JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Systems/Circuits

Ventral tegmental area dopamine neural activity switches simultaneously with rule representations in the medial prefrontal cortex and hippocampus

https://doi.org/10.1523/JNEUROSCI.1670-24.2025

Received: 21 June 2024
Revised: 26 February 2025
Accepted: 8 March 2025

Copyright © 2025 Ding et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes.The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Ventral tegmental area dopamine neural activity switches simultaneously with rule representations in the medial prefrontal cortex and hippocampus

Abbreviated title: Dopamine-CA1-mPFC activity during rule switching

Authors: Mingxin Ding[1], Porter L. Tomsick[2,3], Ryan A. Young[4], Shantanu P. Jadhav[1,4,5]

[1]Graduate Program in Neuroscience, Brandeis University, Waltham, MA 02453, USA

[2]Undergraduate Program in Neuroscience, Brandeis University, Waltham, MA 02453, USA

[3]Department of Neuroscience, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

[4]Department of Psychology, Brandeis University, Waltham, MA, 02453, USA

[5]Volen National Center for Complex Systems, Brandeis University, Waltham, MA, 02453, USA

Contributions: M.D. and S.P.J. conceptualized the study; M.D. and P.L.T. performed the experiments; M.D. analyzed the data; R.A.Y. contributed to analysis packages; M.D. and S.P.J. wrote the manuscript.

Correspondence: Shantanu P. Jadhav (shantanu@brandeis.edu), Brandeis University, 415 South St., MS 062, Waltham, MA 02453

Number of Pages: 42

JNeurosci Accepted Manuscript

Number of Figures: 5 figures

Number of Tables: 1 table

Number of words: Abstract (249 words); Significance statement (119 words); Introduction (657 words); Discussion (1431 words)

Acknowledgements: This work was supported by the National Institutes of Health (R01MH112661) to S.P.J. and a T90 training grant (R90DA033463). We would like to thank Dr. Naoshige Uchida, Dr. Joshua Berke and Dr. Ali Mohebi for the protocol of the opto-tagging experiment, Dr. John Bladon and Audrey Hooker for piloting the W-track rule switching task, Dr. Justin Shin in assisting in implant surgeries, Dr. Wenbo Tang, Dr. Jacob Olson, Dr. Blake Porter and other lab members in providing feedback on the research project.

Conflicts of Interest: Authors report no conflict of interest.

JNeurosci Accepted Manuscript

Abstract

Multiple brain regions need to coordinate activity to support cognitive flexibility and behavioral adaptation. Neural activity in both the hippocampus (HPC) and medial prefrontal cortex (mPFC) is known to represent spatial context and is sensitive to reward and rule alterations. Midbrain dopamine (DA) activity is key in reward seeking behavior and learning. There is abundant evidence that midbrain DA modulates HPC and PFC activity. However, it remains underexplored how these networks engage dynamically and coordinate temporally when animals must adjust their behavior according to changing reward contingencies. In particular, is there any relationship between DA reward prediction change during rule switching, and rule representation changes in mPFC and CA1? We addressed these questions using simultaneous recording of neuronal population activity from the hippocampal area CA1, mPFC and ventral tegmental area (VTA) in male TH-Cre rats performing two spatial working memory tasks with frequent rule switches in blocks of trials. CA1 and mPFC ensembles showed rule-specific activity both during maze running and at reward locations, with mPFC rule coding more consistent across animals compared to CA1. Optogenetically tagged VTA DA neuron firing activity responded to and predicted reward outcome. We found that the correct prediction in DA emerged gradually over trials after rule-switching in coordination with transitions in mPFC and CA1 ensemble representations of the current rule after a rule switch, followed by behavioral adaptation to the correct rule sequence. Therefore, our study demonstrates a crucial temporal coordination between the rule representation in mPFC/CA1, the dopamine reward signal and behavioral strategy.

JNeurosci Accepted Manuscript

Significance Statement

This study examines neural activity in mammalian brain networks that support the ability to respond flexibly to changing contexts. We use a rule-switching spatial task to examine whether the key reward-responsive and predictive dopamine (DA) activity changes in coordination with changes in rule representations in key cognitive regions, the medial prefrontal cortex (mPFC) and hippocampus. We first established distinct rule representations in mPFC and hippocampus, and predictive coding of reward outcomes by DA neuronal activity. We show that the rule-specific DA reward prediction after a rule switch develops in temporal coordination with changes in rule representations in mPFC, eventually leading to behavioral changes. These results thus provide an integrated understanding of reward prediction, cognitive representations of rules and behavioral adaptation.

JNeurosci Accepted Manuscript

Introduction

Behavioral flexibility is critical to survival and for adapting to a changing environment. These functions are frequently driven by changing reward conditions and supported by distributed network processing in the brain to evaluate context and respond appropriately. The hippocampal cognitive map represents spatial context (O'Keefe and Dostrovsky, 1971; O'Keefe, 1978), and hippocampal neuronal activity can also be modulated by rewards and goals (Lee et al., 2012; Gauthier and Tank, 2018; Krishnan et al., 2022). The prefrontal cortex (PFC) serves complementary cognitive functions including working memory, executive control, and context representation (Ragozzino and Kesner, 1998; Yoon et al., 2008; Horst and Laubach, 2009; Durstewitz et al., 2010; Hyman et al., 2012; Urban et al., 2014; Ma et al., 2016). The hippocampus and prefrontal cortex have been demonstrated to coordinate temporally for adaptation of behavioral strategy when experiencing environmental/contextual changes (Guise and Shapiro, 2017; Hasz and Redish, 2020). However, much remains unexplored about how reward signals, acting as major feedback to animals' state/action choices, coordinate with this navigation and memory system for cognitive flexibility and behavioral adaptation. The dopamine (DA) signal from the midbrain is known to be involved in reward prediction error processing (Schultz et al., 1993, 1997) and value estimates (Roesch et al., 2007; Howe et al., 2013; Hamid et al., 2015; Dabney et al., 2020), and is causally linked to learning (Steinberg et al., 2013; Hamid et al., 2015). DA release profile and firing activity has also been reported to ramp up during reward approach and scale positively with reward quantity and probability (Howe et al., 2013; Engelhard et al., 2019; Krausz et al., 2023). How this DA signal changes during rule switches to support behavioral flexibility at short time scales of a few tens of trials over which behavioral change is observed remains less explored.

In this study, we were interested in examining the dynamics of rule and context representations in the hippocampal (area CA1) and prefrontal regions during reward-guided

JNeurosci Accepted Manuscript

behavioral adaptation in response to rule switching, and whether these dynamics are related to ventral tegmental area (VTA) DA neuronal activity. The primary hypotheses we aimed to test are whether changes in reward-associated DA neuronal firing activity are temporally coordinated with changes in rule representations in these cognitive regions as animals adapt to changes in contingencies over a few trials due to rule switching, thus supporting a role in cognitive flexibility. The PFC receives input from and sends output to midbrain DA neurons (Carr and Sesack, 2000; Beier et al., 2015; Kabanova et al., 2015; Morales and Margolis, 2017), and PFC and DA are known to contribute to reinforcement learning in a complimentary manner, with prefrontal signals encoding predictive values and dopamine encoding prediction errors (Lak et al., 2020). In addition, DA projections to dorsal hippocampus have been reported (Gasbarri et al., 1994), and CA1 place fields have been shown to change stability based on reward conditions and DA input (Martig and Mizumori, 2011; McNamara et al., 2014; Krishnan et al., 2022). It remains unknown if and how DA signal changes coordinate with rule and context representation in the CA1 and/or PFC, and the time scale of such coordination if it exists.

To investigate these questions, we implemented a novel rule-switching spatial task for rats in a W/M maze and recorded activity simultaneously from CA1, mPFC and VTA ensembles as animals performed this task. We found single cell and ensemble codes of underlying rules in both the CA1 and mPFC regions. VTA reward signaling and predictive coding of reward outcomes during maze running for the two rules were confirmed. We found that this rule-specific predictive feature of DA spiking activity develops together with correct mPFC and CA1 rule decoding a few trials after a rule switch, followed by a change of behavioral strategy within a few trials. Together, this work establishes that the dynamics of DA spiking activity and mPFC and CA1 rule representation changes occur in a coordinated manner during a rule switching task.

JNeurosci Accepted Manuscript

Materials and Methods

Animals and experimental design

Four adult male TH-Cre rats (450-600 g, 3-7 months, RRID: RRRC_00659) (Witten et al., 2011) were used in the current study for behavior and physiology data. All procedures were conducted under the guidelines of the US National Institutes of Health and approved by the Institutional Animal Care and Use Committee at Brandeis University. Animals were bred in house, kept under a 12h/12h light/dark schedule, with ad libitum food and water access till at least 10 weeks old prior to behavioral training and experiments. After daily handling and habituation to a sleep box (30cm long, 30 cm wide, 50 cm tall), animals were moderately food deprived to 85-90% of their initial weight to motivate reward-seeking behavior on an elevated linear track. Once reaching the criterion of obtaining 60 condensed milk rewards in a 20-min session, animals were allowed free access to food for at least a week before surgery. During surgery, virus AAV5-EF1a-DIO-hChR2(E123T/T159C)-mCherry was injected into VTA and a custom multi-tetrode microdrive with optrodes was implanted into CA1, mPFC and VTA (see **Surgical procedures**). Tetrodes and optrodes were gradually lowered into the target regions in the time span of 3-4 weeks to allow virus expression while animals recovered and retrained on the linear track. Recordings were performed during the process of learning and rule switching on the W track (see **Behavioral paradigm**). We included male rats only for this study due to technical limitations and animal welfare concerns. The weight of our Microdrive implant can be burdensome to female rats, which are typically 250-350 grams as adults, in comparison to 500-600 grams for male rats. Especially for lengthy recording days (4-6 hours) with at least 3-4 rule switches, it is more likely for male rats to meet the task requirements with less fatigue.

Apparatus

JNeurosci Accepted Manuscript

The experimental setup has been previously described (Shin et al., 2023). The linear track and W track were made of aluminum sheets and painted in matte black color. For each training/recording day, the track in use was situated in a dimly lit room with distinct visual features on each side (black curtain, black/white wall, equipment table). Customized reward wells were attached to the end of each track arm and connected to the recording and control system. SpikeGadgets Inc. hardware - an Environmental Control Unit (ECU) that was synchronized with physiology acquisition hardware, and custom stateScript software were used for behavior recording and automated reward dispensary. Choice correctness was determined based on the current maze rule, by the software that tracked well visit sequences. Upon an animal's arrival at the reward port after a correct choice, the animal's nose poke was detected by an infrared beam break sensor, immediately and automatically triggering the delivery of 0.15 mL of evaporated milk from a calibrated pump. Change of rule was implemented by software control, without any human interactions or changes in the environment. Nose poke and reward delivery times were recorded together with camera frames by ECU and electrophysiological signals (see **Data acquisition and processing**).

Behavioral paradigm

Post-recovery animals learned to perform a rule switching task on the W track (80 cm long of each arm, 7 cm wide, see **Figure 1A**). The task consists of two rules with similar structure but different sequence assignments. The first rule is a standard W track alternation task: animals must travel to the center arm (home) to obtain a milk reward when they set out from a side arm (inbound trajectories), whereas they need to alternate between the left and right arms when they start from the center (outbound trajectories) (Jadhav et al., 2012, 2016; Maharjan et al., 2018; Shin et al., 2019). The second rule presents altered identities between the left and center arm. Therefore, the left arm becomes the new home, and animals must alternate between the center and right arms for outbound trajectories. Once animals learned both rules, they were subject to

JNeurosci Accepted Manuscript

rule switching in blocks if performance of the current rule reached 80% correct. Rule switching happened without external cues, requiring the animals to use only reward feedback to deduce the change and switch to the optimal strategy for the new rule. During a recording day, animals ran 3-4 epochs of 20-25 min duration. The run epochs were interleaved by 20-40 min sleep epochs. It typically took animals 6-12 running epochs to learn each rule, and an additional training of at least 3-4 days before they could switch rapidly between rules (3-4 switches/rule blocks per day).

Surgical procedures

Each rat received virus injection and microdrive implantation during the surgery. Microdrive fabrication and implantation procedures were similar to previous reports (Jadhav et al., 2012, 2016; Tang et al., 2017; Shin et al., 2019). Anesthesia was induced by ketamine, xylazine and atropine cocktail and maintained by 0.5% - 2% isoflurane. 500 nL of AAV5-EF1a-DIO-hChR2(E123T/T159C)-mCherry was injected into VTA bilaterally (AP: -5.6 mm, ML: ±1.0 mm, DV: -7.8mm). 24 tetrodes each targeting dorsal CA1 (AP: -3.6 to -4.0 mm, ML: ±2.2 mm, DV: -2.5 mm) and mPFC (AP: +3.0 mm, ML: ±0.9 mm, DV: -2.5 to -3.0 mm), and two optrodes targeting VTA, with each optical fiber surrounded by 8 tetrodes, were encased in a microdrive, and implanted against the surface (for CA1 and mPFC) or 2 mm into the brain (for VTA). Animals received post-operative analgesia and were monitored closely for at least a week before food restriction and behavioral experiments.

Data acquisition and processing

Upon approaching target regions, CA1 was identified by the characteristic EEG features, including sharp-wave ripples (SWRs) and theta modulation. mPFC depth was targeted to anterior cingulate cortex (ACC) and prelimbic (PrL) regions. VTA was identified by finding optogenetically tagged cells (see **Cell type identification** below). Electrophysiological data was

JNeurosci Accepted Manuscript

recorded at 30 kHz using Trodes through a 256-channel headstage (SpikeGadgets, San Francisco, CA). Digital input/output signals of nose pokes and reward delivery were simultaneously recorded. Video recordings of animal behavior were captured at 30 fps with sync'd timestamps to the neural recording. All tetrodes were grounded to a screw above cerebellum. Tetrodes of each brain region were referenced using the group average. Spiking data was bandpass filtered between 600 Hz and 6 kHz, and local field potential (LFP) data was bandpass filtered between 0.5 Hz and 400 Hz and down sampled to 1.5 kHz. Animal's position was tracked using the cameraModule (SpikeGadgets) to identify the red/green LEDs attached to the headstage and verified by the researcher.

Spikes were clustered using MountainSort4 (Chung et al., 2017) and followed by manual inspection and curation in MountainView. Clusters with isolation score >0.9, noise overlap <0.05 and peak signal-to-noise ratio > 2 were accepted and included in this study.

Histology

After recording, animals were put under anesthesia and recording sites were lesioned by passing a current of 30 µA through the electrode tips. Animals were then perfused 1-2 days later using 4% formaldehyde and the brains were kept in 4% formaldehyde and 30% sucrose solution until being sliced into 50-µm sections. The VTA slices were immuno-stained for TH and imaged to verify colocalization of virus expression and anti-TH antibody at optrode implantation sites. The CA1 and mPFC slices were stained with cresyl violet and imaged to verify tetrode locations. Our histology results indicated the span of prefrontal recording sites included the ACC (cg1) and PrL region from bregma +3.0 mm to +2.5 mm (**Figure 1F**). Despite the long-standing inconsistency in naming of rodent prefrontal cortex (Uylings et al., 2003; Vertes, 2006; Laubach et al., 2018; van Heukelum et al., 2020), we felt it was most appropriate to use the term medial prefrontal cortex ('mPFC').

JNeurosci Accepted Manuscript

Cell type identification

Putative DA neurons were identified using optogenetic tagging at least 3 weeks after surgery to allow virus expression. In the first or last sleep epoch of each recording day, a 473-nm light was delivered to the VTA as 5- or 10-ms pulse trains at 1, 4, 10, 20 and 40 Hz to identify TH+ cells expressing ChR2, similar to previous studies (Cohen et al., 2012; Mohebi et al., 2019; Kim et al., 2020). The power of the laser was calibrated to lie within the range of 5-20 mW/mm^2 at the tip of the fiber to avoid spike waveform distortion. The light-evoked spike latency was then tested using Stimulus-Associated spike Latency Test (SALT, Kvitsiani et al., 2013). Units with p value <0.001 were classified as putative DA neurons.

CA1 and mPFC cells were separated into putative pyramidal neurons and interneurons using k-means clustering with parameters including spike width, peak asymmetry and mean firing rate (Barthó et al., 2004; Sirota et al., 2008; Shin and Jadhav, 2024).

Data analysis

Linearization and normalization

Each trial was defined as the time starting from leaving the last reward well, running on track and arriving at the current reward well, until the end of the stay at the current reward well. There are six trajectories in total defined across the two rules. Occupancy of positions of each trajectory was binned into 40 equally sized spatial bins (5-6 cm/bin) on the track during running, and then binned for 100-ms temporal bins at the reward wells. The occupancy was then smoothed by a 5-bin wide Gaussian kernel. The firing activity of each cell is binned and smoothed in the same fashion and firing rates are normalized by occupancy. For dimension

Neurosci Accepted Manuscript

reduction and population activity analyses, firing rates were z-scored for each cell across all task epochs.

Behavior analysis

Behavioral performance of each animal on each recording day was estimated using a previously described state space model (Smith et al., 2004; Kim and Frank, 2009). The average reward rate was summarized for rule 1 to 2 and rule 2 to 1, respectively. For analysis separating performance stages, quantile ranges were set individually for each recording day to have equal number of trials for each performance category (**Figure 3F, 4F**). Behavior strategy was estimated by generating the trajectory transition matrix in a 20-trial block with a 1-trial sliding window. The Pearson's correlation between each transition matrix and the optimal-strategy transition matrix of each rule was computed to determine which rule the animals are currently following (**Figure 3D, 5D**).

Spatial information

Spatial information was calculated according to Skaggs et al., 1992 to estimate the amount of spatial content of each cell's spiking activity:

$$I = \int_x \lambda(x)\, log_2 \frac{\lambda(x)}{\lambda}\, p(x) dx$$

where *I* is the spatial information measure in bits/second, *x* is spatial location of the animal, *p(x)* is the probability of the animal residing at location *x*, *λ(x)* is the mean firing rate of the cell at location *x*, and *λ* is the occupancy-weighted overall mean firing rate of the cell.

Dimensionality reduction with Consistent EmBeddings of high-dimensional Recordings using Auxiliary variables (CEBRA)

JNeurosci Accepted Manuscript

We used a newly developed nonlinear dimensionality reduction method, CEBRA, to uncover consistent and interpretable latent space of neural spiking data conditioned by behavioral variables (Schneider et al., 2023). Rewarded trials with performance better than 60th percentile of the day were selected for training in CEBRA (**Figure 3B, C**). 20% of these trials with good performance and all trials with worse performance were left out for testing. The training trials were balanced for trajectory identities and rules. We used supervised CEBRA-Behavior models for both single- and multi-animal training. Training labels included spatial-temporal bin, trajectory and rule. Spiking data from each brain region was trained separately for individual animals. Parameters for CEBRA training were largely coherent with the example implementation using rat hippocampus data in (Schneider et al., 2023), except for higher batch size and larger number of iterations to improve model performance:

```
model_architecture='offset10-model',
time_offsets=10,
batch_size=2048,
learning_rate=5e-4,
temperature=1,
output_dimension=3,
max_iterations=12000,
distance='cosine',
conditional='time_delta',
num_hidden_units=64
```

JNeurosci Accepted Manuscript

The whole data set was then transformed into 3D embeddings using the trained models. To reveal common features across animals, we took advantage of the multi-animal training from CEBRA. This method allows different dimensions (numbers of cells recorded) from different sessions/animals as inputs and maintains the label space across models. Therefore, the resulting embeddings across animals are directly comparable. For multi-animal training, trial selection, model parameters and subsequent transforming process were held the same as in single-session model training.

Rule decoding

A k-nearest neighbor (kNN) decoder was trained using the 3D embeddings of training data and used to decode the underlying rule for the testing data. Decoding performance of each trial was calculated as the ratio of bins with correct rule assignment (**Figure 3D**). The decoding accuracy curve was then smoothed using a Gaussian kernel. After a rule switch, a change point was identified as the trial from which decoding probability of the current rule was consistently higher than 50%. This process is done for embeddings from each brain region. We categorized the behavioral performance into low, mid and high by splitting the estimated performance into 1/3 of the trials for each day, and reported the proportion of trials with correct rule decoding in **Figure 3F**. To better estimate the timing of rule representation transitions, we used 5-fold validation for training CEBRA models and decoding, resulting in 5 estimates for each animal and each brain region, reflected in **Figure 3E, H, I**.

DA neuron firing rate analyses

Reward responsiveness is defined as showing differences ($p<0.05$) of firing rates at 200ms – 1200ms after the first nose poke between rewarded and unrewarded trials using Wilcoxon rank sum test. For firing rate comparisons across conditions in **Figure 4C**, firing rates are compared

JNeurosci Accepted Manuscript

using Wilcoxon signed rank test for each spatial/temporal bin and Bonferroni corrected for multi-comparison.

The influence of behavioral factors on each DA cell's single-trial firing rates was estimated using multiple linear regression.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

y is the mean firing rate of each trial during either running or the first 3 seconds of the outcome period. **X** is the estimated behavioral factors for each trial, which included: reward, binary outcome, rewarded or unrewarded; reward rate, an estimate of mean reward rate using the last 5 trials' outcome; trajectory reward rate, an estimated reward rate of a certain trajectory, using outcomes of the last 5 times that the animal took the same trajectory. This process is repeated for all DA neurons (n=20), for both the running and reward outcome firing rates. The resulting significant coefficient values and the proportion of significant coefficients were summarized in **Figure 4H**.

JNeurosci Accepted Manuscript

Results

We recorded neuronal activity simultaneously from the dCA1, mPFC (anterior cingulate cortex and prelimbic regions) and VTA in adult male Th-Cre rats (n=4) using tetrodes in mPFC and HPC and optical fibers surrounded by tetrodes in VTA (**Figure 1E-G**, see **Materials and Methods**), while animals performed a non-cued spatial rule switching task (**Figure 1A**). Both rules in the switching task consist of two types of trajectories with different memory demands: inbound trajectories that are rewarding at the home location regardless of the animals' travel history; and outbound trajectories that only reward animals when they choose the different arm from the last non-home visit, requiring working memory. For rule 1 (standard W/-M maze alternation task), the middle arm is the home location and animals alternate between left and right arms for outbound trajectories. For rule 2, the home location is switched to the left arm, and the middle and right arms become outbound destinations. Animals learned the two rules and were subsequently trained to switch between the two rules with solely the feedback of reward outcomes, and no external cues signaling the rule switch (**Figure 1B**). Animals were first trained on Rule 1 and subsequently on Rule 2. After 7-10 days of learning and training, animals could switch between the rules 3-4 times within a recording day. Rule switching was triggered manually during a session after a threshold performance (>80% correct) was achieved on the current rule. Upon unexpected trial outcomes, animals adapted their behavioral choices quickly to obtain more rewards, evident in decreasing numbers of perseverative errors over time (**Figure 1D**). It took 17 trials on average to reflect a Rule 2 to 1 switch in behavior, and 22 trials for Rule 1 to 2 (**Figure 1C**). The Rule 1 to 2 direction was typically more difficult for animals to achieve due to spatial asymmetry of rule 2. During each day of recording, animals ran the tasks in 3-4 20-30 min sessions, which were interleaved by 20-40 min sleep sessions. In the first or last sleep session, opto-tagging was performed to identify putative DA neurons in VTA (**Figure 1H**). An example raster during one trial is shown in **Figure 1I**. Position and speed plots show

JNeurosci Accepted Manuscript

animal motion and immobility at the destination reward well, and raster plots show corresponding patterns in the three regions. Note that CA1 and mPFC firing activity spans the spatial locations on the track. A total of 143 CA1 neurons, 128 mPFC neurons and 53 VTA neurons (20 TH+ neurons) from 4 animals, one recording day for each, were included in this study (**Table 1**).

Table 1: Clustered cell counts by brain region and cell type

Animal	CA1 Pyr	CA1 Int	mPFC Pyr	mPFC Int	VTA DA	VTA non-DA
TH105	34	7	26	4	4	7
TH212	14	5	34	5	8	11
TH510	44	5	30	2	3	6
TH605	31	3	25	2	5	9
Total	123	20	115	13	20	33

CA1 and mPFC neurons show differentiated activity across rules

CA1 and mPFC neurons show spatially modulated firing activity (O'Keefe and Dostrovsky, 1971; O'Keefe, 1978; Hyman et al., 2010; Zielinski et al., 2019), and their features have also been implicated in rule and context representations (Wood et al., 2000; Eschenko and Mizumori, 2007; Griffin et al., 2007; Rich and Shapiro, 2009; Durstewitz et al., 2010; Ferbinteanu et al., 2011; Karlsson et al., 2012; Powell and Redish, 2016; Guise and Shapiro, 2017; Hasz and Redish, 2020). We therefore first asked the question if the firing activity of CA1 and mPFC neurons can reflect the current rule by showing rule-specific activity on trajectories. As the firing activity of VTA neurons showed much weaker spatial modulation (**Figure 2E**), we focused only on CA1 and mPFC for this analysis. Single-unit remapping was investigated using

JNeurosci Accepted Manuscript

358 firing activity on the two common trajectories that lead to reward for correct trials for both rules
359 (*Left-to-Center*: inbound for Rule 1 and outbound for Rule 2, and *Center-to-Left*: outbound for
360 Rule 1 and inbound for Rule 2), as lack of reward is known to destabilize firing activity on mazes
361 (Krishnan et al., 2022). We found cells showing remapped activity across the two rules in both
362 CA1 and mPFC regions (**Figure 2A**; rules are run in alternating blocks with two blocks for each
363 rule). Some cells show rate remapping (**Figure 2A**, first example on left, CA1 excitatory cell),
364 and some cells exhibited relocation of spatial firing to a different part of the track (CA1 and PFC
365 excitatory cell examples). Additionally, reward-associated firing after the same physical
366 trajectory but in different rule contexts also showed robust differences (examples with significant
367 differences in **Figure 2B**). These firing changes are not a result of recording drift over time, as
368 similar patterns were observed both earlier and later in the day, with interleaved rule blocks.
369 Further, trial-wise cell firing vectors also exhibited increases in correlation with stable-
370 performance firing vectors of the current rule as behavioral performance improved, and became
371 less similar (lower correlation) after the rule changed (**Figure 2C**). Overall, we found that 27.3%
372 of hippocampal CA1 cells and 25.8% of mPFC cells show significantly differentiated activity
373 during running based on the current underlying rule, and 14.7% CA1 and 14.8% mPFC neurons
374 during reward. There was no significant difference in the ratio of rule-modulated neurons across
375 regions (**Figure 2D**).

376 **CA1 and mPFC ensemble representations distinguish the rules and show transitions**
377 **during rule switching**

378 We next investigated ensemble activity changes across rules in CA1 and mPFC, and the
379 dynamics of these changes in relationship to behavior. Trial-by-trial population vector similarity
380 was computed and aligned to the first trial after rule switching, shown in the correlation plots in
381 **Figure 3A**. Similar to previous research findings (Hasz and Redish, 2020), the firing pattern of
382 both regions stabilized toward the end of running one rule (seen as increase in correlation with

JNeurosci Accepted Manuscript

neighboring trials within a rule block) and quickly destabilized after changing to the other rule (**Figure 3A**). We adopted a newly developed dimensionality-reduction method CEBRA (Schneider et al., 2023), utilizing contrastive learning to visualize the difference of high-dimensional neural data across rules. For a common trajectory across the two rules (*Left-to-Center*), mPFC 3D embeddings using CEBRA qualitatively showed consistent separation across rules for all animals during stable performance (visualized in **Figure 3B (ii)**, top row). These embeddings used training and test datasets within animals. We also used multi-session training that sampled training data from all animals followed by testing within animals. The embeddings from this method revealed a similar structure across animals, implying animal-invariant features in the mPFC ensemble data for rule representations (**Figure 3B (ii)**, bottom row). CA1 single-session embeddings showed similar separations between rules compared to mPFC. However, much weaker separations were observed in multi-animal embeddings, suggesting a more unique latent space of CA1 rule-specific activity for individual animals (**Figure 3B (i)**). As animals switched from stable performance of one rule to another, we also observed a systematic transition in the manifold space (example mPFC manifold transitions over trials in **Figure 3C**).

In order to quantify these changes, a KNN decoder was trained to decode the underlying rule using single-animal embeddings for each brain region (see **Materials and Methods**). Rule decoding probabilities using CA1 and mPFC data correlated highly with animals' behavior strategy (**Figure 3D, E**). Upon further inspection, behavior strategy showed stronger correlation with mPFC rule decoding than with CA1 decoding (**Figure 3E**). The decoding probability of the current rule increased with behavioral performance for both mPFC and CA1 (**Figure 3F**, mean decoding accuracy for low performance: CA1 0.38, mPFC 0.41; mid performance: CA1 0.66, mPFC 0.71; high performance: CA1 0.97, mPFC 0.96; performance thresholds are reported in **Materials and Methods**). However, when training the decoder using multi-animal embeddings,

JNeurosci Accepted Manuscript

CA1 ensembles performed much worse than mPFC, again implying more consistent rule coding across animals in mPFC and its stronger relevance to behavior than CA1 (**Figure 3F**).

We further examined the timing between neural representation change and behavior strategy shift upon rule switch. The behavior and neural changes align with each other, with neural representations in both CA1 and mPFC transitioned more often before behavior strategy than after (**Figure 3H, I**), with no timing difference seen between CA1 and mPFC ensembles. On average, it took more trials to switch from rule 1 to rule 2, both in the neural representation and behaviorally (**Figure 3G**), potentially due to asymmetric working memory components in rule 2.

VTA DA firing activity reflects and predicts reward outcomes

As midbrain DA plays a crucial role in reward signaling, we aimed to understand VTA DA spiking activity features during rapid rule switches guided by unexpected reward outcomes. We used an opto-tagging strategy in TH-Cre rats, which has been utilized in previous studies (Witten et al., 2011). Opto-tagging was performed during the first or last sleep session for each day of recording to not interfere with activity during task running. We recorded a total of 53 VTA neurons, out of which 20 photo-tagged cells were identified as putative DA neurons (with firing rates, mean ± std, 7.7 ± 3.0 Hz). Consistent with a vast body of literature, putative DA neurons showed increased firing rates when rewards became available, temporally aligned to first poke at reward well, and decreased firing rates during reward omission (**Figure 4A,** median z-scored firing rate for rewarded trials: 0.49, unrewarded trials: -0.76, $p = 5.6\times10^{-5}$ Wilcoxon signed rank test). Non-tagged neurons exhibited a diverse range of activity, including 45.5% (15/33) of neurons responding to reward outcomes (**Figure 4B**). Further, we also observed that DA neurons had higher firing rates during running for trials that lead to rewards in comparison to unrewarded trials (**Figure 4C**, median z-scored firing rate for rewarded trials: 0.11, unrewarded

JNeurosci Accepted Manuscript

trials: -0.060, $p = 1.8\times10^{-4}$). This difference became apparent and significant halfway through running on the trajectory, just after the choice point where the decision was made, similar to previous findings of DA release in striatum, and is likely related to reward expectancy/ uncertainty (Howe et al., 2013). The running speed profile on the track was similar between rewarded and unrewarded trials, suggesting the DA firing rate difference is not caused by a change in motivation or vigor (**Figure 4D**).

Based on the proposed link between DA spiking and reward prediction error, it is predicted that for rewarded trials, as behavioral performance improves, DA firing rates at the reward will decrease and DA firing rates during run will increase; and such trends would be the opposite for unrewarded trials (Watabe-Uchida et al., 2017). However, we observed a strong positive correlation between run and reward firing rates ($r = 0.60$, $p = 6.8\times10^{-148}$), and no obvious relationship between firing rates of either state to performance (**Figure 4E**). We further tested this by separating the performance by trials into low and high performance halves, and found no difference in the firing rates across performance stages for either rewarded or unrewarded trials (n=20 cells, median z-scored firing rate for rewarded trials: run: low performance 0.065, high performance 0.13, p=0.23; outcome: low performance 0.54, high performance 0.44, p=1.00; unrewarded trials: run: low performance -0.072, high performance -0.049, p=0.23; outcome: low performance -0.81, high performance -0.71, p=0.65, Wilcoxon paired rank test, **Figure 4F**).

In addition, we wanted to investigate if the DA reward prediction signal is present for working memory error as compared to rule/perseverative error that occurs after a rule switch. At each time of rule switching, two trajectories became unrewarding under all conditions and two new ones became potentially rewarding when the sequence of arm visits was correct. For example, when rule 1 changes to rule 2, trajectories between the center and the right arms are no longer rewarding. Instead, left-to-right and right-to-left ones yield rewards. We classified the trials that animals ran on the newly unrewarding trajectories as perseverative errors, whereas the trials

JNeurosci Accepted Manuscript

that animals made a wrong outbound choice for the current rule as working memory errors. If the DA firing activity only reflects a model-free system tracking overall reward probability, we would expect a difference only between perseverative errors and correct trials, but not for individual decisions dependent on working memory. We found that the prediction for reward outcome was present in DA firing regardless of the error type, even when comparing error trials to nearest correct trials of the same outbound/inbound trajectory type (**Figure 4G**, median z-scored firing rate difference between rewarded and unrewarded trials: outbound 0.12, $p = 1.0\times10^{-13}$, n = 217 trials; inbound 0.085, $p = 1.8\times10^{-6}$, n = 178 trials).

To summarize the influence of reward conditions and histories on DA firing rates, we fitted multi-regression models for each cell's spiking rates during running and outcome periods, separately. The most significant factor was found to be the current trial's reward outcome (running: mean±sem coefficients: 0.14±0.02, significant for 65% of DA neurons; outcome: 0.95±0.13, 95%), in comparison to a weaker impact of reward history (running: 0.11±0.10, 45%; outcome: -0.13±0.16, 30%) or same-trajectory reward history (running: 0.074±0.058, 30%; outcome: -0.39±0.07, 65%), for both running and reward states (**Figure 4H**). Note that current trial reward outcome is a significant contributor to the approach run firing rate, and not just for reward well firing rate that is expected.

DA reward predicting feature develops after rule switching and coordinates with mPFC rule representation transition before behavior strategy adaptation

Our findings in mPFC/CA1 rule coding and DA reward signals led us to the question -- if the firing rates of DA neurons predictive of reward outcomes develop as animals gather evidence of a rule change. Firing rates upon reward delivery were consistently higher than at reward omission (**Figure 5A (iv)**); however, the difference in DA neuron firing between rewarded and

JNeurosci Accepted Manuscript

unrewarded trials during running only developed a few trials after the rule switch was implemented (**Figure 5A (v)**), and a few trials prior to behavioral strategy switch (**Figure 5A (i)**). We further investigated whether this DA firing activity is coordinated with rule representation changes in CA1 and mPFC. When aligned to the trial where mPFC started to decode the current rule, DA firing rate difference during run between rewarded and unrewarded trials ramped up robustly toward this change point (**Figure 5B**, bottom). The ramp up in DA firing rate difference was qualitatively more robust and consistent for mPFC representation switch rather than when aligned to CA1 representation switch (**Figure 5C**, bottom). Animals' behavior strategies adapted accordingly shortly after (on average 4 trials) the observed mPFC rule representation switch and DA prediction emergence (**Figure 5B,** top). In addition, the DA firing activity difference between rewarded and unrewarded trials during maze running did not stay high as animals reached good, stable performance for the new rule. We observed a positive correlation between the DA firing rate difference and the rate of mPFC decoding probability change, which may suggest a gating mechanism of value/belief update (**Figure 5D**). In summary, the reward predictive/expectancy property of DA firing during running is acquired post rule switch and is coordinated with the changes of mPFC rule representation and behavior strategy.

JNeurosci Accepted Manuscript

Discussion

In the current study, we examined rule-related firing properties in the hippocampus and mPFC, VTA DA spiking activity during rule switching, and the temporal relationships between cognitive rule representation, reward prediction signal and behavior adaptation. The mechanisms of rule representation and reward prediction have been generally examined separately, but it is vital to gain an integrated understanding of dynamics and temporal coordination between these mechanisms for supporting cognitive flexibility and behavioral adaptation. During rule switching, the presence or absence of rewards after specific actions can serve as feedback for organisms to reevaluate their behavior policy. A conflict between belief and outcome leads to a switch in internal rule representations, which can guide correct behavioral actions for the changed context. We therefore reasoned that there must exist some sort of temporal coordination between reward prediction and rule representation dynamics over the course of a few trials after rule switching, which will determine the time course of behavioral adjustments, thus allowing animals to overcome perseveration and adapt to the new context with appropriate actions. We found that mPFC and hippocampal CA1 population activity represented rule context. VTA DA firing activity responded to and predicted reward outcomes. The timings of the switches in mPFC-CA1 rule representation and the emergence of correct reward prediction from VTA were coordinated. Further, these changes in mPFC, CA1 and VTA DA firing activity led the behavioral strategy transition, suggesting coordinated network activity switches to support behavioral adaptation.

We implemented a spatial rule switching task in a W/M maze, probing behavioral flexibility and working memory at the same time in the exact same environment. Rule 1 comprises a traditional W-maze alternation rule with the center arm as home arm as in previous studies (Frank et al., 2000; Jadhav et al., 2012, 2016; Tang et al., 2017; Maharjan et al., 2018; Shin et al., 2019), and the home arm switched to the left arm for Rule 2, requiring the animals to change

JNeurosci Accepted Manuscript

their strategy for optimal trajectory sequences across the rules. Notably, both rules incorporate a spatial working memory component (outbound component), requiring the animals to choose the opposite arm from previous choice when embarking from the home arm. The return-to-home-arm trajectory is the inbound component with a spatial reference memory demand (Kim and Frank, 2009). The nature of the task assigned different memory demands of overlapping trajectories in the two rules, which allowed us to compare animal behavior, CA1/mPFC representations and DA signaling for reward prediction during rule switch.

We first examined the differences of rule representation between the hippocampal CA1 and mPFC. Spiking activity changes associated with rule/context were observed in both brain regions (Guise and Shapiro, 2017; Hasz and Redish, 2020), with single neurons exhibiting remapping on the same trajectories across the two rule contexts. Examining ensemble coding using manifold analysis and decoding to current rule context revealed more robust and consistent coding of rule contexts across animals in mPFC, suggesting that similar neural representations may develop in mPFC to support two rule contexts in the same physical trajectory space on mazes. Rule representations in mPFC appeared to shift over the course of a few to tens of trials after rule switching, prior to observation of a behavioral strategy switch.

We observed more consistency across animals in mPFC ensemble decoding for rules, whereas CA1 ensemble representations were more animal-specific (**Fig. 3B**). In terms of timing, we did not see a significant difference between mPFC and CA1 in terms of neural transition times as compared with behavioral transition times, in terms of number of trials, although there was some variability across regions (**Fig. 3H, I**). However, the DA prediction differences that emerged after rule changes were more robustly aligned to mPFC transition times than to CA1 (**Fig. 5B-C**), and mPFC transition times were more closely aligned to behavioral strategy switch (**Fig. 3E**). Therefore, although there were no significant timing differences between mPFC and CA1, this suggests more consistency in mPFC neural transitions and behavioral transitions.

JNeurosci Accepted Manuscript

We observed that VTA DA neurons participated in reward prediction and this rule-specific feature developed quickly after rule switching. Previous literature has often focused on model-free RL reward prediction error; here, we instead tried to understand how midbrain DA neuronal firing activity changes during a rule switching task that requires working memory. DA neuronal firing rates were highly correlated between running and reward outcome in a trial, which somewhat resembles previous findings in (Coddington and Dudman, 2018). In addition, we did not find a correlation between DA spiking activity and performance at the population level. These results contrast with the prediction based on classical conditioning that firing activity increases toward cue/beginning of a trial and slowly diminishes upon reward delivery (Schultz et al., 1997; Fiorillo et al., 2003; Amo et al., 2022). Rather, our findings suggest that DA signaling may incorporate rule-based and working memory-dependent prediction information, potentially from cortical input.

We found a delay in animals' behavioral strategy adjustment in comparison to rule representation and reward prediction. Animals continued to sample from wrong trajectories, although at a lower frequency, after manifolds shifted to the new rule and DA signal could predict the reward outcome. This can be explained by simply reserving a small chance for random exploration, or perhaps by the theory of active inference (Friston et al., 2012, 2014). In this theory, the dopamine signal minimizes surprise instead of cost. Therefore, sampling from extra trials could help confirm animals' beliefs if correct and eventually stabilize behavioral sequences.

How is information processed in these regions? Does VTA receive reward, expectation or reward prediction error (RPE) information from other regions? Lisman and Grace (2005) suggested that novelty information is conveyed to VTA from the hippocampal formation through an indirect route, possibly with inhibitory afferents from accumbens and ventral pallidum. In return, DA fibers innervate dorsal hippocampus and thus enhance LTP (Huang and Kandel,

JNeurosci Accepted Manuscript

1995; Otmakhova and Lisman, 1996; Li et al., 2003; Rosen et al., 2015; Sayegh et al., 2024). In terms of DA-cortical interactions, VTA receives abundant prefrontal glutamatergic projections (Geisler et al., 2007), and their impact on reward signaling is multifaceted. Input from frontal cortices is critical for conveying predictive and incentive features of a cue (Pan et al., 2021). PFC conveys belief state information to the dopaminergic system and affects RPE computation when hidden states are involved (Babayan et al., 2018; Starkweather et al., 2018). Additionally, Amo et al. (2024) showed glutamatergic input to VTA already carried RPE signal, challenging the classic view of local computation of RPE in VTA by combining different aspects of reward and expectation from glutamatergic and GABAergic inputs (Kawato and Samejima, 2007; Keiflin and Janak, 2015). An intriguing future direction of research could focus on dissecting computational features or components of RPE in these brain regions involved in the reward circuit, particularly when multiple cognitive demands are present.

It is important to differentiate between DA firing activity and DA release in downstream areas. We did not observe an obvious increase in DA spiking activity during goal approach. However, downstream release in striatum has been reported to ramp up towards goal (Howe et al., 2013; Hamid et al., 2015), suggesting partially dissociated activity potentially caused by local modulation (Mohebi et al., 2019). It would be interesting to examine if the working memory-dependent reward prediction signal is present in the DA release profile as in VTA DA spiking activity, and if so, how that influences its downstream striatal and cortical areas.

The exact mechanism of information transfer and cooperation requires further investigation. For example, Fujisawa and Buzsáki (2011) proposed a 4 Hz rhythm orchestrating neuronal activity in the hippocampus, PFC and VTA during a similar working memory task. DA administration in PFC increased HPC-PFC theta coherence and PFC phase locking to CA1 theta (Benchenane et al., 2010). PFC theta sequences selectively coordinate with CA1 theta sequences depicting future trajectory choices (Tang et al., 2021). Therefore, altered DA neuron

JNeurosci Accepted Manuscript

spiking activity might be impacted by the online processes of action evaluation and selection during running. Another possible mechanism is through coordinated reactivation/replay. It has been shown that hippocampal reactivation is associated with increased firing activity of VTA cells as a potential mechanism for memory consolidation (Gomperts et al., 2015). CA1-PFC coordinated replay is biased toward behavioral choice and therefore can potentially participate in working memory and decision making (Shin et al., 2019). This hippocampo-cortical/subcortical temporal coordination during offline states may serve as substrates for value updates, memory consolidation, action evaluation and planning. Our study focused on the behavioral time scale, but investigating simultaneous activity on a shorter time scale (50-200 ms) can potentially uncover detailed mechanisms of inter-regional communication.

Conclusions

To understand the temporal coordination between cognitive systems and reward systems, we recorded from the rat hippocampal CA1, mPFC and VTA during a rapid rule switching task. CA1 and mPFC exhibited rule representation in their population activity. DA neuron firing activity in VTA gradually became predictive of reward outcomes after rule switching, and this predictive feature developed together with the mPFC representation transition to reflect the new rule. These neural representations changed in advance of behavioral adaptation by a few trials. Together, our study revealed a synchronized update of information across mPFC, HPC and VTA that potentially supports behavioral flexibility.

JNeurosci Accepted Manuscript

References

Amo R, Matias S, Yamanaka A, Tanaka KF, Uchida N, Watabe-Uchida M (2022) A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. Nature Neuroscience 2022:1–11.

Amo R, Uchida N, Watabe-Uchida M (2024) Glutamate inputs send prediction error of reward, but not negative value of aversive stimuli, to dopamine neurons. Neuron 112:1001-1019.e6.

Babayan BM, Uchida N, Gershman SJ (2018) Belief state representation in the dopamine system. Nature Communications 2018 9:1 9:1–10.

Barthó P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsáki G (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. J Neurophysiol 92:600–608.

Beier KT, Steinberg EE, Deloach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, Luo L (2015) Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. Cell 162:622–634.

Benchenane K, Peyrache A, Khamassi M, Tierney PL, Gioanni Y, Battaglia FP, Wiener SI (2010) Coherent Theta Oscillations and Reorganization of Spike Timing in the Hippocampal-Prefrontal Network upon Learning. Neuron 66:921–936.

Carr DB, Sesack SR (2000) Projections from the Rat Prefrontal Cortex to the Ventral Tegmental Area: Target Specificity in the Synaptic Associations with Mesoaccumbens and Mesocortical Neurons. Journal of Neuroscience 20:3864–3873.

Chung JE, Magland JF, Barnett AH, Tolosa VM, Tooker AC, Lee KY, Shah KG, Felix SH, Frank LM, Greengard LF (2017) A Fully Automated Approach to Spike Sorting. Neuron 95:1381-1394.e6.

JNeurosci Accepted Manuscript

643 Coddington LT, Dudman JT (2018) The timing of action determines reward prediction signals in
644 identified midbrain dopamine neurons. Nature Neuroscience 2018 21:11 21:1563–1573.

645 Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for
646 reward and punishment in the ventral tegmental area. Nature 2012 482:7383 482:85–88.

647 Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M
648 (2020) A distributional code for value in dopamine-based reinforcement learning. Nature
649 2020 577:7792 577:671–675.

650 Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010) Abrupt Transitions between
651 Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule
652 Learning. Neuron 66:438–448.

653 Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw
654 ND, Tank DW, Witten IB (2019) Specialized coding of sensory, motor and cognitive
655 variables in VTA dopamine neurons. Nature 2019 570:7762 570:509–513.

656 Eschenko O, Mizumori SJY (2007) Memory influences on hippocampal and striatal neural
657 codes: Effects of a shift between task rules. Neurobiol Learn Mem 87:495–509.

658 Ferbinteanu J, Shirvalkar P, Shapiro ML (2011) Memory Modulates Journey-Dependent Coding
659 in the Rat Hippocampus. Journal of Neuroscience 31:9135–9146.

660 Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty
661 by dopamine neurons. Science (1979) 299:1898–1902.

662 Frank LM, Brown EN, Wilson M (2000) Trajectory encoding in the hippocampus and entorhinal
663 cortex. Neuron 27:169–178.

JNeurosci Accepted Manuscript

664 Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ (2014) The
665 anatomy of choice: dopamine and decision-making. Philosophical Transactions of the Royal
666 Society B: Biological Sciences 369.

667 Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan
668 KE, Bestmann S (2012) Dopamine, Affordance and Active Inference. PLoS Comput Biol
669 8:e1002327.

670 Fujisawa S, Buzsáki G (2011) A 4 Hz Oscillation Adaptively Synchronizes Prefrontal, VTA, and
671 Hippocampal Activities. Neuron 72:153–165.

672 Gasbarri A, Packard MG, Campana E, Pacitti C (1994) Anterograde and retrograde tracing of
673 projections from the ventral tegmental area to the hippocampal formation in the rat. Brain
674 Res Bull 33:445–452.

675 Gauthier JL, Tank DW (2018) A Dedicated Population for Reward Coding in the Hippocampus.
676 Neuron 99:179-193.e7.

677 Geisler S, Derst C, Veh RW, Zahm DS (2007) Glutamatergic Afferents of the Ventral Tegmental
678 Area in the Rat. Journal of Neuroscience 27:5730–5743.

679 Gomperts SN, Kloosterman F, Wilson MA (2015) VTA neurons coordinate with the hippocampal
680 reactivation of spatial experience. Elife 4.

681 Griffin AL, Eichenbaum H, Hasselmo ME (2007) Spatial Representations of Hippocampal CA1
682 Neurons Are Modulated by Behavioral Context in a Hippocampus-Dependent Memory Task.
683 Journal of Neuroscience 27:2416–2423.

684 Guise KG, Shapiro ML (2017) Medial Prefrontal Cortex Reduces Memory Interference by
685 Modifying Hippocampal Encoding. Neuron 94:183-192.e8.

JNeurosci Accepted Manuscript

686 Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT,
687 Aragona BJ, Berke JD (2015) Mesolimbic dopamine signals the value of work. Nature
688 Neuroscience 2016 19:1 19:117–126.

689 Hasz BM, Redish AD (2020) Dorsomedial prefrontal cortex and hippocampus represent
690 strategic context even while simultaneously changing representation throughout a task
691 session. Neurobiol Learn Mem 171:107215.

692 Horst NK, Laubach M (2009) The role of rat dorsomedial prefrontal cortex in spatial working
693 memory. Neuroscience 164:444–456.

694 Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM (2013) Prolonged dopamine
695 signalling in striatum signals proximity and value of distant rewards. Nature 2013 500:7464
696 500:575–579.

697 Huang YY, Kandel ER (1995) D1/D5 receptor agonists induce a protein synthesis-dependent
698 late potentiation in the CA1 region of the hippocampus. Proceedings of the National
699 Academy of Sciences 92:2446–2450.

700 Hyman JM, Ma L, Balaguer-Ballester E, Durstewitz D, Seamans JK (2012) Contextual encoding
701 by ensembles of medial prefrontal cortex neurons. Proc Natl Acad Sci U S A 109:5086–
702 5091.

703 Hyman JM, Zilli EA, Paley AM, Hasselmo ME (2010) Working memory performance correlates
704 with prefrontal-hippocampal theta interactions but not with prefrontal neuron firing rates.
705 Front Integr Neurosci 4:1162.

706 Jadhav SP, Kemere C, German PW, Frank LM (2012) Awake hippocampal sharp-wave ripples
707 support spatial memory. Science (1979) 336:1454–1458.

JNeurosci Accepted Manuscript

Jadhav SP, Rothschild G, Roumis DK, Frank LM (2016) Coordinated Excitation and Inhibition of Prefrontal Ensembles during Awake Hippocampal Sharp-Wave Ripple Events. Neuron 90:113–127.

Kabanova A, Pabst M, Lorkowski M, Braganza O, Boehlen A, Nikbakht N, Pothmann L, Vaswani AR, Musgrove R, Di Monte DA, Sauvage M, Beck H, Blaess S (2015) Function and developmental origin of a mesocortical inhibitory circuit. Nature Neuroscience 2015 18:6 18:872–882.

Karlsson MP, Tervo DGR, Karpova AY (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science (1979) 338:135–139.

Kawato M, Samejima K (2007) Efficient reinforcement learning: computational theories, neuroscience and robotics. Curr Opin Neurobiol 17:205–212.

Keiflin R, Janak PH (2015) Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. Neuron 88:247–263.

Kim HGR, Malik AN, Mikhael JG, Bech P, Tsutsui-Kimura I, Sun F, Zhang Y, Li Y, Watabe-Uchida M, Gershman SJ, Uchida N (2020) A Unified Framework for Dopamine Signals across Timescales. Cell 183:1600-1616.e25.

Kim SM, Frank LM (2009) Hippocampal Lesions Impair Rapid Learning of a Continuous Spatial Alternation Task Tanimoto H, ed. PLoS One 4:e5494.

Krausz TA, Comrie AE, Kahn AE, Frank LM, Daw ND, Correspondence JDB, Berke JD (2023) Dual credit assignment processes underlie dopamine signals in a complex spatial environment. Neuron 111:3465–3478.

JNeurosci Accepted Manuscript

729 Krishnan S, Heer C, Cherian C, Sheffield MEJ (2022) Reward expectation extinction
730 restructures and degrades CA1 spatial maps through loss of a dopaminergic reward
731 proximity signal. Nature Communications 2022 13:1 13:1–19.

732 Kvitsiani D, Ranade S, Hangya B, Taniguchi H, Huang JZ, Kepecs A (2013) Distinct behavioural
733 and network correlates of two interneuron types in prefrontal cortex. Nature 498:363–366.

734 Lak A, Okun M, Moss MM, Gurnani H, Farrell K, Wells MJ, Reddy CB, Kepecs A, Harris KD,
735 Carandini M (2020) Dopaminergic and Prefrontal Basis of Learning from Sensory
736 Confidence and Reward Value. Neuron 105:700-711.e6.

737 Laubach M, Amarante LM, Swanson K, White SR (2018) What, If Anything, Is Rodent Prefrontal
738 Cortex? eNeuro 5:315–333.

739 Lee H, Ghim JW, Kim H, Lee D, Jung MW (2012) Hippocampal Neural Correlates for Values of
740 Experienced Events. Journal of Neuroscience 32:15053–15065.

741 Li S, Cullen WK, Anwyl R, Rowan MJ (2003) Dopamine-dependent facilitation of LTP induction
742 in hippocampal CA1 by exposure to spatial novelty. Nature Neuroscience 2003 6:5 6:526–
743 531.

744 Lisman JE, Grace AA (2005) The Hippocampal-VTA Loop: Controlling the Entry of Information
745 into Long-Term Memory. Neuron 46:703–713.

746 Ma L, Hyman JM, Durstewitz D, Phillips AG, Seamans JK (2016) A Quantitative Analysis of
747 Context-Dependent Remapping of Medial Frontal Cortex Neurons and Ensembles. Journal
748 of Neuroscience 36:8258–8272.

749 Maharjan DM, Dai YY, Glantz EH, Jadhav SP (2018) Disruption of dorsal hippocampal –
750 prefrontal interactions using chemogenetic inactivation impairs spatial learning. Neurobiol
751 Learn Mem 155:351–360.

JNeurosci Accepted Manuscript

Martig AK, Mizumori SJY (2011) Ventral tegmental area disruption selectively affects CA1/CA2 but not CA3 place fields during a differential reward working memory task. Hippocampus 21:172–184.

McNamara CG, Tejero-Cantero Á, Trouche S, Campo-Urriza N, Dupret D (2014) Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. Nat Neurosci 17:1658–1660.

Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD (2019) Dissociable dopamine dynamics for learning and motivation. Nature 2019 570:7759 570:65–70.

Morales M, Margolis EB (2017) Ventral tegmental area: Cellular heterogeneity, connectivity and behaviour. Nat Rev Neurosci 18:73–85.

O'Keefe J (1978) The hippocampus as a cognitive map. Oxford University Press.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res 34:171–175.

Otmakhova NA, Lisman JE (1996) D1/D5 Dopamine Receptor Activation Increases the Magnitude of Early Long-Term Potentiation at CA1 Hippocampal Synapses. Journal of Neuroscience 16:7478–7486.

Pan WX, Coddington LT, Dudman JT (2021) Dissociable contributions of phasic dopamine activity to reward and prediction. Cell Rep 36:109684.

Powell NJ, Redish AD (2016) Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. Nature Communications 2016 7:1 7:1–11.

JNeurosci Accepted Manuscript

773 Ragozzino ME, Kesner RP (1998) The Effects of Muscarinic Cholinergic Receptor Blockade in
774 the Rat Anterior Cingulate and Prelimbic/Infralimbic Cortices on Spatial Working Memory.
775 Neurobiol Learn Mem 69:241–257.

776 Rich EL, Shapiro M (2009) Rat Prefrontal Cortical Neurons Selectively Code Strategy Switches.
777 Journal of Neuroscience 29:7208–7219.

778 Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in
779 rats deciding between differently delayed or sized rewards. Nature Neuroscience 2007
780 10:12 10:1615–1624.

781 Rosen ZB, Cheung S, Siegelbaum SA (2015) Midbrain dopamine neurons bidirectionally
782 regulate CA3-CA1 synaptic drive. Nature Neuroscience 2015 18:12 18:1763–1771.

783 Sayegh FJP, Mouledous L, Macri C, Pi Macedo J, Lejards C, Rampon C, Verret L, Dahan L
784 (2024) Ventral tegmental area dopamine projections to the hippocampus trigger long-term
785 potentiation and contextual learning. Nature Communications 2024 15:1 15:1–12.

786 Schneider S, Lee JH, Mathis MW (2023) Learnable latent embeddings for joint behavioural and
787 neural analysis. Nature 2023 617:7960 617:360–368.

788 Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward
789 and conditioned stimuli during successive steps of learning a delayed response task.
790 Journal of Neuroscience 13:900–913.

791 Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science
792 (1979) 275:1593–1599.

793 Shin JD, Jadhav SP (2024) Prefrontal cortical ripples mediate top-down suppression of
794 hippocampal reactivation during sleep memory consolidation. Current Biology 34:2801-
795 2811.e9.

JNeurosci Accepted Manuscript

Shin JD, Tang W, Jadhav SP (2019) Dynamics of Awake Hippocampal-Prefrontal Replay for Spatial Learning and Memory-Guided Decision Making. Neuron.

Shin JD, Tang W, Jadhav SP (2023) Protocol for geometric transformation of cognitive maps for generalization across hippocampal-prefrontal circuits. STAR Protoc 4:102513.

Sirota A, Montgomery S, Fujisawa S, Isomura Y, Zugaro M, Buzsáki G (2008) Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Theta Rhythm. Neuron 60:683–697.

Skaggs WE, Mcnaughton BL, Gothard KM (1992) An Information-Theoretic Approach to Deciphering the Hippocampal Code. Adv Neural Inf Process Syst 5.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. J Neurosci 24:447–461.

Starkweather CK, Gershman SJ, Uchida N (2018) The Medial Prefrontal Cortex Shapes Dopamine Reward Prediction Errors under State Uncertainty. Neuron 98:616-629.e6.

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966–973.

Tang W, Shin JD, Frank LM, Jadhav SP (2017) Hippocampal-Prefrontal Reactivation during Learning Is Stronger in Awake Compared with Sleep States. J Neurosci 37:11789–11805.

Tang W, Shin JD, Jadhav SP (2021) Multiple time-scales of decision-making in the hippocampus and prefrontal cortex. Elife 10.

Urban KR, Layfield DM, Griffin AL (2014) Transient inactivation of the medial prefrontal cortex impairs performance on a working memory-dependent conditional discrimination task. Behavioral Neuroscience 128:639–643.

JNeurosci Accepted Manuscript

818 Uylings HBM, Groenewegen HJ, Kolb B (2003) Do rats have a prefrontal cortex? Behavioural
819 Brain Research 146:3–17.

820 van Heukelum S, Mars RB, Guthrie M, Buitelaar JK, Beckmann CF, Tiesinga PHE, Vogt BA,
821 Glennon JC, Havenith MN (2020) Where is Cingulate Cortex? A Cross-Species View.
822 Trends Neurosci 43:285–299.

823 Vertes RP (2006) Interactions among the medial prefrontal cortex, hippocampus and midline
824 thalamus in emotional and cognitive processing in the rat. Neuroscience 142:1–20.

825 Watabe-Uchida M, Eshel N, Uchida N (2017) Neural Circuitry of Reward Prediction Error. Annu
826 Rev Neurosci 40:373–394.

827 Witten IB, Steinberg EE, Lee SY, Davidson TJ, Zalocusky KA, Brodsky M, Yizhar O, Cho SL,
828 Gong S, Ramakrishnan C, Stuber GD, Tye KM, Janak PH, Deisseroth K (2011)
829 Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-
830 mediated reinforcement. Neuron 72:721–733.

831 Wood ER, Dudchenko PA, Robitsek RJ, Eichenbaum H (2000) Hippocampal neurons encode
832 information about different types of memory episodes occurring in the same location.
833 Neuron 27:623–633.

834 Yoon T, Okada J, Jung MW, Kim JJ (2008) Prefrontal cortex and hippocampus subserve
835 different components of working memory in rats. Learning & Memory 15:97–105.

836 Zielinski MC, Shin JD, Jadhav SP (2019) Coherent Coding of Spatial Position Mediated by
837 Theta Oscillations in the Hippocampus and Prefrontal Cortex. Journal of Neuroscience
838 39:4550–4565.

839

JNeurosci Accepted Manuscript

Fig 1: Behavior and Recording Paradigm.

A. W-track rule-switching behavior. Top: sequence of rule 1, left-center-right-center; bottom: sequence of rule 2, center-left-right-left. Stars denote the home arm in the given rule sequence. Trajectories in shaded boxes are outbound and require working memory. The common trajectories across the two rules are Left-to-Center and Center-to-Left, with different memory demands for the rules.

B. Example performance curve of one animal (estimated mean ± 95% CI). *Left*: Rule 1 learning across behavior epochs/ sessions; *Middle*: Rule 2 learning; *Right*: rapid switching between the two rules within a single day. Note that the rule switch is not indicated by any external cue. The bar on top indicates the current rule (dark red for rule 1, teal for rule 2). Dashed lines in all three panels indicate epoch changes.

C. Average reward rate of behavioral choices based on each rule's reward contingency, for Rule 2 to 1 and Rule 1 to 2 switches separately (n=8 each), aligned to rule switch trials.

D. Probability of perseverative error aligned to rule switch trials, for Rule 2 to 1 and Rule 1 to 2 respectively.

E. Recording setup. Simultaneous recording in dCA1, mPFC and VTA regions during rule switching behavior, with photo-tagging of TH+ neurons in the VTA in TH-Cre animals.

F. Example of Nissl-stained mPFC and CA1 slices. Rectangles indicate regions of lesion marks.

G. Histology in VTA. Red: virus expression of AAV5-EF1a-DIO-hChR2(E123T/T159C)-mCherry; green: antibody staining of TH+ cells; blue: DAPI. White triangles show overlap between virus expression and antibody staining.

JNeurosci Accepted Manuscript

H. VTA spiking responses to laser stimulation at 0 ms. Top: raster plot showing individual spikes aligned to each stimulation onset. Bottom: probability of spiking. Left: example of an opto-tagged neuron (p=0); right: example non-tagged VTA neuron (p=0.30, Stimulus-Associated spike Latency Test, SALT).

I. Top: spike raster during an example trial starting with the animal leaving the last reward location, running a trajectory to the next reward, followed by immobility and consumption at new reward location. Green: CA1 cells; orange: mPFC; blue: VTA. Middle: linear distance to reference reward location. Bottom: movement speed. Dashed line: time of first nose poke at the destination reward location.

Fig 2: Remapping of single cells in CA1 and mPFC across rules.

A. Illustrative examples of spatial representations in CA1 and mPFC showing remapping on the common trajectories performed across the two rules, Left-to-Center or Center-to-Left. (i). 2D place field at stable performance of rule 1 (*top*) and rule 2 (*bottom*) in interleaved blocks (first block for each rule showed on Left and second block shown on Right, numbers in circles denote order of blocks), with peak firing rate on top. Numbers in circles denote sequence of rule block; (ii). linearized firing fields of the cells shown above (mean ± standard error).

B. Illustrative examples showing firing rate change of neurons at the same well locations across the two rules, aligned to reward onset for the same common trajectory across rules (Left-to-Center or Center-to-Left).

C. Correlation of single cell firing patterns during individual trials to average firing patterns during stable performance of a given rule (> 60^{th} percentile of the performance of the

JNeurosci Accepted Manuscript

day) during rule switching blocks. Note the change in correlations that occur after rule switch.

D. Proportion of remapped cells for each region. During running: CA1 27.3% (39/143, 32 Pyr, 7 Int), mPFC 25.8% (33/128, 26 Pyr, 7 Int), χ^2=0.077, p=0.78; At reward location: CA1 14.7% (21/143, 13 Pyr, 8 Int), mPFC 14.8% (19/128, 18 Pyr, 1 Int), χ^2=0.0013, p=0.97, Fisher's exact test for proportions of remapped cells between regions.

E. Distribution of spatial information encoded by individual cells for each brain region. CA1: median 2.3 bits/sec, n=143; mPFC: median 1.3 bits/sec, n=128; VTA: median 0.23 bits/sec, n=53. CA1-mPFC p=0.0010, mPFC-VTA p=6.2×10^{-8}, CA1-VTA p=7.7×10^{-17}, Kruskal-Wallis test with multi-comparison correction.

Fig 3: mPFC and CA1 ensembles encode rule representations and transitions

A. Correlation of population activity vectors aligned to rule switch (dashed black line) in a trial wise manner for CA1 and mPFC. Note the weakened stability shown by lower correlations following rule switches in both regions. Areas of stable correlations are denoted by grey squares.

B. CA1 and mPFC population activity show differential representation of the same trajectory across rules. Populations activity space shown by dimensionality reduction using CEBRA (see **Methods**). A particular 3D view is illustrated, with arbitrary units for axes. (i) CA1 embeddings of the example common trajectory across rules (Left to Center, schematic on bottom right). *Top*: 3D embedding from models trained using individual animal's data; bottom: model trained with multi-session CEBRA with data from multiple animals to capture common underlying structure across animals. Start and End

JNeurosci Accepted Manuscript

denote beginning and end of the same trajectory across the two rules. (ii) Embeddings of the same trajectory but for mPFC.

C. An illustrative example showing that the mPFC population representation of the common trajectory shifts in the manifold space (shift denoted by arrow) during rule transition over the course of trials. Trajectory representations are grouped in 10 trials per color.

D. Decoding probability of each rule from an example animal shows CA1 (*top*) and mPFC (*bottom*) neural activity transitions from representing one rule to the other (Rule 1 in dark red, Rule 2 in teal) over a few to tens of trials after each rule switching (rule switching is denoted by dashed black vertical lines). Dashed colored curves show estimated behavior strategy similarity to the optimized behavior for each rule.

E. CA1 and mPFC rule decoding probabilities are highly correlated with behavior strategy. Mean Pearson r: between CA1 and behavior 0.64±0.02, between mPFC and behavior 0.70±0.02. Dots and dashed lines show correlation between behavior and each decoding estimate (n=20, 4 animals, 5 estimates each, see **Methods**). p values for all behavior-rule decoding correlations are less than 0.001. Behavior correlation with mPFC rule decoding probabilities was higher than with CA1 decoding (p=0.0047, paired t test).

F. Rule decoding accuracy (proportion of trials decoded correctly) in each region, grouped by behavioral performance. *Top*: decoding accuracy of single-animal models (low performance: CA1, 0.38, mPFC, 0.41, p=0.33 for CA1 vs mPFC; mid performance: CA1, 0.66, mPFC, 0.71, p=0.12; high performance: CA1, 0.97, mPFC, 0.96, p=0.31, two-proportion z-test). *Bottom*: decoding accuracy using multi-animal models (low performance: CA1, 0.47, mPFC, 0.35, $p=1.3\times10^{-4}$; mid performance: CA1, 0.61, mPFC, 0.64, p=0.27; high performance: CA1, 0.79, mPFC, 0.93, $p=9.9\times10^{-12}$). Dots and dashed lines show decoding accuracy for each animal under each condition (n=4 animals).

JNeurosci Accepted Manuscript

G. Behavior strategy similarity (top), probabilities of mPFC rule decoding (middle) and CA1 rule decoding (bottom) aligned to rule switching trials, shown separately as Rule 2 to Rule 1 (left) and Rule 1 to Rule 2 (right) switches. Behavior transitions occurred 17 (median, IQR: 13-29) trials after the Rule1 -> Rule2 switch, with mPFC and CA1 neural representations slightly leading behavior. For the switch type of Rule 2 -> Rule 1, behavior started to reflect the new rule after 22 (median, IQR: 20.75-28.5) trials, with mPFC leading in contrast to CA1 lagging.

H. Correlation between the number of trials taken for behavior switch and neural representation switch, combined across rule switch types. Dashed lines indicate diagonal lines where the timing of neural transitions equals that of behavior switches. Left: mPFC and behavior, Pearson r=0.62, $p=3.4\times10^{-10}$; right: CA1 and behavior, Pearson r=0.46, $p=1.9\times10^{-5}$.

I. Distribution of timing differences between behavior switch and neural representation switch. Behavior relative to mPFC transitions: median +4.0 (IQR: -3.75-10.5) trials, n=80, p=0.0044; behavior relative to CA1 transitions: median +8.0 (IQR: 1-15) trials, n=78, p=0.0040, Wilcoxon signed rank test. For box plots, boxes, whiskers and circles indicate quartile, 1.5×IQR and outliers, respectively.

Fig 4: DA spiking activity mirrors reward prediction error.

A. Illustrative examples showing that DA neuronal firing activity at reward wells signals robust differences between rewarded and unrewarded outcomes. Left: example photo-tagged TH+ neurons; right: example non-DA neurons. Line: trial-averaged firing rate; shaded area: standard error.

JNeurosci Accepted Manuscript

B. Percentage of cell types and reward responses in VTA. RR: reward responsive at 200-1200ms after first nose pokes. DA RR 36% (19/53), DA non-RR 2% (1/53), non-DA RR 28% (15/53), non-DA non-RR 34% (18/53).

C. (i) Population response of putative DA neurons to reward outcomes during reward approach and at reward well. Bar on top indicates spatial/temporal bins with $p<0.05$ after Bonferroni correction using paired rank tests. Vertical black dotted line is first nose poke/ reward at the reward well, also corresponding to transition from spatial to temporal bins. Spatial bins during approach to reward well are from 1-40 (5-6 cm per bin, 200-240 cm in total), and temporal bins after arrival at reward well are from 41-70 (100 ms per bin, 3 seconds in total). (ii) average z-scored firing rates of all putative DA cells during rewarded vs unrewarded trials. Left: during run, starting from leaving the last reward/nose poke to the beginning of the current trial's reward/nose poke, rewarded median=0.11, unrewarded median = -0.060, $p = 1.8\times10^{-4}$; right: reward time (200ms-1200ms after the first nose poke), rewarded median=0.49, unrewarded median = -0.76, $p = 5.6\times10^{-5}$. Wilcoxon signed-rank test, n=20 cells.

D. Control for running speed during trial. No significant differences are seen in running speed between rewarded and unrewarded trials ($p = 0.12$).

E. Strong correlation of DA population firing rates during reward approach running and at reward outcome ($r = 0.60$, $p = 6.8\times10^{-148}$). Colors from light to dark indicate performance from low to high. Red: error trials; blue: correct trials.

F. DA firing rates show no difference between low- and high-performance period. Top: rewarded trials (median z-scored firing rate during running: low performance 0.065, high performance 0.13, p=0.23; during reward outcome: low performance 0.54, high performance 0.44, p=1.00); bottom: unrewarded trials. (run: low performance -0.072,

JNeurosci Accepted Manuscript

high performance -0.049, p=0.23; outcome: low performance -0.81, high performance -0.71, p=0.65, Wilcoxon paired rank test, n=20 cells).

G. Difference in DA firing rates for rewarded vs unrewarded trials during running is present for both inbound trials (median difference 0.085, $p = 1.8\times10^{-6}$, n = 178 trials) and outbound trials (median difference 0.12, $p = 1.0\times10^{-13}$, n = 217 trials).

H. Multiple regression for DA neuron firing rates during run or reward outcome using the following parameters: current-trial reward, reward rate of the last 5 trials, reward rate the same trajectory for the last 5 visits. *Top*: mean significant coefficients. Current-trial reward: run: $\beta = 0.14\pm0.02$ (mean±sem), $p=2.4\times10^{-4}$; outcome: $\beta = 0.95\pm0.13$, $p=1.6\times10^{-4}$; reward rate: run: $\beta = 0.11\pm0.10$, p=0.36; outcome: $\beta = -0.13\pm0.16$, p=0.56; same-trajectory reward rate: run: $\beta = 0.074\pm0.058$, p=0.16; outcome: $\beta = -0.39\pm0.07$, $p=2.4\times10^{-4}$, Wilcoxon signed rank test. *Bottom*: Proportions of DA neurons modulated by each factor. Run: current reward 65% vs reward rate 45%: $\chi^2=1.6$, p=0.20; current reward 65% vs same-trajectory reward rate 30%: $\chi^2=4.9$, p=0.027; reward rate 45% vs same-trajectory reward rate 30%: $\chi^2=0.96$, p=0.33. Outcome: current reward 95% vs reward rate 30%: $\chi^2=18.0$, $p=2.2\times10^{-5}$; current reward 95% vs same-trajectory reward rate 65%: $\chi^2=5.6$, p=0.018; reward rate 30% vs same-trajectory reward rate 65%: $\chi^2=4.9$, p=0.027. Fisher's exact test, n=20 cells.

Fig 5: DA spiking activity relationship with neural manifold change and rule switching behavior.

A. Behavior strategy (i), mPFC (ii) and CA1 (iii) decoding probability changes aligned to rule switching trials. (iv-v). Difference in DA neuron firing rates for rewarded vs unrewarded trials at well locations (iv) and during running (v), plotted as a function of

JNeurosci Accepted Manuscript

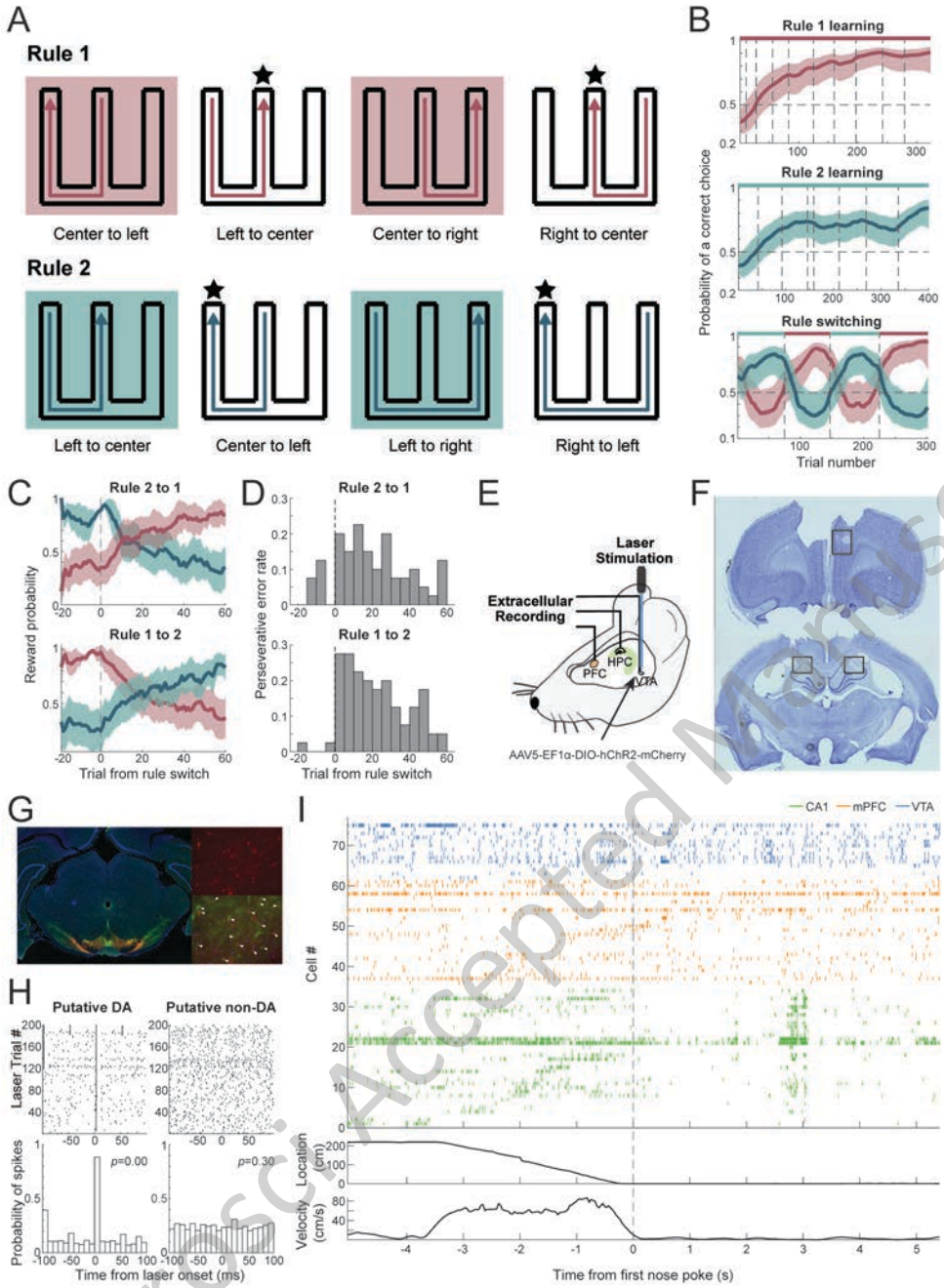
trials aligned to beginning of rule switch. Black bars indicate significant difference between rewarded and unrewarded trials with $p<0.05$ after Bonferroni correction. Vertical black dotted line across sub-panels denotes behavioral strategy switch trial.

B. Behavior strategy, CA1/mPFC decoding probability changes and DA firing rate differences between rewarded and unrewarded trials, aligned to decoded rule representation switch in mPFC (vertical orange dotted line). Top: Behavior strategy similarity to optimized behavior of current and previous rules. Animals' behavior adapted to the new rule on average 4 (IQR: -3.75-10.5) trials after mPFC showed correct rule decoding. Middle: CA1 and mPFC decoding accuracy of current rule. Bottom: DA firing rate difference for rewarded vs unrewarded trials during running. The Black bars above indicate trials showing significant difference with $p<0.05$ after Bonferroni correction.

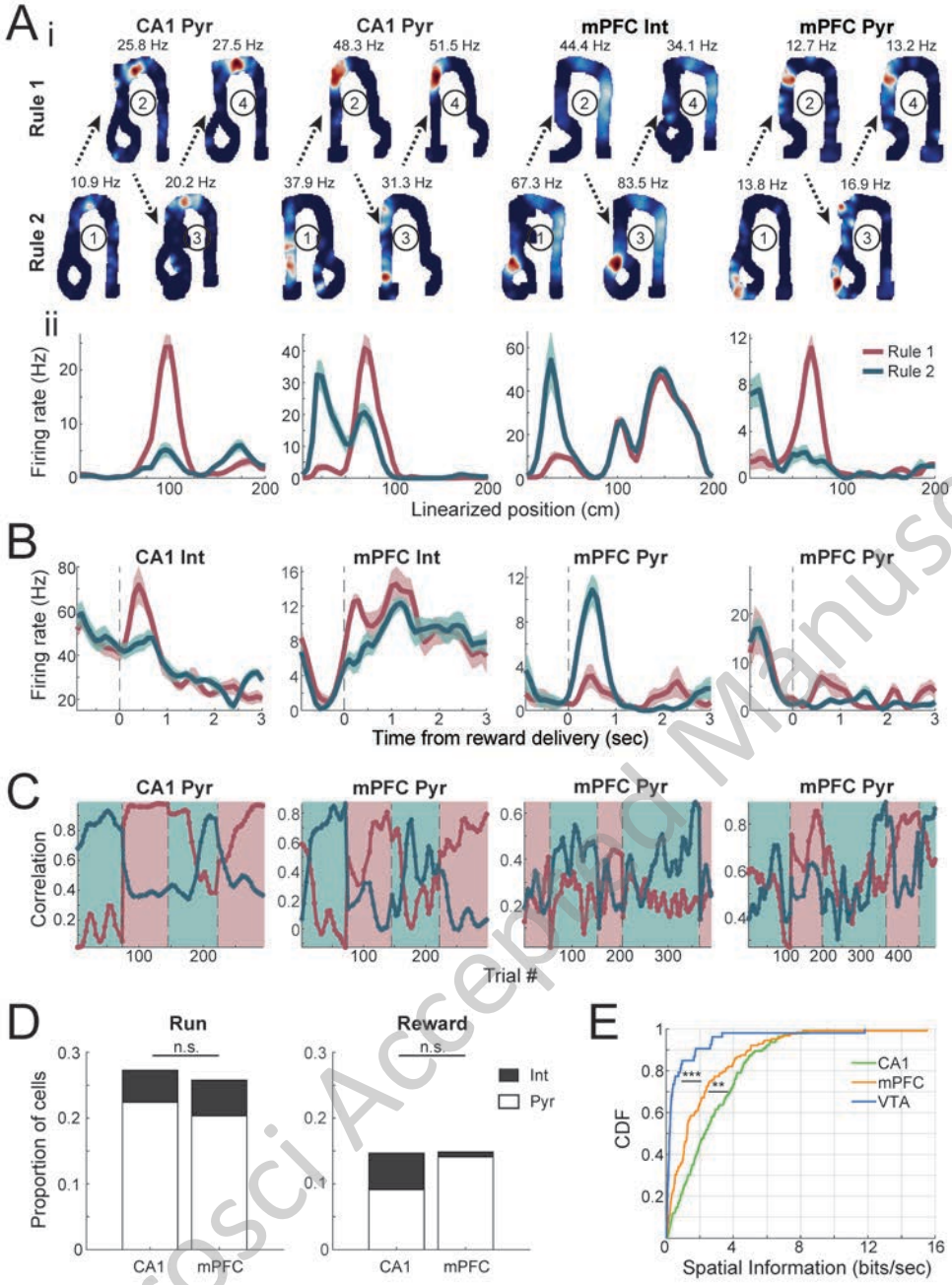
C. Same as B but aligned to CA1 rule representation switch (vertical green dotted line).

D. DA firing activity difference between rewarded and unrewarded trials is correlated with changing rate of mPFC manifold (Pearson correlation $r = 0.16$, $p = 2.8\times10^{-9}$, $n=1361$ trials).

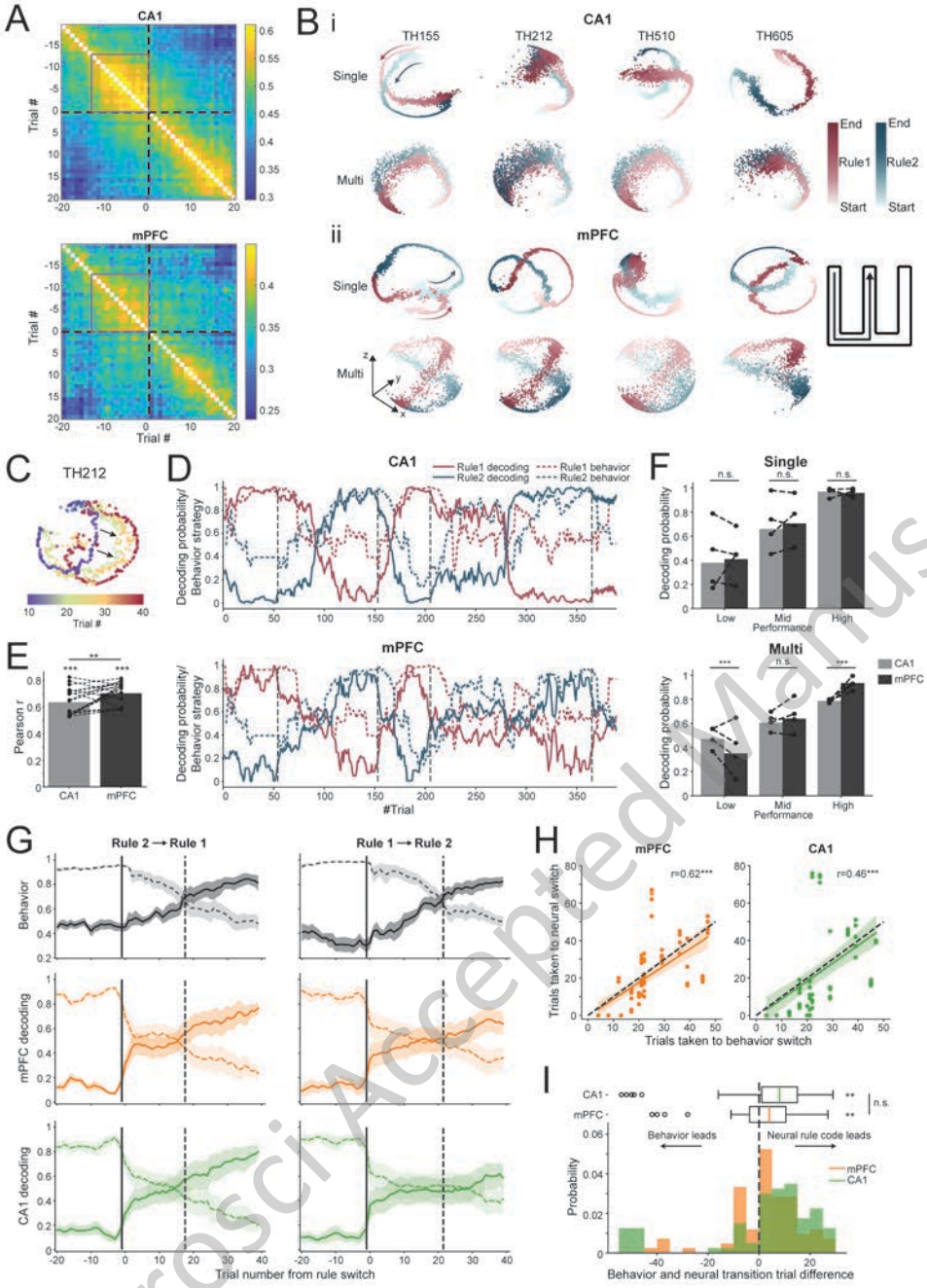
JNeurosci Accepted Manuscript



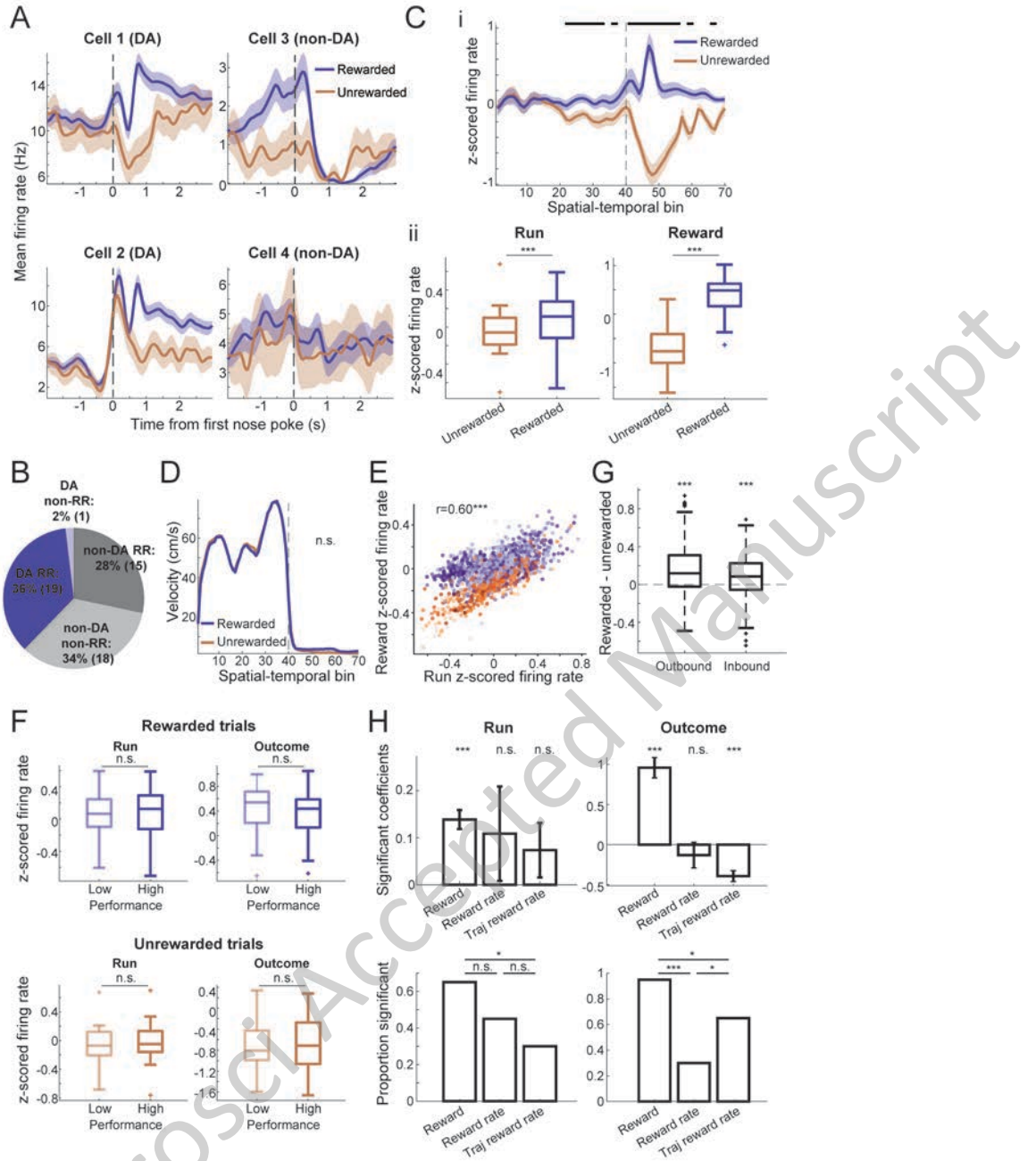
JNeurosci Accepted Manuscript



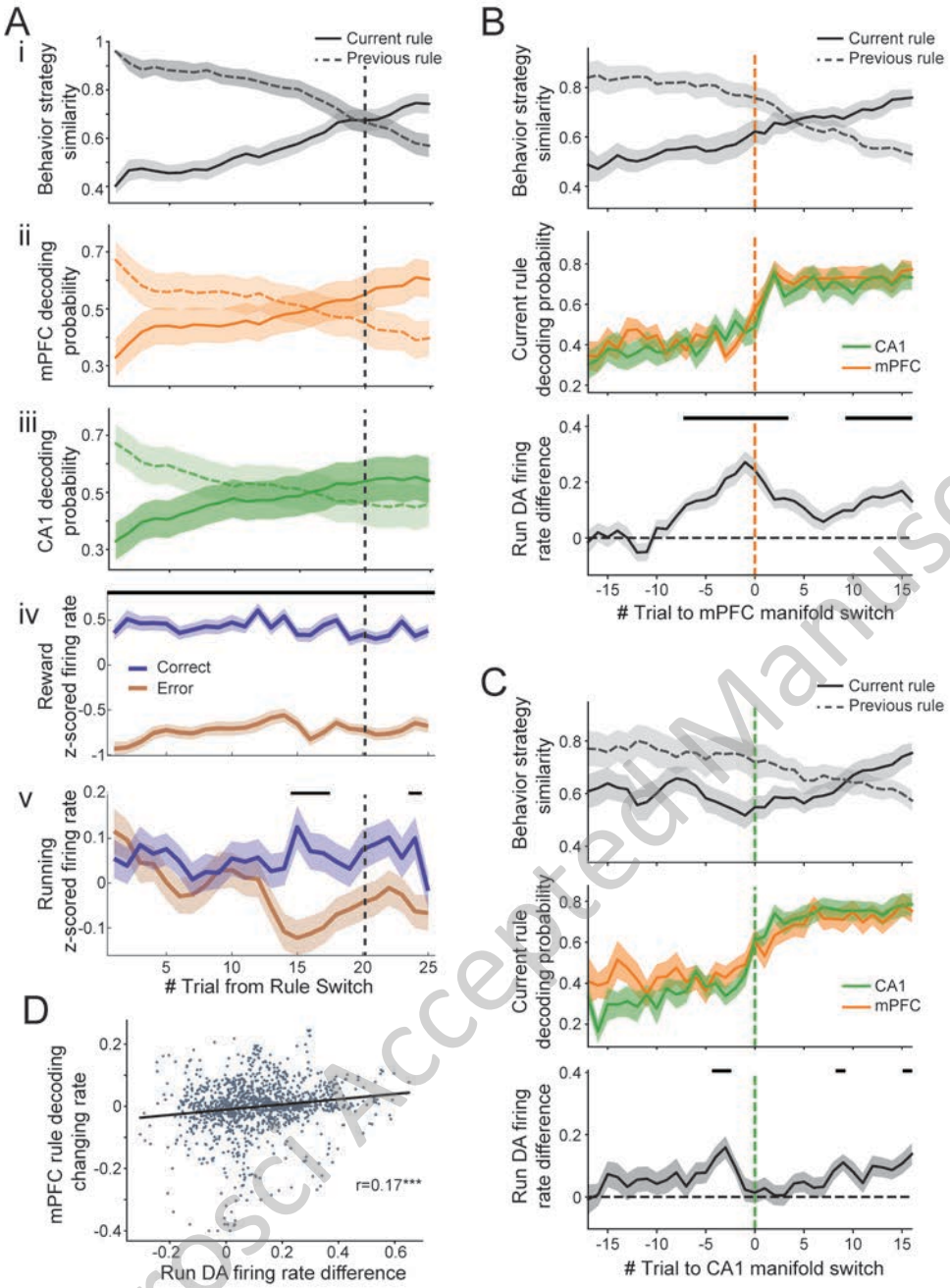
JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript