bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

How do dopamine neurons resolve a tradeoff between performance and energy?

Minryung R. Song[1], Sang Wan Lee[1*]

[1]Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

[*]Corresponding author:

E-mail: sangwan@kaist.ac.kr (SWL)

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

How do dopamine neurons resolve a tradeoff between performance and energy?

Abstract

The reinforcement learning (RL) theory accounts for the two types of dopamine activity patterns: phasic and ramping. However, we still lack an understanding of when and why dopamine activity transitions between the two modes. This study examines a theoretical possibility that dopamine neurons ramp up, though it is a physiologically demanding job, to distribute cognitive resources in both space and time, while they transition to an energy-efficient phasic firing mode to concentrate cognitive resources on learning reward-predicting cues. These results lend better insight into how animals find tradeoffs between task performance maximization and resource consumption minimization.

Introduction

Midbrain dopamine neurons have been found to be important in RL. The dopamine neurons phasically respond to rewards and reward-predicting cues, the size of which reflects the reward prediction error (RPE) [1–6]. Phasic dopamine activity has been found to drive RL [7–10], and dopamine in the striatum modulates corticostriatal synaptic plasticity [11–13].

Based on these findings, computational efforts have been made to account for the role of dopamine in RL. The models have often considered dopamine activity as the RPE and well-simulated phasic dopamine responses to experimentally inserted cues and rewards [1,14–16]. RL models have also provided an explanation for the gradually increasing dopamine activity as animals approach a reward by considering internal spatial representation, the temporal decay of dopamine-dependent synaptic potentiation, the uncertainty of action timing or discounted vigor or by assuming dopamine as a value signal [17–20].

However, the patterns of dopamine activity may change, rather than being fixed to phasic or ramping patterns. A recent study observed that dopamine activity transitions from ramping to phasic as learning proceeds and transitions from phasic to ramping when reward values change

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[21]. No RL model has accounted for in which condition dopamine activity transitions between ramping and phasic. We hypothesize that the dopamine transitions might support the re-allocation of cognitive resources during RL.

Dopamine modulates resource allocation

Efficient resource allocation is as important as reward maximization for the survival of animals. Resource allocation is tightly linked with locomotion, attention and learning [22]. To use their biological resources efficiently, animals should quickly learn which stimuli are highly informative regarding rewards and be particularly responsive to more informative stimuli. Conversely, to quickly learn and respond to key stimuli, animals should allocate their resources to reward-informative stimuli. Animals approach, learn, exploit, explore and pay attention to (potentially) important stimuli and changes that occur to these stimuli.

A potential neural substrate for resource allocation is the midbrain dopamine neurons. Previous studies have found that a bilateral dopamine lesion reduces the possibility of high-effort, high-reward choices, while increasing the possibility of low-effort, low-reward choices [23,24]. In the striatum, dopamine inhibition decreases the initiation and maintenance of instrumental response, while dopamine excitation raises the likelihood of movement initiation and accelerates locomotion and biases action selection [25–29]. Enhanced dopamine levels have been found to increase bias toward the exploration of novel choices as opposed to the exploitation of learned options [30–32]. Dopamine also drives the learning of reward-informative stimuli [33–35]. Dopamine neurons are necessary for the attention signal in the amygdala, which accelerates RL [36,37]. Prefrontal dopamine is involved in working memory maintenance and has been suggested to reflect cognitive effort [38,39]. These findings suggest that, in its target regions, dopamine guides which stimuli the agent learns to respond to in the environment.

This case raises the theoretical possibility that midbrain dopamine may also be sensitive to changes in effective task dimensionality, which refers to the number of environmental stimuli or features associated with a task [40–42]. In addition to rewards and reward-predicting cues, previous studies have found phasic dopamine excitation to novel stimuli [35,43]. Dopamine

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neurons are sensitive to changes in reward features or values [44,45]. Dopamine also responds to stimuli that are weakly or potentially related to rewards [34,46–49]. Regardless of which stimulus evoked phasic dopamine activity, phasic dopamine activity would promote learning and locomotion in the striatum and elsewhere. Taken together, these observations suggest the possibility that one of the fundamental features of dopamine is to manage cognitive and motor resources efficiently so as to detect changes in effective task dimensionality, to learn which stimuli are more informative of rewards and to initiate a behavioral response to learned stimuli.

Ramping and phasic dopamine depends on effective task dimensionality

Animals do not perceive every stimulus in an environment, but they tend to recognize stimuli with saliency that exceed a certain threshold level (Fig 1A). Although experimentally inserted cues (e.g. tones, light) are usually more salient than others (e.g. wells, floor), pseudo-conditioning or generalization indicates that the latter are also informative of the reward [47–49]. Previous studies have suggested that the more salient a stimulus is, the more readily it should be learned [40,50,51].

To investigate the influence of effective task dimensionality on dopamine activity during RL, we considered a situation in which both weakly informative stimuli and highly informative, salient stimuli (key stimuli in Fig 1B) were present. We ran a simulation with a temporal difference (TD) learning model with an eligibility trace, the standard version known to well account for dopaminergic activity during RL [14,52]. Frequent exposure to stimuli in conditions of highly effective task dimensionality may shorten the effective time window for learning during which previously experienced stimuli affect task performance. An eligibility trace is useful to implement such an effect. Similarly to the Pearce-Hall model, we made more salient cues to be learned faster than less salient cues [50,53].

When effective task dimensionality is high and the distribution of stimulus value is widely dispersed (Fig 1C), the learning agent finds difficulty screening candidates for a key stimuli. In such cases, it broadly distributes cognitive resources. Our model simulation suggested that a ramping RPE appears in this situation. As learning proceeds, the number of candidates for key stimuli decreases (i.e., decreasing effective task dimensionality), and the value of key stimuli

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

increases, which favors a focused allocation of resources to the key stimuli. Our model predicted that phasic RPE occurs in this condition (Fig 1D).

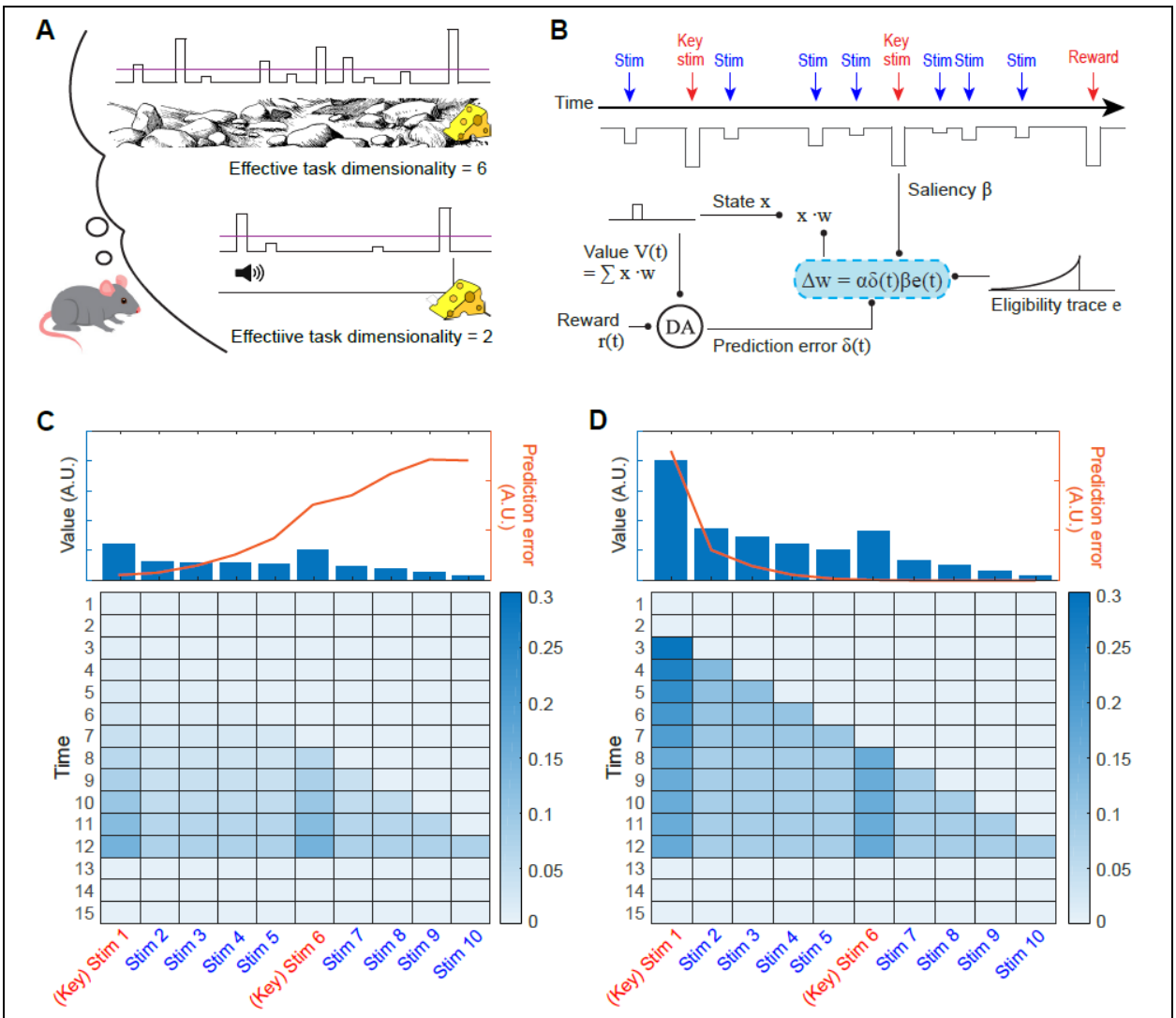

Fig 1. Dopamine modulates resource allocation depending on task dimensionality

(A) Conceptual example of high and low effective task dimensionality conditions. (B) Model structure; value distribution in conditions of high (C) and low (D) effective task dimensionality. The bottom plot shows the weights of each stimulus over time. The top plot shows the prediction error and the weight of each stimulus summed across time.

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To further examine this hypothesis, we tested whether the model replicates previous empirical findings. First, we tested whether the model reproduced dopamine activity in a low-effective task dimensionality experiment (Fig 2A-D). In this experiment, rats were placed in a simple, small chamber [14]. Two consecutive tone cues deterministically predicted a liquid reward for which animals only had to lick a spout. During early training, dopamine neurons showed strong phasic excitation to the reward, whereas the strong dopamine excitation was transferred to the initial experimental cue during late training. Regardless of the learning stage, the omission of the second experimental cue resulted in a larger phasic response to the reward. All the dopamine activity patterns were successfully simulated by the model.

Next, we tested whether the model reproduced the ramping dopamine activity observed in Howe et al. (2013) (Fig 2E, F) [54]. In this experiment, rats had to travel through a large T-maze to earn a reward. The first and second tone cues indicated the start of each trial and which arm to visit to receive the reward, respectively. This experiment is more complicated than the previous one. For example, background stimuli—such as decreasing the distance between the animal and the corner of the T-maze—may provide subsidiary information to guide the animal's behavior in a large maze, thereby increasing the effective task dimensionality. To implement a high effective task dimensionality, we increased the decay rate for the eligibility trace. The model successfully replicated the findings regarding ramping dopamine (Fig 2G).

According to our simulation, it appears that both the non-experimental cues increasing the effective task dimensionality and the fast decay of the eligibility trace are necessary for the ramping RPE. Although the rapid decay of the eligibility trace alone contributes to a ramping trend of RPE to some extent, the RPE inevitably peaked at the intermediate experimental cue (Fig 2H), which contradicts earlier findings that dopamine activity ramps up during the early stages of learning. Non-experimental cues were required for smoothly ramping dopamine activity because these factors cause an increase in the effective task dimensionality and a dispersion of cognitive resources. The magnitude of the RPE protrusion at the intermediate cue was dependent on the saliency contrast between experimental and non-experimental cues (Fig 2I).

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

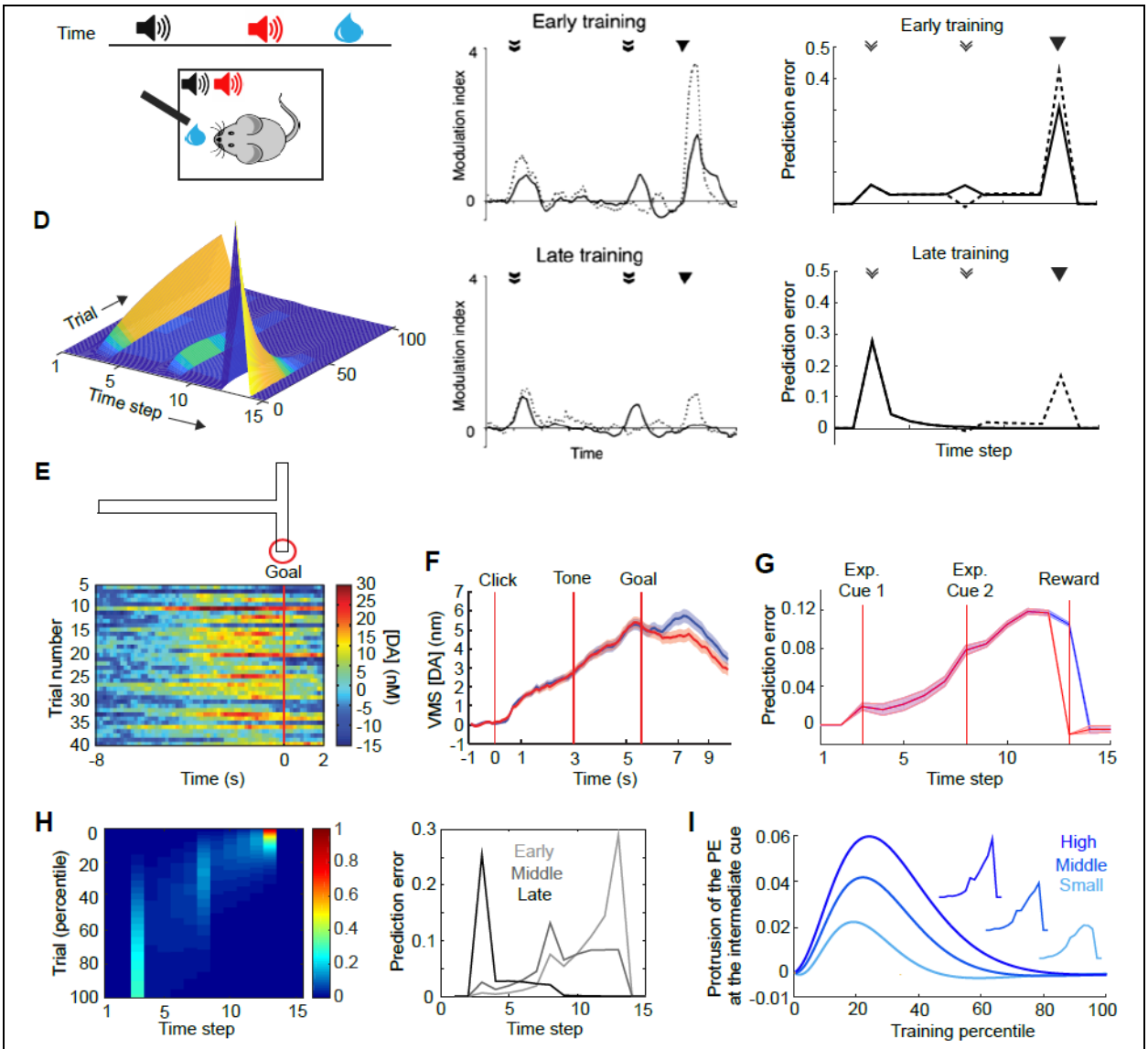

Fig 2. Phasic and ramping dopamine in conditions of low and high effective task dimensionality.

(A) The experimental paradigm of Pan et al. (2005). (B) The empirical results from Pan et al. (2005) during early and late training. (C) The simulation results of the model during early and late training. The agent was assumed to have completed training when the values of the first experimental cue converged. To clearly demonstrate how the shape of the RPE signal changed during the learning, the 10th, 20th and 80th percentiles of training were used to simulate early, middle and late training stages. λ, α, and γ were 0.9, 0.005 and 0.9, respectively. The β of the experimental stimuli were 2, and the β of other stimuli were 1. (D) The simulation results

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the model throughout learning. (E-F) The experimental results of Howe et al. (2013). (G) The model behavior during early training. All other values were the same as (C). (H) The behavior of a model that does not have non-experimental stimuli. The RPE signal of the model throughout learning (left). The RPE signal of the model during early, middle and late training (right). λ of 0.5 was used. (I) As the saliency contrast between experimental and non-experimental cues increased, the RPE at the intermediate experimental cue protruded accordingly. The level of the protrusion of the RPE at the intermediate experimental cue was measured as the RPE at the intermediate experimental cue minus the average of the RPE immediately before and after the experimental cue. The inset shows the RPE trajectory during early training.

Transition from ramping to phasic as RL focuses resources on a few important cues

The model predicted that the RPE peaks at reward onset in the early stage of learning, but as learning proceeds, the peak time point gradually transitions to the onset of an initial cue (Fig 3 B, D). This simulation result is consistent with previous experimental result [21]. In the study, rats were trained to press two different levers to collect a reward, and this training continued for a few days after their performance reached an asymptote. The authors observed that the ramping dopamine activity appeared during early training but that dopamine activity peaked at the first cue after extended training (Fig 3A, C).

Unpublished data also support our hypothesis. In this experiment, rats freely choose one of the two arms of a modified T-maze (Fig 3E inset). The reward probability of one arm was higher than that of the other, which remained constant within a block of 35-45 trials. The reward probability values were reversed across four blocks without any sensory cues indicating this change. Fig 3E shows dopamine activity around the reward onset. As the learning proceeded, the dopamine activity before the reward onset decreased, while phasic excitation to the reward diminished and moved to the earlier time points. This finding is fully consistent with our simulation results (Fig 3B, D).

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

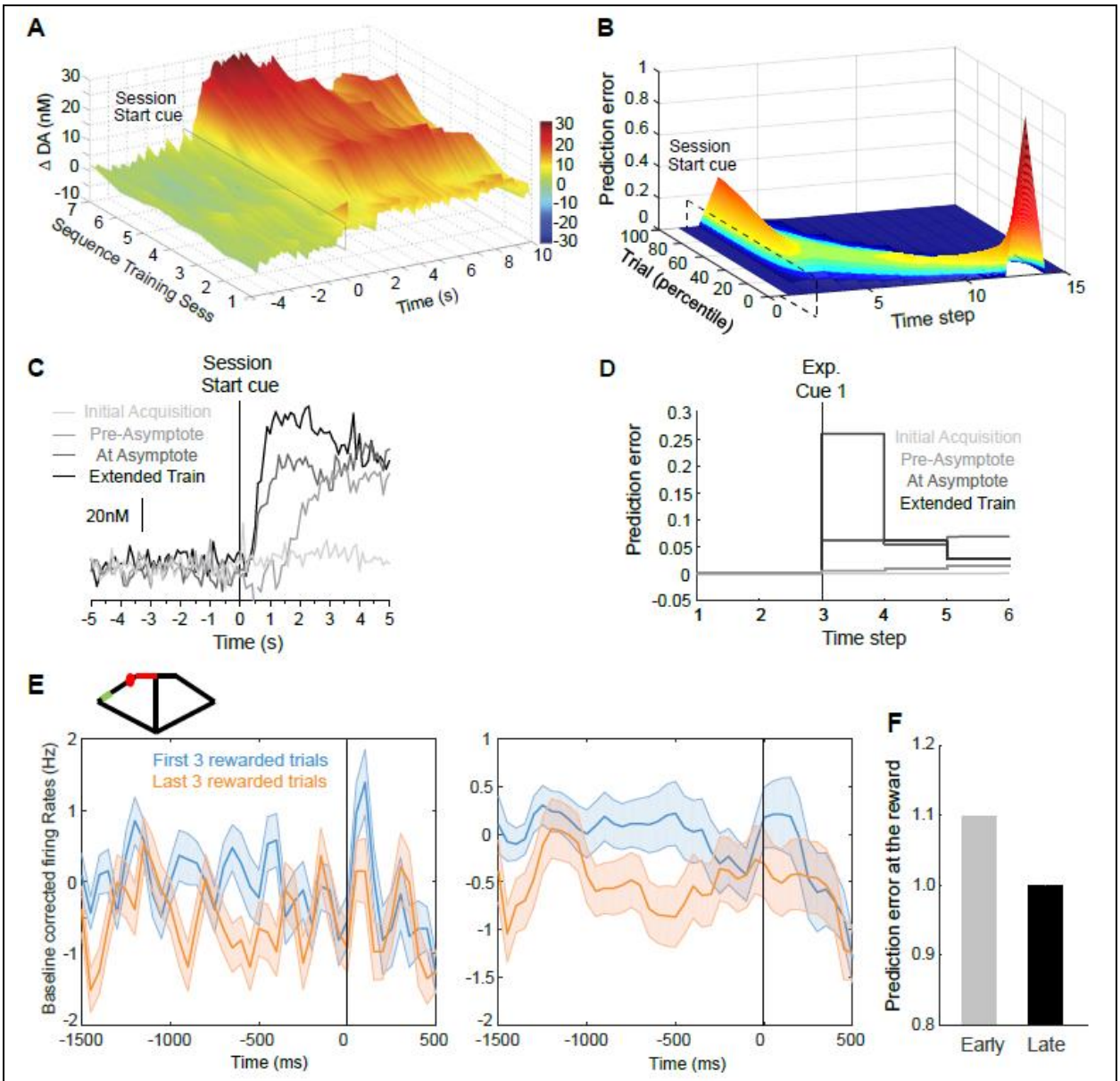

Fig 3. Dopamine transitions from ramping to phasic as RL decreases the effective task dimensionality.

(A) and (C) show the experimental results of Collins et al. (2016). (B) and (D) show the simulation results. The 1st, 10th, 30th and 80th percentiles were considered the initial acquisition, pre-asymptote, at asymptote and extended training, respectively. The 30th percentile was chosen as the asymptote because the PE of the model peaks at the first cue around the 30th percentile. The parameter values were the same as in Fig 2G. (E) Dopamine

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

activity at the beginning and at the end of each block (4 rats, 62 dopamine cells). The inset shows the experimental apparatus. Firing rates in the red region of the inset were normalized by subtracting the mean firing rate in the green region. The right plot shows smoothened PSTHs on the left plot. (F) The model prediction of the RPE at reward delivery when the size of the reward was doubled during early or late training.

Our simulation results support the view that a longer training time is required until dopamine activity transitions from ramping to phasic when effective task dimensionality is relatively high. Since the ramping RPE appeared during early training, around five times more trials were needed until the RPE peaked at the initial key stimulus (Fig 3B). The reason this transition was not observed in Howe et al. (2013) [54] might be that the animals were not trained enough for their performance to reach an asymptote (see their Figure 4e) unlike Collins et al. (2016) [21]. Even in an experiment much simpler than that of Howe et al. (2013), hundreds of trials were required until dopamine excitation to the reward were suppressed [14].

As learning proceeded, RPEs around the reward onset diminished, which causes a weaker sensitivity to the change in the reward value (Fig 3F). This suggests that extended training eliciting a dopamine transition from ramping to phasic makes learned responses habitual and less susceptible to environmental changes [55]. Switching to a phasic activity pattern during RL would allow the learning agent to focus cognitive and motor resources on key stimuli at the expense of reduced behavioral flexibility.

Transition from phasic to ramping upon changes in reward value

Learning and underlying behavior are motivated by reward value. A change in the reward earned by an animal's behavioral policy motivates it to re-learn which stimuli predict reward, and this process inevitably requires a re-allocation of resources. Our hypothesis, which concerns the role of ramping and phasic dopamine activity in resource allocation during RL, predicts that ramping activity will reappear upon changes in reward values. Both previous experimental results and our model simulation support this view. Collins et al. (2016) found that ramping activity reappears after doubling the reward value (Fig 4A), which was replicated in our model (Fig 4B) [21].

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

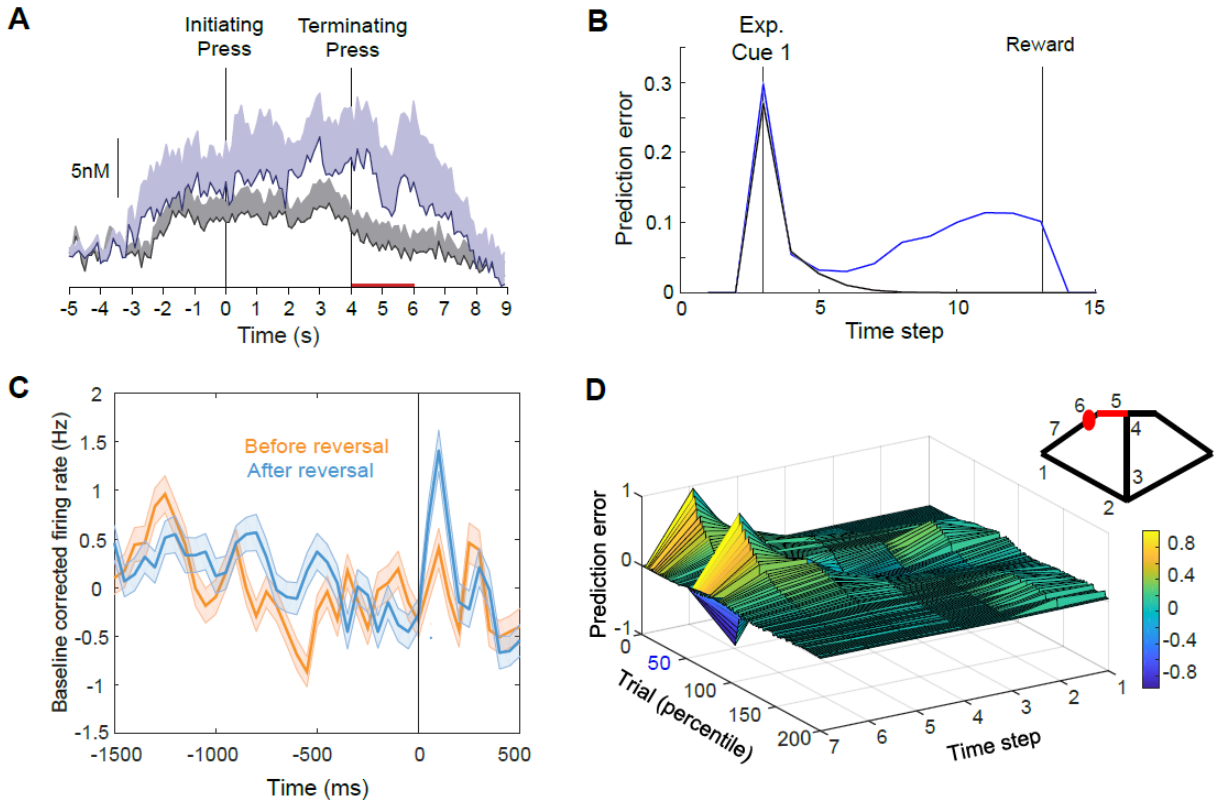

Fig 4. Dopamine transitions from phasic to ramping upon value changes.
(A) The experimental results of Collins et al. (2016). (B) The model behavior. After extended training, the ramping activity disappeared (gray). The ramping activity reappeared when the size of the reward was doubled (blue). (C) Dopamine activity before and after reversal (4 rats, 62 dopamine cells). (D) Simulation results of model behavior. The reward was delivered to the left arm until the 50th trial and to the right arm from the 51st trial onward. α , γ, and the temperature of the softmax function were 0.1, 0.9 and 2, respectively.

A similar effect was found in the unpublished data (Fig 4C). At the end of each block, the phasic excitation to reward disappeared, and dopamine activity began to peak at earlier time steps. After the reward probabilities of the left and right arms were reversed, however, the phasic excitation to reward reappeared, and dopamine activity before the reward onset increased. Because no experimental stimulus was used in this experiment, we used a basic Q-learning model for simulation. The simulation indicates that ramping RPE gradually decayed in the first block but re-

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

appeared after the reversal (Fig 4D). Overall, we showed that transitions from phasic to ramping activity reflect widespread resource distribution that allows for rapid adaptation to potential changes in the stimulus value.

Conclusions

In this paper, we tested the idea that dopamine transitions between ramping and phasic activities during RL reflect efficient resource allocation while the agent is learning to maximize reward. Both the simulation and experimental results suggest that dopamine activity transitions from ramping to phasic as the RL agent narrows down the candidate stimuli of the task to decrease the effective task dimensionality. The opposite occurred when the agent had to re-learn the candidate stimuli by increasing the effective task dimensionality. This affords insight into a more fundamental question: How does the animal resolve the tradeoff between prediction performance and efficient resource management?

Although efficient resource management is as important as reward maximization for animals' survival, efficient resource management has received relatively little attention. Recent empirical studies have found an increasingly diverse repertoire of dopamine activity, which has presented questions railing against the presumed role of dopamine [22,42,54,56–58]. Our hypothesis examining the role of dopamine in resource allocation explains why dopamine responds to novel and intense stimuli that do not encode the RPE [6,43]. RL has been found to be very inefficient in reducing task dimensionality [59]. Further work on the role of dopamine in resource allocation may shed light on how dopamine arbitrates RL and dimensionality reduction.

Acknowledgment

We thank Dr. Min Whan Jung and Dr. Sue-Hee Huh for their generosity in allowing us to use their unpublished data for this paper. We also thank Jung Hwan Shin for pre-processing the unpublished data.

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Schultz W, Dayan P, Montague R. A Neural Substrate of Prediction and Reward. Science. 1997;275: 1593–1599. doi:10.1126/science.275.5306.1593
2. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005;47: 129–41. doi:10.1016/j.neuron.2005.05.020
3. Watabe-Uchida M, Eshel N, Uchida N. Neural Circuitry of Reward Prediction Error. Annu Rev Neurosci. 2017;40: 373–394. doi:10.1146/annurev-neuro-072116-031109
4. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature. Nature Publishing Group; 2012;482: 85–88. doi:10.1038/nature10754
5. Hart AS, Rutledge RB, Glimcher PW, Phillips PEM. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. J Neurosci. 2014;34: 698–704. doi:10.1523/JNEUROSCI.2489-13.2014
6. Schultz W. Dopamine reward prediction-error signalling: a two-component response. Nat Rev Neurosci. 2016; doi:10.1038/nrn.2015.26
7. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci. Nature Publishing Group; 2013;16: 966–973. doi:10.1038/nn.3413
8. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, et al. Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat Neurosci. 2017;20: 735–742. doi:10.1038/nn.4538
9. Frank M, Seeberger L, O'Reilly R. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science. 2004;306: 1940–3. doi:10.1126/science.1102941
10. Chang CY, Esber GR, Marrero-Garcia Y, Yau H-J, Bonci A, Schoenbaum G. Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat Neurosci. 2016;19: 111–6. doi:10.1038/nn.4191
11. Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. Science. 2008;321: 848–851. doi:10.1126/science.1160575
12. Yagishita S, Hayashi-Takagi A, Ellis-Davies GCR, Urakubo H, Ishii S, Kasai H. A critical time window for dopamine actions on the structural plasticity of dendritic spines. Science.

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2014;345: 1616–1620. doi:10.1126/science.1255514

13. Wickens JR. Synaptic plasticity in the basal ganglia. Behav Brain Res. 2009;199: 119–28. doi:10.1016/j.bbr.2008.10.030
14. Pan W-X, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci. 2005;25: 6235–42. doi:10.1523/JNEUROSCI.1478-05.2005
15. Pan W-X, Schmidt R, Wickens JR, Hyland BI. Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. J Neurosci. 2008;28: 9619–31. doi:10.1523/JNEUROSCI.0255-08.2008
16. Starkweather CK, Babayan BM, Uchida N, Gershman SJ. Dopamine reward prediction errors reflect hidden-state inference across time. Nat Neurosci. 2017;20: 581–589. doi:10.1038/nn.4520
17. Gershman SJ. Dopamine ramps are a consequence of reward prediction errors. Neural Comput. 2014;26: 467–71. doi:10.1162/NECO_a_00559
18. Morita K, Kato A. Striatal dopamine ramping may indicate flexible reinforcement learning with forgetting in the cortico-basal ganglia circuits. Front Neural Circuits. 2014;8: 36. doi:10.3389/fncir.2014.00036
19. Lloyd K, Dayan P. Tamping Ramping: Algorithmic, Implementational, and Computational Explanations of Phasic Dopamine Signals in the Accumbens. PLoS Comput Biol. 2015;11: 1–34. doi:10.1371/journal.pcbi.1004622
20. Kato A, Morita K. Forgetting in Reinforcement Learning Links Sustained Dopamine Signals to Motivation. PLoS Comput Biol. 2016;12: 1–41. doi:10.1371/journal.pcbi.1005145
21. Collins AL, Greenfield VY, Bye JK, Linker KE, Wang AS, Wassum KM. Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. Sci Rep. Nature Publishing Group; 2016;6: 1–15. doi:10.1038/srep20231
22. Berke JD. What does dopamine mean? Nat Neurosci. Springer US; 2018;21: 787–793. doi:10.1038/s41593-018-0152-y
23. Cousins MS, Atherton A, Turner L, Salamone JD. Nucleus accumbens dopamine depletions alter relative response allocation in a T- maze cost/ benefit task. Behav Brain Res. 1996;74:

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

189–197. doi:10.1016/0166-4328(95)00151-4

24. Le Heron C, Plant O, Manohar S, Ang YS, Jackson M, Lennox G, et al. Distinct effects of apathy and dopamine on effort-based decision-making in Parkinson's disease. Brain. 2018;141: 1455–1469. doi:10.1093/brain/awy110
25. Da Silva JA, Tecuapetla F, Paixão V, Costa RM. Dopamine neuron activity before action initiation gates and invigorates future movements. Nature. Nature Publishing Group; 2018;554: 244–248. doi:10.1038/nature25457
26. Howe MW, Dombeck DA. Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature. Nature Publishing Group; 2016;535: 505–510. doi:10.1038/nature18942
27. Kravitz A V, Tye LD, Kreitzer AC. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat Neurosci. 2012;15: 816–8. doi:10.1038/nn.3100
28. Howard CD, Li H, Geddes CE, Jin X. Dynamic Nigrostriatal Dopamine Biases Action Selection. Neuron. Elsevier Inc.; 2017;93: 1436–1450.e8. doi:10.1016/j.neuron.2017.02.029
29. du Hoffmann J, Nicola SM. Activation of Dopamine Receptors in the Nucleus Accumbens Promotes Sucrose-Reinforced Cued Approach Behavior. Front Behav Neurosci. 2016;10: 144. doi:10.3389/fnbeh.2016.00144
30. Kayser AS, Mitchell JM, Weinstein D, Frank MJ. Dopamine, locus of control, and the exploration-exploitation tradeoff. Neuropsychopharmacology. Nature Publishing Group; 2015;40: 454–462. doi:10.1038/npp.2014.193
31. Costa VD, Tran VL, Turchi J, Averbeck BB. Dopamine modulates novelty seeking behavior during decision making. Behav Neurosci. 2014;128: 556–566. doi:10.1037/a0037128
32. Beeler J, Daw N. Tonic dopamine modulates exploitation of reward learning. Front Behav …. 2010;4: 1–14. doi:10.3389/fnbeh.2010.00170
33. Chang CY, Gardner M, Di Tillio MG, Schoenbaum G. Optogenetic Blockade of Dopamine Transients Prevents Learning Induced by Changes in Reward Features. Curr Biol. Elsevier Ltd.; 2017;27: 3480–3486.e3. doi:10.1016/j.cub.2017.09.049
34. Sadacca BF, Jones JL, Schoenbaum G. Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. Elife. 2016;5.

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

doi:10.7554/eLife.13665

35. Lak A, Stauffer WR, Schultz W. Dopamine neurons learn relative chosen value from probabilistic rewards. Elife. 2016;5: 1–19. doi:10.7554/eLife.18044
36. Roesch MR, Esber GR, Li J, Daw ND, Schoenbaum G. Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. Eur J Neurosci. 2012;35: 1190–1200. doi:10.1111/j.1460-9568.2011.07986.x
37. Esber GR, Roesch MR, Bali S, Trageser J, Bissonette GB, Puche AC, et al. Attention-related pearce-kaye-hall signals in basolateral amygdala require the midbrain dopaminergic system. Biol Psychiatry. Elsevier Inc.; 2012;72: 1012–1019. doi:10.1016/j.biopsych.2012.05.023
38. Westbrook A, Braver TS. Dopamine Does Double Duty in Motivating Cognitive Effort. Neuron. Elsevier; 2016;89: 695–710. doi:10.1016/j.neuron.2016.07.020
39. Jacob SN, Ott T, Nieder A. Dopamine Regulates Two Classes of Primate Prefrontal Neurons That Represent Sensory Signals. J Neurosci. 2013;33: 13724–13734. doi:10.1523/JNEUROSCI.0210-13.2013
40. Nasser HM, Calu DJ, Schoenbaum G, Sharpe MJ. The dopamine prediction error: Contributions to associative models of reward learning. Front Psychol. 2017;8: 1–17. doi:10.3389/fpsyg.2017.00244
41. Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y. Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron. Elsevier Inc.; 2017;93: 451–463. doi:10.1016/j.neuron.2016.12.040
42. Lau B, Monteiro T, Paton JJ. The many worlds hypothesis of dopamine prediction error: implications of a parallel circuit architecture in the basal ganglia. Curr Opin Neurobiol. Elsevier Ltd; 2017;46: 241–247. doi:10.1016/j.conb.2017.08.015
43. Menegas W, Babayan BM, Uchida N, Watabe-Uchida M. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. Elife. 2017;6: 1–26. doi:10.7554/eLife.21886
44. Takahashi YK, Batchelor HM, Liu B, Khanna A, Morales M, Schoenbaum G. Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. Neuron. Elsevier Inc.; 2017;95: 1395–1405.e3. doi:10.1016/j.neuron.2017.08.025
45. Collins AGE, Frank MJ. Surprise! Dopamine signals mix action, value and error. Nat

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Neurosci. 2016;19: 3–5. doi:10.1038/nn.4207

46. Kim HF, Ghazizadeh A, Hikosaka O. Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. Cell. Elsevier Inc.; 2015;163: 1165–1175. doi:10.1016/j.cell.2015.10.063
47. Sheafor PJ. "Pseudoconditioned" jaw movements of the rabbit reflect associations conditioned to contextual background cues. J Exp Psychol Anim Behav Process. 1975;1: 245–60. Available: http://www.ncbi.nlm.nih.gov/pubmed/1185111
48. Sheafor PJ, Gormezano I. Conditioning the rabbit's (Oryctolagus cuniculus) jaw-movement response: US magnitude effects on URs, CRs, and pseudo-CRs. J Comp Physiol Psychol. 1972;81: 449–56. Available: http://www.ncbi.nlm.nih.gov/pubmed/4649185
49. Schultz W. Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. Neuron. Elsevier Inc.; 2011;69: 603–17. doi:10.1016/j.neuron.2011.02.014
50. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND. Differential roles of human striatum and amygdala in associative learning. Nat Neurosci. Nature Publishing Group; 2011;14: 1250–1252. doi:10.1038/nn.2904
51. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI. Rational regulation of learning dynamics by pupil-linked arousal systems. Nature Neuroscience. 2012. pp. 1040–1046. doi:10.1038/nn.3130
52. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 1996;16: 1936–1947. doi:10.1.1.156.635
53. Pearce JM, Hall G. A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned But Not of Unconditioned Stimuli. Psychol Rev. 1980;87: 532–552. Available: https://pdfs.semanticscholar.org/bfde/4e5ddaa6a968df9984b557baccd7cb38fb82.pdf
54. Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature. Nature Publishing Group; 2013;500: 575–9. doi:10.1038/nature12475
55. Wickens JR, Horvitz JC, Costa RM, Killcross S. Dopaminergic mechanisms in actions and habits. J Neurosci. 2007;27: 8181–3. doi:10.1523/JNEUROSCI.1671-07.2007
56. Lammel S, Ion DI, Roeper J, Malenka RC. Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. Neuron. Elsevier Inc.; 2011;70: 855–

bioRxiv preprint doi: https://doi.org/10.1101/381103; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

62. doi:10.1016/j.neuron.2011.03.025

57. Menegas W, Babayan BM, Uchida N, Watabe-Uchida M. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. Elife. 2017;6: 1–26. doi:10.7554/eLife.21886

58. Pignatelli M, Bonci A. Role of Dopamine Neurons in Reward and Aversion: A Synaptic Plasticity Perspective. Neuron. Elsevier Inc.; 2015;86: 1145–1157. doi:10.1016/j.neuron.2015.04.015

59. Sutton RS, Barto AG. Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning) [Internet]. A Bradford Book; 1998. Available: http://www.amazon.com/Reinforcement-Learning-Introduction-Adaptive-Computation/dp/0262193981