Modeling hidden cognitive states reveals acute and chronic effects of fentanyl on decision-making

Authors
Zhenlong Zhang[1], Patricia H. Janak[2,3,4], Eric Garr[2,*]

Affiliations
[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA
[2]Department of Psychological & Brain Sciences, Krieger School of Arts & Sciences, Johns Hopkins University, Baltimore, MD, 21218, USA
[3]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD, 21218, USA
[4]Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA

*Corresponding author (egarr1@jhu.edu)

Abstract

The cognitive mechanisms underlying behavior are often dynamic, shifting gradually or abruptly over time scales spanning years, to weeks, to minutes. Whether drug-induced changes in learning and decision-making follow similarly dynamic patterns remains unclear. To address this, we apply a reinforcement learning model to choice data from rats performing a two-step task for oral fentanyl and sucrose rewards. The model contains a set of agents with their own learning and decision-making rules that differentially influence choice, and, critically, each agent's contribution to choice is allowed to vary across latent states that fluctuate over time. Using a dimensionality reduction method to align latent states across subjects, we identified three distinct states reflecting mixtures of goal-directed, habitual, and novelty-driven strategies. We found that acute fentanyl reward increased the frequency of transitions out of a goal-directed state into a habit-driven state, while chronic fentanyl exposure selectively diminished goal-directed control within a habit-dominant state, independent of sex. Together, these results demonstrate that fentanyl reshapes both the dynamics and cognitive architecture of decision-making, underscoring the utility of latent-state modeling combined with dimensionality reduction for uncovering drug-driven cognitive changes.

Introduction

The habit theory of addiction posits that drug-seeking becomes involuntary over time, with drug-paired stimuli eliciting drug-seeking directly and bypassing goal-directed assessment of future outcomes (Robinson & Berridge, 2025). This framework has come under scrutiny because it has remained unclear whether maladaptive drug seeking is explained predominantly by habit formation or an alternative mechanism (Hogarth, 2020). Embedded within this skepticism is the issue of how habits are commonly measured, with the standard procedure being instrumental conditioning of a single action followed by outcome devaluation and then testing in extinction (e.g. Adams, 1982; Giovanniello et al., 2025; Thrailkill & Bouton, 2015). The drawbacks of this method are relevant for the translational validity of habits: real-world drug seeking rarely takes place in the absence of choice, and rarely during extinction (Vandaele & Ahmed, 2021).

We recently devised an experiment to overcome these limitations and ask whether drug seeking is habitual during a reinforced choice setting (Garr et al., 2025). Rats were trained to perform a two-step task to earn either oral fentanyl or sucrose rewards across many sessions, and then given a brief number of alternating sessions with fentanyl or sucrose. The two-step task allows for separate measurements of habit and goal-directed action by analyzing the frequency with which subjects will repeat a choice as a function of recent trial events (Daw et al., 2011; Miller et al., 2017). An index of goal-directed action can be computed as the degree of model-based (MB) choice, where choice is guided by knowledge of how often an action transitions to a given state as well as the reward probabilities within each state. Two indices of habit can also be computed: one termed model-free (MF) choice, where choice is based on whether an action was recently rewarded without accounting for the route connecting action to reward, and another termed perseveration, where a choice is repeated regardless of prior trial events. We found that the expression of habit depended on the history of fentanyl exposure and sex: while female rats given extensive training with fentanyl showed high degrees of MF bias and perseveration that carried over to fentanyl and sucrose seeking, females given brief fentanyl training showed a high degree of perseveration that was specific to fentanyl seeking.

While these findings are informative, a more thorough exploration of the data would benefit from computational modelling. Modelling has been part of the two-step task since its inception (Daw et al., 2011), and the utility of modelling comes from its ability to go beyond surface-level behavioral summaries and instead estimate latent cognitive variables that govern choice. When modelling choice data from the two-step task, it is common to specify a set of learning and decision rules, and then evaluate which one most closely matches trial-to-trial choices. In this way, choices are seen as the product of a single algorithm, or a weighted hybrid of algorithms, whose parameters remain constant through time. However, recent work has developed a method for identifying time-varying latent states that give rise to distinct decision-making strategies (Calhoun et al., 2019; Ashwood et al., 2022). This approach involves mapping perceptual inputs to a common behavioral output, where the mapping between inputs and output are allowed to vary across hidden states that must be inferred by the experimenter using an unsupervised method. This approach was recently extended to replace perceptual inputs with cognitive

variables generated from reinforcement learning models (Venditto et al., 2024). Applying this approach to choice data from rats performing a two-step task, it was shown that a three-state model characterizes choice behavior in a manner superior to a single-state model. By observing how the state probabilities changed over the course of a session, combined with examining the agent weights within each state, the authors were able to infer that behavior follows a common trajectory: initial exploration of the state transitions, followed by a predominantly model-based strategy, and then a more disengaged perseverative state toward the end of the session.

In the present study, we applied this modelling framework, termed a mixture-of-agent hidden Markov model (MoA-HMM), to investigate how fentanyl reinforcement history shapes decision-making strategies in rats. We used an MoA-HMM in which each latent state is defined by a unique combination of weights on reinforcement learning agents (MB, MF, perseveration, side bias, and transition preference). To solve the problem of how to best align the states across subjects, we turned to dimensionality reduction to identify a set of common low-dimensional components that could be used to sort latent states and compare them across rats in a principled, data-driven way. We found that a three-state model was the best fit for the majority of rats, and each state was defined by a distinct set of agent weights. In addition, we found that acute fentanyl reward altered how often rats transitioned out of a goal-directed state, while chronic fentanyl diminished the influence of MB choice in a state-dependent manner.

Methods

Data set

The data analyzed here have been reported previously (Garr et al., 2025). Long-Evans rats were initially trained for 22-30 daily sessions to perform a two-step decision-making task to earn oral fentanyl (25 µg/ml; 9 males, 8 females) or sucrose solution (100 mg/ml; 10 males, 8 females). During the task, rats were required to initiate trials by holding their nose in a center magazine and then making a choice between left and right nose-poke ports. Choices triggered the insertion of a lever in the opposite side of the chamber, and pressing the lever delivered probabilistic reward. One lever was always set to 0.8 reward probability and the other to 0.2, and the probabilities switched in blocks of 20-35 trials. Each nose port was predominantly associated with different levers, with transition probabilities fixed at 0.8 and 0.2 (common and rare transitions, respectively). Reward size was fixed at 0.05 ml. During each session, the fentanyl group was free to earn up to 150 rewards, while the sucrose group was limited to the average number earned by the fentanyl group during the previous session. This was done to equate the average number of rewards between groups.

Following training, all rats received alternating sessions with fentanyl and sucrose rewards (6 sessions per reward, 12 total). Sessions with the unfamiliar reward occurred in an altered context with lemon scent and honeycomb textured floors. The maximum number of sucrose rewards per session per rat was set to the number of fentanyl rewards earned during the previous session.

Analyses were conducted on the final six sessions of alternating rewards (i.e. 3 sessions per reward type).

Reinforcement learning model

Choice behavior was modeled using a mixture-of-agents hidden Markov model (MoA-HMM). Five agents simultaneously and independently updated choice values on every trial according to their own rules (see Venditto et al., 2024 for rationale behind learning rules).

Model-based

$$Q_{MB}(y) \leftarrow \begin{cases} (1 - \alpha_{MB})Q_{MB}(y) + r_t, & y = y_t(common), y \neq y_t(rare) \\ (1 - \alpha_{MB})Q_{MB}(y), & y \neq y_t(common), y = y_t(rare) \end{cases}$$

where α_{MB} is the model-based learning rate and r_t is the trial outcome (1 for reward and -1 for omission). *y* and *t* are choice options and trials, respectively.

Model-free

$$Q_{MF}(y) \leftarrow \begin{cases} (1 - \alpha_{MF})Q_{MF}(y) + r_t, & y = y_t \\ (1 - \alpha_{MF})Q_{MB}(y), & y \neq y_t \end{cases}$$

where α_{Mf} is the model-free learning rate.

Perseveration

$$Q_{persev}(y) \leftarrow \begin{cases} (1 - \alpha_{persev})Q_{persev}(y) + 1, & y = y_t \\ (1 - \alpha_{persev})Q_{persev}(y), & y \neq y_t \end{cases}$$

where α_{persev} is the perseveration learning rate.

Side bias

$$Q_{bias}(y) \leftarrow \begin{cases} 1, & y = left \\ -1, & y = right \end{cases}$$

Transition preference

$$Q_{TP}(y) \leftarrow \begin{cases} (1 - \alpha_{TP})Q_{TP}(y) + 1, & y = y_t(common), y \neq y_t(rare) \\ (1 - \alpha_{TP})Q_{TP}(y), & y \neq y_t(common), y = y_t(rare) \end{cases}$$

where α_{TP} is the transition preference learning rate.

These learning rules were adapted from Venditto et al. (2024), where each agent was assigned different names than we use here: MB reward (MB), MF reward (MF), MF choice (perseveration), bias (side bias), and MB choice (transition preference).

Choice probabilities are determined by passing each agent's weighted choice value through a softmax function that is conditioned on a latent state *z*:

$$p(y|Q,z) = \frac{\exp(\sum_A \beta_A^z Q_A(y))}{\sum_{y'} \exp(\sum_A \beta_A^z Q_A(y'))}$$

where β_A^z is the weight assigned to agent *A* in state *z*. The learning rate parameters were constant across states. Latent states were defined by an initial state probability π and a state transition probability matrix *P*. Models were fit using expectation maximization, which generated trial-by-trial probabilities for each latent state (Venditto et al., 2024).

The MoA-HMM was fit using maximum a posteriori estimation. Model parameters (θ) included the initial latent state probabilities (π), the state transition probability matrix (*P*), state-specific agent weights (β), and learning rates (α_{MB}, α_{MF}, α_{persev}, α_{TP}). Agent weights, initial state probabilities, and transition probabilities were initialized uniformly over [0,1], and learning rates were initialized uniformly over [.05, .95].

Model fitting was performed by maximizing a regularized log-likelihood:

$$\mathcal{L}(\theta) = \sum_t \log p\,(D_t|\theta) \; + \; \lambda\|\theta\|^2$$

where $p(D_t|\theta)$is the probability of the observed choice on trial t, and λ=0.01is the L2 penalty coefficient. Optimization stopped when the objective change reached below 1e-5, or after 200 iterations. Model fitting was done separately for fentanyl and sucrose sessions, concatenating over sessions. The log-likelihood was used to calculate the Akaike information criterion (AIC), which estimates the model fit:

$$\text{AIC} = 2k - 2\mathcal{L}(\theta)$$

where *k* = number of free parameters. The number of free parameters was determined by the number of initial state probabilities (π), the number of cells in the transition probability matrix *P*, the four learning rates (α_{MB}, α_{MF}, α_{persev}, α_{TP}), and the five agent weights (β's).

Tensor decomposition

Tensor component analysis (TCA) was performed using the tensor toolbox for MATLAB. A four-dimensional matrix was constructed using the data generated from a three-state model fit: a 35 x 6 x 5 x 9 matrix (subjects x sessions x agents x session time bins). Sessions were divided

into 9 time bins of equal numbers of trials per individual session because each rat performed different numbers of trials. Each data point in the matrix represented a subject-, session-, and bin-specific agent weight. Bin-specific agent weights were obtained by multiplying the agent weight in state *z* by the mean probability of state *z* in a given bin, and then averaging over states. States were averaged together so that state labels would not influence the TCA output.

We fit a tensor CANDECOMP/PARAFAC decomposition model to identify a set of low-dimensional components describing variability along each of the four axes. Model fits were performed 5000 times, each one with a different random seed, for each of 9 components. To determine the optimal number of components, we computed a normalized reconstruction error and a similarity score for each iteration of model fitting (Willams et al., 2018). The reconstruction error reflects how accurately the tensor decomposition can reconstruct the data and decreases with the number of components. The similarity score reflects how similar each iteration of tensor decomposition is to the one with the lowest reconstruction error, and gives a sense of how fickle the model is with regard to the initialization parameters.

Aligning state labels

To align the state labels across subjects and sessions, the TCA output was used to construct an agent x component matrix, where the number of components is equivalent to the number of states. This was done individually for each rat and each session type (fentanyl or sucrose) by multiplying together the agent loadings, mean session loadings, subject loadings, and λ (scaling factor similar to each component's explained variance). The resulting agent x component matrix generated from TCA was compared to the agent x state matrix generated from the MoA-HMM by building a cost matrix:

$$cost_{i,j} = \sum_{k=1}^{n} \left(MoA_{k,i} - TCA_{k,j}\right)^2$$

where element (*i*,*j*) is the squared difference between the *i*-th column of the MoA agent x state matrix and the *j*-th column of the TCA agent x component matrix, summed over all rows *k*. We then used the Munkres assignment algorithm (Munkres, 1957) to find which permutation of the agent x state matrix minimized the cost.

Statistical analysis

Statistical analyses were conducted using mixed-model analysis of variance (ANOVA) with a Type I error rate of 0.05. Within-subject variables (e.g. session reward) were always evaluated jointly with between-subject variables (i.e. training group and sex). Significant interactions were followed up with simple main effects tests. Agent weights were first multiplied by their respective session-wide state probabilities before being passed through ANOVAs to control for the relative frequency of agent weights.

Results

A multi-latent state model describes rat decision-making but presents a problem for state alignment

To probe whether a multi-latent state model could describe rat decision-making during performance of a two-step task, we fit a series of MoA-HMM models while varying the number of latent states (Venditto et al., 2024). Within each state, there are five reinforcement learning agents competing for control of choice: MB, MF, perseveration, side bias, and transition preference (TP). The latter captures a preference for common vs. rare state transitions, a metric of novelty preference. The influence of each agent on trial-by-trial choice is quantified by their weights, which are allowed to vary between states. A 3-state model provided the best fit, with 91% of rats showing an improvement in model fit compared to a single state model (**Figure 1A**). Reward type (fentanyl or sucrose) did not affect how model fit varied as a function of number of states (reward x state, $p = 0.918$), although sucrose sessions showed an overall superior fit compared to fentanyl sessions (reward, $p = 0.036$). There were no interactions with training group or sex.

The latent states are assigned arbitrary labels, and sorting the states is necessary to compare them across subjects. In an initial attempt to sort the states, we followed the steps from Venditto and colleagues (2024). Specifically, we labeled state 1 as the state with the highest initial probability per session, while states 2 and 3 were sorted in descending order of MB weights. Unlike Venditto et al. (2024), who also fit an MoA-HMM model to data from rats performing a two-step task, we observed minimal common state dynamics from our rat data: state 1 started with the highest probability and decreased slightly, but remained relatively high throughout each session, while state 2 probability increased slightly (**Figure 1B**; state, $p = .006$; state x bin, $p < .001$). This does not mean state dynamics were non-existent (**Figure S1**), but that it was difficult to extract common dynamics across individuals. Furthermore, although averaging state probabilities into bins and across subjects gives the impression of a highly probabilistic set of states, the majority of individual trials were composed of a high probability state. Specifically, the proportion of trials containing a state probability greater than 0.66 (the threshold for dominance) was 63%, on average (21% per state). State probability was influenced by the type of instrumental reward (reward x state, $p = 0.025$), with state 1 being slightly more probable during sucrose versus fentanyl sessions ($p = 0.004$; mean difference = 0.05). State transition probabilities were not affected by reward type ($p = 0.564$), but there was an overall pattern to state transitions (**Figure 1C**; previous state x current state, $p < .001$). Specifically, there was a tendency to avoid remaining in the same state from trial to trial, and state 3 was more likely to transition to state1 than to state 2 ($p < 0.001$).

[Insert Figure 1 here]

To understand the defining features of each state, we examined the state-varying agent weights. The states were largely defined by differences in MB weights, which was partly by design (agent x state, $p < 0.001$). State 1 was associated with an intermediate MB weight, state 2 with a high MB weight, and state 3 with a negative MB weight (**Figure 1D**; p's < 0.001). The MF weights also differed by state (state 1 > states 2 and 3; p's < .012), as did transition preference weights (state 2 < state 3; $p = .015$). We also found that MB weights were affected by reward and sex in a state-dependent manner (reward x sex x state, $p = .044$), with females showing greater MB weights during sucrose versus fentanyl sessions in states 1 and 2 (**Figure S2**). Overall, this sorting scheme revealed a set of states distinguished by a strong reliance on reward-driven habits (state 1), goal-directed planning (state 2), and transition-driven habits (state 3), with fentanyl reward acutely driving down goal-directed planning in females in a state-dependent manner.

An alternative method for aligning latent states

This sorting scheme imposes structure such that the states are easily defined, but because they are defined in an arbitrary way, it raises questions about whether the results are meaningful. To approach state sorting in a more data-driven way, we turned to tensor component analysis (TCA). TCA is an unsupervised method that identifies a set of low-dimensional components that describe variability along a set of axes, and unlike the commonly used principal component analysis, it can take as input high-dimensional data arrays (Drieu et al., 2025; Williams et al., 2018). TCA can potentially be used to help align the latent states across rats so that they reflect shared underlying structure. We fed TCA a high-dimensional matrix consisting of subject-, session-, time bin-, and agent-specific weights (see *Methods*). State labels were removed by averaging over states, as our goal was to use the resulting low-dimensional components to align state labels later, and we therefore did not wish to bias the results using the original arbitrary state labels. TCA requires identifying the optimal number of components by consulting two metrics: the reconstruction error and the mean similarity score (**Figure 2A**). Examination of these metrics indicated that three components was optimal. This finding is convenient because aligning the states with the TCA output requires the number of components to match the number of states (see below).

The TCA output gives component-specific loadings for each dimension of the input matrix: individual subjects, individual sessions, the five agents, and the session time bins (**Figure 2B**). The agent loadings provided a picture of how each component can be defined: component 1 was defined by a high MB loading and a modest MF loading, component 2 was dominated by a high MB loading, and component 3 was defined by high MF and perseveration loadings and a strongly negative side bias loading. Interestingly, fentanyl and sucrose sessions did not show much session-to-session variation but did show different loadings on each component: while fentanyl sessions showed a higher loading on component 3, sucrose sessions showed a higher loading on component 2 (**Figure 2B**).

[Insert Figure 2 here]

The three components identified by TCA were then used to sort the latent states from the MoA-HMM model. This was accomplished by finding the state permutation that maximized the similarity between the agent x state matrix and the agent x component matrix for each individual subject and session type (see *Methods*; see also **Figure 3A,B**).

Aligning latent states with shared underlying dynamics produces a rich pattern of state-dependent decision-making mechanisms

Using the newly sorted latent states, we ran statistical tests on the time-varying state probabilities (**Figure 3C**), the state transition matrices (**Figure 3D**), and the state-varying agent weights (**Figure 3E**) to first probe for any effects that generalized across session reward, prior training history, and sex. Aside from a small increase and decrease in state 1 and 2 probabilities, respectively, in the middle of the session (state x bin, $p = 0.045$), there were no systematic changes in the state probabilities across time. Nor was there any one dominant state (state, $p = 0.899$). Once again, although state probabilities were time invariant and roughly uniform when averaged across rats, the majority of individual trials were composed of a high probability state (trials with a state probability > 0.66: 63% in total; 22% for state 1, 21% for state 2, 20% for state 3).

The state transition matrices once again showed a strong tendency to avoid remaining in the same state from trial to trial (previous state x current state, $p = 0.019$). The state-varying agent weights also showed a distinct pattern (agent x state, $p = 0.026$). State 1 was distinguished by a strong reliance on habits, with perseveration dominating ([perseveration > MF] > all other agents, p's < 0.044). State 2 was associated with a mix of habits and goal-directed control (MB ≈ MF ≈ perseveration, p's > 0.133). State 3 was very similar to state 1 with one exception: while state 1 was associated with a negative weight on the transition preference agent, implying a preference to explore the rare transition states, state 3 was associated with a positive weight (state 1 < state 3, $p = 0.032$). Taken together, this sorting scheme revealed a set of states distinguished by a mix of habits and novelty preference (state 1), a mix of habits and goal-directed planning (state 2), and a mix of habits and novelty avoidance (state 3). This sorting scheme also produced greater within-state variances of agent weights compared to the more arbitrary method of sorting states (method, $p < .001$), which confirms that using TCA to sort latent states can reveal a richer pattern of state-dependent learning and decision-making.

[Insert Figure 3 here]

Acute fentanyl increases transitions out of goal-directed states, while chronic fentanyl decreases model-based weights in a state-dependent manner

Next, we looked for any effects of drug—both acute and chronic, evaluated by effects of session reward and training group, respectively. Time-varying state probabilities were not affected by these variables (p's > 0.103). However, the instrumental reward affected how rats moved between states (reward x previous state x current state, $p = .011$). The state that showed the most

changes in transitions was state 2, the more goal-directed state (**Figures 3E and 4A**). During sucrose sessions, the transition probability between state 2 and all other states was roughly uniform. But during fentanyl sessions, rats showed a strong preference for transitioning out of this more goal-directed state and into the predominantly perseverative state 1 (s2 → s1: fentanyl > sucrose, $p = 0.041$; s2 → s2: fentanyl < sucrose, $p = 0.037$). This implies that fentanyl reward reduces the tendency to persist within the more goal-directed state compared to sucrose.

[Insert Figure 4 here]

When examining agent weights, we found that fentanyl reinforcement history, as represented by group, had a significant effect on MB weights in a state-dependent manner (state x group, $p = 0.011$; **Figure 4B**). Specifically, rats given extensive prior training with fentanyl showed a low MB weight in state 1 compared to rats given extensive training with sucrose. To confirm that this finding holds up in the empirical choice data, we regressed a binary MB coder against trial-by-trial state probabilities. The MB coder was constructed by assigning a 1 to all cases where the rat repeated choices following reward-common and omission-rare trials and switched choices following reward-rare and omission-common trials, while a 0 was assigned to all other cases. We found a pattern of coefficients that qualitatively matched the pattern of state-dependent MB weights (**Figure 4C**). A comparison of state 1 coefficients between training groups came up short of statistical significance ($p = 0.077$), which is not surprising given that MB weights compete with other agents for control of behavior. These results imply that an extensive history with fentanyl drives down the influence of MB decision-making in a state-dependent manner. Notably, we previously found that chronic fentanyl attenuates MB-related behavior in a session-wide manner only for females (Garr et al., 2025). Here, we did not find that sex interacted with the group- and state-dependent change in MB weights (**Figure S3**; $p = 0.541$), implying that chronic fentanyl induces a cognitive impairment broader than previously thought.

Discussion

Modelling learning and decision-making as occurring within discrete latent states can help with avoiding model misspecification, allowing more accurate interpretations of the cognitive mechanisms underlying behavior (Urai, 2025). In this study, we applied an MoA-HMM to choice data from rats performing a two-step task for either sucrose or fentanyl, aiming to uncover how latent cognitive states and fentanyl reinforcement history interact to shape behavior. We found that a three-state model best captured behavioral patterns for the majority of subjects. By using TCA to align latent states across rats in a data-driven manner, we were able to expose a rich structure in the dynamics of decision-making—one that was obscured by a more conventional, arbitrary state-sorting method. Specifically, each state was characterized by a unique profile of reinforcement learning agent weights that revealed broadly different mixtures of goal-directed, habitual, and novelty seeking strategies.

A major finding was that fentanyl reward acutely destabilized goal-directed states, while chronic fentanyl suppressed MB control in a state-dependent manner. During sessions reinforced with

fentanyl, rats were more likely to transition out of the more goal-directed state and into a habit-driven state. This suggests that fentanyl acutely reduces the "stickiness" of MB control, perhaps by impairing the stability of computations needed to sustain planning-based strategies. In contrast, chronic fentanyl exposure did not alter state dynamics, but rather suppressed MB weights within one of the habit dominant states. While our original set of findings emphasized a global suppression of MB-related behavior specifically in females given extensive fentanyl training (Garr et al., 2025), here we show that there exists a latent cognitive state for which chronic fentanyl impairs MB control across both sexes.

A particularly striking finding in the current study was the strong tendency for rats to avoid remaining in the same latent state from trial to trial. This is not an artifact of how the states were sorted, since consistently small diagonal elements in the state transition matrix are impervious to the sorting scheme. This pattern is consistent with our previous finding that rats integrate prior trial information over a severely limited time window (Garr et al., 2025; see Figure S1), but contrasts with prior findings in which state transitions were not as stochastic and choices integrated a broader range of past trials (Ashwood et al., 2022; Venditto et al., 2024). There are several potential explanations for this discrepancy. One possible contributor is the type of reward: our study involved oral fentanyl. Even in sucrose sessions, the interleaved drug context and history of fentanyl exposure, whether brief or prolonged, may have carried over to affect behavioral stability. Another important difference between studies lies in the amount of training given. The rats from the data set analyzed in the current study (Garr et al., 2025) were given a total of 34-42 session (training period and alternating reward phases combined), each with a limited number of rewards to control for pharmacological thresholds. In contrast, the rats in the data set analyzed by Venditto and colleagues (Miller et al., 2017) were often trained for more than 100 sessions. It is conceivable that a longer training period induces stereotyped behavioral routines, and that is what possibly accounts for the difference in the pattern of state transitions.

One question that arises from the current analysis is how to infer from behavioral data, without post-hoc modelling, the cognitive state that gives rise to fentanyl-induced changes in MB control. If the latent cognitive states identified in this study are largely time-invariant and uniform in their probabilities, does that mean the expression of MB-impaired behavior by chronic fentanyl is completely unpredictable? Averaging over time bins and across individual rats obscures the fact that the majority of individual trials contained a high degree of certainty regarding state occupancy. For example, the state that gave rise to a group difference in MB weights (i.e. state 1) surpassed the threshold as the dominant state on 22% of trials. Although these trials appear to be randomly distributed in time, one clue for identifying when they occur could lie in rats' tendencies to repeat choices following rare transitions. Recall that state 1 could be distinguished from the other states by a negative weight on the transition preference agent—indicative of a bias toward repeating choices following rare transitions. A supplemental analysis showed that repeating choices after rare transitions could positively predict state 1 probabilities in 60% of rats, while the predictive accuracy fell for states 2 and 3 to 37% and 46%, respectively. These differences in proportions did not reach statistical significance ($\chi(2) = 3.74$, $p = 0.15$), which means that the preference for rare transitions cannot reliably be used to decode

the state—consistent with the negative transition preference weight being just one defining feature of state 1. It is possible that pairing a behavioral metric of transition preference with other behavioral measures beyond choice (e.g. movement kinematics from video recordings) may improve decoding of states.

Finally, this work contributes methodologically by demonstrating how tensor decomposition can be integrated with latent state modeling to solve the problem of state alignment across subjects. While hidden Markov models provide powerful tools for capturing time-varying behavioral modes, they produce arbitrary state labels that prevent comparisons across individuals. Our use of TCA to define a common low-dimensional space allowed us to take advantage of shared underlying behavioral dynamics to align the states. The fact that this approach yielded greater within-state variance of agent weights (relative to arbitrary sorting) further validates its utility, showing that data-driven alignment methods can uncover richer, more heterogeneous cognitive patterns. We wish to note that an alternative and simpler method for sorting states is to find the permutation that minimizes the cross-subject variance of agent weights. In the context of our data set, this method would require either averaging over fentanyl and sucrose sessions or defining states separately for fentanyl and sucrose sessions without any common reference. In contrast, TCA allows us to define states separately for each session type with a common reference (i.e. the low-dimensional components).

In summary, this study shows that rat decision-making during a two-step task can be parsed into multiple, temporally fluid latent states, each characterized by distinct contributions of learning systems. Fentanyl, whether acutely or chronically administered, alters both the dynamics of transitioning between these states and the cognitive architecture within them. These findings underscore the power of latent-state modeling in behavioral neuroscience and provide new insight into how addictive drugs reshape learning and decision-making.

Funding
This work was supported by National Institutes of Health grants F32DA054767 (EG) and R01DA035943 (PHJ).

Competing interests
The authors have no competing interests to declare.

Author contributions
ZZ and EG analyzed the data. EG wrote the manuscript with input from ZZ and PHJ.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *34*(2), 77–98.

Ashwood, Z. C., Roy, N. A., Stone, I. R., Urai, A. E., Churchland, A. K., Pouget, A., & Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), 201–212.

Calhoun, A. J., Pillow, J. W., & Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. *Nature Neuroscience*, *22*(12), 2040–2049.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Drieu, C., Zhu, Z., Wang, Z., Fuller, K., Wang, A., Elnozahy, S., & Kuchibhotla, K. (2025). Rapid emergence of latent knowledge in the sensory cortex drives learning. *Nature*, *641*(8064), 960–970.

Garr, E., Cheng, Y., Dong, A., & Janak, P. H. (2025). Fentanyl reinforcement history has sex-specific effects on multi-step decision-making. *BioRxiv*.

Giovanniello, J. R., Paredes, N., Wiener, A., Ramírez-Armenta, K., Oragwam, C., Uwadia, H. O., Yu, A. L., Lim, K., Pimenta, J. S., Vilchez, G. E., Nnamdi, G., Wang, A., Sehgal, M., Reis, F. M. C. V., Sias, A. C., Silva, A. J., Adhikari, A., Malvaez, M., & Wassum, K. M. (2025). A dual-pathway architecture for stress to disrupt agency and promote habit. *Nature*, *640*(8059), 722–731.

Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. *Neuropsychopharmacology*, *45*(5), 720–735.

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, *20*(9), 1269–1276.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*(1).

Robinson, T. E., & Berridge, K. C. (2024). The Incentive-Sensitization Theory of Addiction 30 Years On. *Annual Review of Psychology*, 29–58.

Thrailkill, E. A., & Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*(1), 69–80.

Urai, A. E. (2025). Structure uncovered: understanding temporal variability in perceptual decision-making. *Trends in Cognitive Sciences*.

Vandaele, Y., & Ahmed, S. H. (2021). Habit, choice, and addiction. *Neuropsychopharmacology*, *46*(4), 689–698.

Venditto, S. J. C., Miller, K. J., Brody, C. D., & Daw, N. D. (2024). Dynamic reinforcement learning reveals time-dependent shifts in strategy during reward learning. *eLife*, *13*, 1–49.

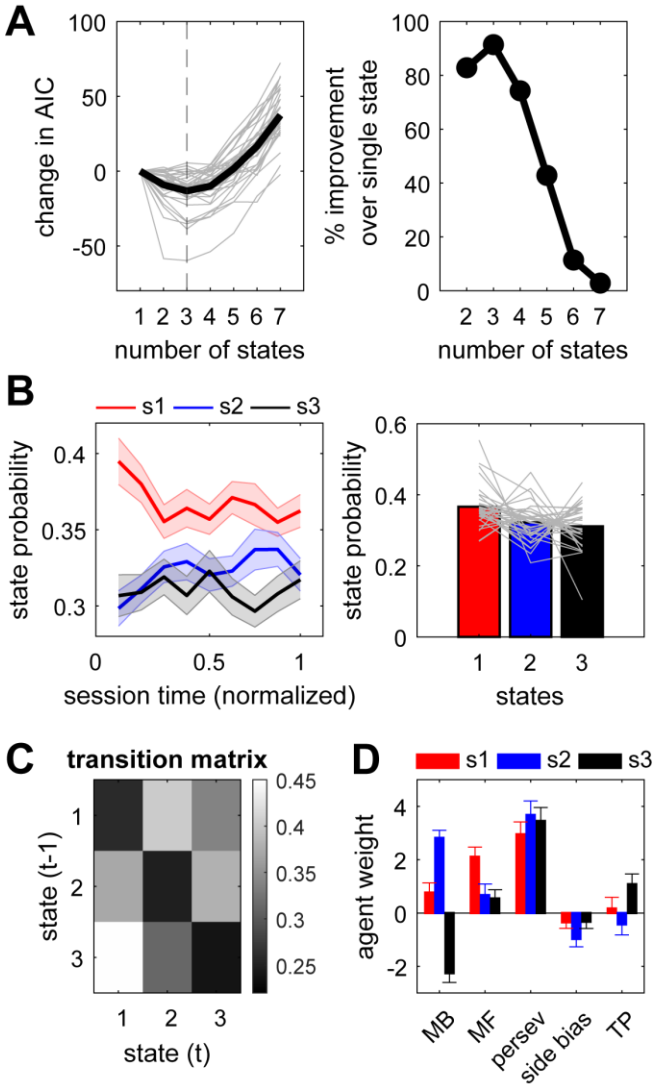

Figure 1. MoA-HMM fit and initial state sorting **(A)** Left: AIC scores are plotted relative to a single state model. Grey lines are individual rats. Black line is the mean. A 3-state model yields the lowest mean AIC score (lower values indicate better fit). Right: Percentage of rats that show smaller AIC score relative to the single state model as a function of number of states. **(B)** Left: State probabilities over normalized session time. Right: mean state probabilities over time. Grey lines are individual rats. (**C**) State transition matrix. Each cell indicates the probability of transitioning from a state on trial *t*-1 to another state on trial *t*. (**D**) Mean agent weights are shown for each latent state. All data in this figure are averaged across fentanyl and sucrose sessions.

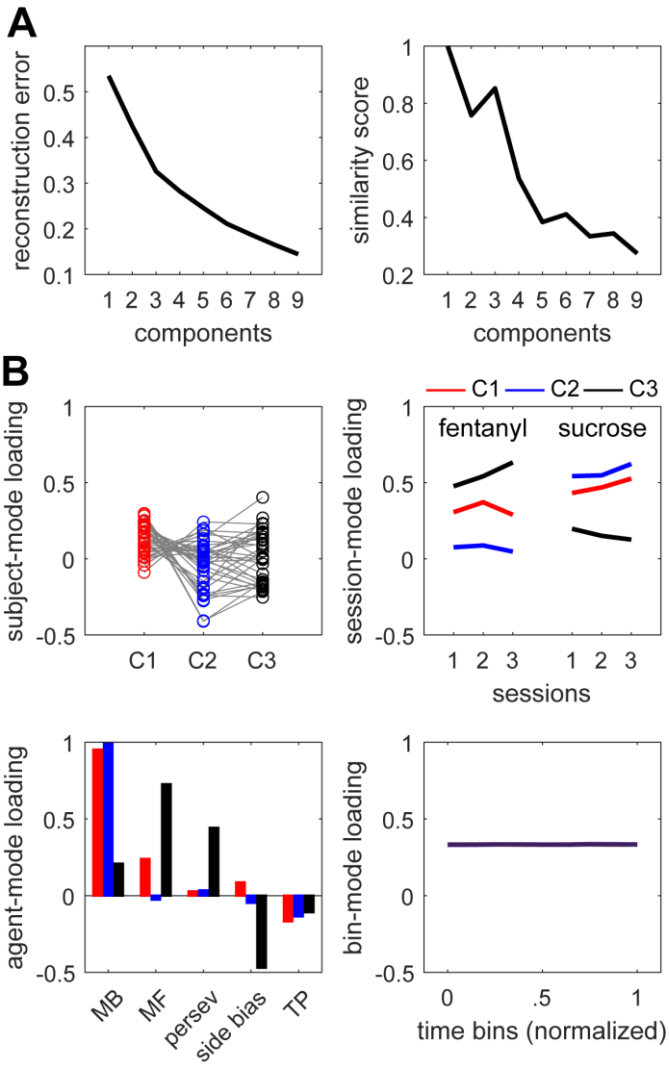

Figure 2. Tensor component analysis results. (**A**) Reconstruction errors and similarity scores indicate that the optimal number of low-dimensional components is 3. The deceleration in reconstruction error slows beyond 3 components, and the similarity score drops sharply after 3. (**B**) Loadings across the 3 components shown separately for subjects, sessions, agents, and bins.

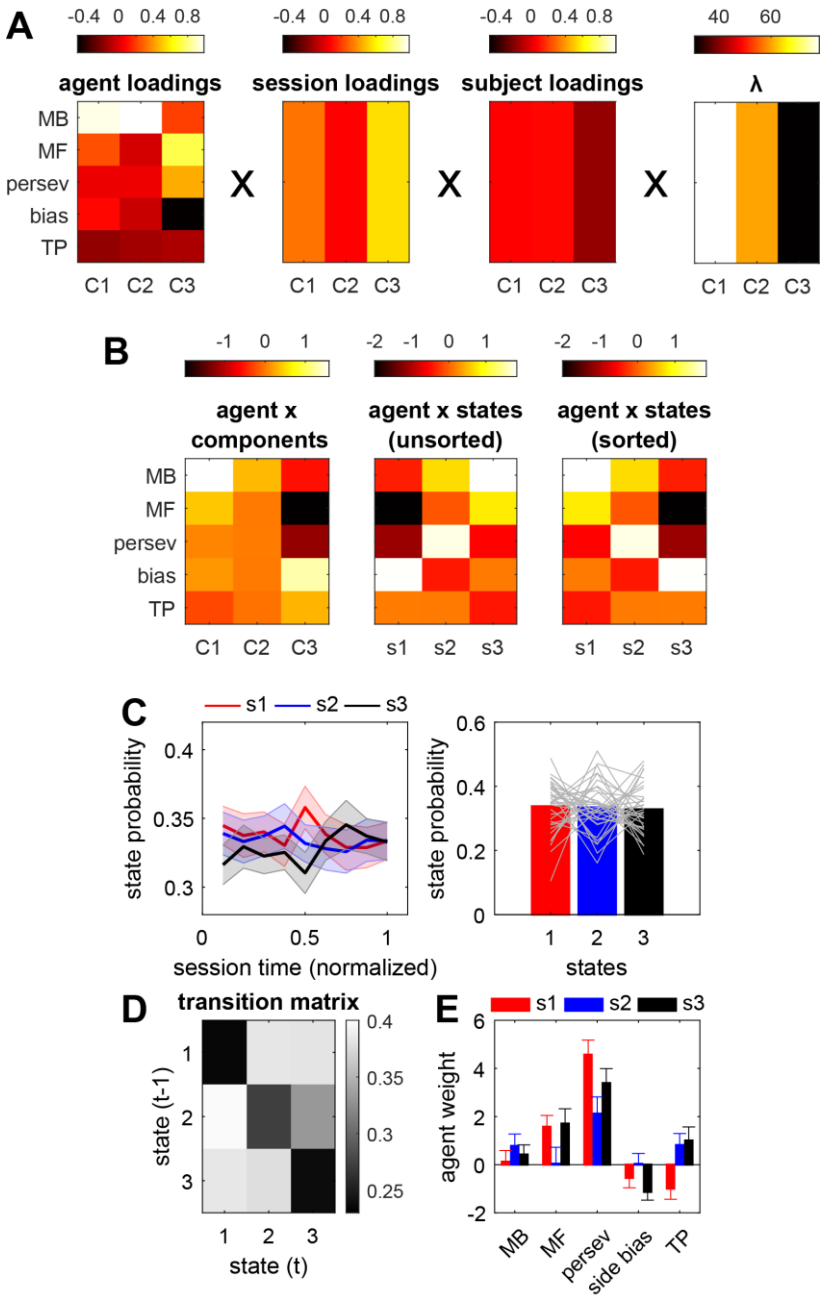

Figure 3. Latent state alignment. (**A**) Illustration of method for combining tensor component loadings for one example rat (fentanyl session). The agent-mode loadings are multiplied by the mean session loadings, the unique subject loadings, and λ. This yields a 5 x 3 matrix (agent x component), shown in **B**. (**B**) Left: agent x component matrix produced by **A**. Middle: Raw agent x state matrix from the MoA-HMM model. Right: the sorted agent x state matrix that most closely aligns with the agent x component matrix on the left. (**C**) State probabilities over normalized session time. Right: mean state probabilities over time. States are aligned according to TCA. (**D**) State transition matrix when states are aligned according to TCA. (**E**) Mean agent weights are shown for each latent state. States are aligned according to TCA. Data in **C**-**E** are averaged across fentanyl and sucrose sessions.

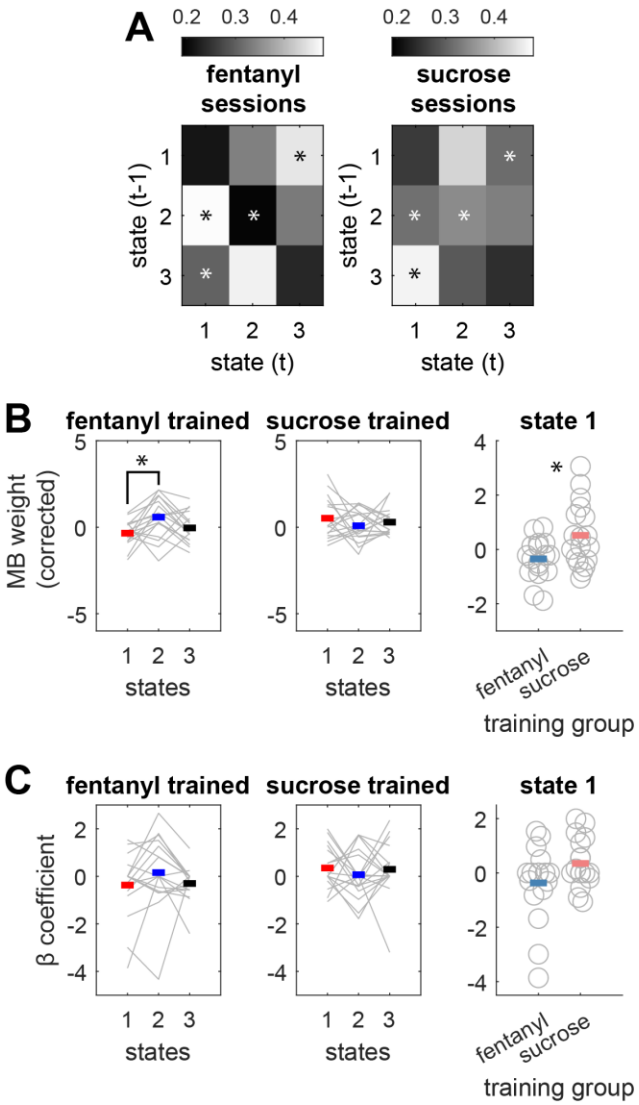

Figure 4. Drug effects after state alignment. (**A**) Transition matrices broken down by fentanyl and sucrose sessions. Asterisks indicate significant difference between session type. (**B**) MB weights over states, shown separately for groups given either extensive prior fentanyl training or sucrose training. Weights are multiplied by their respective session-wide probabilities. (**C**) Regression coefficients where a binary MB coder was predicted by trial-by-trial state probabilities.

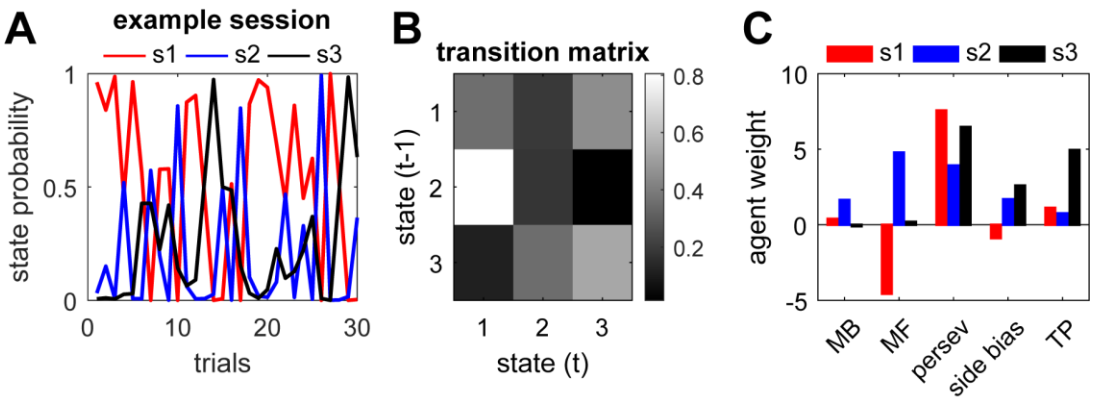

Figure S1. Example rat. (**A**) Example session showing state probabilities over time. The session is truncated for clarity. (**B**) State transition probability matrix. (**C**) Agent weights per state.

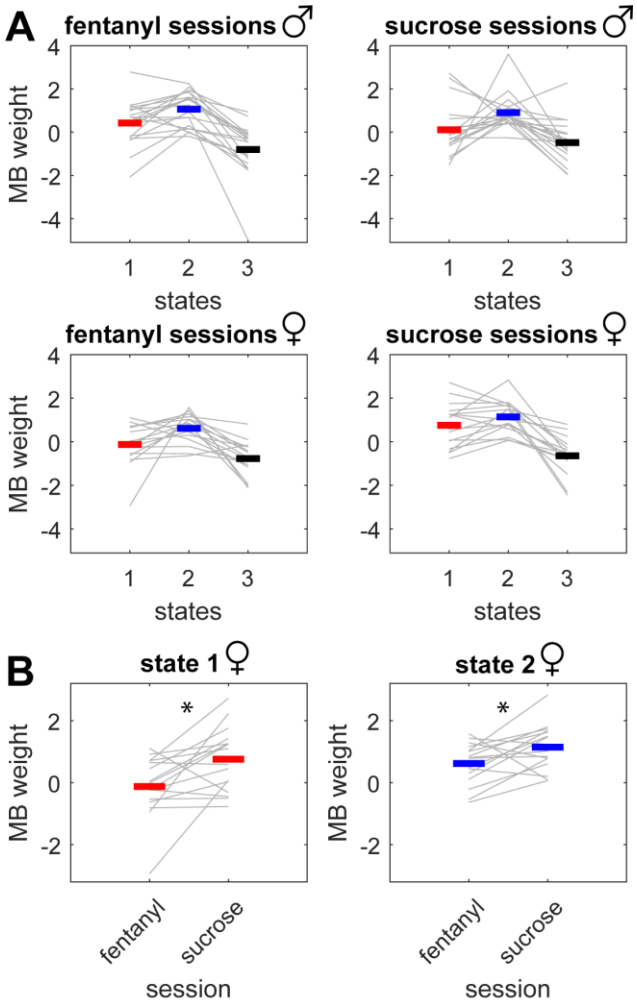

Figure S2. Significant reward x sex x state interaction under the state sorting scheme used by Venditto et al. (2024). (**A**) Each plot shows the state-dependent MB weights (multiplied by their respective state probabilities) separately for male (top) and female (bottom) rats during fentanyl (left) and sucrose (right) sessions. (**B**) Simple main effects tests identified the source of the interaction as arising from greater MB weights during states 1 and 2 for females reinforced with sucrose vs. fentanyl.

Figure S3. State 1 MB weights shown separately by sex and training group. Weights are multiplied by sate 1 probabilities.