, and we can now conclude that gradual improvements had an important role in this, more so than sudden marked changes. The pattern of introductions within sessions is even more striking—the majority of states were introduced at the very start of a session. This resonates with previous findings about change points in behavior occurring at session boundaries[16].

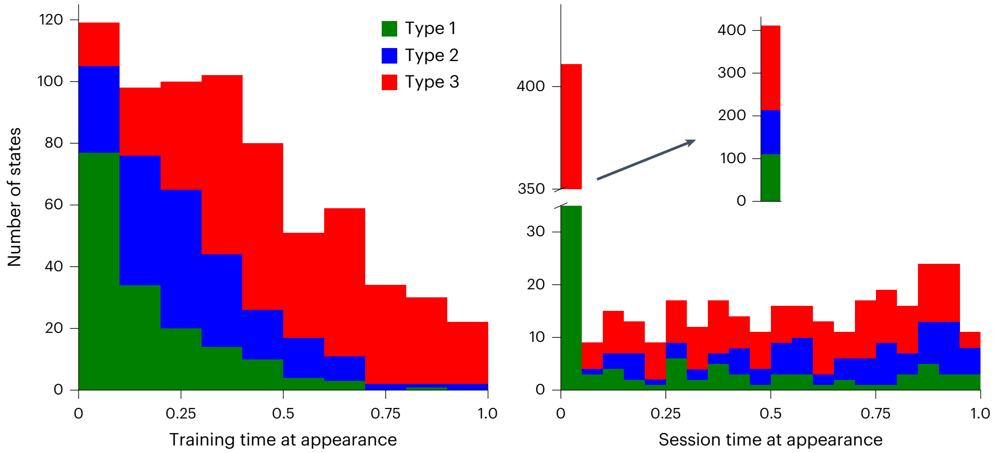


Fig. 7 | Histograms of all state introductions. The first state of every animal (which necessarily occurred on the first trial of the first session) was excluded. State introductions are shown across all of the training sessions (left) and within sessions (right). We color by state type (green, blue and red for types 1, 2 and 3, respectively) and normalize the entire length of training of an animal, as well as all individual sessions, onto the range between 0 and 1 for comparison purposes. The inset on the right plot shows the bar of the first time bin uninterrupted.

Interindividual differences and variability. So far, we have highlighted general patterns during learning, but perhaps even more salient than these similarities was the wide-ranging variability across animals. Such differences are already visible in many of the plots above. Biases in type 1 states spanned the entire range of possible response patterns. Similarly, type 2 PMFs appeared to be randomly biased toward one side or the other, or, rarely, symmetric. We were particularly surprised to find no regularity between type 1 and type 2 biases. Of the 56 mice in which type 2 onset occurred suddenly, 31 had expressed the same direction of bias (average choice from the PMF being more than 5% away from 50%) as the new type 2 state in any previous type 1 state, whereas 25 had not (two-sided binomial test for whether the proportion of previously expressed biases differs from 0.5 gives $P = 0.504$). Thus, we were unable to predict future biases of the animal from its stage 1 biases.

The number of sessions mice required to learn varied greatly, spanning an order of magnitude. Surprisingly, many animals with a large number of sessions were fitted by a small number of states, which changed considerably via the slow process, as exemplified in Extended Data Fig. 1 (notably, our recovery analyses indicate that the model can cope effectively with long training trajectories, as described in Methods). We revisit this issue in the 'Discussion'.

The number of sessions spent in the different stages was similarly highly variable. To gain insight into the factors underlying the learning steps between the stages, we analyzed the correlations between the number of sessions spent in them. The simplex plot does not strongly indicate any patterns. We quantify this as follows: duration of stage 1 to stage 2—Pearson's $r = 0.21$, $P = 0.015$; stage 1 to stage 3—Pearson's $r = 0.04$, $P = 0.685$; stage 2 to stage 3—Pearson's $r = 0.14$, $P = 0.095$ ($n = 134$ mice). Notably, the main chunks of training time, stages 1 and 3, show no correlation whatsoever. A speedy understanding of the basic contingency of the task, therefore, did not tend to go along with the ability (or will) to perfect this behavior quickly, suggesting that they required different competencies. The strongest correlation exists between stages 1 and 2, which makes sense insofar as they were both concerned with discovering how to make use of the stimulus information.

Beyond the training sessions analyzed here, the mice underwent a further phase ('biased block training', in which left or right stimuli dominated in blocks of 20–100 trials). Consistent with our other results, the length of this phase also turned out not to positively correlate with the total prebias training duration, nor with any of the stage durations. At most, there was a negative correlation between the overall bias training time and the stage 3 duration (see Supplementary Results for details).

Discussion

We presented a highly flexible model that describes the stages of learning from the very first day an animal interacts with a task until it becomes an expert. Using it on the shaping sessions of the IBL decision-making task, we showcased a number of useful capabilities of this approach. It allowed us to distinguish fast, abrupt transitions in behavior, and slower, gradual ones. Learning on this task decomposed into the following three distinct stages, through which almost all animals went: initial, undifferentiated and often biased behavior; partial, one-sided understanding of task contingencies; and, finally, full understanding of the task. While these broad-stroke characteristics were consistent across mice, and indeed resonate with recent results from other tasks[33,34], the details of behavior in these stages differed considerably across the population. Similarly, the way they progressed through these stages differed widely in both its duration and the composition of the sudden and gradual steps.

We found only a weak correlation between the time it took individual mice to progress through some of the behavioral stages, suggesting that they had to draw upon largely different skills to learn the requirements of the task. Similarly, animals expressed varying, largely uncorrelated, biases across the stages of learning. They might therefore have different sources—in stage 1, when the mice paid no attention to the stimulus, biases might be motoric; in stage 2, they could have been an expression of the side that individual mice happened to notice first as being informative; in stage 3, they might have stemmed from differences in sensory acuity. Beyond the initial training considered here, the duration of the subsequent biased block training of the animals did not exhibit positive correlations to the training phase durations (as elaborated in Supplementary Results). This again shows that learning was influenced by a large number of factors in our setting.

We originally expected that mice who took many sessions to train would be characterized by many states. However, although recovery analyses show that the model can cope effectively with long trajectories, this was not always the case. Instead, we often saw that few states took a long way via the slow process, from uninformed to proficient (Extended Data Fig. 1). It will be important to assess the underlying nature of these states and their progression by tracking neural data through the course of learning.

It is important to note that our model does not require as large a dataset as we used. Individuals were fitted by themselves, the model proved flexible enough to accommodate considerably different numbers of training sessions, and our cross-validation indicates that the fits are not critically sensitive to hyperparameter selection, the only part which made use of all subjects combined. Nevertheless, our modeling approach does have a number of limitations. First, the setting of the slow change variance parameter, which determined how much the behavior of a state could change from one session to the next, has a critical role in steering the trade-off between introducing a new state versus adapting an existing one. We optimized this parameter in terms of cross-validation performance for the entire population (Methods). However, the magnitude of slow changes may depend on the individual or vary across training time, and thus, a more differentiated treatment might be appropriate. Furthermore, slow changes may also occur within a session[28], which could be incorporated into the model by adding additional time points at which weights can change. This might well be necessary to apply our method usefully to the sort of more rapid, continuous changes that occur within a single session. Another desirable extension would be to allow the duration distributions to change over sessions. As training progresses, an animal might, for instance, be able to use a highly performant state for longer. Similarly, a dynamic transition matrix and dynamic initial state distribution might better capture the evolution of state usage across training.

The model may be extended by making the states predict additional observations, as binary choice behavior may limit the power to distinguish between behavioral modes. One obvious possibility is the reaction times of the animal's choices; in principle, this would only require adding a suitable distribution to produce times for each state (for example, from a drift diffusion decision-making process[35,36]). It would likely be necessary to make the distributions dynamic, as the reaction times improve with training. Other possibilities include pupil dilation or even body posture[37].

Previous work using an HMM-based approach discovered demotivated states in behavior during the first 90 unbiased trials per session in the subsequent biased block training[21]. The prevalence of sizable blocks of trials during which the animal performs at a decreased level will, if left unaccounted for, lead to confounded estimates of model parameters and a flawed understanding of the animal's current skill development stage, making it an integral component of a good behavioral model of this task. We also find such states, characterized by reduced sensitivity to the contrast feature of at least one side, and a strong bias in extreme cases, leading to higher-than-normal lapse rates on strong contrasts. However, these were not as pervasive as might have been expected from ref. 21. For us, a majority of sessions were dominated by a single state. The model sometimes acknowledged the dip in performance of the animals at the ends of sessions for tens of trials with a separate state (as shown in Fig. 2 on multiple sessions). We analyze aspects of these trials in 'Posterior predictive checks'. However, frequently, we just see a decrease in the prevalence of all sufficiently represented states. The main source of behavioral variability in our data came from learning and other large jumps in psychometric space; therefore, the model used its capacity to capture these.

Besides task acquisition, our approach to capturing behavioral evolution, which has conceptual relations to those used in the animal conditioning[38,39], structure learning[40] and motor learning[8,41] literatures, should be well suited to model other progressive changes, such as those occurring during ageing[42]. Furthermore, our framework can be flexibly adapted to other cases of long-run learning. For instance, it is possible to tune the model to capture minute changes within sessions rather than broad-stroke states across sessions, as here, by adjusting the propensity to infer new states for small changes in behavior. Equally, the modular resampling procedure of the model allows it to be adapted to different kinds of observations, for example, multinomial or Gaussian, by simply swapping out the inference mechanism of this component (although only some distributions are convenient for the gradual dynamics). We therefore hope that the tool we developed here will enable a wide range of researchers to study behavioral development in a systematic and revealing manner.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-025-02130-x.

References

1. Maier, N. R. Reasoning and learning. *Psychol. Rev.* **38**, 332 (1931).
2. Breland, K. & Breland, M. A field of applied animal psychology. *Am. Psychol.* **6**, 202–204 (1951).
3. Krueger, K. A. & Dayan, P. Flexible shaping: how learning in small steps helps. *Cognition* **110**, 380–394 (2009).
4. Köhler, W. in *Readings in the History of Psychology* (ed. Dennis, W.) 497–505 (Appleton-Century-Crofts, 1948).
5. Epstein, R., Kirshnit, C. E., Lanza, R. P. & Rubin, L. C. 'Insight' in the pigeon: antecedents and determinants of an intelligent performance. *Nature* **308**, 61–62 (1984).
6. Rescorla, R. A. in *Classical Conditioning II: Current Research and Theory* (eds Black A. H. & Prokasy, W. F.) 64–99 (Appleton-Century-Crofts, 1972).
7. Moore, S. & Kuchibhotla, K. V. Slow or sudden: re-interpreting the learning curve for modern systems neuroscience. *IBRO Neurosci. Rep.* **13**, 9–14 (2022).
8. Luft, A. R. & Buitrago, M. M. Stages of motor skill learning. *Mol. Neurobiol.* **32**, 205–216 (2005).
9. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proc. Natl Acad. Sci. USA* **101**, 13124–13131 (2004).
10. Piaget, J. *The Origins of Intelligence in Children* (W. W. Norton & Co, 1952).
11. The International Brain Laboratory et al. Standardized and reproducible measurement of decision-making in mice. *eLife* **10**, 63711 (2021).
12. Kastner, D. B. et al. Spatial preferences account for inter-animal variability during the continual learning of a dynamic cognitive task. *Cell Rep.* **39**, 110708 (2022).
13. Akiti, K. et al. Striatal dopamine explains novelty-induced behavioral dynamics and individual variability in threat prediction. *Neuron* **110**, 3789–3804 (2022).
14. Dayan, P., Roiser, J. P. & Viding, E. in *Psychiatry Reborn: Biopsychosocial Psychiatry in Modern Medicine* (eds Savulescu, J. et al.) 213–228 (Oxford University Press, 2020).
15. Smith, A. C. et al. Dynamic analysis of learning in behavioral experiments. *J. Neurosci.* **24**, 447–461 (2004).
16. Papachristos, E. B. & Gallistel, C. R. Autoshaped head poking in the mouse: a quantitative analysis of the learning curve. *J. Exp. Anal. Behav.* **85**, 293–308 (2006).
17. Jang, A. I. et al. The role of frontal cortical and medial-temporal lobe brain areas in learning a Bayesian prior belief on reversals. *J. Neurosci.* **35**, 11751–11760 (2015).
18. Maggi, S. et al. Tracking subject's strategies in behavioural choice experiments at trial resolution. *eLife* **13**, 86491 (2024).
19. Durstewitz, D., Vittoz, N. M., Floresco, S. B. & Seamans, J. K. Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* **66**, 438–448 (2010).
20. Song, M., Baah, P. A., Cai, M. B. & Niv, Y. Humans combine value learning and hypothesis testing strategically in multi-dimensional probabilistic reward learning. *PLOS Comput. Biol.* **18**, 1010699 (2022).
21. Ashwood, Z. C. et al. Mice alternate between discrete strategies during perceptual decision-making. *Nat. Neurosci.* **25**, 201–212 (2022).

22. Calhoun, A. J., Pillow, J. W. & Murthy, M. Unsupervised identification of the internal states that shape natural behavior. *Nat. Neurosci.* **22**, 2040–2049 (2019).
23. Gershman, S. J. & Blei, D. M. A tutorial on Bayesian nonparametric models. *J. Math. Psychol.* **56**, 1–12 (2012).
24. Heald, J. B., Lengyel, M. & Wolpert, D. M. Contextual inference underlies the learning of sensorimotor repertoires. *Nature* **600**, 489–493 (2021).
25. Johnson, M. J. & Willsky, A. S. Bayesian nonparametric hidden semi-markov models. *J. Mach. Learn. Res.* **14**, 673–701 (2013).
26. Beal, M., Ghahramani, Z. & Rasmussen, C. The infinite hidden Markov model. In *Proc. Advances in Neural Information Processing Systems* (eds Dietterich, T. et al.) Vol. 14 (MIT, 2001).
27. Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. Hierarchical Dirichlet processes. *J. Am. Stat. Assoc.* **101**, 1566–1581 (2006).
28. Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D. & Pillow, J. W. Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron* **109**, 597–6106 (2021).
29. Miller, K. J., Botvinick, M. M. & Brody, C. D. From predictive models to cognitive models: separable behavioral processes underlying reward learning in the rat. Preprint at *bioRxiv* https://doi.org/10.1101/461129 (2021).
30. Beron, C. C., Neufeld, S. Q., Linderman, S. W. & Sabatini, B. L. Mice exhibit stochastic and efficient action switching during probabilistic decision making. *Proc. Natl Acad. Sci. USA* **119**, 2113961119 (2022).
31. Findling, C. et al. Brain-wide representations of prior information in mouse decision-making. *Nature* **645**, 192–200 (2025).
32. Berditchevskaia, A., Cazé, R. D. & Schultz, S. R. Performance in a go/nogo perceptual task reflects a balance between impulsive and instrumental components of behaviour. *Sci. Rep.* **6**, 27389 (2016).
33. Dekker, R. B., Otto, F. & Summerfield, C. Curriculum learning for human compositional generalization. *Proc. Natl Acad. Sci. USA* **119**, 2205582119 (2022).
34. Liebana, S. et al. Dopamine encodes deep network teaching signals for individual learning trajectories. *Cell* **188**, 3789–3805.e33 (2025).
35. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
36. Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
37. Wiltschko, A. et al. Mapping sub-second structure in mouse behavior. *Neuron* **88**, 1121–1135 (2015).
38. Gershman, S. J. & Niv, Y. Exploring a latent cause theory of classical conditioning. *Learn. Behav.* **40**, 255–268 (2012).
39. Lloyd, K. & Leslie, D. S. Context-dependent decision-making: a simple Bayesian model. *J. R. Soc. Interface* **10**, 20130069 (2013).
40. Teng, T., Li, K. & Zhang, H. Bounded rationality in structured density estimation. In *Proc. 37th Conference on Neural Information Processing Systems* (eds Oh, A. et al.) Vol. 36, 25211–25237 (Curran Associates, 2023).
41. Heald, J. B., Wolpert, D. M. & Lengyel, M. The computational and neural bases of context-dependent learning. *Ann. Rev. Neurosci.* **46**, 233–258 (2023).
42. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U. & Bäckman, L. Memory aging and brain maintenance. *Trends Cogn. Sci.* **16**, 292–305 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

International Brain Laboratory

Sebastian A. Bruijns[1,2], Kcénia Bougrova[3], Inês C. Laranjeira[3], Petrina Y. P. Lau[4], Guido T. Meijer[3], Nathaniel J. Miska[5], Jean-Paul Noel[6], Alejandro Pan-Vazquez[7], Noam Roth[8], Karolina Z. Socha[4], Anne E. Urai[9] & Peter Dayan[1,2]

Methods

In this section, the data source is described briefly, followed by a detailed explanation of the infinite hidden semi-Markov model. Inference for the logistic regression observation distributions is then covered, with a focus on the resampling steps. Together, these components make up the full diHMM. The aggregation of generated samples is then explained, addressing challenges such as label switching and multimodality to define clear states. The process for assigning states and their PMFs to the three types is described in 'Psychometric type classification' section. Finally, validation analyses are presented, including cross-validation for parameter and prior selection, model ablations, posterior predictive checks, and recovery of generative models.

Ethics statement

All procedures and experiments were carried out in accordance with the local laws and approval was obtained from the following the relevant institutions: the Animal Welfare Ethical Review Body of University College London (P1DB285D8); the Institutional Animal Care and Use Committees of Cold Spring Harbor Laboratory (1411117; 19.5), Princeton University (1876-20) and University of California at Berkeley (AUP-2016-06-8860-1); the University Animal Welfare Committee of New York University (18-1502); the Portuguese Veterinary General Board (0421/0000/0000/2016-2019).

Animals and behavioral data

The data we used were collected under the IBL protocol, as described in detail in ref. 11 and its accompanying materials. The study subjects were female and male C57BL6/J mice, aged 3–7 months, which were cohoused whenever possible. Mice were kept in a 12-h light/12-h dark cycle and fed a diet containing 5–6% fat and 18–20% protein. No statistical methods were used to predetermine our sample size, but the IBL represents a large-scale approach to data collection and offers an exceptionally large dataset of learning trajectories (covering more individuals than the studies on learning by, for example, refs. 12,13). There was no blinding of experimenters, as there were no experimental groups. The stimulus sides and strengths that animals were presented with were independently drawn for each session (although the debiasing protocol could affect these probabilities, and weaker contrasts were introduced in a performance-dependent manner). When using Pearson's *r* to quantify correlation, the data distribution was assumed to be normal, but this was not formally tested.

Infinite hidden semi-Markov model

We start by describing the diHMM, focusing on Bayesian inference over its random variables. Following ref. 25, we use Gibbs sampling, an MCMC algorithm, to realize an iterative resampling scheme over the model components, including the PMFs of the hidden states and the assignments of the individual trials onto those states. For this purpose, all distributions are paired up with conjugate priors in this section, to enable simple resampling steps. The posterior distribution is ultimately represented by a collection of samples, with every component being assigned an explicit value in each sample.

We first describe all the relevant random variables, using the iterator notation from Supplementary Table 1.

The technical backbone of an infinite HMM is a hierarchical Dirichlet process. At the top of the hierarchy of this process is the prototypical transition vector

$$\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma), \tag{2}$$

where GEM (named after Griffiths, Engen and McCloskey) is a Dirichlet process without a base distribution, a pure stick-breaking process that samples a probability vector over infinitely many elements (which will be states in our case). The concentration parameter γ probabilistically determines the size of the individual sticks and, therefore, the number of practically relevant states, with higher γ encouraging more states. We put a vague Gamma prior on γ, making it, and thereby the propensity to introduce new states, part of the inference as well, with γ ~ Gamma(0.01, 0.01).

At the next level, we sample the transition vectors, a classical HMM component, $\boldsymbol{\pi}_i$, of the individual states i. These are tied together via $\boldsymbol{\beta}$, which is used as the base distribution for a second Dirichlet process

$$\boldsymbol{\pi}_i \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}), \qquad i = 1, 2, \dots, L, \tag{3}$$

$$\boldsymbol{\pi}_\mathbf{0} \sim \mathrm{GEM}(3). \tag{4}$$

α is another concentration parameter and determines how closely the $\boldsymbol{\pi}_i$ are related to $\boldsymbol{\beta}$. Sampling the individual state transition vectors from this common source formalizes an overall kind of state popularity. The higher α, the more like $\boldsymbol{\beta}$ is $\boldsymbol{\pi}_i$, $\forall i$, and so the more the bias in the frequency of state i' in the particular sample $\boldsymbol{\beta}$ will be reflected in the transitions from i to i', and so the more popular i' will be overall. We put another vague Gamma prior on it, α ~ Gamma(0.01, 0.01). The initial state distribution $\boldsymbol{\pi}_0$ is drawn entirely separately, with a concentration parameter of 3 as a trade-off between allowing new states but not encouraging the invention of new states at the start of sessions.

For our inference scheme, we make use of the weak-limit approximation, which puts an upper limit $L = 15$ on the number of states, rather than using the full infinite process. This simplifies the resampling scheme, while still behaving similarly to an infinite HMM if L is sufficiently large. Across the entire population, there were only three mice with 12 states, after applying our hierarchical state clustering procedure ('Aggregation and interpretation of chains' section); all other mice used fewer states. Furthermore, the minimum fraction of trials captured in states (as described further below) is 99.38% (mean = 99.97%), justifying the choice of $L = 15$ (although a higher limit would possibly allow us to capture motivational fluctuations better). In particular, we still perform inference over the realized state complexity. In the weak-limit framework, equations (2)–(4) turn into L-dimensional Dirichlet distributions

$$\boldsymbol{\beta} \sim \mathrm{Dir}(\gamma/L, \dots \gamma/L), \tag{5}$$

$$\boldsymbol{\pi}_i \sim \mathrm{Dir}(\alpha\boldsymbol{\beta}_1, \dots, \alpha\boldsymbol{\beta}_L), \qquad i = 1, 2, \dots, L, \tag{6}$$

$$\boldsymbol{\pi}_0 \sim \mathrm{Dir}(3/L, \dots, 3/L). \tag{7}$$

The transition structure within a session is given by

$$z_{n,1} \sim \boldsymbol{\pi}_0, \tag{8}$$

$$z_{n,s} \sim \boldsymbol{\pi}_{z_{n,s-1}}, \tag{9}$$

where $z_{n,s} \in \{1\dots L\}$ is an indicator for the sth state within a session n (which does not align with the trial number), and $\boldsymbol{\pi}_0$ is the initial state distribution.

Given the transition vectors, the workings of the hidden semi-Markov model are fairly standard, except that the duration distributions are specified explicitly rather than being drawn from a geometric distribution (as in a regular HMM). We therefore prohibit self-transitions, which makes a data-augmentation scheme for resampling necessary, as described in ref. 25. Nevertheless, as in a standard HMM, durations are statistically independent of the target state of transitions. Durations are drawn from a negative-binomial distribution, with state-specific random variables, coming from their own priors

$$r_i \sim U(5, 6, 7, \dots, 704), \qquad i = 1, 2, \dots, L, \tag{10}$$

$$p_i \sim \mathrm{Beta}(1,1), \tag{11}$$

$$d_{n,s} \sim \mathrm{NB}(r_{z_{n,s}}, p_{z_{n,s}}). \tag{12}$$

Note the difference between state names i, which hold for the entire model, and the session-specific state counters s, which can be used to find the current state name via the indicator $z_{n,s}$. We chose a uniform prior over a large range of numbers for the possible values of r, to enable long durations, but excluded small values for r (in particular, $r = 1$ would give the geometric distribution). Small values of r encourage transitions after a very small number of trials, which would capture the statistics of the presentation of left and right stimuli by the experimenter rather than the longer-lasting states that we sought. Using cross-validation, we ensured that enabling larger values of r did not benefit the fits.

States stay active and generate observations for as long as the drawn duration indicates

$$t_n(s) = \sum_{k=1}^{k<s} d_{n,k}, \tag{13}$$

$$x_{n,t_n(s)+1:t_n(s)+d_s} = z_{n,s} \tag{14}$$

$$P(y_{n,t} = R) = \mathrm{sig}\left(\mathbf{f}_{n,t} \times \mathbf{w}_{x_{n,t},n}\right), \tag{15}$$

where we defined $t_n(s)$ to return the trial on which the sth state of a session n starts, which allows for the definition of $x_{n,t}$, the state on any given trial t. We denote the logistic sigmoid function as sig. This takes the dot product between the state weights $\mathbf{w}_{s,n}$ (which we discuss in the next section) and the input features of the current trial $\mathbf{f}_{n,t}$ and produces the probability over the observation $y_{n,t}$. The binary response variable y has 0 representing a leftward, and 1 a rightward choice. See also Supplementary Fig. 2 for a visual summary of these variables.

We summarize this collection of variables as

$$\Theta = \left\{\gamma, \alpha, \beta, \boldsymbol{\pi}_0, \left\{\boldsymbol{\pi}_i, r_i, p_i, \{\mathbf{w}_{i,n}\}_{n=1}^{N}\right\}_{i=1}^{L}, \left\{\{x_{n,t}\}_{n=1}^{N_t}\right\}_{t=1}^{T}\right\},$$

where N is the total number of sessions. The connections between these variables are visualized in the form of a graphical model in Supplementary Fig. 3. The result of inference is a set of samples $\{\Theta_j\}_{j=1}^{J}$. Each sample is a full instantiation of the listed random variables, which we can treat as a posterior representation. Gibbs sampling works by iteratively sampling each variable from its distribution, given all other variables in the model. After updating all variables, the result is one new sample within the MCMC chain. Details on how to resample the individual components can be found in ref. 25.

Dynamic logistic regression prior and sampling

Gibbs sampling resamples each random variable conditioned on all others. Thus, inference over the observation distributions of the states is separate from almost all the rest of the model, only using the information as to which trial is currently assigned to which state. We drop the explicit state dependence i in $\mathbf{w}_{i,t}$ for this section, but it is important to keep in mind that this sampling scheme is applied to every state individually, with each state s being influenced only by trials for which $x_{n,t} = s$ in the current sample. We implement slow changes in the characteristics of the states by putting a Gaussian random walk prior on the weights $\mathbf{w}_n$, allowing for modest change across session boundaries, parameterized by the variance σ. We choose a diffuse initial distribution for the weights and use cross-validation to select the intersession variance $\sigma = 0.04$ (we performed cross-validation on a range of small values, to limit the state adaptation process to small changes)

$$\mathbf{w}_1 \sim \mathcal{N}(0, 8I), \tag{16}$$

$$\mathbf{w}_{n+1} \sim \mathcal{N}(\mathbf{w}_n, \sigma I), \tag{17}$$

where I denotes the identity matrix. If a state has no trial assigned to it in a particular session, its weights are held fixed during the next transition, preventing states from morphing radically during a prolonged absence.

Inference for the logistic regression weights is performed using Pólya-Gamma data-augmentation, which allows for efficient inference in settings with binomial likelihoods[43,44], because it is not possible to choose a conjugate prior. We review the relevant computations here; for a full treatment, we refer to ref. 45. In the first step of the resampling scheme, we sample pseudo-observations. This uses a Pólya-Gamma distribution PG, by first sampling $\omega_n \sim \mathrm{PG}(b_n, \psi_n)$, where $\psi_n = \mathbf{f}_n \times \mathbf{w}_n$ is the dot product of features and weights, and b_n is the total number of times this exact instantiation of features was observed in session n. However, the same state is associated with more than just one specific instantiation of features (that is, including contrasts of different strengths and sides and different response histories). To handle this, we treat a single session as multiple different time points, but prevent weight changes between time points that belong to the same session. In this way, the observations from different features within the same session are effectively aggregated. To complete the pseudo-observation generation, we need $\kappa_n = a_n - b_n/2$, where a_n is the number of rightward answers observed for the current ψ_n under consideration. Now $z_n = \kappa_n/\omega_n$ can be treated as if they were drawn from $\mathcal{N}(\psi_n, 1/\omega_n)$.

This data-augmentation serves the purpose of having the $\mathbf{w}_n$ emit observations with Gaussian noise (after combination with the features $\mathbf{f}_n$ into ψ_n). Because the prior on $\mathbf{w}$ is a Gaussian random walk, this places inference in the well-studied realm of Kalman filtering. To resample the $\mathbf{w}_n$, we use the forward filter backward sample algorithm[46,47], which filters forward through all the observations using a Kalman filter, then samples the sequence of $\mathbf{w}_n$ backwards through time. A single resampling step, therefore, consists of first drawing the Pólya-Gamma variables to create pseudo-observations, then using them to sample the $\mathbf{w}_n$ using the forward filter backward sample algorithm.

We consider four features for the logistic regression—the contrast on the left side, the contrast on the right side, an exponentially weighted history over all previous choices and a bias. Separating the features for left and right contrast allows the sensitivities to the two sides to be different. Because the notional contrast values do not match the psychophysical difficulty of the contrasts (100% and 50% are both virtually equally easy to perceive, not a factor of 2 apart), we apply a transformation to have a better alignment. For this, we follow ref. 28 and use a tanh transformation, mapping the actual contrast c onto the input $\tilde{c}$ for our logistic regression through $\tilde{c} = \tanh(pc)/\tanh(p)$, where we follow their recommendation and set $P = 5$, which scales the steepness of the transformation. This maps the contrasts (1, 0.5, 0.25, 0.125, 0.0625, 0) onto (1, 0.987, 0.848, 0.555, 0.302, 0).

The regressor for previous answers, enabling perseveration as a strategy, proved to be beneficial for cross-validated performance. It is associated with the famous law of exercise[48,49] and has also been found to be exhibited by the mice in the asymptotic regime that arises after the sessions that we are presently analyzing[31]. The same analyses showed no general statistical support for a regressor sensitive to the interaction between past choice and past reward, as would be reflected, for instance, in win-stay, lose-shift behavior. We implement the perseveration regressor as an exponentially weighted sum over all past trials. We found that weighting previous trials with an exponentially decaying filter with a smoothing factor of 0.25 worked best (although slightly different parameter settings have almost equal cross-validation performance). Thus, we compute this feature on session n and trial m as such

$$\frac{1}{Z}\sum_{k=1}^{m-1} \exp(-0.25\,k)\,(2\,y_{n,m-k} - 1), \tag{18}$$

where $Z = \sum_{k=1}^{m-1} \exp(-0.25\,k)$ is a normalization constant, such that the entire exponential filter adds to 1. The transformation $2\,y - 1$ serves to encode responses as −1 and 1, for the purpose of having the perseverative feature sway the current response appropriately. Therefore, this feature reaches its maximal value of 1 if all previous responses were rightward and −1 if they were all leftward, putting it on the same scale as the other features. Timeout trials, where the animal did not respond before 60 s had passed, while skipped for the logistic regression of responses, are taken into account for the previous answer regressor, encoded as 0.

Aggregation and interpretation of chains

We generally generated 48,000 samples from each of 16 chains (with different starting points), discarding the first 4,000 as burn-in. We assessed convergence of the chains using the classical measure $\hat{R}$[50] and generated more samples by continuing each chain if necessary (although not all animals ever reached a sufficiently low $\hat{R}$ score, we excluded 12 animals for this reason). $\hat{R}$ compares intrachain and interchain variability of bespoke, state-independent features of the chains. To detect differences in the variances of the chains and other problems, which $\hat{R}$ is known to miss, we also used folded-$\hat{R}$ and rank-normalized-$\hat{R}$ [51]. We reduced the memory cost by thinning the chain, using only every 25th sample (we did this purely for memory reasons, not because it is necessary for MCMC algorithms[52]). For a first pass, we sought to discard chains that differed substantially from other chains in the explored region in parameter space, either because they never reached the relevant parts of it or because they spent disproportionate amounts of time in some modes over others. This is a known problem for MCMC algorithms in multimodal environments and can be mitigated by taking nonmixed chains and combining them via stacking[53]. However, because our goal here is not prediction, we still want to focus on finding and visualizing the most important modes of the posterior, which we did by combining the (possibly not perfectly mixed) chains, and considering the regions of probability space in which they collectively spent the most time. Given the slow transitions between different modes, we also did not split our individual MCMC chains when computing $\hat{R}$, as the two halves of the chains were often too different.

As scalars underlying $\hat{R}$, we used the concentration parameters α and γ, as they are independent of states. We also included general properties of the fit—the number of trials assigned to the state with the most trials, the second-most trials and the overall numbers of states with more than 20% and more than 10% of trials assigned to them (we chose multiple cutoffs to gain information about the fit at different levels of resolution). By greedily discarding the chains that increase $\hat{R}$ the most, we reduced the number of chains under consideration from 16 to at least 8. For this, we considered all features and all variants of $\hat{R}$ (normal, folded, rank-normalized) at once, so we were minimizing the maximum over all these $\hat{R}$s. We only further processed the chains when $\hat{R} < 1.05$, which is more conservative than some recommendations, but, in light of the strong multimodality, more lenient than the newest ones[51].

However, it is still not trivial to extract information from the remaining chains given the multimodality. There are two main sources of multimodality, which are as follows: (1) genuine uncertainty in the usage of states or the exact setup of the random variables of the states, and (2) mode equivalence with permuted labels (for example, state $i = 1$ in the first chain might explain roughly the same set of trials as state $i = 2$ in the second). Although the second source makes evaluating the results more complicated, it is in fact just the sampling scheme working correctly, as there is nothing special about the particular state labels—solutions with permuted state labels are functionally equivalent. For the same reason, even within a single chain, a relatively consistent set of trials might be explained by one label for some part of the chain, but by a different label in another. Indeed, we frequently observed this kind of label switching, where one state completely took over the trials of another within a few sampling steps. In the limit of infinitely many samples, we can expect any trial to have a uniform distribution over the state label assigned to it; the only important question is which other trials were usually accounted for by the same state as the given trial within suitably similar samples.

To formalize the necessary abstraction from direct state assignments, we computed co-occupancy matrices C^j for each sample j. C^j is a matrix of size $T \times T$, with T being the total number of trials across all sessions of a mouse, whose t, mth entry reports whether trials t and m (for convenience, dropping the additional session label) used the same state in sample j

$$C^j_{t,m} = \mathbb{1}(x_t = x_m). \tag{19}$$

We used these co-occupancy matrices as a basis for the following two different processing steps: (1) at a coarser resolution across trials, we applied dimensionality reduction to find posterior modes; (2) at full resolution, we averaged C^j across similar samples j to derive a matrix that describes the mutual affiliation of trials, allowing us to overcome the labeling issues. Both steps are reminiscent of representational similarity analysis[54], in that, instead of comparing two samples directly, we compare state co-occurrence within the samples.

In principle, to explore the posterior, we could have flattened each C^j into a $\mathbf{T}^2$ vector and applied principal components analysis (PCA). However, there were too many trials per mouse (of the order of 15,000) to do this at full resolution, so we binned the trials into 200 bins, ignoring session boundaries, and then used the Wasserstein distance to measure state co-occurrence between the bins. That is, we define modified matrices C'^j as

$$C'^j_{t,m} = \sum_{i=1}^{L} 1 - |p^j_{t,i} - p^j_{m,i}|, \tag{20}$$

where $p^j_{t,i}$ is the proportion of trials in bin t, which is assigned to state i in sample j. C'^j reduces to C^j for bins comprising a single trial. We then plotted individual samples in the first three dimensions of the PCA space arising from flattened versions of C'^j, as shown in Supplementary Fig. 4.

In doing this, we found that the posterior for a number of animals wanders itinerantly between different modes, reflecting true uncertainty. These modes are distinct solutions and should not be blended. To isolate them, we performed Gaussian density estimation in the 3D PCA space to identify the ones that were most prevalent, as the regions of highest estimated density. We used this clustering to select samples $j \in \mathcal{J}^\eta$ that were sufficiently similar as to comprise an individual mode $\mathcal{J}^\eta$. For now, we did this by hand; however, the process could be made more formal by fitting a mixture of Gaussians to the posterior and then selecting samples around the means of the Gaussians with sufficiently large mixture weights. We selected at least 400 samples from a mode to form a representative collection.

Next, we sought to understand how trials within that mode were co-assigned to states. To do this, we averaged the co-occurrences $C^\eta = \frac{1}{|\mathcal{J}^\eta|}\sum_{j \in \mathcal{J}^\eta} C^j$ and treated $\tilde{C}^\eta = 1 - C^\eta$ as a distance matrix, where trials were close if they shared a state in most samples in the mode. We then performed hierarchical clustering on $\tilde{C}^\eta$, using as a cluster distance $d(\upsilon, \nu) = \max(\tilde{C}^\eta_{\upsilon[k],\nu[l]}), k \in \upsilon; l \in \nu$, which took as the distance between clusters the maximum distance between any two trials in the clusters υ and ν. The result of the hierarchical clustering was a tree on the individual trials; cutting this tree at a certain level leads to a specific clustering. Thus, cutting at, say, 0.6 means that we only have clusters in which every trial was explained by the same state in at least 40% (1 − 0.6) of the samples. For our plots, we cut at 0.95, which empirically returned good results. Although this meant that trials needed to use the same state in only 5% of samples to be in one cluster, most trials were assigned to the same state much more frequently (Supplementary Fig. 5). This also shows a number of alternative clusterings from different thresholds, demonstrating that there is little change across a wide range of thresholds—the 95% threshold leads to 8 states with 100% trial coverage, an 80% cutoff leads to 9 states and 99.92% coverage, a 50% cutoff gives 12 states with 98.77% coverage and, finally, a threshold at 20% gives 15 states and 95.27% coverage. We can thus see that low criteria led to trials becoming unassigned and some states splitting apart, which is why we chose a rather high cutoff. A further verification that the procedure and its threshold gave a faithful representation of the collection of samples comes from comparing the overall solution against individual solutions from single samples. Empirically, these did indeed align. Our later recovery analyses also used this approach.

The states we show are therefore defined at heart by sets of trials. To compute the PMFs of such a set, we first considered a single MCMC sample and noted which states it assigns to the trials within this set on a session-by-session basis (although each individual trial only had one state assigned in a single sample, for the whole set of trials, it usually will not just have been a single state, due to random fluctuations, but mostly a single state). We turned the psychometric weights of these states into PMFs, over which we then averaged (in a weighted manner, considering how often any state occurred in the set of trials). For a single sample, this resulted in an average PMF of that state for each session. This then got averaged across samples within a cluster (evenly over all selected samples of a mode) to obtain the ultimate PMFs of this state.

To determine how closely a single trial is connected to its assigned state, we averaged the proportions of samples in which it was in the same state as all the other trials assigned to this state. That is, for a given trial t, we took a row of the consistency matrix C_t^η and considered only the entries corresponding to other trials within the state under consideration. We then averaged over those entries, yielding the average proportion of co-assignment. We think of this as a proxy of the posterior over which state a trial is assigned to, and we show it in Fig. 3.

Psychometric type classification

We observed by eye that the PMFs that the model found for the behavioral states had a tendency to fall into one of the following three characteristic classes: flat (type 1), half-tuned (type 2) and fully tuned (type 3). However, the boundaries between the classes were blurry, so we sought an objective distinction, recognizing its inevitable arbitrariness. Note that a state may change its type through the slow process; it is thus a session-dependent classification.

The measure we used in the main paper is the mean reward rate implied by the PMF on easy trials (100% and 50%), ignoring the effects of perseveration (and the debiasing protocol). We chose the reward rate because this tends to grow as the animals proceed from ignorance to competence. We chose to assess only the easy trials because early PMFs were not defined on the lower contrasts (because these stimuli were not presented), and including more difficult contrasts can lead to lower reward rates for more broadly defined PMFs, even when they are better on easy contrasts. Supplementary Fig. 6 shows the distribution of such reward rates across all states. It is apparent that there is a rather clear grouping of PMFs with reward rates below 0.6, defining type 1. The boundary between types 2 and 3 is somewhat less evident, implying that edge cases will be hard to assign. The threshold reward rate of 0.78 served reasonably well, as evidenced in Fig. 4. We further split type 2 PMFs on whether they were left-biased, right-biased or symmetric. We considered a PMF symmetric if its error rate on 100% leftwards contrasts was within 10 percentage points of the error rate on 100% rightwards contrasts.

Cross-validation and ablations

Our model contains a number of free parameters that we set using a cross-validation procedure. We used this most notably for the variance σ of the normal distribution over how much the logistic weights of a state can change from session to session and the decay constant of the exponential filter over previous actions, which are fixed parameters that are not inferred during the inference procedure. This inference procedure is itself guided by priors, which we set to be vague, exerting minimal influence upon the ultimate posterior. However, their precise setting can nevertheless also be evaluated via cross-validation. This applies to the two gamma distributions over the concentration parameters α and γ and the priors over the parameters of the states' duration distributions. Cross-validation also allowed us to verify that our usage of the weak-limit $L = 15$ did not hurt our model fits, and that including a win-stay lose-switch feature, indicating which side was or would have been rewarded on the previous trial, was not beneficial in capturing animal choices during learning.

We used a tenfold cross-validation scheme, randomly masking 10% of trials on each session. Because we were not interested in the details of the fits, we only ran one chain of 10,000 samples for each parameter combination and cross-validation fold we wanted to test and evaluated the quality of the fit through the summed negative log-likelihood on the last 4,000 samples on the held-out trials, which was sufficient for a stable estimation of the held-out log-likelihood. Despite this time-saving strategy, there were too many combinations of parameters to check exhaustively, so we used a manual heuristic search over promising combinations, finding an optimal setting and verifying that any relevant deviations from it only lowered the negative log-likelihood (Supplementary Fig. 7, left). As another measure to save computation, we only evaluated two folds of each animal for each parameter setting, but because we evaluated our model on 154 mice (this was before exclusions due to missing sessions or too low $\hat{R}$), we still evaluated on a substantial number of folds in total.

We tested the perseveration decay constant over the set of values (0.15, 0.2, 0.25, 0.3, 0.35, 0.4), the variance σ over the set (0.01, 0.02, 0.03, 0.04, 0.06, 0.12, 0.24), representing the small range that we found desirable for a consistent state identity, as well as some larger values to ensure that they did not outperform smaller variances. The search also included a larger support for the r parameter of the duration distribution (running from $r = 2$ to $r = 905$) and different settings of the α and γ concentration priors, which were independently varied over the set ((0.1, 0.1), (0.01, 0.01), (0.001, 0.001)).

Many of the parameter configurations yielded comparably high performance. Of note, the parameter setup closest to the selected model simply allows higher r values in the duration distribution, representing a strict extension of the model that, however, does not improve fit. When studying the correlations across two different parameter settings, but within the same animal and the same cross-validation fold, we found extremely strong correlations, with only slight offsets from the identity line and a small handful of outliers accounting for the differences. This provided evidence that the fits were fundamentally the same, and different mice did not significantly benefit from different settings, allowing us to simply take the best among many good settings and proceed with it for the population-wide fit. These settings were the ones specified throughout the study—perseveration = 0.25, $\sigma = 0.04$, $r_i \sim U(5, 6, 7, \ldots, 704)$ and both α and $\gamma \sim \text{Gamma}\ (0.01, 0.01)$.

In addition to finding the best parameters for our fit, we also used this approach to ablate the most important model components, verifying that all aspects of the model were necessary to provide as good a fit as possible within our framework (Supplementary Fig. 7, right). In particular, we tested the best parameter setting we found, but did not allow for change in weights between sessions (effectively removing the slow process of the model), both with 3 states (thus emulating the work described in ref. 21, although with duration distributions) and with the usual upper bound of 15 states. Allowing for 15 states but no slow process led to only somewhat worse performance than the full model (Supplementary Fig. 7, right—'15 states, no slow proc.'), but did so at the cost of significantly increasing the usage of short-lived states. We tested this by considering how many states explained more than x% of trials of an individual animal (which can be read directly from the cross-validation samples, not requiring the sample aggregation procedure described previously). The full model makes more use of highly prevalent states that explain more than 20% of trials—1.7 ± 0.53 (mean ± s.d.) per animal versus 1.07 ± 0.67 of such states for a model without the slow process (two-sided Mann–Whitney *U* test, $U = 21858.5$, $P < 1 \times 10^{-30}$, effect size = 1.18 (standardized mean difference with s.d. over full model state number), $n = 154$ mice), but fewer overall states, such as any that explain more than 2% of trials—5.16 ± 1.62 versus 9.13 ± 2.14 (two-sided Mann–Whitney *U* test, $U = 87773.5$, $P < 1 \times 10^{-73}$, effect size = 2.45, $n = 154$ mice). Thus, while the removal of the slow process can mostly be made up for by an increased reliance on new states (for which our model has plenty of capacity), the slow process benefits the fits by tying together highly similar trials across short timescales, rather than arbitrarily separating them when behavior gradually changes too much to be accommodated by a single state.

We also allowed only one state (including the slow process), removing the notion of multiple states from the fit (Supplementary Fig. 7, right—'1 state'). This model performed, perhaps surprisingly well, but because a session is usually dominated by a single state, a single adaptable state may perform somewhat well. We tested whether a win-stay lose-switch (WSLS) feature, indicating which choice was or would have been rewarded on the last trial, was beneficial, which it was not (Supplementary Fig. 7, right—'Best + WSLS'), and whether the perseveration feature could be removed, which it could not (labeled 'No perseveration'). Finally, we also tested the improvement due to the duration distributions (which replaced the implicit geometric duration distribution of an HMM; Supplementary Fig. 7, right, 'No duration (exp. only)'). This test proved somewhat problematic within our framework, as restricting the model to implement durations through the transition matrix led many of the posteriors to settle on an unsatisfying solution. In this solution, states were extremely strongly biased leftwards or rightwards and rapidly alternated, depending on the choice of the animal. Such a model has, of course, almost no predictive power on held-out trials. This is seemingly a consequence of the hierarchical nature of the transition matrix—if we often transition into a state (and without duration distributions, we have a state transition after every single trial, with most of them being self-transitions), it becomes generally attractive in the iHMM framework, encouraging transition distributions that are much closer to uniform than one would expect for a reasonable notion of temporally extended states. We thus implemented geometric distributions that prefer longer states by fixing $r = 1$, but biasing the prior over p. We performed another small cross-validation sweep and present here the best model found in this way.

Posterior predictive checks

To identify any mismatches between our modeling assumptions and actual behavior, we performed posterior predictive checks using multiple test statistics. The goal of this analysis was to determine whether responses generated solely from posterior samples reproduced the behavioral trends observed in the actual data. We simulated behavior for each session of an animal by taking each sample from our selected posterior mode, initializing with the state that was the actual state on the first trial for that sample and then generating responses. We needed to initialize with the true state, because the model uses a static initial state distribution $\boldsymbol{\pi}_0$, so a random initialization would lead to an unstructured mix of proficient and inexperienced behavior. However, after initializing the first state, the model ran completely independently—we drew a duration from the duration distribution of that state, using posterior parameters, randomly sampled a next state from the transition matrix once a state ended and sampled responses from the observation distribution of the current state, given the current features. These features included the contrast that was presented on that trial and a recomputed perseveration feature based on the choices of the current run of the simulation (so notably not the perseveration feature based on the choices of the animal). This unguided generation of behavior thus represents a very stringent test of the posterior fit.

We visualized the results by plotting actual behavior in relation to the distribution created by simulating behavior three separate times with each sample (because we use at least 400 samples from a mode, this equates to >1,200 simulations). As metrics of interest, we chose the percentage of correct choices in a session and the percentage of rightward choices for each contrast. We plot the accuracy of a single individual (the mouse of Fig. 2) in Supplementary Fig. 8a and the PMF on the last session of that animal in Supplementary Fig. 8b. As we can see here, behavior simulated from the posterior generally provides both a tight as well as accurate estimate around the true behavior.

To summarize the relationship between true behavior and the simulated distribution across the population, we calculated the percentiles of the empirical values within the simulated distribution, visualized in Supplementary Fig. 8c,d. In an ideal case, the histograms over these percentiles would be uniform, indicating that the posterior provides an unbiased and calibrated estimate for the true behavior. This is not quite true here—we can see that accuracy has a modest tendency to be overestimated (that is, the true accuracy tends to fall onto lower percentiles of the simulated distribution). As mentioned in the 'Discussion', behavior often degrades toward the end of a session (almost by necessity, as it is one of the session termination criteria), but this was not always acknowledged with a separate state by the model, perhaps because behavior degrades in a gradual and inconsistent manner across sessions. We suggest this as an interesting direction for a possible extension of our framework, by combining the states with a mechanism for change on a shorter time scale, similar to the work described in ref. 28. However, implementing this in a way that keeps states distinct and has them retain their identity over long time periods seems challenging, in the face of motivational changes that occur gradually but can change behavior quite notably on the order of tens of trials. Note that the overestimation of accuracy also occurs on sessions on which the model does ultimately include a state that reflects a substantial reduction in performance. This happens because the model sometimes fails to appropriately transition to this worse state (given that it is only a descriptive model with no foresight of when a session ends). Thus, accuracy in free simulations can be too great.

While the percentage of rightwards choices across contrasts forms a seemingly uniform distribution, splitting the histogram over the different contrasts reveals that there is a modeling assumption that biases the estimates for the different contrasts somewhat, as shown in Supplementary Fig. 8e. Most notably, for the 100% contrasts, the model underestimates how accurate the animals are (by overestimating the % rightwards choices on leftwards contrasts and vice versa). Note, however, that the insets for these contrasts show that the actual deviation is very small. Somewhat more subtly, the opposite occurs for the respective 50% contrasts. These deviations arise from the psychophysical transform we borrowed from refs. 21,28, namely the tanh transformation on the raw contrast values. The 100% and 50% contrasts are mapped onto very similar values (1 and 0.987, respectively), strongly coupling the percentage of rightward choices for the two contrasts, requiring them to take on almost the same value. This is intuitively desirable—allowing a smoothing over the different contrast strengths and reducing the number of parameters in our logistic regression (using a general 'leftwards sensitivity', rather than having a separate parameter for each contrast). While 100% and 50% are very different in terms of absolute value, they are both highly visible, meaning their difference from a psychophysical perspective is rather minor[55]. Nevertheless, as it turns out, some mice can occasionally exhibit rather different behavior on the two contrasts (Supplementary Fig. 8e, insets), leading to an underestimation for the stronger contrast and an overestimation on the weaker one.

The 0% contrast plot, on the other hand, exemplifies a posterior predictive check without such reservations—there is no noticeable bias, and the posteriors appear correctly calibrated. The predictive checks thus serve as an important tool to study the limitations of our modeling approach, highlighting that degrading behavior is not fully captured by the model and that the smoothing over contrasts imposes some structure onto the PMFs that biases the performance estimates. To study further the effect these biases in the model have upon the fits, we analyzed the magnitude of the bias imposed (Supplementary Fig. 16). As we can see, most of the differences fall within a close range around the posterior mean.

As a proof of concept, we refitted the model with a different PMF parameterization to see whether this could address the observed issue. This alternative parameterization was inspired by another line of work that uses neural networks to capture animal behavior on the IBL task. Using this, we mapped the contrast strengths (1, 0.5, 0.25, 0.125, 0.0625, 0) onto (1, 0.899, 0.705, 0.416, 0.207, 0) for the logistic regression, whereas the tanh transformation mapped onto (1, 0.987, 0.848, 0.555, 0.302, 0). The results of repeating the posterior predictive checks on a representative random sample of mice ($n = 84$) using this new parameterization are shown in Supplementary Fig. 9. This reveals that the tension between the predictive distributions on 100% and 50% contrasts was mostly caused by the PMF parameterization, rather than by the model itself. Because the fits under this new PMF did not qualitatively differ from fits under the old parameterization, we did not redo our analyses, but accepted this as evidence for the suitability of the model and fitting procedure. The remaining slight tension between the predictions on strong contrasts might be caused by changes in the perceptual sensitivities of the animals during learning, which is an interesting avenue to pursue in further studies of learning.

Model recovery

We tested the model and our inference procedures by fitting to data for which the ground truth was available. For this, we instantiated all the random variables of the model to specific values and generated responses from it. This was performed for multiple different variable settings to assess the accuracy of the fitting procedure in all relevant regimes and using input data (that is, contrast sequences) from actual training trajectories. The data generated this way were processed exactly as those from the IBL mice.

We paid particular attention to assessing the strength of the inductive biases of the inference procedure—particularly in terms of the number of states it inferred (given that this could be potentially unbounded, within our weak-limit approximation) and the degree of change between sessions (because slow and fast state changes could interact). We tested multiple settings in which all the data were actually generated from a single state, to test whether the model would incorrectly split behavior into multiple states. In one setting, the psychometric weights of the state stayed constant throughout all sessions. In another, the weights gradually evolved from poor performance to proficiency (at constant steps of a magnitude that corresponds to a variance of 0.0311; the variance of the fitting procedure was fixed to 0.03). Both fits recovered their ground truth successfully, explaining virtually all trials with a single state, as can be seen for the example of the changing state in Supplementary Fig. 10. We also tried a variation of the latter situation, in which the psychometric weights changed in (proportionally smaller) steps on every single trial, rather than all at once at a session boundary (as the model assumes). This, too, was recovered by the model with only one state (which we consider the best possible solution, given that the generative process was outside the model class).

We also successfully recovered settings from 2 to 9 states, with and without session-to-session variation on the weights, with strongly varying trial proportions between the different states (Supplementary Fig. 14) and of varying overall training lengths (particularly to test whether long training trajectories lead the model to impose fewer states, making more use of the slow process), as seen in Supplementary Fig. 11. The model was also tested on a setting with completely implausible PMFs, but with the added difficulty of having a larger number of states active within each session (Supplementary Fig. 15). This, too, was captured accurately. These successful recoveries suggest that the model can uncover states that truly correlate with distinct modes of animal behavior.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Please follow the instructions at https://int-brain-lab.github.io/iblenv/notebooks_external/data_download.html to download the data used in this article. Our public code contains a script to download the dataset.

Code availability

The code used for this analysis, as well as installation instructions for the necessary packages, can be found at https://github.com/SebastianBruijns/diHMM.

References

43. Polson, N. G., Scott, J. G. & Windle, J. Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Am. Stat. Assoc.* **108**, 1339–1349 (2013).
44. Linderman, S. W., Johnson, M. J. & Adams, R. P. Dependent multinomial models made easy: stick breaking with the Pólya-Gamma augmentation. In *Proc. 29th International Conference on Neural Information Processing Systems* (eds Cortes, C. et al.) Vol. 2, 3456–3464 (MIT, 2015).
45. Windle, J., Carvalho, C. M., Scott, J. G. & Sun, L. Efficient data augmentation in dynamic models for binary and count data. Preprint at https://arxiv.org/abs/1308.0774 (2013).
46. Carter, C. K. & Kohn, R. On Gibbs sampling for state space models. *Biometrika* **81**, 541–553 (1994).
47. Frühwirth-Schnatter, S. Data augmentation and dynamic linear models. *J. Time Ser. Anal.* **15**, 183–202 (1994).
48. Thorndike, E. L. *Animal Intelligence: Experimental Studies* (Macmillan Press, 1911).
49. Gershman, S. J. Origin of perseveration in the trade-off between reward and complexity. *Cognition* **204**, 104394 (2020).
50. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
51. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: an improved rˆ for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16**, 667–718 (2021).
52. Link, W. A. & Eaton, M. J. On thinning of chains in MCMC: thinning of MCMC chains. *Methods Ecol. Evol.* **3**, 112–115 (2012).
53. Yao, Y., Vehtari, A. & Gelman, A. Stacking for non-mixing Bayesian computations: the curse and blessing of multimodal posteriors. *J. Mach. Learn. Res.* **23**, 1–45 (2022).
54. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
55. Fechner, G. T. *Elemente der Psychophysik* Vol. 2 (Breitkopf & Härtel, 1860).

Acknowledgements

We thank J. Pillow, A. Churchland, A. Pouget, C. Tessereau, Z. Ashwood, N. Roy, I. Murray, S. Lindermann and the Behavioral Analysis and Theory working groups of the International Brain Lab for discussions. This work was funded by the Simons Foundation (to S.A.B. and P.D. under ID 552343 and the IBL in general), the

Max Planck Society (to S.A.B. and P.D.), the Wellcome Trust (to S.A.B. (IBL) and P.D. (ID 216324)) and the Alexander von Humboldt Foundation (to P.D.).

Author contributions

S.A.B. and P.D. initiated the project and developed the modeling framework. K.B., I.C.L., P.Y.P.L., G.T.M., N.J.M., J.-P.N., A.P.-V, N.R., K.Z.S., A.E.U. and the IBL as a whole collected the data. S.A.B. and P.D. analyzed the data and wrote the paper.

Funding

Open access funding provided by Max Planck Society.

Competing interests

The authors declare no competing interests.

Additional information

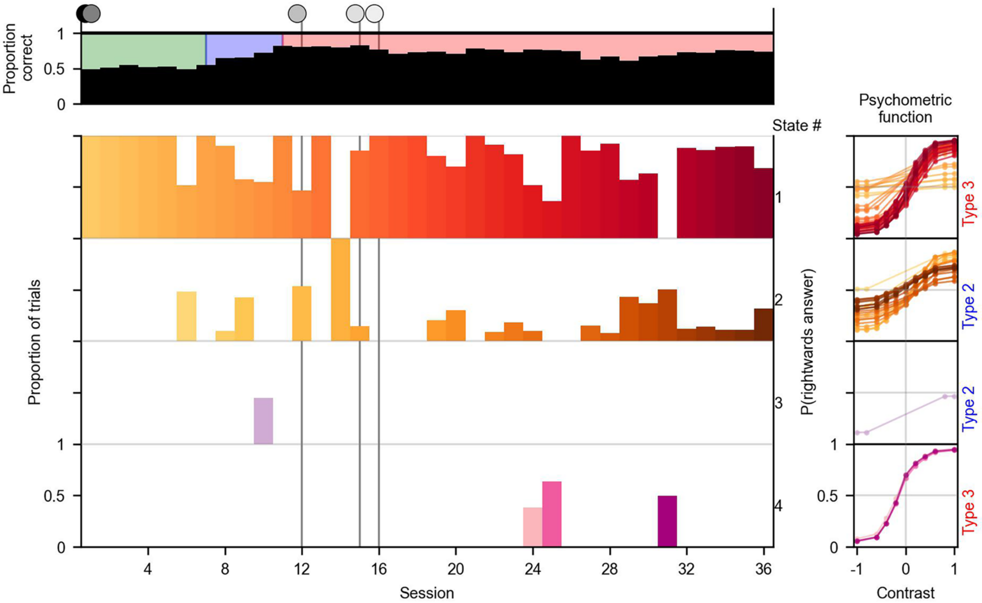
Extended data is available for this paper at https://doi.org/10.1038/s41593-025-02130-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-025-02130-x.

Correspondence and requests for materials should be addressed to Sebastian A. Bruijns.

Peer review information *Nature Neuroscience* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41593-025-02130-x



Extended Data Fig. 1 | Model fit to a mouse with a larger number of sessions. Using the plotting conventions of Fig. 2, this depicts the states and corresponding psychometric functions identified in a mouse that required 36 sessions to learn. This illustrates the counterintuitive phenomenon that long training trajectories were sometimes fitted by the model using a small number of states. Some of these states, particularly state 1 in this example, underwent substantial changes through the slow process, spanning the range of uninformed to proficient behavior (note that the type labels to the right of the PMFs are determined by the highest type reached by a state; thus, state 1 is labeled as type 3 even though it began as type 1).

nature portfolio

Corresponding author(s): Sebastian Bruijns

Last updated by author(s): Oct 2, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

□ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement

□ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

□ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.

□ ☒ A description of all covariates tested

□ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

□ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

□ ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
Give P values as exact values whenever suitable.

□ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

□ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

□ ☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

https://github.com/int-brain-lab/iblrig contains the code running on data collection rigs, precise task protocol identifiers are listed within the dataset.

Data analysis

Our analysis code can be found at the repository linked below. We make use of Markov chain Monte-Carlo algorithms, from these packages:

pybasicbayes (original version='0.2.4', our version linked below)
pyhsmm (original version='0.1.6', our version linked below)
pypolyagamma (version='1.2.3')

Code availability:

The analysis code with installation instructions is deposited at https://github.com/SebastianBruijns/diHMM

This uses:

https://github.com/SebastianBruijns/sab_pybasicbayes a modified version of https://github.com/mattjj/pybasicbayes

https://github.com/SebastianBruijns/sab_pyhsmm a modified version of https://github.com/mattjj/pyhsmm

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Please follow these: https://int-brain-lab.github.io/iblenv/notebooks_external/data_download.html instructions to download the data used in this article. Use for example the following code snippet to download the data using Python.

```
from one.api import ONE
import re

# use password as indicated on the website
one = ONE(base_url='https://openalyx.internationalbrainlab.org', password='*****')

regexp = re.compile(r'Subjects/\w*/((\w|-)+)/_ibl')
datasets = one.alyx.rest('datasets', 'list', tag='2023_Q4_Bruijns_et_al')

# extract subject names
subjects = [regexp.search(ds['file_records'][0]['relative_path']).group(1) for ds in datasets]
# reduce to list of unique names
subjects = list(set(subjects))

for subject in subjects:
    trials = one.load_aggregate('subjects', subject, '_ibl_subjectTrials.table')
    training = one.load_aggregate('subjects', subject, '_ibl_subjectTraining.table')
    # save data
```

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

□ Life sciences ☒ Behavioural & social sciences □ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Mice were taught a perceptual decision-making task: On each trial, a patch of black bars was presented on a white background, on either the right or left side of a screen. Mice used a wheel to indicate which side the contrast was on, for a water reward if correct. By modulating the strength of the contrast, a trial could be made more or less difficult. Mice were only presented easy contrasts at the start, more difficult contrasts were introduced as performance improved. This gave us a quantitative experimental study.
Research sample	We analysed 134 C57BL6/J mice aged 3-7 months obtained from Jackson Laboratory or Charles River. We used the publicly available

Research sample

IBL dataset and included all subjects. We did not therefore determine the data collection ourselves, but relied on an existing, exceptionally large data set. In particular, the number of individuals is larger than that used by the studies of e.g. Kastner et al. (2022) or Akiti et al. (2022).

Sampling strategy

Our sampling strategy was convenience/exhaustive. To our knowledge we used all mice which trained under the standard IBL protocol without any manipulations, but we did not make entirely sure that none were missed. We did not specifically leave out any appropriate mice, but we did exclude mice which had incomplete training trajectories (missing sessions for whatever reason).

Data collection

Data was collected using the IBL rig (https://github.com/int-brain-lab/iblrig) setup, in particular mouse responses were recorded via computer. All the details can be found in the paper describing the experiment setup: https://elifesciences.org/articles/63711. Researchers were not blind to experimental condition, as there were no conditions. Researchers were effectively blind to the study hypothesis, as hypotheses were formed during model construction, which was mostly after data collection had concluded.

Timing

Samples were collected beginning on the 3rd of November 2019 and ending on the 8th of April 2022.

Data exclusions

12 subjects were excluded from the analysis because the R^hat metric was too bad (above 1.05) on their chains, as described in the paper. The R^hat metric quantifies how much the chains vary from one another, and indicate poor convergence. We also excluded mice with any missing training sessions.

Non-participation

We analysed mice which completed training, in that sense there were no dropouts.

Randomization

There were no experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	□	Antibodies
☒	□	Eukaryotic cell lines
☒	□	Palaeontology and archaeology
□	☒	Animals and other organisms
☒	□	Clinical data
☒	□	Dual use research of concern
☒	□	Plants

Methods

n/a	Involved in the study	
☒	□	ChIP-seq
☒	□	Flow cytometry
☒	□	MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

C57BL6/J mice aged 3-7 months obtained from Jackson Laboratory or Charles River.

Wild animals

Study did not involve wild animals.

Reporting on sex

Sex was not considered in this study. We wanted to consider learning in general.

Field-collected samples

Study did not involve samples collected from the field.

Ethics oversight

All procedures and experiments were carried out in accordance with the local laws and following approval by the relevant institutions: the Animal Welfare Ethical Review Body of University College London [P1DB285D8]; the Institutional Animal Care and Use Committees of Cold Spring Harbor Laboratory [1411117; 19.5], Princeton University [1876-20], and University of California at Berkeley [AUP-2016-06-8860-1]; the University Animal Welfare Committee of New York University [18-1502]; and the Portuguese Veterinary General Board [0421/0000/0000/2016-2019].

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.