

1 **Title:** The neurocognitive role of working memory load when Pavlovian motivational control
2 affects instrumental learning

3

4

5 **Authors:**

6 Heesun Park¹, Hoyoung Doh¹, Harhim Park¹, Woo-Young Ahn^{1,2}

7

8

9 ¹Department of Psychology, Seoul National University, Seoul, Korea 08826

10 ²Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Korea 08826

11

12

13 **Corresponding author:**

14 Woo-Young Ahn, Ph.D.

15 Department of Psychology

16 Seoul National University

17 Seoul, Korea 08826

18 Tel: +82-2-880-2538, Fax: +82-2-877-6428. E-mail: wahn55@snu.ac.kr

19

20 Number of pages: 41

21 Number of Figures & Tables: 5/1

22 Number of words in the abstract: 270

23 Number of words in the introduction: 1014

24 Number of words in the discussion: 2223

25

26 **Conflict of Interest:** The authors declare no competing financial interests.

1 **Abstract**

2 Humans and animals learn optimal behaviors by interacting with the environment. Research
3 suggests that a fast, capacity-limited working memory (WM) system and a slow, incremental
4 reinforcement learning (RL) system jointly contribute to instrumental learning. Situations that
5 strain WM resources alter several decision-making processes and the balance between
6 multiple decision-making systems: under WM loads, learning becomes slow and incremental,
7 while reward prediction error (RPE) signals become stronger; the reliance on computationally
8 efficient learning increases as WM demands are balanced against computationally costly
9 strategies; and action selection becomes more random. Meanwhile, instrumental learning is
10 known to interact with Pavlovian learning, a hard-wired system that motivates approach to
11 reward and avoidance of punishment. However, the neurocognitive role of WM load during
12 instrumental learning under Pavlovian influence remains unknown, while conflict between the
13 two systems sometimes leads to suboptimal behavior. Thus, we conducted a functional
14 magnetic resonance imaging (fMRI) study (N=49) in which participants completed an
15 instrumental learning task with Pavlovian–instrumental conflict (the orthogonalized go/no-go
16 task); WM load was manipulated with dual-task conditions. Behavioral and computational
17 modeling analyses revealed that WM load compromised learning by reducing the learning rate
18 and increasing random choice, without affecting Pavlovian bias. Model-based fMRI analysis
19 revealed that WM load strengthened RPE signaling in the striatum. Moreover, under WM load,
20 the striatum showed weakened connectivity with the ventromedial and dorsolateral prefrontal
21 cortex when computing reward expectations. These results suggest that the limitation of
22 cognitive resources by WM load decelerates instrumental learning through the weakened
23 cooperation between WM and RL; such limitation also makes action selection more random,
24 but it does not directly affect the balance between instrumental and Pavlovian systems.

1 Introduction

2 The process of learning about the environment from experience and making adaptive
3 decisions involves multiple neurocognitive systems, among which reinforcement learning (RL)
4 and working memory (WM) systems are known to significantly contribute to learning (Collins
5 & Frank, 2012; Huys et al., 2021; Rmus et al., 2021). RL processes facilitate “incremental”
6 learning from the discrepancy between actual and predicted rewards, known as reward
7 prediction error (RPE); RL is regarded as a slow but steady process (Sutton & Barto, 2018).
8 Dopaminergic activity in the basal ganglia conveys RPEs (Bornstein & Daw, 2011; Khamassi
9 et al., 2005; Montague et al., 1996; Niv, 2009; Schultz, 1997, 1998; Schultz et al., 1997), and
10 human imaging studies have found that blood-oxygen-level-dependent (BOLD) signals in the
11 striatum are correlated with RPEs (Garrison et al., 2013; O’Doherty et al., 2004; O’Doherty et
12 al., 2003).

13 In addition to RL, WM is a crucial component in learning. In particular, WM allows the
14 rapid learning of actions via retention of recent stimulus-action-outcome associations, while
15 RL constitutes a slow learning process (Collins, Ciullo, et al., 2017; Collins, 2018; Collins &
16 Frank, 2012; Yoo & Collins, 2022). WM can also offer various inputs to RL, such as reward
17 expectations (Collins & Frank, 2018) and models of the environment (Dayan, 2009; Dolan &
18 Dayan, 2013; Tanaka et al., 2008; Valentin et al., 2007) as well as complex states and actions
19 (Collins & Shenhav, 2021; Rmus et al., 2021). In the brain, the WM system is presumably
20 associated with sustained neural activity throughout the dorsolateral prefrontal cortex (dlPFC)
21 and prefrontal cortex (PFC) (Baddeley & Hitch, 1974; Barbey et al., 2013; Curtis & D’Esposito,
22 2003; Funahashi, 2006; Funahashi & Kubota, 1994; Rottschy et al., 2012).

23 Because RL and WM cooperate to promote successful learning, the deterioration of
24 either system can alter the learning and balance between the two systems. In particular,
25 increasing WM load during learning and decision-making can lead to various consequences
26 through the depletion of WM resources. For example, first, instrumental learning becomes
27 slow and incremental under WM load (Collins, 2018; Collins, Albrecht, et al., 2017; Collins &

1 Frank, 2012; McDougle & Collins, 2020). Limited resources in the WM system cause WM
2 contribution to decline while the RL contribution increases, causing learning to occur more
3 slowly and strengthening the RPE signal in the brain (Collins, Ciullo, et al., 2017; Collins &
4 Frank, 2018). Second, among the multiple RL systems that use varying degrees of WM
5 resources, the demands of WM can be balanced against computationally costly strategies.
6 Otto et al. demonstrated that under WM load, the reliance on computationally efficient model-
7 free learning was increased, compared with model-based learning (Otto et al., 2013). Lastly,
8 limited WM resources may cause action selection to become more random and inconsistent.
9 Different values must be compared to inform decision-making during the action selection stage
10 (Rangel et al., 2008), but several studies have reported that WM load may interrupt these
11 processes without affecting valuation itself (Franco-Watkins et al., 2006, 2010; Olschewski et
12 al., 2018).

13 While reductions of WM resources substantially alter instrumental learning, another
14 important factor known to shape instrumental learning is the Pavlovian system. Through the
15 motivation of hard-wired responses, such as active responses to appetitive cues and inhibitory
16 responses toward aversive cues (Dickinson & Balleine, 2002; Mackintosh, 1983; Wasserman
17 et al., 1974; Wasserman & Miller, 1997), the Pavlovian system may facilitate certain
18 instrumental behaviors and impede others. This bias in instrumental learning is known as
19 Pavlovian bias (Breland & Breland, 1961; Dayan et al., 2006; Hershberger, 1986; Williams &
20 Williams, 1969). Pavlovian bias is generally presumed to be associated with maladaptive
21 behaviors such as substance use disorder and compulsivity-related disorders (Everitt &
22 Robbins, 2005; Garbusow et al., 2014, 2016; Glasner et al., 2005; Lüscher et al., 2020).

23 Although it is well known that the enhancement of WM load alters instrumental learning
24 in several ways, it remains unclear how WM load changes instrumental learning when it is
25 under Pavlovian influence. To investigate this relationship, we conducted a functional
26 magnetic resonance imaging (fMRI) study in which participants completed an instrumental

1 learning task that involved Pavlovian–instrumental conflicts (Guitart-Masip et al., 2012), with
2 and without additional WM load.

3 We tested the following three hypotheses. First, if the role of WM in learning is
4 unaffected by Pavlovian influence, WM load will lead to slower learning and increased striatal
5 RPE signals, consistent with previous findings (Collins, 2018; Collins, Ciullo, et al., 2017;
6 Collins & Frank, 2012, 2018). Second, if WM load leads to a computational trade-off between
7 Pavlovian and instrumental learning, as model-free and model-based learning (Otto et al.,
8 2013), WM load will enhance Pavlovian bias because the Pavlovian system is known to
9 require fewer resources and to be computationally efficient as an evolutionarily embedded
10 system that learns values as a function of cues, regardless of actions (Dayan et al., 2006). We
11 also presumed that neural signaling associated with Pavlovian bias would increase under WM
12 load. We focused on regions of the basal ganglia, specifically the striatum and substantia
13 nigra/ventral tegmental area (SN/VTA), which are considered important in Pavlovian bias
14 (Boer et al., 2018; Chowdhury et al., 2013; Frank et al., 2004; Guitart-Masip et al., 2012;
15 Guitart-Masip, Duzel, et al., 2014). Third, if the contribution of WM to consistent action
16 selection remains consistent, WM load will cause action selection to become more random,
17 as in previous studies (Franco-Watkins et al., 2006, 2010; Olschewski et al., 2018). We tested
18 whether the value comparison signal in the brain would decrease under WM load because
19 consistent action selection may be associated with the extent to which value difference
20 information is utilized during the decision-making process (Gläscher & O’Doherty, 2010;
21 Rangel et al., 2008).

22 Our behavioral and computational modeling results revealed that Pavlovian bias did
23 not increase under WM load, while learning decelerated and action selection became
24 increasingly random; these findings supported hypotheses 1 and 3 but not 2. Increased striatal
25 RPE signaling suggests that the increased contribution of RL and decreased contribution of
26 WM may explain slower learning. Further analyses revealed weakened connectivity between
27 the striatal and prefrontal regions under WM load, suggesting diminished cooperation between
28 the WM and RL systems.

1 Results

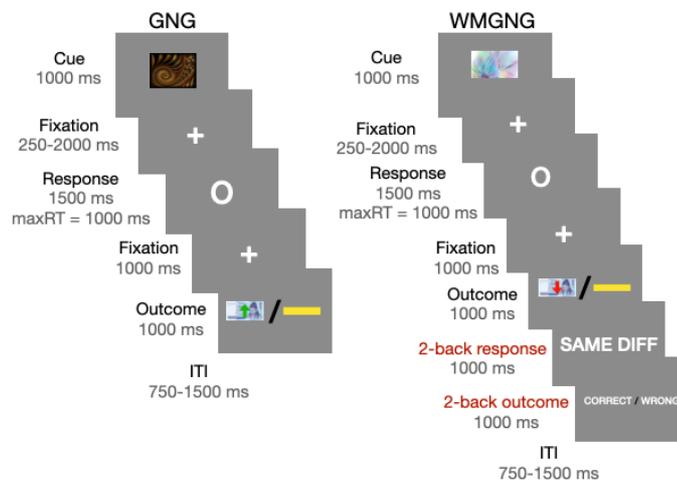
2 The participants (N=56) underwent fMRI imaging while performing an instrumental
3 learning task under a control condition and a WM load condition (**Figure 1**). In the control
4 condition, they participated in the orthogonalized go/no-go (GNG) task (Guitart-Masip et al.,
5 2012), a model-free learning task that contained Pavlovian–instrumental conflicts. In the WM
6 load condition, a 2-back task was added to the GNG task; the modified task was named the
7 working memory go/no-go (WMGNG) task (see Materials and Methods for more detail).

8

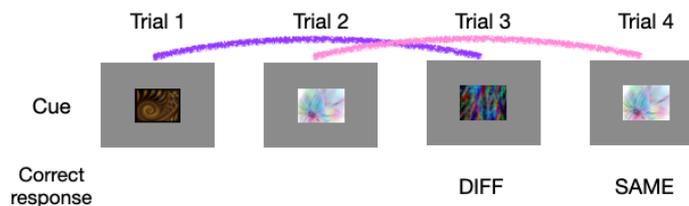
(A) Orthogonalized go/no-go task - 4 trial types



(B) Trial timelines of GNG and WMGNG tasks



(C) 2-back task in WMGNG task



9

10 **Figure 1.** The GNG and WMGNG tasks. (A) In both tasks, four fractal cues indicated the combination
11 of action (go/no-go) and valence at the outcome (win/loss). (B) In each trial, a fractal cue was presented,
12 followed by a variable delay. After the delay, actions were required in response to a circle, and

1 participants had to decide whether to press a button. After an additional brief delay, the probabilistic
2 outcome was presented, indicating monetary reward (green upward arrow on a ¥1000 bill) or monetary
3 punishment (red downward arrow on a ¥1000 bill). A yellow horizontal bar indicated no win or loss. In
4 the WMGNG task, the original GNG task was followed by a 2-back response and 2-back outcome
5 phases. (C) The participants were asked to indicate whether the cue in the current trial was identical to
6 the cue in the two preceding trials. Here, because the cue in trial 3 differed from the cue in trial 1, “DIFF”
7 was the correct response. Similarly, because the cue in trial 4 was identical to the cue in trial 3, “SAME”
8 was the correct response. The lines mark two cues for comparison: the purple line indicates that the
9 cues differ, while the pink line indicates that the cues are identical.

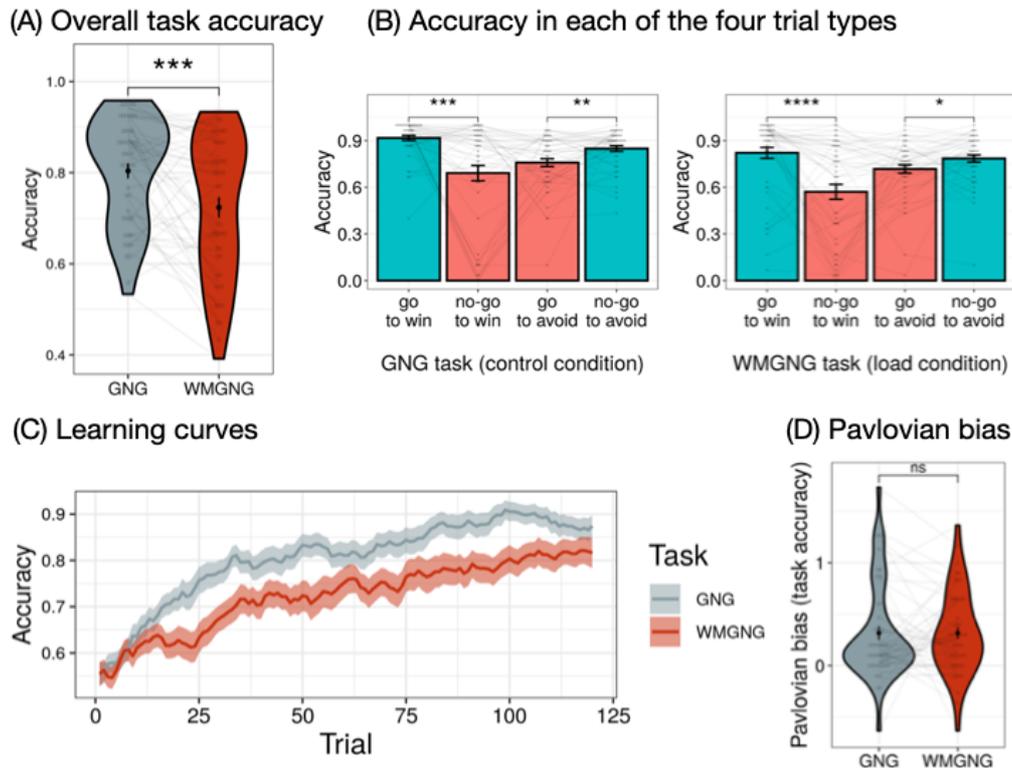
10 **Task performance: Decreased performance and learning speed under WM load**

11 Comparison of overall task accuracy between the two tasks confirmed that our dual-
12 task manipulation with a 2-back task successfully imposed WM load. Participants performed
13 better in the GNG task ($M=0.80$, $SD=0.12$) than in the WMGNG task ($M=0.72$, $SD=0.16$), as
14 illustrated in **Figure 2A** (paired t-test, $t(48)=3.86$, $p<0.001$, $d=0.55$). We also confirmed that
15 participants exhibited go bias and Pavlovian bias in both tasks, thus replicating the findings of
16 earlier studies (Adams et al., 2020; Betts et al., 2020; Boer et al., 2018; Ereira et al., 2021;
17 Guitart-Masip, Economides, et al., 2014; Guitart-Masip et al., 2012; Perosa et al., 2020;
18 Richter et al., 2014, 2021). Two-way ANOVA on accuracy, with the factors action (go/no-go)
19 and valence (reward/punishment) as repeated measures for both tasks, revealed a main effect
20 of action ($F(48)=6.05$, $p=0.018$, $\eta^2=0.03$ in GNG task, $F(48)=9.44$, $p=0.003$, $\eta^2=0.04$ in
21 WMGNG task) and action by valence interaction ($F(48)=22.43$, $p<0.001$, $\eta^2=0.12$ in the GNG
22 task, $F(48)=30.59$, $p<0.001$, $\eta^2=0.10$ in the WMGNG task); it showed no effect of valence
23 ($F(48)=0.00$, $p=0.99$, $\eta^2=0.00$ in the GNG task, $F(48)=2.77$, $p=0.103$, $\eta^2=0.01$ in the WMGNG
24 task). In both tasks (**Figure 2B**), participants exhibited superior performances in “go to win”
25 and “no-go to avoid losing” conditions (i.e., Pavlovian-congruent conditions; blue columns)
26 than in “no-go to win” and “go to avoid losing” trials (i.e., Pavlovian-incongruent conditions;
27 red columns). Specifically, in the GNG task, accuracy was higher in the “go to win” ($M=0.92$,
28 $SD=0.12$) than “no-go to win” condition ($M=0.69$, $SD=0.35$) (paired t-test, $t(48)=4.13$, $p<0.001$,

1 $d=0.59$), and in the “no-go to avoid losing” ($M=0.85$, $SD=0.13$) than in the “go to avoid losing”
2 condition ($M=0.76$, $SD=0.18$) (paired t-test, $t(48)=3.29$, $p=0.002$, $d=0.47$). Similarly, in the
3 WMGNG task, accuracy was higher in the “go to win” ($M=0.82$, $SD=0.25$) than in the “no-go
4 to win” condition ($M=0.57$, $SD=0.34$) (paired t-test, $t(48)=4.82$, $p<0.001$, $d=0.69$), and in the
5 “no-go to avoid losing” ($M=0.79$, $SD=0.16$) than in the “go to avoid losing” condition ($M=0.72$,
6 $SD=0.19$) (paired t-test, $t(48)=2.51$, $p=0.015$, $d=0.36$).

7 Next, we tested the effect of WM load on learning speed (hypothesis 1, **Figure 2C**).
8 While the learning curves indicated that participants learned during both tasks, the learning
9 curve was slower in the WMGNG task than in the GNG task (i.e., WM load reduced learning
10 speed and overall accuracy). To test the effect of WM load on Pavlovian bias (hypothesis 2,
11 **Figure 2D**), we quantified Pavlovian bias by subtracting the accuracy in Pavlovian-
12 incongruent conditions (“no-go to win” and “go to avoid losing”) from accuracy in Pavlovian-
13 congruent conditions (“go to win” and “no-go to avoid losing”), then compared it between the
14 two tasks. No significant difference in Pavlovian bias was observed between the GNG and
15 WMGNG tasks, confirming that WM load did not affect Pavlovian bias.

16



1

2 **Figure 2.** Task performance (N=49). (A) Task accuracies (mean percentages of correct responses) in
3 the GNG and WMGNG tasks show that participants performed better in the GNG task than in the
4 WMGNG task. (B) Accuracy in each of the four trial types between the two tasks demonstrated that
5 participants performed better in “go to win” and “no-go to avoid losing” trials (Pavlovian-congruent, blue)
6 than in “no-go to win” and “go to avoid losing” trials (Pavlovian-incongruent, red). (C) The learning curve
7 (i.e., the increase in accuracy across trials) was slower in the WMGNG task than in the GNG task. Note
8 that moving average smoothing was applied with filter size 5 to remove the fine variation between time
9 steps. Lines indicate group means and ribbons indicate \pm standard errors of the mean. (D) Pavlovian
10 bias was calculated by subtracting accuracy in Pavlovian-incongruent conditions (“no-go to win” + “go
11 to avoid losing”) from accuracy in Pavlovian-congruent conditions (“go to win” + “no-go to avoid losing”).
12 No significant difference in Pavlovian bias was observed between the GNG and WMGNG tasks. (A)-
13 (B), (D) Dots indicate group means and error bars indicated \pm standard errors of the mean. Gray dots
14 indicate individual accuracies; lines connect a single participant’s performances. Asterisks indicate the
15 results of pairwise t-tests. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

1 **Computational modeling: WM load influences learning rate and irreducible noise**

2 We used a computational modeling approach to test the three hypotheses. For this
3 purpose, we developed eight nested models that assumed different learning rate, Pavlovian
4 bias, or irreducible noise parameters under WM load. These models were fitted to the data
5 using hierarchical Bayesian analysis, then compared using the leave-one-out information
6 criterion (LOOIC), where a lower LOOIC value indicated better out-of-sample predictive
7 accuracy (i.e., better fit) (Vehtari et al., 2017). Importantly, the use of computational modeling
8 allowed us to test hypothesis 3 regarding whether WM load would increase random choices;
9 this would have not been possible if we had performed behavioral analysis alone.

10 Based on earlier studies (Cavanagh et al., 2013; Guitart-Masip et al., 2012), we
11 constructed a baseline model (model 1) that used a Rescorla-Wagner updating rule and
12 contained learning rate (ϵ), Pavlovian bias, irreducible noise, go bias, and separate
13 parameters for sensitivity to rewards and punishments (Materials and Methods). In the model,
14 state-action values are updated with the prediction error; learning rate (ϵ) modulates the
15 impact of the prediction error. Reward/punishment sensitivity (ρ) scales the effective size of
16 outcome values. Go bias (b) and cue values weighted by Pavlovian bias (π) are added to the
17 value of go choices. Here, as the Pavlovian bias parameter increases, the go tendency
18 increases under the reward condition whereas the go tendency is reduced under the
19 punishment condition; this results in an increased no-go tendency. Computed action weights
20 are used to estimate action probabilities, and irreducible noise (ξ) determines the extent to
21 which information about action weights is utilized to make decisions. As irreducible noise
22 increases, action probabilities will be less reflective of action weights, indicating that action
23 selection will become increasingly random.

24 In models 2, 3, and 4, we assumed that WM load affects only one parameter. For
25 example, in model 2, a separate Pavlovian bias parameter (π_{wm}) was assumed for the WM
26 load condition. Models 3 and 4 assumed different learning rates (ϵ_{wm}) and irreducible noise
27 (ξ_{wm}) parameters in their respective WM load conditions. In models 5, 6, and 7, we assumed

1 that WM load would affect two parameters: model 5 had different Pavlovian bias (π_{wm}) and
2 learning rate (ϵ_{wm}); model 6 had different Pavlovian bias (π_{wm}) and irreducible noise (ξ_{wm}); and
3 model 7 had different learning rate (ϵ_{wm}) and irreducible noise (ξ_{wm}). Finally, model 8 was the
4 full model, in which all three parameters were assumed to be affected by WM load.

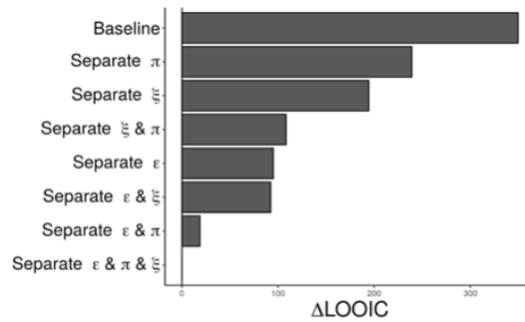
5 The full model (model 8) was the best model (**Figure 3A; Table S2**). In other words, it
6 demonstrated that participant behavior could be best explained when separate parameters
7 were included for Pavlovian bias, learning rate, and irreducible noise parameters. Next, we
8 analyzed the parameter estimates of the best-fitting model; we focused on comparing the
9 posterior distributions of the parameters that were separately fitted in the two tasks (**Figure**
10 **3B**). The parameters were considered credibly different from each other if the 95% highest
11 density intervals (HDI) of the two distributions showed no overlap (Kruschke, 2014). **Figure**
12 **3B** illustrates that Pavlovian bias was not credibly different between the two tasks, consistent
13 with the lack of support for hypothesis 2 (Pavlovian bias) in the behavioral results. Conversely,
14 the learning rate was credibly lower, while irreducible noise was credibly greater in the
15 WMGNG than in the GNG task. These results support hypothesis 1 (i.e., WM load will reduce
16 learning rate) and hypothesis 3 (i.e., WM load will increase random choices). While the best
17 model was the full model that assumed separate Pavlovian bias in the two tasks, no credible
18 group difference was observed between these parameters. This is presumably because the
19 full model was able to capture individual variations among participants (**Figure S3**), despite
20 the lack of credible difference in the group-level estimates between the two tasks. As expected,
21 the 95% HDIs of go bias, reward sensitivity, and punishment sensitivity did not include zero,
22 indicating that the participants exhibited go bias and reward/punishment sensitivity (see
23 Supplementary Material for the posterior distributions of individual parameters; **Figure S2-S5**).

24 To further compare choice randomness between the two tasks, we examined the
25 extent to which choices were dependent on value discrepancies between the two options. We
26 first plotted the percentage of go choices for the GNG and WMGNG tasks by varying the
27 quantiles of differences in action weight between the “go” and “no-go” actions ($W_{go} - W_{nogo}$)
28 (**Figure 4A**). The trial-by-trial action weights were extracted from the best-fitting model. Higher

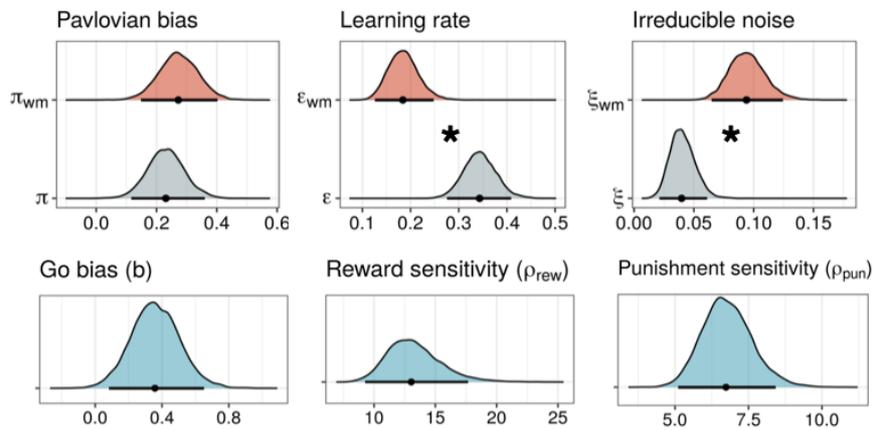
1 quantiles corresponded to a greater “go” action weight than “no-go” action weight. Overall, the
2 go ratio increased from the first to the tenth quantile, indicating that the value differences
3 between the “go” and “no-go” actions affected participants’ choices. This result further
4 illustrates the difference between the two tasks: the increase in the go ratio was steeper in the
5 GNG task than in the WMGNG task. In particular, the go ratio significantly differed between
6 the two tasks for the first ($t(48)=-3.59$, $p=0.001$, $d=0.51$), second ($t(48)=-3.23$, $p=0.002$,
7 $d=0.46$), third ($t(48)=-2.55$, $p=0.014$, $d=0.36$), eighth ($t(48)=2.95$, $p=0.005$, $d=0.42$), and tenth
8 ($t(48)=2.76$, $p=0.008$, $d=0.39$) quantiles. Thus, under WM load, participants were less
9 sensitive to the significant value difference between “go” and “no-go”.

10 To compare these patterns in a different way and further explore the extent to which
11 performance was dependent on choice difficulty, we plotted accuracies for the two tasks and
12 for different quantiles of the absolute value differences ($|W_{go} - W_{nogo}|$; **Figure 4B**). We assumed
13 that the choices would become easier when the absolute value difference was increased
14 because a small value difference makes it difficult to choose between options. Overall, the
15 accuracy increased from the first to the tenth quantile, indicating that participants performed
16 better as the choices became easier. This result further illustrates the difference between the
17 two tasks: the increase in accuracy was steeper in the GNG task than in the WMGNG task.
18 Specifically, the accuracy significantly differed between the two tasks for the fifth ($t(48)=4.12$,
19 $p<0.001$, $d=0.59$), sixth ($t(48)=2.95$, $p=0.005$, $d=0.42$), seventh ($t(48)=2.44$, $p=0.018$, $d=0.35$),
20 eighth ($t(48)=3.13$, $p=0.003$, $d=0.45$), ninth ($t(48)=2.87$, $p=0.006$, $d=0.41$), and tenth
21 ($t(48)=2.55$, $p=0.014$, $d=0.36$) quantiles. Thus, participants performed worse in the WM load
22 condition than in the control condition when choices were easier. Overall, **Figure 4**
23 demonstrates that WM load reduced the effect of the value difference on participants,
24 indicating increased choice randomness.

(A) Model comparison



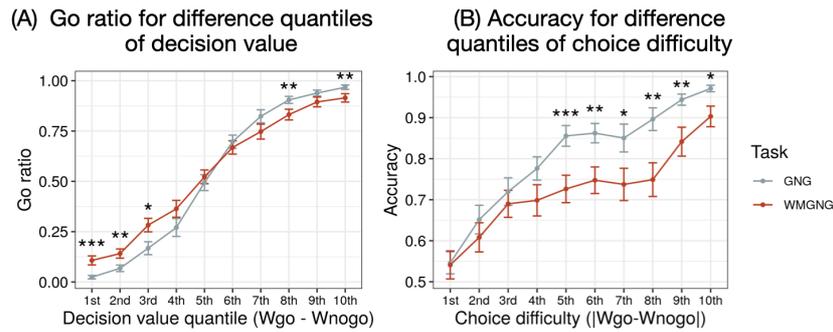
(B) Posterior distributions of the group-level parameters



1

2 **Figure 3.** Model comparison results and posterior distribution of the group-level parameters of the best-
 3 fitting model (N=49). (A) Relative LOOIC difference indicates the difference in LOOIC between the best-
 4 fitting model and each of the other models. The best-fitting model was the full model, which assumed
 5 separate Pavlovian bias, learning rate, and irreducible noise in GNG and WMGNG tasks. Lower LOOIC
 6 indicates better model fit. (B) Posterior distributions of group-level parameters from the best-fitting
 7 model. Learning rate and irreducible noise estimates were credibly different in the GNG and WMGNG
 8 tasks, while Pavlovian bias estimates were not. Dots indicate medians and bars indicate 95% HDIs.
 9 Asterisks indicate that the 95% HDIs of the two parameters' posterior distributions do not overlap (i.e.,
 10 differences are credible).

11



1

2 **Figure 4.** Choice consistency (N=49). (A) Mean percentage of go choices for different quantiles of
3 action weight differences ($W_{go} - W_{nogo}$) between “go” and “no-go” choices, where higher quantiles
4 indicate higher decision values for “go” choices. Under WM load, the increase in go ratio according to
5 quantile was less steep. (B) Mean accuracies for different quantiles of absolute value differences ($|W_{go}$
6 $- W_{nogo}|$), where higher quantiles indicate larger value differences between two options or easier choices.
7 Under WM load, the increase in accuracy according to quantile was less steep. (A)-(B) Dots are group
8 means, and error bars are \pm standard errors of the mean. Asterisks show the results of pairwise t-tests.
9 **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

10 **Larger RPE signals in the striatum and weakened connectivity with prefrontal regions** 11 **under WM load**

12 Behavioral analysis revealed that WM load caused learning to occur more slowly but
13 did not affect Pavlovian bias. The computational approach confirmed that the learning rate
14 decreased; Pavlovian bias did not change under the load; and WM load led to increased
15 choice randomness. Here, we sought to investigate the underlying neural correlates of these
16 effects of WM load on learning rate, Pavlovian bias, and random action selection. First, we
17 hypothesized that RPE signaling in the striatum would increase under WM load (Collins, Ciullo,
18 et al., 2017; Collins & Frank, 2018). We conducted a model-based fMRI analysis using RPE
19 as a regressor derived from the best-fitting model (see Materials and Methods for the full
20 general linear models (GLMs) and regressor specifications). The RPE signal in the a priori
21 striatum region of interest (ROI) was significantly greater in the WMGNG task than in the GNG
22 task (contrast: RPE in WMGNG > RPE in GNG, MNI space coordinates $x = 13, y = 14, z = -3,$
23 $Z = 3.96, p < 0.05$ small-volume corrected (SVC), **Figure 5A, Table S4**). This supports

1 hypothesis 1, which predicts an increased contribution of the RL system and decreased
2 contribution of the WM system, to learning under WM load. We also tested hypothesis 2
3 regarding Pavlovian bias, but we found no main effect of Pavlovian bias between the GNG
4 and WMGNG tasks (WMGNG > GNG [Pavlovian-congruent > Pavlovian-incongruent]) within
5 the striatum or SN/VTA ($p < 0.05$ SVC, a priori ROIs). Note that previous studies showed no
6 significant result for the same contrast (Pavlovian-congruent > Pavlovian-incongruent) within
7 the same regions (Guitart-Masip et al., 2012). With regard to hypothesis 3 concerning random
8 choices, we observed no main effect of WM load on random choice (WMGNG > GNG [W_{chosen}
9 - W_{unchosen}]) within the ventromedial prefrontal cortex (vmPFC; $p < 0.05$ SVC, a priori ROI).
10 See Supplementary Material for further details regarding these findings (**Table S5**).

11 Increased RPE signals under WM load may indicate reduced WM contribution and
12 increased RL contribution to learning because of the load, suggesting diminished cooperation
13 between the two systems for learning. Therefore, we conducted a psychophysiological
14 interaction (PPI) analysis (Friston et al., 1997) using the gPPI toolbox (McLaren et al., 2012)
15 to test whether functional connectivity between areas associated with RL and WM systems
16 would weaken under WM load. Specifically, we explored differences between the two tasks in
17 terms of functional coupling between the striatum, which showed increased RPE signaling
18 under WM load during the feedback phase, and other regions when computing reward
19 expectations. The striatum showed significantly decreased connectivity with the vmPFC (MNI
20 space coordinates $x = 13, y = 56, z = 0, Z = -4.90, p < 0.05$ whole-brain cluster-level familywise
21 error rate (FWE)) and dlPFC (MNI space coordinates $x = -20, y = 63, z = 23, Z = -4.24, p <$
22 0.05 whole-brain cluster-level FWE, **Figure 5B, Table S6**) in the WMGNG task, compared
23 with the GNG task.

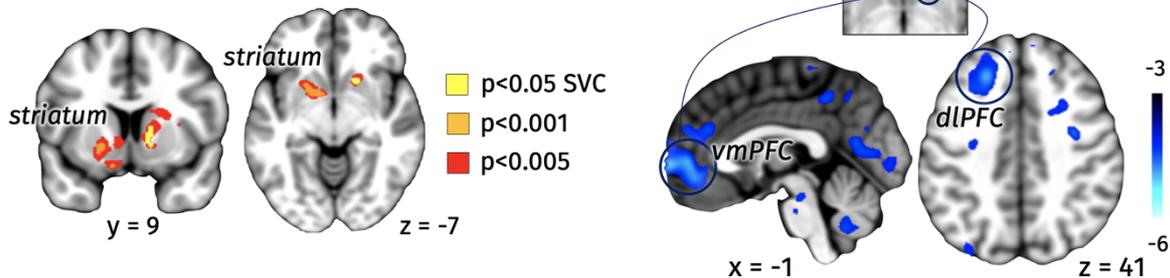
(A) Model-based fMRI result

(B) PPI result

RPE (WMGNG > GNG)

Anticipation phase (WMGNG > GNG)

Seed: striatum



1

2 **Figure 5.** fMRI results (N=44). (A) In the predefined ROI-based analysis, RPE signaling in the striatum
3 was stronger in the WMGNG task than in the GNG task. Effects that were significant at $p < 0.05$ (SVC)
4 are shown in yellow. (B) Functional connectivity between the striatum (seed region, top) and prefrontal
5 regions, including vmPFC (bottom left) and dlPFC (bottom right), was weaker in the WMGNG task than
6 in the GNG task when computing reward expectation ($p < 0.05$, whole-based cluster-level FWE).
7 Overlays are shown with a threshold of $p < 0.001$ (uncorrected). Color scale indicates t-values.

8 Discussion

9 In this study, our main objective was to elucidate the neurocognitive effects of WM load
10 on instrumental learning that involves Pavlovian–instrumental conflicts. We hypothesized that
11 under WM load, 1) learning rate would decrease and RPE signals would become stronger, 2)
12 Pavlovian bias would increase, and 3) action selection would become increasingly random.
13 First, we found that the limitation of WM resources according to WM load led to a decrease in
14 the learning rate and increases in striatal RPE signals. The striatum, which subsequently
15 showed stronger RPE signals under WM load, demonstrated weakened functional
16 connectivity with prefrontal regions including the dlPFC and vmPFC, during reward prediction.
17 WM load also increased random action selection. However, Pavlovian bias did not increase
18 under WM load, suggesting that WM load did not affect the balance between Pavlovian and
19 instrumental systems.

1 **Decreased contribution of the WM system and increased contribution of the RL system** 2 **under WM load**

3 The effect of WM load on instrumental learning remained consistent despite Pavlovian
4 bias. In particular, our behavioral analysis revealed a deceleration in learning speed under
5 WM load (**Figure 2C**); modeling analysis confirmed that WM load reduced learning rate
6 (**Figure 3**).

7 These findings can be attributed to the reduced contribution of the WM system and
8 increased contribution of the striatal RL system, consistent with previous findings that WM
9 improves learning efficiency (in parallel with RL), as well as reward prediction precision in RL
10 processes. First, as a rapid and immediate learning system, WM learns in parallel with the
11 slow and incremental RL system by directly storing associations between states and actions
12 (Collins, 2018; Collins, Ciullo, et al., 2017; Collins & Frank, 2012; Tsujimoto & Sawaguchi,
13 2004; Yoo & Collins, 2022). Specifically, WM and RL systems compete with each other based
14 on their reliability in a given situation. Under WM load, the fast and capacity-limited WM system
15 becomes less reliable; thus, the slow and incremental RL system supersedes the WM system,
16 causing learning to occur more slowly and incrementally (Collins, 2018; Collins, Albrecht, et
17 al., 2017; Collins & Frank, 2012). Second, RL computations themselves are intertwined with
18 WM; WM can represent feed reward expectations to the RL system (Ballard et al., 2011; Kahnt
19 et al., 2011; D. Lee & Seo, 2007; Wallis & Miller, 2003) and improve reward prediction
20 precision, which can reduce RPE and improve learning efficiency (Collins, Ciullo, et al., 2017;
21 Collins & Frank, 2018). In our study, WM load, which limited the contribution of WM, may have
22 increased the striatal RL contribution while reducing the accuracy of RL reward computation.
23 Consistent with this interpretation, we found that RPE signaling in the striatum – a marker of
24 RL – was strengthened under WM load (**Figure 5A**). This is consistent with previous findings
25 that RPE-associated neural signals were increased under higher WM load (Collins, Ciullo, et
26 al., 2017; Collins & Frank, 2018).

1 Overall, these findings suggest that WM load led to reduced cooperation between RL
2 and WM by interrupting and reducing the contribution of WM. This notion is further supported
3 by the finding that the striatum showed weakened functional connectivity with the dlPFC during
4 reward prediction under WM load (**Figure 5A**). Taken together, these findings suggest that
5 WM load may have weakened the interplay between WM in the dlPFC and RL in the striatum
6 during the value estimation process, which subsequently led to stronger RPE signals.
7 However, further research is necessary to demonstrate the directionality of functional
8 connectivity between the two systems during reward prediction; frontostriatal connectivity is
9 reportedly bidirectional, such that the striatum may also provide prefrontal regions with inputs
10 that relate to reward information (Park et al., 2010; Pasupathy & Miller, 2005).

11 Notably, we observed weakened connectivity between the vmPFC and the striatum.
12 The vmPFC has been identified as a critical neural correlate of value-based decision-making;
13 it integrates reward predictions (Kahnt et al., 2011), represents value signals or decision value
14 (Economides et al., 2014; Lim et al., 2011; O'Doherty, 2011; Smith et al., 2010), and affects
15 reward anticipation/processing in the striatum (Hiser & Koenigs, 2018; Pujara et al., 2016).
16 Our findings suggest that value integration through the cortico-striatal loop was also weakened
17 under WM load.

18 **No effect of WM load on Pavlovian bias**

19 Contrary to our hypothesis, WM load did not influence Pavlovian bias. Behavioral and
20 modeling results showed that Pavlovian bias did not significantly differ between the GNG and
21 WMGNG tasks (**Figure 2D**; **Figure 3**), while fMRI analysis revealed that neural signaling
22 associated with Pavlovian bias did not significantly differ between the two tasks (**Table S6**).
23 These findings indicate that the brain did not exhibit greater reliance on the computationally
24 efficient system under WM load, in contrast to the results of previous studies (Otto et al., 2013).
25 We identified two possible explanations for this discrepancy. First, instrumental and Pavlovian
26 learning require similar amounts of WM resources; second, the WM system may not be
27 involved in modulating the balance between Pavlovian and instrumental systems.

1 In the first potential explanation, the amounts of WM resources may be similar for
2 Pavlovian and instrumental learning (especially model-free learning), in contrast to model-
3 based and model-free learning. Model-based learning system constructs an internal model to
4 compute the values of actions; thus, it requires greater WM resources to compute and retain
5 the model online (Balleine & O’doherly, 2010; Daw et al., 2005, 2011; Dolan & Dayan, 2013;
6 Keramati et al., 2011). However, the model-free system simply uses the action-reward
7 association history to compute action values (i.e., “cached values”), and does not require the
8 internal model (Balleine & O’doherly, 2010; Daw et al., 2005; Dickinson, 1985). Pavlovian
9 learning is similar to model-free learning but differs in terms of the dimensions for value
10 learning—the Pavlovian system learns state-outcome associations, while the instrumental
11 system learns state-action-outcome associations (Dayan et al., 2006; Dorfman & Gershman,
12 2019). Therefore, the difference in WM demands between model-based and model-free
13 learning could be significantly greater than the difference between model-free instrumental
14 and Pavlovian learning. In our task, in particular, the instrumental learning was model-free;
15 both instrumental and Pavlovian systems were required to learn the associations without prior
16 information. Thus, the difference in WM demands may not have been sufficient to trigger a
17 trade-off between the two learning systems. Rather than depending more on Pavlovian
18 learning which has little computational benefit in our task, the participants may simply have
19 compromised overall learning.

20 In the second potential explanation, WM resources may be unimportant with respect
21 to modulating the Pavlovian–instrumental interaction, despite earlier studies’ suggestions to
22 the contrary. Several studies have proposed that prefrontal WM control systems are crucial
23 for controlling Pavlovian bias. Electroencephalogram studies demonstrated that midfrontal
24 theta oscillations are important for controlling Pavlovian bias (Cavanagh et al., 2013; Swart et
25 al., 2018), suggesting top-down prefrontal control over Pavlovian bias (Cavanagh et al., 2013).
26 Furthermore, recruitment of the inferior frontal gyrus (IFG) is involved in appropriate response
27 inhibition, helping to overcome Pavlovian bias (Guitart-Masip et al., 2012). Finally, there is
28 indirect evidence that administration of levodopa, which increases dopamine levels, reduced

1 Pavlovian influences on instrumental learning; such a reduction was speculated to result from
2 increased dopamine levels in the PFC, which may have facilitated the operation of prefrontal
3 WM functions (Guitart-Masip, Economides, et al., 2014). A related finding suggested that
4 genetic determinants of prefrontal dopamine function may be important in overcoming
5 Pavlovian bias (Richter et al., 2021).

6 While the results of the present study appear to contradict these findings, several
7 complex possibilities exist. In particular, although previous findings implied the involvement of
8 prefrontal mechanisms (e.g., model-based prefrontal control (Cavanagh et al., 2013) and WM
9 (Guitart-Masip, Duzel, et al., 2014; Guitart-Masip, Economides, et al., 2014)) in controlling the
10 Pavlovian system, they did not directly suggest active recruitment of the prefrontal WM system.
11 First, while Cavanagh et al. speculated that midfrontal theta power could be indicative of
12 “model-based top-down prefrontal control” (Cavanagh et al., 2013), a subsequent study by
13 Swart et al. suggested that midfrontal theta signals could only be involved in the detection of
14 conflict by signaling “the need for control” (Cavanagh & Frank, 2014; Swart et al., 2018), rather
15 than being a source of direct control. Next, the IFG showed an increased BOLD response only
16 in the “no-go” condition (Guitart-Masip et al., 2012), implying that the IFG is important for
17 “inhibitory” motor control (i.e., as a brake (Aron et al., 2014)); it does not participate in active
18 maintenance or representation of goal-directed behaviors including both “go” and “no-go,”
19 which would be more closely associated with WM (Levy & Goldman-Rakic, 2000; Petrides,
20 2000; Rottschy et al., 2012). Finally, elevated dopamine levels should be cautiously
21 interpreted as improvements in prefrontal WM function (Guitart-Masip, Economides, et al.,
22 2014). While dopamine has been shown to enable successful cognitive control in the prefrontal
23 cortex, it may have three roles: gating behaviorally relevant sensory signals; maintaining and
24 manipulating information in WM to guide goal-directed behavior; and relaying motor
25 information to premotor areas for action preparation (Ott & Nieder, 2019). Moreover, distinct
26 mechanisms have been known to modulate the influence of dopamine on WM in the PFC
27 through distinct types of dopamine receptors (Ott & Nieder, 2019). Thus, there may be several
28 ways to interpret the observation that dopamine level (Guitart-Masip, Economides, et al., 2014)

1 or function (Richter et al., 2021) was associated with the modulation of Pavlovian influences.
2 Considerable research is needed to fully understand the mechanisms by which dopamine
3 levels affect Pavlovian bias. Alternatively, the role of prefrontal WM in controlling Pavlovian
4 bias may not require vast resources. It may only be responsible for signaling a need for control
5 (Swart et al., 2018), promoting response inhibition (Guitart-Masip et al., 2012), or influencing
6 subcortical areas (e.g., the striatum and subthalamic nucleus (Albrecht et al., 2016; Cools,
7 2016)).

8 **Increased random choices under WM load**

9 Another notable finding was that random choice increased under WM load. Our
10 modeling analysis revealed that irreducible noise parameter estimates were greater in the
11 WMGNG task than in the GNG task (**Figure 3**), suggesting increased random action selection
12 under WM load. Further analysis using the modeling outputs revealed that participants'
13 choices were less affected by the relative value difference between the “go” and “no-go”
14 actions under WM load (**Figure 4A**). Moreover, analysis using the absolute difference between
15 the two options (**Figure 4B**) revealed that the increase in accuracy became smaller as the
16 absolute difference increased (i.e., the choice became easier). Both findings suggest that WM
17 involvement led to an increase in random choices, regardless of value comparison and choice
18 difficulty.

19 Our findings are broadly consistent with the results of previous studies concerning the
20 role of WM and prefrontal regions in action selection and execution (Barrouillet et al., 2007;
21 Dalley et al., 2004; Granon et al., 1994; Oberauer, 2019; Ridderinkhof et al., 2004; Ripke et
22 al., 2012; Seo et al., 2012; Szmalec et al., 2005). In particular, several studies have
23 demonstrated that the interruption of WM function via WM load could increase the frequency
24 of random choices in value-based decision-making tasks (Franco-Watkins et al., 2006, 2010;
25 Olschewski et al., 2018). Additionally, transcranial direct current stimulation, a brain
26 stimulation method, over the left PFC led to increased random action selection during an RL
27 task, suggesting that the prefrontal WM component influenced action selection (Turi et al.,

1 2015). Furthermore, the importance of WM in action selection during learning tasks is
2 supported by the indirect evidence that individual differences in WM capacity were correlated
3 with appropriate exploratory action selection in multi-armed bandit tasks (Laureiro-Martinez et
4 al., 2019). Overall, the reduced availability of WM resources because of WM load in our study
5 may have compromised the participants' abilities to actively represent their current goals and
6 actions, leading to reduced WM control over consistent choice based on value computation.

7 No significant neural correlates were identified with respect to the increased random
8 choices. We assumed that random action selection would be associated with the reduced
9 sensitivity to value difference or value comparison between the two options ("go" and "no-go")
10 (Gläscher & O'Doherty, 2010); thus, we hypothesized that value comparison signals would
11 decrease under WM load. Contrary to our hypothesis, no significant differences in value
12 comparison signaling in ROIs were observed between GNG and WMGNG tasks. There are
13 several possible explanations for this null finding. Our assumption of value sensitivity may not
14 be the source of the random choice observed here. Alternatively, subsequent attentional lapse
15 (Master et al., 2020; Nassar & Frank, 2016) or value-independent noise (Talmi et al., 2009)
16 may have led to inconsistent action selection despite the presence of value comparison
17 signals. Further research is necessary to distinguish these possibilities.

18 In summary, the present study has shown that WM load compromises overall learning
19 by reducing learning speed via weakened cooperation between RL and WM; it also increases
20 random action selection without affecting the balance between Pavlovian and instrumental
21 systems. To our knowledge, this is the first study to investigate the neurocognitive effect of
22 WM load during interactions between Pavlovian and instrumental systems. By investigating
23 how learning and decision-making using different systems are altered in the presence of WM
24 load and by linking such behaviors to their underlying neural mechanisms, this study
25 contributes to our understanding of how distinct cognitive components interact with each other
26 and synergistically contribute to learning. Because impairments in learning, balance among
27 multiple systems, and action selection have been reported in various neurological and

1 psychiatric disorders (Huys et al., 2016, 2021), our findings represent an important step toward
2 improved understanding of various symptoms.

3 **Materials and methods**

4 **Participants**

5 Fifty-six adults participated in this study (34 women; 24.5 ± 3.6 years old). All
6 participants were healthy, right-handed; they had normal or corrected-to-normal visual acuity.
7 They were screened prior to the experiment to exclude individuals with a history of
8 neurological, or psychiatric illness. All participants provided written informed consent, and the
9 study protocol was approved by the Institutional Review Board of Seoul National University.

10 The behavioral analysis included 49 participants (29 women; 24.3 ± 3.3 y.o); the fMRI
11 analysis included 44 participants (27 women; 24.2 ± 3.3 y.o). Four participants were excluded
12 because of technical issues; one participant was excluded because they slept during the task.
13 Two participants were excluded because of poor performance in the 2-back task since the
14 results in the dual-task paradigm could only be valid and interpretable when participants
15 actually performed both tasks. The accuracy cutoff was 0.575, a value that rejects the null
16 hypothesis that participants would randomly choose one of two options. After assessment of
17 preprocessed image quality, five participants were excluded from the fMRI analysis because
18 of head movements in the scanner, which can systematically alter brain signals; four out of
19 these five were excluded because the mean framewise displacement exceeded 0.2 mm (Gu
20 et al., 2015), while the remaining one was excluded after visual assessment of carpet plots
21 (Power, 2017).

22 **Experimental design and task**

23 The experiment was performed in two blocks: one contained the original GNG task
24 (Guitart-Masip et al., 2012) and one contained the GNG task paired with the 2-back task as a
25 secondary task. The order of task completion was counterbalanced among participants. The

1 GNG and WMGNG tasks consisted of two blocks (four blocks in total); each block consisted
2 of 60 trials. Therefore, each task contained 120 trials (240 trials in total). Participants
3 underwent fMRI while performing the tasks for approximately 50 min, with a short (~60 s)
4 break after each set of 60 trials. Before scanning, participants performed 20 practice trials
5 each of GNG task and WMGNG task to help them become accustomed to the task structure
6 and response timing. Participants received additional compensation based on their accuracy
7 in the two tasks, along with the standard participation fee at the end of the experiment.

8 *Orthogonalized go/no-go (GNG) task*

9 Four trial types were implemented depending on the nature of the fractal cue (**Figure**
10 **1A**): press a button to gain a reward (go to win); press a button to avoid punishment (go to
11 avoid losing); do not press a button to earn a reward (no-go to win); and do not press a button
12 to avoid punishment (no-go to avoid losing). The meanings of fractal images were randomized
13 among participants.

14 Each trial consisted of three phases: fractal cue presentation, response, and
15 probabilistic outcome. **Figure 1B** illustrates the trial timeline. In each trial, participants were
16 presented with one of four abstract fractal cues for 1000 ms. After a variable interval drawn
17 from a uniform probability distribution within the range of 250-2000 ms, a white circle was
18 displayed on the center of the screen for 1000 ms. When the circle appeared, participants
19 were required to respond by pressing a button or not pressing a button. Next, the outcome
20 was presented for 1000 ms: a green arrow pointing upwards on a ₩1000 bill indicated
21 monetary reward, a red arrow pointing downwards on a ₩1000 bill indicated monetary
22 punishment, and a yellow horizontal bar indicated no reward or punishment.

23 The outcome was probabilistic; thus, 80% correct responses and 20% incorrect
24 responses resulted in the best outcome. Participants were instructed that the outcome would
25 be probabilistic; for each fractal image, the correct response could be either “go” or “no-go,”
26 and they would have to learn the correct response for each cue through trial and error. The
27 task included 30 trials for each of the four trial types (120 trials in total). Trial types were
28 randomly shuffled throughout the duration of the task.

1 *Orthogonalized go/no-go + 2-back (WMGNG) task*

2 In the WM load condition, the GNG task was accompanied by a 2-back task to induce
3 WM load. The combined task was named the WMGNG task; each trial had 2-back response
4 and 2-back outcome phases after the GNG task (fractal cue, response, and probabilistic
5 outcome). Participants were required to indicate whether the cue in the current trial was
6 identical to the cue presented in the two previous trials. For example, as shown in **Figure 1C**,
7 the cue in the third trial differs from the cue in the first trial (two trials prior); thus, participants
8 should respond “different” by pressing button after responding to the reinforcement learning
9 task. In the fourth trial, they should respond “same.” The positions of “SAME” and “DIFF” were
10 randomized among participants.

11 **Computational modeling**

12 *Baseline RL model with Pavlovian bias*

13 We adopted a previously implemented version of an RL model (Guitart-Masip et al.,
14 2012) that can model Pavlovian bias and choice randomness as well as learning rate. In our
15 baseline model, we assumed no difference in parameters between the control and load
16 conditions.

17 Expected values $Q(a_t, s_t)$ were calculated for each action a , “go” or “no-go”, on each
18 stimulus s (i.e., four trial types of the task) on each trial t . $Q(a_t, s_t)$ was determined by
19 Rescorla-Wagner or delta rule updating:

$$20 \quad Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \epsilon(\rho r_t - Q_{t-1}(a_t, s_t))$$

21 where ϵ is the learning rate. The learning rate (ϵ) is a step size of learning (Sutton & Barto,
22 2018) that modulates how much of the prediction error, a teaching signal, is incorporated into
23 the value update.

24 Rewards, neutral outcomes, and punishments were entered in the model through $r_t \in$
25 $\{-1, 0, 1\}$, where ρ reflects the weighting (and effect sizes) of rewards and punishments. In all
26 models, ρ could be different for rewards and punishments (ρ_{rew} for gain, ρ_{pun} for loss).

1 Action weights $W(a_t, s_t)$ were calculated from Q values, and the Pavlovian and go
2 biases:

$$3 \quad W_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b + \pi V_t(s_t) & \text{if } a = \text{go} \\ Q_t(a_t, s_t), & \text{else} \end{cases}$$

4 where b was added to the value of go, while the expected value on the current state $V_t(s_t)$
5 was weighted by π and added to the value of go choices. $V_t(s_t)$ was computed as follows:

$$6 \quad V_t(s_t) = V_{t-1}(s_t) + \epsilon(\rho r_t - V_{t-1}(s_t)).$$

7 If the Pavlovian bias parameter (π) is positive, it increases the action weight of “go” in
8 the reward conditions because $V_t(s_t)$ is positive. In the punishment conditions, positive π
9 decreases the action weight of “go” because $V_t(s_t)$ is negative.

10 Action probabilities were dependent on these action weights $W(a_t, s_t)$, which were
11 passed through a squashed softmax (Sutton & Barto, 2018):

$$12 \quad P(a_t, s_t) = \left[\frac{\exp[W(a_t, s_t)]}{\sum_{a'} \exp[W(a', s_t)]} \right] (1 - \xi) + \frac{\xi}{2}$$

13 where ξ was the irreducible noise in the decision rule; it was free to vary between 0 and 1 for
14 all models. The irreducible noise parameter explains the extent to which information about
15 action weights is utilized in making a choice. As the irreducible noise increases, the influence
16 of the difference between the action weights is reduced, indicating that action selection
17 becomes random.

18 *Additional models*

19 To test our hypotheses regarding the effects of WM load on parameters, we
20 constructed seven additional nested models assuming different Pavlovian biases (π), learning
21 rate (ϵ), and irreducible noise (ξ) under WM load (**Table 1**). Model 1 is the baseline model.
22 Model 2 assumed a separate Pavlovian bias parameter (π) for the WM load condition.
23 Similarly, models 3 and 4 assumed different learning rates (ϵ) and irreducible noises (ξ) in the
24 WMGNG block, respectively. To address the possibility that two of the three parameters would
25 be affected by the WM load, we constructed three additional models with eight free parameters:
26 model 5 with different Pavlovian bias (π) and learning rate (ϵ); model 6 with different Pavlovian

1 bias (π) and irreducible noise (ξ); and model 7 with different learning rate (ϵ) and irreducible
 2 noise (ξ). Finally, we constructed the full model, which assumed that all of these three
 3 parameters would be affected by WM load, leading to nine free parameters.

4

5 **Table 1.** Free parameters of all models

Model No.	Model	# of parameters
1	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi$	6
2	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \pi_{wm}$	7
3	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \epsilon_{wm}$	7
4	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \xi_{wm}$	7
5	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \pi_{wm}, \epsilon_{wm}$	8
6	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \pi_{wm}, \xi_{wm}$	8
7	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \epsilon_{wm}, \xi_{wm}$	8
8	$\epsilon, \rho_{rew}, \rho_{pun}, b, \pi, \xi, \pi_{wm}, \epsilon_{wm}, \xi_{wm}$	9

6

7 **Procedures for model fitting and model selection**

8 Model parameters were estimated using hierarchical Bayesian analysis (HBA). Group-
 9 level distributions were assumed to be normally distributed, with mean and standard deviation
 10 parameters set as two free hyperparameters. We employed weakly informative priors to
 11 minimize the influences of those priors on the posterior distributions (Ahn et al., 2017;
 12 Kruschke, 2014). Additionally, for parameter estimation, the Matt trick was used to minimize
 13 the dependence between group-level mean and standard deviation parameters; it also
 14 facilitated the sampling process (Papaspiliopoulos et al., 2007). Moreover, bounded
 15 parameters such as learning rates and irreducible noise ($\in [0, 1]$) were estimated within an
 16 unconstrained space; they were then probit-transformed to the constrained space, thus
 17 maximizing MCMC efficiency within the parameter space (Ahn et al., 2017; Wetzels et al.,
 18 2010).

1 We ran four independent chains with 4000 samples each, including 2000 warm-up
2 samples (i.e., burn-in) to ensure that the parameters converged to the target distributions.
3 Four chains were run to ensure that the posterior distributions were not dependent on initial
4 starting points (Vehtari et al., 2019). We visually checked convergence to target distributions
5 by observing trace plots (**Figure S1**) and computing the R statistics - a measure of
6 convergence across chains (Gelman & Rubin, 1992). R statistics were < 1.1 for all models,
7 indicating that the estimated parameter values converged to their target posterior distributions
8 (**Table S1**).

9 Models were compared using the LOOIC, which is an information criterion calculated
10 from the leave-one-out cross-validation (Vehtari et al., 2017). This method is used to estimate
11 the out-of-sample predictive accuracy of a fitted Bayesian model for model comparison and
12 selection. The LOOIC is computed using the log-likelihood evaluated from posterior
13 distributions or simulations of the parameters. The R package loo (Vehtari et al., 2017), which
14 provides an interface for the approximation of leave-one-out cross-validated log-likelihood,
15 was used to estimate the LOOIC for each model. Lower LOOIC values indicated better fit.

16 **fMRI scans: acquisition and protocol**

17 fMRI was performed on the same scanner (Siemens Tim Trio 3 Tesla) using a 32-
18 channel head coil across all participants. A high-resolution T1-weighted anatomical scan of
19 the whole brain resolution was also acquired for each participant (TR = 2300ms, TE = 2.36ms,
20 FOV = 256mm, 1mm×1mm×1mm) to enable spatial localization and normalization. The
21 participant's head was positioned with foam pads to limit head movement during acquisition.
22 Functional data was acquired using echo-planar imaging (EPI) in four scanning sessions
23 containing 64 slices (TR = 1500ms, TE = 30ms, FOV = 256mm, 2.3mm × 2.3mm × 2.3mm).
24 For the GNG task, functional imaging data were acquired in two separate 277-volume runs,
25 each lasting about 7.5 min. For the WMGNG task, data were acquired in two separate 357-
26 volume runs, each lasting about 9.5 min.

1 **fMRI scans: general linear models**

2 Preprocessing was performed using fMRIPrep 20.2.0 (Esteban et al., 2018, 2019;
3 RRID:SCR_016216), which is based on Nipype 1.5.1 (Gorgolewski et al., 2011; Gorgolewski
4 et al., 2018; RRID:SCR_002502). Details of preprocessing with fMRIPrep are provided in
5 Supplementary Material. Subsequently, images were smoothed using a 3D Gaussian kernel
6 (8mm FWHM) to adjust for anatomical differences among participants. BOLD-signal image
7 analysis was then performed using SPM12 [<http://www.fil.ion.ucl.ac.uk/spm/>] running on
8 MATLAB v9.5.0.1067069(R2018b).

9 We built participant-specific GLMs, including all runs – two runs for the GNG block and
10 two runs for the WMGNG block – and calculated contrasts to compare the two blocks at the
11 individual level. The first-level model included six movement regressors to control the
12 movement-related artifacts as nuisance regressors. Linear contrasts at each voxel were used
13 to obtain participant-specific estimates for each effect. These estimates were entered into
14 group-level analyses, with participants regarded as random effects, using a one-sample t-test
15 against a contrast value of 0 at each voxel. The group-level model included covariates for
16 gender, age, and the task order. For all GLM analyses, we conducted ROI analysis; the results
17 were corrected for multiple comparisons using small volume correction (SVC) within ROIs.

18 *GLM1 (Hypothesis 1):* GLM1 was used to test hypothesis 1: RPE signaling in the
19 striatum would be increased under WM load. Therefore, GLM was implemented by the model-
20 based fMRI approach and included the following regressors: (1) cue onset of “go to win” trials,
21 (2) cue onset of “no-go to win” trials, (3) cue onset of “go to avoid losing” trials, (4) cue onset
22 of “no-go to avoid losing” trials, (5) target onset of “go” trials, (6) target onset of “no-go” trials,
23 (7) outcome onset, (8) outcome onset parametrically modulated by the trial-by-trial RPEs, and
24 (9) wait onset (i.e., inter-trial interval). The regressor of interest was “RPE”; we compared the
25 main effect of RPE between two tasks ($RPE(8)_{WMGNG} - RPE(8)_{GNG}$). RPE regressors were
26 calculated by subtracting the expected values (Q) from the outcome for each trial. Here, the
27 outcome was the product of feedback multiplied by reward/punishment sensitivity. The a priori

1 ROI was the striatum, which is widely known to process RPE (Chase et al., 2015; Garrison et
2 al., 2013; O'Doherty et al., 2003). The striatum ROI was constructed anatomically by
3 combining the AAL3 definitions of bilateral caudate, putamen, olfactory bulb, and nucleus
4 accumbens.

5 *GLM2 (Hypothesis 2):* GLM2 was used to test hypothesis 2: neural responses
6 associated with Pavlovian bias would be increased under WM load. Specifically, the GLM
7 examined whether the difference between the anticipatory response to fractal cues in
8 Pavlovian-congruent trials and Pavlovian-incongruent trials was greater in the WMGNG task
9 than in the GNG task in regions associated with Pavlovian bias. Therefore, GLM included the
10 following regressors: (1) cue onset of “go to win” trials, (2) cue onset of “no-go to win” trials,
11 (3) cue onset of “go to avoid losing” trials, (4) cue onset of “no-go to avoid losing” trials, (5)
12 target onset of “go” trials, (6) target onset of “no-go” trials, (7) outcome onset of win trials, (8)
13 outcome onset of neutral trials, (9) outcome onset of loss trials, (10) wait onset (i.e., inter-trial
14 interval). We compared the main effect of Pavlovian bias (Pavlovian-congruent trials -
15 Pavlovian-incongruent trials) between two tasks ($[(1) + (4) - ((2) + (3))]_{\text{WMGNG}} - [(1) + (4) - ((2) +$
16 $(3))]_{\text{GNG}}$). The a priori ROIs were the striatum and SN/VTA, which are considered important in
17 Pavlovian bias (Boer et al., 2018; Chowdhury et al., 2013; Frank et al., 2004; Guitart-Masip et
18 al., 2012; Guitart-Masip, Duzel, et al., 2014). The striatum ROI was constructed anatomically
19 by combining the AAL3 definitions of bilateral caudate, putamen, olfactory bulb, and nucleus
20 accumbens. Furthermore, the SN/VTA was constructed by combining the AAL3 definitions of
21 bilateral SN and VTA.

22 *GLM3 (Hypothesis 3):* GLM3 was used to test hypothesis 3: value comparison signals
23 would decrease under WM load. GLM3 was also implemented with a model-based fMRI
24 approach: (1) cue onset of all trials, (2) cue onset parametrically modulated by the trial-by-trial
25 decision values ($W_{\text{chosen}} - W_{\text{unchosen}}$), (3) target onset of “go” trials, (4) target onset of “no-go”
26 trials, (5) outcome onset, and (6) wait onset (i.e., inter-trial interval). Decision value regressors
27 were calculated by subtracting the action weights of the unchosen option (W_{unchosen}) from the
28 action weights of the chosen option (W_{chosen}). We then compared the main effect of decision

1 value between two blocks ($(2)_{\text{WMGNG}} - (2)_{\text{GNG}}$). The a priori ROI for GLM3 was the vmPFC, which
2 was suggested as a region that encodes the relative chosen value ($W_{\text{chosen}} - W_{\text{unchosen}}$)
3 (Boorman et al., 2009; S. W. Lee et al., 2014). Here, ROI masks were created by drawing a
4 sphere with a diameter of 10 mm around the peak voxel reported in the previous studies ($[-$
5 $6, 48, -8]$ for vmPFC (Boorman et al., 2009)).

6 *PPI analysis:* In addition to GLMs, we used PPI analysis to test whether WM load led
7 to reduced cooperation between WM and RL systems for learning (Collins, Ciullo, et al., 2017;
8 Collins & Frank, 2018) by using PPI analysis. Here, to examine differences between the two
9 blocks in terms of functional coupling between the prefrontal areas and the area computing
10 RPE after choices, we performed PPI analysis using the gPPI toolbox (McLaren et al., 2012);
11 the physiological variable was the time course of the striatum in the anticipation phase, and
12 the psychological variable was the effect of WM load during the same phase. As a seed region
13 (i.e., a physiological variable), the cluster striatum ROI (peak MNI space coordinates $x = 13,$
14 $y = 14, z = -3$) was derived from the results of GLM1, which revealed stronger RPE signaling
15 in the WMGNG task than in the GNG task during the feedback phase. The entire time series
16 throughout the experiment was extracted from each participant in the striatum ROI. To create
17 the PPI regressor, these normalized time series were multiplied by task condition vectors for
18 the anticipation phase, which consisted of the cue representation and fixation phases as in
19 GLM1. A GLM with PPI regressors of the seed region was thus generated together with
20 movement regressors. The effects of PPI for each participant were estimated in the individual-
21 level GLM; the parameter estimates represented the extent to which activity in each voxel was
22 correlated with activity in the striatum during the anticipation phase. The contrast was
23 constructed by subtracting activity during the anticipation phase in the GNG task from activity
24 in the WMGNG task (WMGNG vs. GNG in the anticipation phase). Individual contrast images
25 for functional connectivity were then computed and entered into one-sample t-tests in a group-
26 level GLM together with nuisance covariates (i.e., gender, age, and task order). Whole-brain
27 cluster correction was applied for PPI analysis.

28

1 **Data availability**

2 Behavioral data and codes for behavioral analysis and modeling are available at
3 https://github.com/CCS-Lab/project_WMDM_public. Unthresholded t maps are available on
4 Neurovault: <https://identifiers.org/neurovault.collection:12860>.

5 **Acknowledgments**

6 This work was supported by the Basic Science Research Program through the National
7 Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT, & Future
8 Planning (NRF-2018R1C1B300731315 and NRF-2018R1A4A102589113), the National
9 Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.
10 2021M3E5D2A0102249311), the Creative-Pioneering Researchers Program through Seoul
11 National University, and the BK21 FOUR program [5199990314123] to W.-Y.A.

13 **References**

- 14 Adams, R. A., Moutoussis, M., Nour, M. M., Dahoun, T., Lewis, D., Illingworth, B., Veronese,
15 M., Mathys, C., Boer, L. de, Guitart-Masip, M., Friston, K. J., Howes, O. D., & Roiser,
16 J. P. (2020). Variability in Action Selection Relates to Striatal Dopamine 2/3 Receptor
17 Availability in Humans: A PET Neuroimaging Study Using Reinforcement Learning
18 and Active Inference Models. *Cerebral Cortex*, 30(6), 3573–3589.
19 <https://doi.org/10.1093/cercor/bhz327>
- 20 Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of
21 Reinforcement Learning and Decision-Making With the hBayesDM Package.
22 *Computational Psychiatry*, 1, 24–57. https://doi.org/10.1162/cpsy_a_00002
- 23 Albrecht, M. A., Waltz, J. A., Cavanagh, J. F., Frank, M. J., & Gold, J. M. (2016). Reduction
24 of Pavlovian Bias in Schizophrenia: Enhanced Effects in Clozapine-Administered
25 Patients. *PLoS ONE*, 11(4), e0152781. <https://doi.org/10.1371/journal.pone.0152781>
- 26 Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal
27 cortex: one decade on. *Trends in Cognitive Sciences*, 18(4), 177–185.
- 28 Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and*
29 *Motivation*, 8, 47–89. [https://doi.org/10.1016/s0079-7421\(08\)60452-1](https://doi.org/10.1016/s0079-7421(08)60452-1)
- 30 Ballard, I. C., Murty, V. P., Carter, R. M., MacInnes, J. J., Huettel, S. A., & Adcock, R. A.

- 1 (2011). Dorsolateral Prefrontal Cortex Drives Mesolimbic Dopaminergic Regions to
2 Initiate Motivated Behavior. *Journal of Neuroscience*, 31(28), 10340–10346.
3 <https://doi.org/10.1523/jneurosci.0895-11.2011>
- 4 Balleine, B. W., & O’doherly, J. P. (2010). Human and rodent homologues in action control:
5 corticostriatal determinants of goal-directed and habitual action.
6 *Neuropsychopharmacology*, 35(1), 48–69.
- 7 Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to
8 human working memory. *Cortex*, 49(5), 1195–1205.
- 9 Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and
10 Cognitive Load in Working Memory. *Journal of Experimental Psychology: Learning,*
11 *Memory, and Cognition*, 33(3), 570–585. <https://doi.org/10.1037/0278-7393.33.3.570>
- 12 Betts, M. J., Richter, A., Boer, L. de, Tegelbeckers, J., Perosa, V., Baumann, V., Chowdhury,
13 R., Dolan, R. J., Seidenbecher, C., Schott, B. H., Düzel, E., Guitart-Masip, M.,
14 Krauel, K. (2020). Learning in anticipation of reward and punishment: perspectives
15 across the human lifespan. *Neurobiology of Aging*, 96, 49–57.
16 <https://doi.org/10.1016/j.neurobiolaging.2020.08.011>
- 17 Boer, L. de, Axelsson, J., Chowdhury, R., Riklund, K., Dolan, R. J., Nyberg, L., Bäckman, L.,
18 & Guitart-Masip, M. (2018). Dorsal striatal dopamine D1 receptor availability predicts
19 an instrumental bias in action learning. *Proceedings of the National Academy of*
20 *Sciences*, 116(1), 201816704. <https://doi.org/10.1073/pnas.1816704116>
- 21 Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How
22 Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor
23 of Alternative Courses of Action. *Neuron*, 62(5), 733–743.
24 <https://doi.org/10.1016/j.neuron.2009.05.014>
- 25 Bornstein, A. M., & Daw, N. D. (2011). Multiplicity of control in the basal ganglia:
26 computational roles of striatal subregions. *Current Opinion in Neurobiology*, 21(3),
27 374–380.
- 28 Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*,
29 16(11), 681–684. <https://doi.org/10.1037/h0040090>
- 30 Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., & Frank, M. J. (2013). Frontal
31 Theta Overrides Pavlovian Learning Biases. *The Journal of Neuroscience*, 33(19),
32 8541–8548. <https://doi.org/10.1523/jneurosci.5754-12.2013>
- 33 Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control.
34 *Trends in Cognitive Sciences*, 18(8), 414–421.
35 <https://doi.org/10.1016/j.tics.2014.04.012>
- 36 Chase, H. W., Kumar, P., Eickhoff, S. B., & Dombrovski, A. Y. (2015). Reinforcement
37 learning models and their neural correlates: An activation likelihood estimation meta-
38 analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 15(2), 435–459.
39 <https://doi.org/10.3758/s13415-015-0338-7>

- 1 Chowdhury, R., Guitart-Masip, M., Lambert, C., Dolan, R. J., & Düzel, E. (2013). Structural
2 integrity of the substantia nigra and subthalamic nucleus predicts flexibility of
3 instrumental learning in older-age individuals. *Neurobiology of Aging*, 34(10), 2261–
4 2270.
- 5 Collins, A. G. E. (2018). The Tortoise and the Hare: Interactions between Reinforcement
6 Learning and Working Memory. *Journal of Cognitive Neuroscience*, 30(10), 1422–
7 1432. https://doi.org/10.1162/jocn_a_01238
- 8 Collins, A. G. E., Albrecht, M. A., Waltz, J. A., Gold, J. M., & Frank, M. J. (2017). Interactions
9 Among Working Memory, Reinforcement Learning, and Effort in Value-Based
10 Choice: A New Paradigm and Selective Deficits in Schizophrenia. *Biological*
11 *Psychiatry*, 82(6), 431–439. <https://doi.org/10.1016/j.biopsych.2017.05.017>
- 12 Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load
13 strengthens reward prediction errors. *Journal of Neuroscience*, 37(16), 4332–4342.
- 14 Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working
15 memory, not reinforcement learning? A behavioral, computational, and neurogenetic
16 analysis. *European Journal of Neuroscience*, 35(7), 1024–1035.
17 <https://doi.org/10.1111/j.1460-9568.2011.07980.x>
- 18 Collins, A. G. E., & Frank, M. J. (2018). Within- and across-trial dynamics of human EEG
19 reveal cooperative interplay between reinforcement learning and working memory.
20 *Proceedings of the National Academy of Sciences*, 115(10), 201720963.
21 <https://doi.org/10.1073/pnas.1720963115>
- 22 Collins, A. G. E., & Shenhav, A. (2021). Advances in modeling learning and decision-making
23 in neuroscience. *Neuropsychopharmacology*, 1–15. [https://doi.org/10.1038/s41386-](https://doi.org/10.1038/s41386-021-01126-y)
24 [021-01126-y](https://doi.org/10.1038/s41386-021-01126-y)
- 25 Cools, R. (2016). The costs and benefits of brain dopamine for cognitive control. *Wiley*
26 *Interdisciplinary Reviews: Cognitive Science*, 7(5), 317–329.
- 27 Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during
28 working memory. *Trends in Cognitive Sciences*, 7(9), 415–423.
- 29 Dalley, J. W., Cardinal, R. N., & Robbins, T. W. (2004). Prefrontal executive and cognitive
30 functions in rodents: neural and neurochemical substrates. *Neuroscience &*
31 *Biobehavioral Reviews*, 28(7), 771–784.
- 32 Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based
33 influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204–
34 1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- 35 Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal
36 and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12),
37 1704–1711.
- 38 Dayan, P. (2009). Goal-directed control and its antipodes. *Neural Networks*, 22(3), 213–219.
- 39 Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the

- 1 discipline of the will. *Neural Networks*, 19(8), 1153–1160.
2 <https://doi.org/10.1016/j.neunet.2006.03.002>
- 3 Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy.
4 *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*,
5 308(1135), 67–78.
- 6 Dickinson, A., & Balleine, B. (2002). *The role of learning in the operation of motivational*
7 *systems*.
- 8 Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2), 312–325.
- 9 Dorfman, H. M., & Gershman, S. J. (2019). Controllability governs the balance between
10 Pavlovian and instrumental action selection. *Nature Communications*, 10(1), 5826.
11 <https://doi.org/10.1038/s41467-019-13737-7>
- 12 Economides, M., Guitart-Masip, M., Kurth-Nelson, Z., & Dolan, R. J. (2014). Anterior
13 Cingulate Cortex Instigates Adaptive Switches in Choice by Integrating Immediate
14 and Delayed Components of Value in Ventromedial Prefrontal Cortex. *The Journal of*
15 *Neuroscience*, 34(9), 3340–3349. <https://doi.org/10.1523/jneurosci.4313-13.2014>
- 16 Ereira, S., Pujol, M., Guitart-Masip, M., Dolan, R. J., & Kurth-Nelson, Z. (2021). Overcoming
17 Pavlovian bias in semantic space. *Scientific Reports*, 11(1), 3416.
18 <https://doi.org/10.1038/s41598-021-82889-8>
- 19 Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I.,
20 Erramuzpe, A., Kent, J. D. and Goncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P.,
21 Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep. *Software*.
22 <https://doi.org/10.5281/zenodo.852659>
- 23 Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J.
24 D., Goncalves, M., DuPre, E., Snyder, M., & others. (2019). fMRIPrep: a robust
25 preprocessing pipeline for functional MRI. *Nature Methods*, 16(1), 111–116.
- 26 Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction:
27 from actions to habits to compulsion. *Nature Neuroscience*, 8(11), 1481–1489.
28 <https://doi.org/10.1038/nn1579>
- 29 Franco-Watkins, A. M., Pashler, H., & Rickard, T. C. (2006). Does Working Memory Load
30 Lead to Greater Impulsivity? Commentary on Hinson, Jameson, and Whitney (2003).
31 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 443–
32 447. <https://doi.org/10.1037/0278-7393.32.2.443>
- 33 Franco-Watkins, A. M., Rickard, T. C., & Pashler, H. (2010). Taxing Executive Processes
34 Does Not Necessarily Increase Impulsive Decision Making. *Experimental*
35 *Psychology*, 57(3), 193–201. <https://doi.org/10.1027/1618-3169/a000024>
- 36 Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: cognitive
37 reinforcement learning in parkinsonism. *Science*, 306(5703), 1940–1943.
- 38 Friston, K., Buechel, C., Fink, G., Morris, J., Rolls, E., & Dolan, R. J. (1997).
39 Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3),

- 1 218–229.
- 2 Funahashi, S. (2006). Prefrontal cortex and working memory processes. *Neuroscience*,
3 139(1), 251–261.
- 4 Funahashi, S., & Kubota, K. (1994). Working memory and prefrontal cortex. *Neuroscience*
5 *Research*, 21(1), 1–11.
- 6 Garbusow, M., Schad, D. J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S. P., Steinacher,
7 B., Kathmann, N., Geurts, D. E., Sommer, C., & others. (2016). Pavlovian-to-
8 instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol
9 dependence. *Addiction Biology*, 21(3), 719–731.
- 10 Garbusow, M., Schad, D. J., Sommer, C., Jünger, E., Sebold, M., Friedel, E., Wendt, J.,
11 Kathmann, N., Schlagenhauf, F., Zimmermann, U. S., Heinz, A., Huys, Q. J. M., &
12 Rapp, M. A. (2014). Pavlovian-to-Instrumental Transfer in Alcohol Dependence: A
13 Pilot Study. *Neuropsychobiology*, 70(2), 111–121. <https://doi.org/10.1159/000363507>
- 14 Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A
15 meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*,
16 37(7), 1297–1310. <https://doi.org/10.1016/j.neubiorev.2013.03.023>
- 17 Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple
18 sequences. *Statistical Science*, 7(4), 457–472.
- 19 Gläscher, J. P., & O’Doherty, J. P. (2010). Model-based approaches to neuroimaging:
20 combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary*
21 *Reviews: Cognitive Science*, 1(4), 501–510.
- 22 Glasner, S. V., Overmier, J. B., & Balleine, B. W. (2005). The role of Pavlovian cues in
23 alcohol seeking in dependent and nondependent rats. *Journal of Studies on Alcohol*,
24 66(1), 53–61.
- 25 Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., &
26 Ghosh, S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data
27 processing framework in Python. *Frontiers in Neuroinformatics*, 5, 13.
28 <https://doi.org/10.3389/fninf.2011.00013>
- 29 Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P.,
30 Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault,
31 B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F.,
32 ... Ghosh, S. (2018). Nipype. Software. <https://doi.org/10.5281/zenodo.596855>
- 33 Granon, S., Vidal, C., Thinus-Blanc, C., Changeux, J.-P., & Poucet, B. (1994). Working
34 memory, response selection, and effortful processing in rats with medial prefrontal
35 lesions. *Behavioral Neuroscience*, 108(5), 883.
- 36 Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., & Bassett, D.
37 S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings*
38 *of the National Academy of Sciences*, 112(44), 13681–13686.
- 39 Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in

- 1 decision making. *Trends in Cognitive Sciences*, 18(4), 194–202.
2 <https://doi.org/10.1016/j.tics.2014.01.003>
- 3 Guitart-Masip, M., Economides, M., Huys, Q. J. M., Frank, M. J., Chowdhury, R., Duzel, E.,
4 Dayan, P., & Dolan, R. J. (2014). Differential, but not opponent, effects of l-DOPA
5 and citalopram on action learning with reward and punishment.
6 *Psychopharmacology*, 231(5), 955–966. <https://doi.org/10.1007/s00213-013-3313-4>
- 7 Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J.
8 (2012). Go and no-go learning in reward and punishment: Interactions between affect
9 and effect. *NeuroImage*, 62(1), 154–166.
10 <https://doi.org/10.1016/j.neuroimage.2012.04.024>
- 11 Hershberger, W. A. (1986). An approach through the looking-glass. *Animal Learning &*
12 *Behavior*, 14(4), 443–451.
- 13 Hiser, J., & Koenigs, M. (2018). The Multifaceted Role of the Ventromedial Prefrontal Cortex
14 in Emotion, Decision Making, Social Cognition, and Psychopathology. *Biological*
15 *Psychiatry*, 83(8), 638–647. <https://doi.org/10.1016/j.biopsych.2017.10.030>
- 16 Huys, Q. J. M., Browning, M., Paulus, M. P., & Frank, M. J. (2021). Advances in the
17 computational understanding of mental illness. *Neuropsychopharmacology*, 46(1), 3–
18 19. <https://doi.org/10.1038/s41386-020-0746-4>
- 19 Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from
20 neuroscience to clinical applications. *Nature Neuroscience*, 19(3), 404–413.
21 <https://doi.org/10.1038/nn.4238>
- 22 Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2011). Decoding different roles for
23 vmPFC and dlPFC in multi-attribute decision making. *Neuroimage*, 56(2), 709–715.
- 24 Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual
25 and the goal-directed processes. *PLoS Computational Biology*, 7(5), e1002055.
- 26 Khamassi, M., Lachèze, L., Girard, B., Berthoz, A., & Guillot, A. (2005). Actor–Critic models
27 of reinforcement learning in the basal ganglia: from natural to artificial rats. *Adaptive*
28 *Behavior*, 13(2), 131–148.
- 29 Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*.
- 30
31 Laureiro-Martinez, D., Brusoni, S., Tata, A., & Zollo, M. (2019). The Manager’s
32 Notepad: Working Memory, Exploration, and Performance. *Journal of Management*
33 *Studies*, 56(8), 1655–1682. <https://doi.org/10.1111/joms.12528>
- 34 Lee, D., & Seo, H. (2007). Mechanisms of reinforcement learning and decision making in the
35 primate dorsolateral prefrontal cortex. *Annals of the New York Academy of Sciences*,
36 1104(1), 108–122.
- 37 Lee, S. W., Shimojo, S., & O’Doherty, J. P. (2014). Neural Computations Underlying
38 Arbitration between Model-Based and Model-free Learning. *Neuron*, 81(3), 687–699.
39 <https://doi.org/10.1016/j.neuron.2013.11.028>

- 1 Levy, R., & Goldman-Rakic, P. S. (2000). Segregation of working memory functions within
2 the dorsolateral prefrontal cortex. *Executive Control and the Frontal Lobe: Current*
3 *Issues*, 23–32.
- 4 Lim, S.-L., O’Doherty, J. P., & Rangel, A. (2011). The Decision Value Computations in the
5 vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention.
6 *The Journal of Neuroscience*, 31(37), 13214–13223.
7 <https://doi.org/10.1523/jneurosci.1246-11.2011>
- 8 Lüscher, C., Robbins, T. W., & Everitt, B. J. (2020). The transition to compulsion in addiction.
9 *Nature Reviews Neuroscience*, 21(5), 247–263. [https://doi.org/10.1038/s41583-020-](https://doi.org/10.1038/s41583-020-0289-z)
10 [0289-z](https://doi.org/10.1038/s41583-020-0289-z)
- 11 Mackintosh, N. J. (1983). *Conditioning and associative learning*. Clarendon Press Oxford.
- 12 Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., & Collins, A. G. E. (2020).
13 Disentangling the systems contributing to changes in learning during adolescence.
14 *Developmental Cognitive Neuroscience*, 41, 100732.
15 <https://doi.org/10.1016/j.dcn.2019.100732>
- 16 McDougale, S. D., & Collins, A. G. E. (2020). Modeling the influence of working memory,
17 reinforcement, and action uncertainty on reaction time and choice during instrumental
18 learning. *Psychonomic Bulletin & Review*, 1–20. [https://doi.org/10.3758/s13423-020-](https://doi.org/10.3758/s13423-020-01774-z)
19 [01774-z](https://doi.org/10.3758/s13423-020-01774-z)
- 20 McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-
21 dependent psychophysiological interactions (gPPI): A comparison to standard
22 approaches. *NeuroImage*, 61(4), 1277–1286.
23 <https://doi.org/10.1016/j.neuroimage.2012.03.068>
- 24 Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic
25 dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*,
26 16(5), 1936–1947.
- 27 Nassar, M. R., & Frank, M. J. (2016). Taming the beast: extracting generalizable knowledge
28 from computational models of cognition. *Current Opinion in Behavioral Sciences*, 11,
29 49–54.
- 30 Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*,
31 53(3), 139–154. <https://doi.org/10.1016/j.jmp.2008.12.005>
- 32 Oberauer, K. (2019). Working Memory and Attention – A Conceptual Analysis and Review.
33 *Journal of Cognition*, 2(1), 36. <https://doi.org/10.5334/joc.58>
- 34 O’Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004).
35 Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*,
36 304(5669), 452–454.
- 37 O’Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed
38 action selection. *Annals of the New York Academy of Sciences*, 1239(1), 118–129.
39 <https://doi.org/10.1111/j.1749-6632.2011.06290.x>

- 1 O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal
2 Difference Models and Reward-Related Learning in the Human Brain. *Neuron*, *38*(2),
3 329–337. [https://doi.org/10.1016/s0896-6273\(03\)00169-7](https://doi.org/10.1016/s0896-6273(03)00169-7)
- 4 Olschewski, S., Rieskamp, J., & Scheibehenne, B. (2018). Taxing Cognitive Capacities
5 Reduces Choice Consistency Rather Than Preference: A Model-Based Test. *Journal*
6 *of Experimental Psychology: General*, *147*(4), 462–484.
7 <https://doi.org/10.1037/xge0000403>
- 8 Ott, T., & Nieder, A. (2019). Dopamine and Cognitive Control in Prefrontal Cortex. *Trends in*
9 *Cognitive Sciences*, *23*(3), 213–234. <https://doi.org/10.1016/j.tics.2018.12.006>
- 10 Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning:
11 dissecting multiple reinforcement-learning systems by taxing the central executive.
12 *Psychological Science*, *24*(5), 751–761.
- 13 Papaspiliopoulos, O., Roberts, G. O., & Sköld, M. (2007). A general framework for the
14 parametrization of hierarchical models. *Statistical Science*, 59–73.
- 15 Park, S. Q., Kahnt, T., Beck, A., Cohen, M. X., Dolan, R. J., Wrase, J., & Heinz, A. (2010).
16 Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence.
17 *Journal of Neuroscience*, *30*(22), 7749–7753.
- 18 Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the
19 prefrontal cortex and striatum. *Nature*, *433*(7028), 873–876.
- 20 Perosa, V., Boer, L. de, Ziegler, G., Apostolova, I., Buchert, R., Metzger, C., Amthauer, H.,
21 Guitart-Masip, M., Düzel, E., & Betts, M. J. (2020). The Role of the Striatum in
22 Learning to Orthogonalize Action and Valence: A Combined PET and 7 T MRI Aging
23 Study. *Cerebral Cortex*, *30*(5), 3340–3351. <https://doi.org/10.1093/cercor/bhz313>
- 24 Petrides, M. (2000). The role of the mid-dorsolateral prefrontal cortex in working memory.
25 *Experimental Brain Research*, *133*(1), 44–54.
- 26 Power, J. D. (2017). A simple but useful way to assess fMRI scan qualities. *NeuroImage*,
27 *154*(NeuroImage 10 1999), 150–158.
28 <https://doi.org/10.1016/j.neuroimage.2016.08.009>
- 29 Pujara, M. S., Philippi, C. L., Motzkin, J. C., Baskaya, M. K., & Koenigs, M. (2016).
30 Ventromedial Prefrontal Cortex Damage Is Associated with Decreased Ventral
31 Striatum Volume and Response to Reward. *The Journal of Neuroscience*, *36*(18),
32 5047–5054. <https://doi.org/10.1523/jneurosci.4236-15.2016>
- 33 Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the
34 neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7),
35 545–556. <https://doi.org/10.1038/nrn2357>
- 36 Richter, A., Boer, L. de, Guitart-Masip, M., Behnisch, G., Seidenbecher, C. I., & Schott, B. H.
37 (2021). Motivational learning biases are differentially modulated by genetic
38 determinants of striatal and prefrontal dopamine function. *Journal of Neural*
39 *Transmission*, 1–16. <https://doi.org/10.1007/s00702-021-02382-4>

- 1 Richter, A., Guitart-Masip, M., Barman, A., Libeau, C., Behnisch, G., Czerney, S., Schanze,
2 D., Assmann, A., Klein, M., Düzel, E., Zenker, M., Seidenbecher, C. I., & Schott, B.
3 H. (2014). Valenced action/inhibition learning in humans is modulated by a genetic
4 variant linked to dopamine D2 receptor expression. *Frontiers in Systems*
5 *Neuroscience*, 8, 140. <https://doi.org/10.3389/fnsys.2014.00140>
- 6 Ridderinkhof, K. R., Wildenberg, W. P. V. D., Segalowitz, S. J., & Carter, C. S. (2004).
7 Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action
8 selection, response inhibition, performance monitoring, and reward-based learning.
9 *Brain and Cognition*, 56(2), 129–140.
- 10 Ripke, S., Hübner, T., Mennigen, E., Müller, K. U., Rodehacke, S., Schmidt, D., Jacob, M. J.,
11 & Smolka, M. N. (2012). Reward processing and intertemporal decision making in
12 adults and adolescents: The role of impulsivity and decision consistency. *Brain*
13 *Research*, 1478, 36–47. <https://doi.org/10.1016/j.brainres.2012.08.034>
- 14 Rmus, M., McDougale, S. D., & Collins, A. G. (2021). The role of executive function in
15 shaping reinforcement learning. *Current Opinion in Behavioral Sciences*, 38, 66–73.
16 <https://doi.org/10.1016/j.cobeha.2020.10.003>
- 17 Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., &
18 Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-
19 based meta-analysis. *NeuroImage*, 60(1), 830–846.
20 <https://doi.org/10.1016/j.neuroimage.2011.11.050>
- 21 Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current*
22 *Opinion in Neurobiology*, 7(2), 191–197.
- 23 Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of*
24 *Neurophysiology*, 80(1), 1–27.
- 25 Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and
26 Reward. *Science*, 275(5306), 1593–1599.
27 <https://doi.org/10.1126/science.275.5306.1593>
- 28 Seo, M., Lee, E., & Averbeck, B. B. (2012). Action selection and action value in frontal-
29 striatal circuits. *Neuron*, 74(5), 947–960.
- 30 Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010).
31 Distinct Value Signals in Anterior and Posterior Ventromedial Prefrontal Cortex.
32 *Journal of Neuroscience*, 30(7), 2490–2495. <https://doi.org/10.1523/jneurosci.3319-09.2010>
33
- 34 Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- 35 Swart, J. C., Frank, M. J., Määttä, J. I., Jensen, O., Cools, R., & Ouden, H. E. M. den.
36 (2018). Frontal network dynamics reflect neurocomputational mechanisms for
37 reducing maladaptive biases in motivated action. *PLoS Biology*, 16(10), e2005979.
38 <https://doi.org/10.1371/journal.pbio.2005979>
- 39 Szmalec, A., Vandierendonck, A., & Kemps, E. (2005). Response selection involves

- 1 executive control: Evidence from the selective interference paradigm. *Memory &*
2 *Cognition*, 33(3), 531–541.
- 3 Talmi, D., Dayan, P., Kiebel, S. J., Frith, C. D., & Dolan, R. J. (2009). How humans integrate
4 the prospects of pain and reward during choice. *Journal of Neuroscience*, 29(46),
5 14617–14626.
- 6 Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating consequences: brain
7 systems that encode the causal effects of actions. *Journal of Neuroscience*, 28(26),
8 6750–6755.
- 9 Tsujimoto, S., & Sawaguchi, T. (2004). Neuronal representation of response–outcome in the
10 primate prefrontal cortex. *Cerebral Cortex*, 14(1), 47–55.
- 11 Turi, Z., Mittner, M., Opitz, A., Popkes, M., Paulus, W., & Antal, A. (2015). Transcranial direct
12 current stimulation over the left prefrontal cortex increases randomness of choice in
13 instrumental learning. *Cortex*, 63, 145–154.
- 14 Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates
15 of goal-directed learning in the human brain. *Journal of Neuroscience*, 27(15), 4019–
16 4026.
- 17 Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using
18 leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–
19 1432. <https://doi.org/10.1007/s11222-016-9696-4>
- 20 Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2019). Rank-
21 normalization, folding, and localization: An improved \widehat{R} for assessing
22 convergence of MCMC. *ArXiv Preprint ArXiv:1903.08008*.
- 23 Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital
24 prefrontal cortex during performance of a reward preference task. *European Journal*
25 *of Neuroscience*, 18(7), 2069–2081. [https://doi.org/10.1046/j.1460-](https://doi.org/10.1046/j.1460-9568.2003.02922.x)
26 [9568.2003.02922.x](https://doi.org/10.1046/j.1460-9568.2003.02922.x)
- 27 Wasserman, E. A., Franklin, S. R., & Hearst, E. (1974). Pavlovian appetitive contingencies
28 and approach versus withdrawal to conditioned stimuli in pigeons. *Journal of*
29 *Comparative and Physiological Psychology*, 86(4), 616–627.
30 <https://doi.org/10.1037/h0036171>
- 31 Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning?
32 *Annual Review of Psychology*, 48(1), 573–607.
- 33 Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2010). Bayesian
34 parameter estimation in the Expectancy Valence model of the Iowa gambling task.
35 *Journal of Mathematical Psychology*, 54(1), 14–27.
- 36 Williams, D. R., & Williams, H. (1969). AUTO-MAINTENANCE IN THE PIGEON:
37 SUSTAINED PECKING DESPITE CONTINGENT NON-REINFORCEMENT 2.
38 *Journal of the Experimental Analysis of Behavior*, 12(4), 511–520.
- 39 Yoo, A. H., & Collins, A. G. E. (2022). How Working Memory and Reinforcement Learning

1 Are Intertwined: A Cognitive, Neural, and Computational Perspective. *Journal of*
2 Cognitive Neuroscience, 34(4), 551–568. https://doi.org/10.1162/jocn_a_01808