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

Short Title: Working memory load and instrumental learning under Pavlovian influence

Authors:

Heesun Park[1*], Hoyoung Doh[1*], Eunhwi Lee[1], Harhim Park[1], Woo-Young Ahn[1,2]

[1]Department of Psychology, Seoul National University, Seoul, Korea 08826

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

*These authors contributed equally.

Corresponding Author:

Woo-Young Ahn, Ph.D.

Department of Psychology

Seoul National University

Seoul, Korea 08826

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

Number of pages: 44

Number of Figures & Tables: 5/1

Number of words in the abstract: 230

Number of words in the introduction: 910

Number of words in the discussion: 2234

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

Abstract

Research suggests that a fast, capacity-limited working memory (WM) system and a slow, incremental reinforcement learning (RL) system jointly contribute to instrumental learning. Thus, situations that strain WM resources alter instrumental learning: under WM loads, learning becomes slow and incremental, the reliance on computationally efficient learning increases, and action selection becomes more random. It is also suggested that Pavlovian learning influences people's behavior during instrumental learning by providing hard-wired instinctive responses including approach to reward predictors and avoidance of punishment predictors. However, it remains unknown how constraints on WM resources affect instrumental learning under Pavlovian influence. Thus, we conducted a functional magnetic resonance imaging (fMRI) study (N=49) in which participants completed an instrumental learning task with Pavlovian–instrumental conflict (the orthogonalized go/no-go task) both with and without extra WM load. Behavioral and computational modeling analyses revealed that WM load reduced the learning rate and increased random choice, without affecting Pavlovian bias. Model-based fMRI analysis revealed that WM load strengthened RPE signaling in the striatum. Moreover, under WM load, the striatum showed weakened connectivity with the ventromedial and dorsolateral prefrontal cortex when computing reward expectations. These results suggest that the limitation of cognitive resources by WM load promotes slow and incremental learning through the weakened cooperation between WM and RL; such limitation also makes action selection more random, but it does not directly affect the balance between instrumental and Pavlovian systems.

Author Summary

Among multiple decision-making systems of humans and animals, the Pavlovian system is known for promoting automatic and instinctive behaviors. Understanding the Pavlovian influence on decision-making can offer valuable insights into the mechanism of impulsive and addictive behaviors. Previous studies suggested that prefrontal executive control can be important in regulating the Pavlovian influence. We tested if reducing cognitive resources available for executive control modulates the Pavlovian influence, by adding WM load to an instrumental learning task where Pavlovian influence is beneficial in some conditions but detrimental in others. Contrary to our expectation, constraining cognitive resources with WM load failed to significantly change the contribution of Pavlovian system. Nonetheless, with behavioral analysis and computational modeling, we revealed that WM load promotes slower learning and makes decisions noisier in an instrumental learning task with Pavlovian components. FMRI analysis revealed that WM load strengthens the RPE signaling in striatum upon observing the outcome and weakens the functional connectivity between the prefrontal cortex and the striatum before making a decision. The current study contributes to understanding how cognitive resource constraints alter learning and decision-making under Pavlovian influence as well as the neural mechanisms of those effects.

Introduction

The process of learning about the environment from experience and making adaptive decisions involves multiple neurocognitive systems, among which reinforcement learning (RL) and working memory (WM) systems are known to significantly contribute to learning [1–3]. RL processes facilitate “incremental” learning from the discrepancy between actual and predicted rewards, known as reward prediction error (RPE); RL is regarded as a slow but steady process [4]. Dopaminergic activity in the basal ganglia conveys RPEs [5–11], and human imaging studies have found that blood-oxygen-level-dependent (BOLD) signals in the striatum are correlated with RPEs [12–14].

In addition to RL, WM is a crucial component in learning. In particular, WM allows the rapid learning of actions via retention of recent stimulus-action-outcome associations, while RL constitutes a slow learning process [1,15–17]. WM can also offer various inputs to RL, such as reward expectations [18] and models of the environment [19–22] as well as complex states and actions [3,23]. In the brain, the WM system is presumably associated with sustained neural activity throughout the dorsolateral prefrontal cortex (dlPFC) and prefrontal cortex (PFC) [24–29].

Because RL and WM cooperate to promote successful learning, restricting either system can alter learning and the balance between the two systems. In particular, increasing WM load during learning and decision-making can lead to various consequences through the depletion of WM resources. For example, first, instrumental learning becomes slow and incremental under WM load [1,16,30,31]. Limited resources in the WM system cause WM contribution to decline while the RL contribution increases, causing learning to occur more slowly and strengthening the RPE signal in the brain [15,18]. Second, among the multiple RL systems that use varying degrees of WM resources, the demands of WM can be balanced against computationally costly strategies. Otto et al. demonstrated that under WM load, the reliance on computationally efficient model-free learning was increased, compared with model-based learning [32]. Lastly, limited WM resources may cause action selection to

93 become more random and inconsistent. Different values must be compared to inform decision-
94 making during the action selection stage [33], but several studies have reported that WM load
95 may interrupt these processes without affecting valuation itself [34–36].

96 While reductions of WM resources substantially alter instrumental learning, another
97 important factor known to shape instrumental learning is the Pavlovian system. Through the
98 motivation of hard-wired responses, such as active responses to appetitive cues and inhibitory
99 responses toward aversive cues [37–40], the Pavlovian system may facilitate certain
100 instrumental behaviors and impede others. This bias in instrumental learning is known as
101 Pavlovian bias [41–44]. In a widely-used paradigm, Pavlovian bias emerges when the cues
102 used in a instrumental learning task trigger Pavlovian conditioned responses (e.g., active
103 responses to reward cues regardless of action requirements), leading to an increase or
104 decrease in task performance [45,46]. Many computational models of the phenomenon
105 assume that decision-makers consider both instrumentally trained responses and Pavlovian
106 conditioned responses when deciding which response to give in each trial. This decision-
107 making procedure balances the instrumental and Pavlovian learning systems, where the
108 instrumental system is more flexible but computationally more expensive while the Pavlovian
109 system is more rigid but computationally cheaper. This is because the instrumental system
110 learns stimulus-response-outcome associations, whereas the Pavlovian system only learns
111 stimulus-outcome associations and has hard-wired responses for the learned associations.
112 Pavlovian bias is generally presumed to be associated with maladaptive behaviors such as
113 substance use disorder and compulsivity-related disorders [47–51].

114 Although it is well known that increasing WM load alters instrumental learning in
115 several ways, it remains unclear how WM load changes instrumental learning when it is under
116 Pavlovian influence. To investigate this relationship, we conducted a functional magnetic
117 resonance imaging (fMRI) study in which participants completed an instrumental learning task
118 that involved Pavlovian–instrumental conflicts [45], with and without extra WM load.

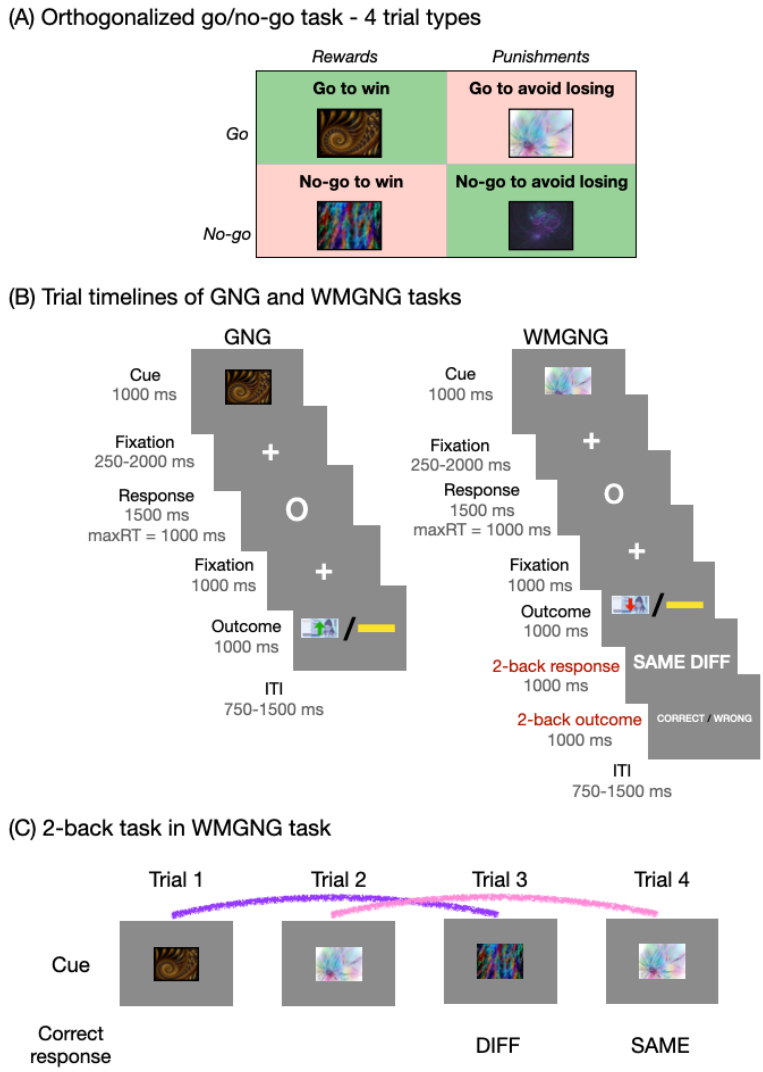
We tested the following three hypotheses. First, if the role of WM in learning is unaffected by Pavlovian influence, WM load will lead to slower learning and increased striatal RPE signals, consistent with previous findings [1,15,16,18]. Second, if WM load leads to a computational trade-off between Pavlovian and instrumental learning, like the one between model-free and model-based learning [32], WM load will enhance Pavlovian bias because the Pavlovian system is known to require fewer resources and to be computationally efficient as an evolutionarily embedded system that learns values as a function of cues, regardless of actions [42]. We also presumed that neural signaling associated with Pavlovian bias would increase under WM load. We focused on regions of the basal ganglia, specifically the striatum and substantia nigra/ventral tegmental area (SN/VTA), which are considered important in Pavlovian bias [45,46,52–54]. Third, if the contribution of WM to consistent action selection remains consistent, WM load will cause action selection to become more random, as in previous studies [34–36]. We tested whether the value comparison signal in the brain would decrease under WM load because consistent action selection may be associated with the extent to which value difference information is utilized during the decision-making process [33,55].

Our behavioral and computational modeling results revealed that Pavlovian bias did not increase under WM load, while learning became slower and more incremental and action selection became more random. We observed increased striatal RPE signaling which was in line with our hypothesis that the contribution of RL will increase under WM load. Further analyses revealed weakened connectivity between the striatal and prefrontal regions under WM load, suggesting diminished cooperation between the WM and RL systems.

Results

The participants (N=56) underwent fMRI imaging while performing an instrumental learning task under a control condition and a WM load condition (**Fig 1**). In the control condition, they participated in the orthogonalized go/no-go (GNG) task [45], a learning task that

146 contained Pavlovian–instrumental conflicts. In the WM load condition, a 2-back task was
147 added to the GNG task; the modified task was named the working memory go/no-go (WMGNG)
148 task (see Materials and Methods for more detail).
149



150

151 **Fig 1**. The GNG and WMGNG tasks. (A) In both tasks, four fractal cues indicated the combination of
152 action (go/no-go) and valence at the outcome (win/loss). (B) In each trial, a fractal cue was presented,
153 followed by a variable delay. After the delay, actions were required in response to a circle, and
154 participants had to decide whether to press a button. After an additional brief delay, the probabilistic
155 outcome was presented, indicating monetary reward (green upward arrow on a ₩1000 bill) or monetary
156 punishment (red downward arrow on a ₩1000 bill). A yellow horizontal bar indicated no win or loss. In
157 the WMGNG task, the original GNG task was followed by a 2-back response and 2-back outcome
158 phases. (C) The participants were asked to indicate whether the cue in the current trial was identical to

the cue in the two preceding trials. Here, because the cue in trial 3 differed from the cue in trial 1, “DIFF” was the correct response. Similarly, because the cue in trial 4 was identical to the cue in trial 3, “SAME” was the correct response. The lines mark two cues for comparison: the purple line indicates that the cues differ, while the pink line indicates that the cues are identical.

Task performance: Decreased performance and learning speed under WM load

Imposing extra WM load with a 2-back task led to a decrease in task accuracy. Participants performed better in the GNG task (M=0.80, SD=0.12) than in the WMGNG task (M=0.72, SD=0.16), as illustrated in **Fig 2A** (paired t-test, $t(48)=3.86$, $p<0.001$, $d=0.55$). Participants’ performance decreased both in the Pavlovian-congruent ("go to win” and “no-go to avoid losing”) and Pavlovian-incongruent (“no-go to win” and “go to avoid losing”) conditions (**Fig S1A**).

We also confirmed that participants exhibited go bias and Pavlovian bias in both tasks, thus replicating the findings of earlier studies [45,52,56–62]. Two-way ANOVA on accuracy, with the factors action (go/no-go) and valence (reward/punishment) as repeated measures for both tasks, revealed a main effect 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 WMGNG task) and action by valence interaction ($F(48)=22.43$, $p<0.001$, $\eta^2=0.12$ in the GNG task, $F(48)=30.59$, $p<0.001$, $\eta^2=0.10$ in the WMGNG task); it showed no effect of valence ($F(48)=0.00$, $p=0.99$, $\eta^2=0.00$ in the GNG task, $F(48)=2.77$, $p=0.103$, $\eta^2=0.01$ in the WMGNG task). In both tasks (**Fig 2B**), participants exhibited superior performances in “go to win” and “no-go to avoid losing” conditions (i.e., Pavlovian-congruent conditions; blue columns) than in “no-go to win” and “go to avoid losing” trials (i.e., Pavlovian-incongruent conditions; red columns). Specifically, in the GNG task, accuracy was higher in the “go to win” (M=0.92, SD=0.12) than “no-go to win” condition (M=0.69, SD=0.35) (paired t-test, $t(48)=4.13$, $p<0.001$, $d=0.59$), and in the “no-go to avoid losing” (M=0.85, SD=0.13) than in the “go to avoid losing” condition (M=0.76, SD=0.18) (paired t-test, $t(48)=3.29$, $p=0.002$, $d=0.47$). Similarly, in the WMGNG task, accuracy was higher in the “go to win” (M=0.82, SD=0.25) than in the “no-go 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 “no-go to avoid losing” ($M=0.79$, $SD=0.16$) than in the “go to avoid losing” condition ($M=0.72$, $SD=0.19$) (paired t-test, $t(48)=2.51$, $p=0.015$, $d=0.36$).

Next, we tested the hypothesis that WM load would decrease learning speed (**Fig 2C**). While the learning curves indicated that participants learned during both tasks, the learning curve was shallower in the WMGNG task than in the GNG task (i.e., WM load reduced learning speed and overall accuracy).

To test the hypothesis that WM load would increase Pavlovian bias (**Fig 2D**), we quantified Pavlovian bias by subtracting the accuracy in Pavlovian-incongruent conditions (“no-go to win” and “go to avoid losing”) from the accuracy in Pavlovian-congruent conditions (“go to win” and “no-go to avoid losing”). No significant difference in Pavlovian bias was observed between the GNG (M=0.32, SD=0.47) and WMGNG (M=0.32, SD=0.40) tasks (paired t-test, t(48)=-0.02, p=0.986, d=0.00). However, participants could have been slower in learning the Pavlovian cue-outcome associations under WM load, as the instrumental learning became slower under WM load. Thus, it should be examined if participants expressed similar levels of Pavlovian bias in the GNG and WMGNG tasks after they learned the cue-outcome associations in both tasks. To this end, we plotted the temporal development of Pavlovian bias across trials (**Fig S2**). We observed a delayed peak in the WMGNG compared to the GNG task, which indicates that acquiring Pavlovian associations could have taken more time under WM load. Nonetheless, we observed similar levels of Pavlovian bias in both tasks after the initial peak. Thus, we concluded that our data do not show noticeable changes in Pavlovian bias under WM load.

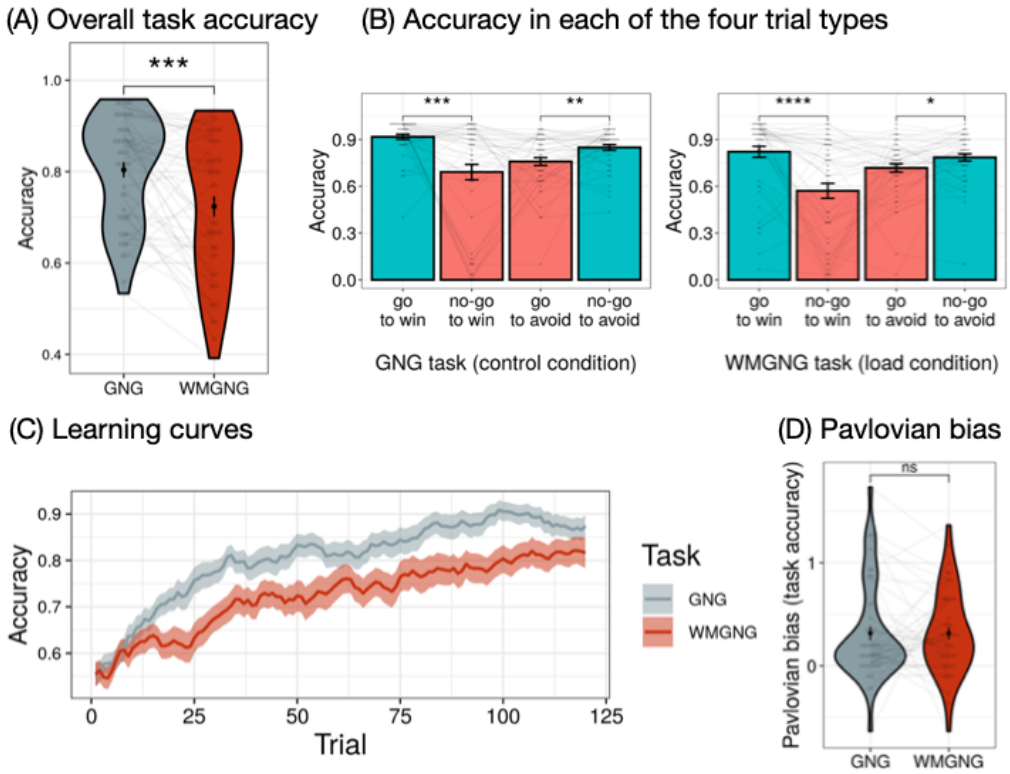


Fig 2. Task performance (N=49). (A) Task accuracies (mean percentages of correct responses) in the GNG and WMGNG tasks show that participants performed better in the GNG task than in the WMGNG task. (B) Accuracy in each of the four trial types between the two tasks demonstrated that participants performed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

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

226 We used a computational modeling approach to test the three hypotheses. For this
227 purpose, we developed eight nested models that assumed different learning rate, Pavlovian
228 bias, or irreducible noise parameters under WM load. These models were fitted to the data
229 using hierarchical Bayesian analysis, then compared using the leave-one-out information
230 criterion (LOOIC), where a lower LOOIC value indicates better out-of-sample predictive
231 accuracy (i.e., better fit) [63]. Importantly, the use of computational modeling allowed us to test
232 our hypothesis that WM load would increase random choices; this would have not been
233 possible if we had performed behavioral analysis alone.

234 Based on earlier studies [45,64], we constructed a baseline model (model 1) that used
235 a Rescorla-Wagner updating rule and contained learning rate (ε), Pavlovian bias, irreducible
236 noise, go bias, and separate parameters for sensitivity to rewards and punishments (Materials
237 and Methods). In the model, state-action values are updated with the prediction error; learning
238 rate (ε) modulates the impact of the prediction error. Reward/punishment sensitivity (ρ) scales
239 the effective size of outcome values. Go bias (b) and cue values weighted by Pavlovian bias
240 (π) are added to the value of go choices. Here, as the Pavlovian bias parameter increases,
241 the go tendency increases under the reward condition whereas the go tendency is reduced
242 under the punishment condition; this results in an increased no-go tendency. Computed action
243 weights are used to estimate action probabilities, and irreducible noise (ξ) determines the
244 extent to which information about action weights is utilized to make decisions. As irreducible
245 noise increases, action probabilities will be less reflective of action weights, indicating that
246 action selection will become more random.

247 In models 2, 3, and 4, we assumed that WM load affects only one parameter. For
248 example, in model 2, a separate Pavlovian bias parameter (π_{wm}) was assumed for the WM
249 load condition. Models 3 and 4 assumed different learning rates (ε_{wm}) and irreducible noise
250 (ξ_{wm}) parameters in their respective WM load conditions. In models 5, 6, and 7, we assumed
251 that WM load would affect two parameters: model 5 had different Pavlovian bias (π_{wm}) and

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

The full model (model 8) was the best model (**Fig 3A**; **Table S2**). In other words, participant behavior could be best explained when separate parameters were included for Pavlovian bias, learning rate, and irreducible noise parameters. Next, we analyzed the parameter estimates of the best-fitting model; we focused on comparing the posterior distributions of the parameters that were separately fitted in the two tasks (**Fig 3B**). The parameters were considered credibly different from each other if the 95% highest density intervals (HDI) of the two distributions showed no overlap [65]. **Fig 3B** illustrates that Pavlovian bias was not credibly different between the two tasks, consistent with our behavioral results that failed to show a change in Pavlovian bias under WM load. Conversely, the learning rate was credibly lower, while irreducible noise was credibly greater in the WMGNG than in the GNG task. These results support our hypotheses that WM load would reduce learning rate and that it would increase random choices. While the best model was the full model that assumed separate Pavlovian bias in the two tasks, no credible group difference was observed between these parameters. This is presumably because the full model was able to capture individual variations among participants (**Fig S5**), despite the lack of credible difference in the group-level estimates between the two tasks. As expected, the 95% HDIs of go bias, reward sensitivity, and punishment sensitivity did not include zero, indicating that the participants exhibited go bias and reward/punishment sensitivity (see Supporting Information for the posterior distributions of individual parameters; **Fig S4-S7**).

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

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

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

Next, we examined if our model predicts the observed decrease in task performance under WM load both in the Pavlovian-congruent and Pavlovian-incongruent conditions. **Fig S1B** shows that our model indeed predicts the lower task performance in both types of conditions. This result is in line with our result that choice randomness increased under WM load. We can expect that an increase in randomness would result in a lower accuracy unless

the accuracy was below the chance level in the first place. While the accuracy decrease could also be associated with the lower learning rate under WM load, we need to be cautious in this interpretation because lower learning rates might instead increase the accuracy by making learning more robust against noise.

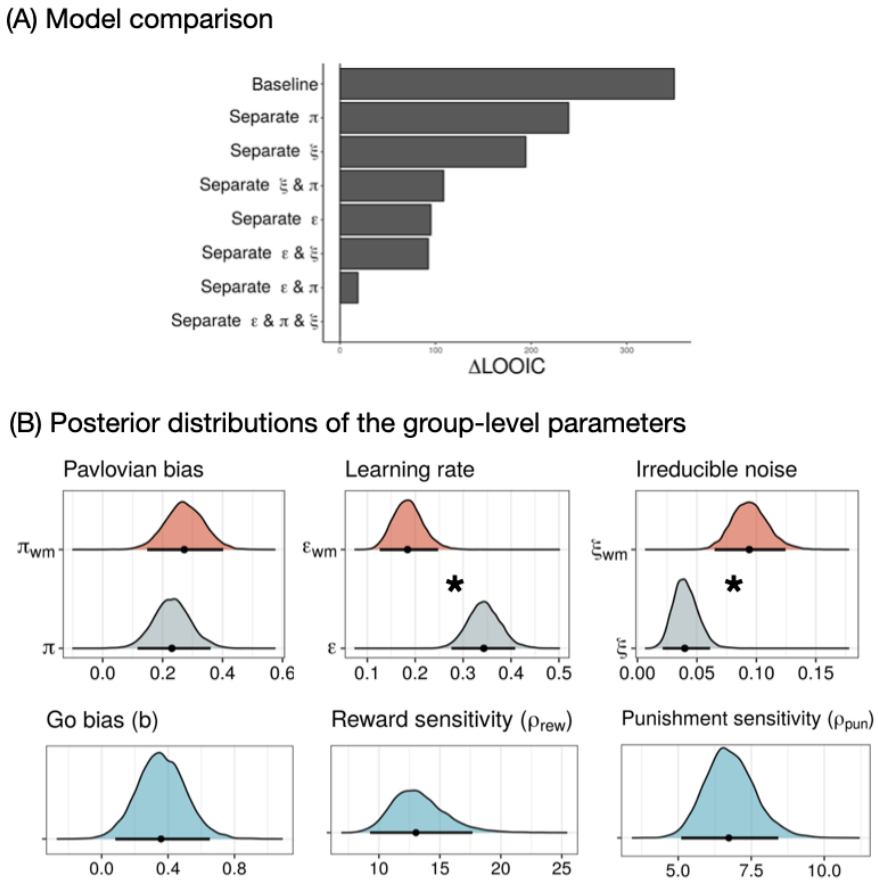


Fig 3. Model comparison results and posterior distribution of the group-level parameters of the best-fitting model (N=49). (A) Relative LOOIC difference indicates the difference in LOOIC between the best-fitting model and each of the other models. The best-fitting model was the full model, which assumed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

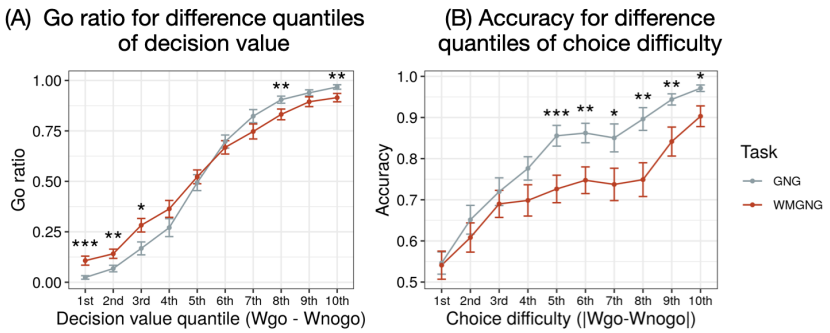


Fig 4. Choice randomness (N=49). (A) Mean percentage of go choices for different quantiles of action weight differences (W_{go} - W_{nogo}) between "go" and "no-go" choices, where higher quantiles indicate higher decision values for "go" choices. Under WM load, the increase in go ratio according to quantile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

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

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

346 increase the contribution of the RL system and decrease the contribution of the WM system.
347 We also tested the hypothesis that WM load would lead to a greater brain activation associated
348 with Pavlovian bias, but we found no main effect of Pavlovian bias between the GNG and
349 WMGNG tasks (WMGNG > GNG [Pavlovian-congruent > Pavlovian-incongruent]) within the
350 striatum or SN/VTA ($p < 0.05$ SVC, a priori ROIs). Note that previous studies showed no
351 significant result for the same contrast (Pavlovian-congruent > Pavlovian-incongruent) within
352 the same regions [45]. As a test of our hypothesis that WM load would increase choice
353 randomness, we examined whether the WM load led to less brain activation associated with
354 value comparison (which would indicate a decreased contribution of the value comparison
355 signal and thus relatively more contribution of randomness). We observed no main effect of
356 WM load on random choice (WMGNG > GNG [W_{chosen} - $W_{unchosen}$]) within the ventromedial
357 prefrontal cortex (vmPFC; $p < 0.05$ SVC, a priori ROI). See Supporting Information for further
358 details regarding these findings (**Table S5**).

359 Increased RPE signals under WM load may indicate reduced WM contribution and
360 increased RL contribution to learning because of the load, suggesting diminished cooperation
361 between the two systems for learning. Therefore, we conducted a psychophysiological
362 interaction (PPI) analysis [66] using the gPPI toolbox [67] to test whether functional
363 connectivity between areas associated with RL and WM systems would weaken under WM
364 load. Specifically, we explored differences between the two tasks in terms of functional
365 coupling between the striatum, which showed increased RPE signaling under WM load during
366 the feedback phase, and other regions when computing reward expectations. The striatum
367 showed significantly decreased connectivity with the vmPFC (MNI space coordinates $x = 13$,
368 $y = 56$, $z = 0$, $Z = -4.90$, $p < 0.05$ whole-brain cluster-level familywise error rate (FWE)) and
369 dlPFC (MNI space coordinates $x = -22$, $y = 63$, $z = 23$, $Z = -4.24$, $p < 0.05$ whole-brain cluster-
370 level FWE, **Fig 5B**, **Table S6**) in the WMGNG task, compared with the GNG task.

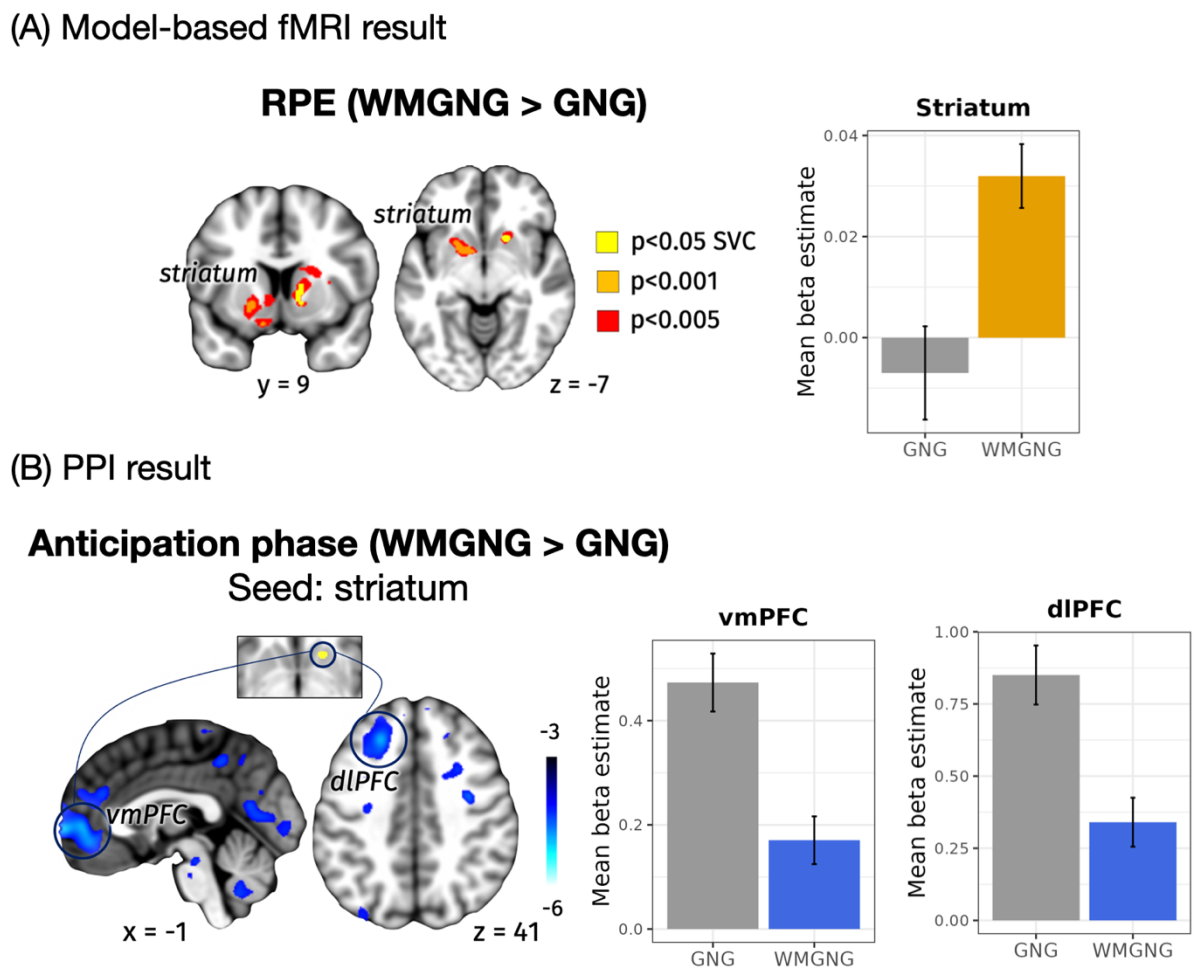


Fig 5. fMRI results (N=44). (A) In the predefined ROI-based analysis, RPE signaling in the striatum was stronger in the WMGNG task than in the GNG task. The left figure shows significant regions at p < 0.05 (SVC) in yellow. The right figure shows an increased mean beta estimate in the striatum under WM load. (B) The left figure shows that functional connectivity between the striatum (seed region, top) and prefrontal regions, including vmPFC (bottom left) and dlPFC (bottom right), was weaker in the WMGNG task than in the GNG task when computing reward expectation (p < 0.05, whole-brain cluster-level FWE). The right figures show decreased mean beta estimates in vmPFC and dlPFC under WM load. In all figures, error bars are means ± standard errors of the means. Overlays are shown with a threshold of p < 0.001 (uncorrected). Color scale indicates t-values.

Discussion

In this study, our main objective was to elucidate the neurocognitive effects of WM load on instrumental learning that involves Pavlovian–instrumental conflicts. We hypothesized that

under WM load, 1) learning rate would decrease and RPE signals would become stronger, 2) Pavlovian bias would increase, and 3) action selection would become more random. First, we found that the limitation of WM resources with WM load led to a decrease in the learning rate and increases in striatal RPE signals. The striatum, which subsequently showed stronger RPE signals under WM load, demonstrated weakened functional connectivity with prefrontal regions including the dlPFC and vmPFC, during reward prediction. WM load also increased random action selection. However, Pavlovian bias did not increase under WM load, suggesting that WM load did not affect the balance between Pavlovian and instrumental systems.

Decreased contribution of the WM system and increased contribution of the RL system under WM load

The effect of WM load on instrumental learning remained consistent despite Pavlovian bias. In particular, our behavioral analysis revealed a shallower learning curve under WM load (**Fig 2C**); modeling analysis confirmed that WM load reduced learning rate (**Fig 3**). We also found that RPE signaling in the striatum was strengthened under WM load (**Fig 5A**), consistent with previous findings [15,18].

These findings support theoretical explanations that WM load reduces the contribution of the WM system and increases the contribution of the RL system. In one such explanation, the rapid and immediate WM system learns in parallel with the slow and incremental RL system by directly storing the observed associations between states and actions [1,15–17,68]. WM and RL systems compete with each other based on their reliability in a given situation. Under WM load, the fast and capacity-limited WM system becomes less reliable; thus, the slow and incremental RL system supersedes the WM system, causing learning to occur more slowly and incrementally [1,16,30]. Such a shift toward slow, incremental learning can be adaptive because it is more reliable under resource-limited conditions and robust against noise in stationary environments [69]. Another explanation posits that RL computations are intertwined with WM; WM feeds reward expectations to the RL system [70–73] to improve reward prediction precision, thereby reducing RPEs and improving learning efficiency [15,18].

412 WM load constrains the WM system, which leads to a lower precision in reward prediction and
413 thus greater RPEs.

414 These ideas are further supported by our finding that the striatum showed weakened
415 functional connectivity with the dlPFC during reward prediction under WM load (**Fig 5A**). This
416 result is in line with theoretical predictions (1) that the WM system would become unreliable
417 and thus its contribution to instrumental learning would decrease and (2) that that the reward
418 prediction information delivered from the WM to the RL system would become less precise
419 and the RL system would rely less on it. However, further research is necessary to
420 demonstrate the directionality of functional connectivity between the two systems during
421 reward prediction; frontostriatal connectivity is reportedly bidirectional, such that the striatum
422 may also provide prefrontal regions with inputs that relate to reward information [74,75].

423 Notably, we observed weakened connectivity between the vmPFC and the striatum.
424 The vmPFC has been identified as a critical neural correlate of value-based decision-making;
425 it integrates reward predictions [71], represents value signals or decision value [76–79], and
426 affects reward anticipation/processing in the striatum [80,81]. Future research is needed to
427 clarify which function of vmPFC was influenced by our WM load manipulation.

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

429 Contrary to our hypothesis, WM load did not influence Pavlovian bias. Behavioral and
430 modeling results showed that Pavlovian bias did not significantly differ between the GNG and
431 WMGNG tasks (**Fig 2D; Fig 3; Fig S2**), while fMRI analysis revealed that neural signaling
432 associated with Pavlovian bias did not significantly differ between the two tasks (**Table S6**).
433 These findings indicate that the brain did not exhibit greater reliance on the computationally
434 efficient system under WM load, in contrast to the results of previous studies [32]. We identified
435 two possible explanations for this discrepancy. First, instrumental and Pavlovian learning in
436 our task require similar amounts of WM resources; second, the WM system may not be related
437 to modulating the balance between Pavlovian and instrumental systems.

In the first potential explanation, the amounts of WM resources necessary might have been similar for Pavlovian and instrumental learning in our task. Otto et al. showed that WM resource constraints promote the use of model-free compared to model-based system [32]. In their task, the model-based system required WM resources to retain the internal model of the task, which was its transition structure [20,82–85], whereas the model-free system did not maintain any model of the task [82,83,86]. Thus, WM resource requirement were quite different between the two systems. In our paradigm, on the other hand, the WM resource requirements might not have been similar between the Pavlovian and the instrumental system. The Pavlovian system learned state-outcome associations, while the instrumental system learned state-action-outcome associations [42,87]. The difference in WM resource requirement was only the extra one dimension for the instrumental system. Instead of relying more on the Pavlovian system, which might only provide little boost in resource efficiency in our task, the participants may simply have favored slow, incremental learning, which requires small cognitive resources.

In the second potential explanation, WM resources may be unimportant with respect to modulating the Pavlovian–instrumental interaction, despite earlier studies' suggestions to the contrary. Several studies have proposed that prefrontal WM control systems are crucial for controlling Pavlovian bias. Electroencephalogram studies demonstrated that midfrontal theta oscillations are important for controlling Pavlovian bias [64,88], suggesting top-down prefrontal control over Pavlovian bias [64]. Furthermore, recruitment of the inferior frontal gyrus (IFG) is involved in appropriate response inhibition, helping to overcome Pavlovian bias [45]. Finally, there is indirect evidence that administration of levodopa, which increases dopamine levels, reduced Pavlovian influences on instrumental learning; such a reduction was speculated to result from increased dopamine levels in the PFC, which may have facilitated the operation of prefrontal WM functions [59]. A related finding suggested that genetic determinants of prefrontal dopamine function may be important in overcoming Pavlovian bias [62].

While the results of the present study appear to contradict these findings, several complex possibilities exist. In particular, although previous findings implied the involvement of prefrontal mechanisms (e.g., model-based prefrontal control [64] and WM [46,59] in controlling the Pavlovian system, they did not directly suggest active recruitment of the prefrontal WM system. First, while Cavanagh et al. speculated that midfrontal theta power could be indicative of “model-based top-down prefrontal control” [64], a subsequent study by Swart et al. suggested that midfrontal theta signals could only be involved in the detection of conflict by signaling “the need for control” [88,89], rather than being a source of direct control. Next, the IFG showed an increased BOLD response only in the “no-go” condition [45], implying that the IFG is important for “inhibitory” motor control (i.e., as a brake [90]); it does not participate in active maintenance or representation of goal-directed behaviors including both “go” and “no-go,” which would be more closely associated with WM [29,91,92]. Finally, elevated dopamine levels should be cautiously interpreted as improvements in prefrontal WM function [59]. While dopamine has been shown to enable successful cognitive control in the prefrontal cortex, it may have three roles: gating behaviorally relevant sensory signals; maintaining and manipulating information in WM to guide goal-directed behavior; and relaying motor information to premotor areas for action preparation [93]. Moreover, distinct mechanisms have been known to modulate the influence of dopamine on WM in the PFC through distinct types of dopamine receptors [93]. Thus, there may be several ways to interpret the observation that dopamine level [59] or function [62] was associated with the modulation of Pavlovian influences. Considerable research is needed to fully understand the mechanisms by which dopamine levels affect Pavlovian bias. Alternatively, the role of prefrontal WM in controlling Pavlovian bias may not require vast resources. It may only be responsible for signaling a need for control [88], promoting response inhibition [45], or influencing subcortical areas (e.g., the striatum and subthalamic nucleus [94,95]).

Our result help refine the understanding of elevated Pavlovian bias in individuals suffering from certain mental disorders [48,49,96]. Researchers often assume the increased Pavlovian bias in clinical populations is, at least in part, a consequence of their impaired

executive control, which hinders their goal-directed behavior [47,51]. However, the causal relationship between executive control and the Pavlovian influence on behavior has not been actively investigated. In this study, we showed that constraining executive control with WM load did not increase the Pavlovian bias. It suggests that executive control deficits observed in clinical populations may not directly contribute to their heightened Pavlovian bias. Instead, clinical populations might perceive reward or punishment (or both) as more motivationally salient than healthy individuals, which leads to stronger urges for Pavlovian responses. Nonetheless, one should be cautious in this interpretation because the executive control impairments in clinical populations might be more severe than (or qualitatively different from) the ones elicited with WM load in our study.

One limitation of our study design is that the acquisition of Pavlovian cue-outcome associations and the expression of Pavlovian bias are not clearly separated. A more sophisticated way of comparing Pavlovian bias would first guarantee that participants acquired the cue-outcome associations to similar degrees in both tasks. Although we conducted a similar analysis (**Fig S2**), it is not a complete solution to the issue because we cannot separate the two components with the data from our task. Future studies can address this issue by using a between-subject design, where the extra WM load are applied to only one group of participants after all participants perform enough trials of the GNG task. Another way would be to use the Pavlovian-instrumental transfer paradigm, which is similar to our paradigm but separates the Pavlovian learning, instrumental learning, and Pavlovian-instrumental transfer phases. Experimenters can manipulate the WM load during the transfer phase to examine the WM load effect on the expression of Pavlovian bias.

Increased random choices under WM load

Another notable finding was that random choice increased under WM load. Our modeling analysis revealed that irreducible noise parameter estimates were greater in the WMGNG task than in the GNG task (**Fig 3**), suggesting increased random action selection under WM load. Further analysis using the modeling outputs revealed that participants'

520 choices were less affected by the relative value difference between the “go” and “no-go”
521 actions under WM load (**Fig 4A**). Moreover, analysis using the absolute difference between
522 the two options (**Fig 4B**) revealed that the increase in accuracy became smaller as the
523 absolute difference increased (i.e., the choice became easier). Both findings suggest that WM
524 involvement led to an increase in random choices, regardless of value comparison and choice
525 difficulty.

526 Our findings are broadly consistent with the results of previous studies concerning the
527 role of WM and prefrontal regions in action selection and execution [97–104]. In particular,
528 several studies have demonstrated that the interruption of WM function via WM load could
529 increase the frequency of random choices in value-based decision-making tasks [34–36].
530 Additionally, transcranial direct current stimulation, a brain stimulation method, over the left
531 PFC led to increased random action selection during an RL task, suggesting that the prefrontal
532 WM component influenced action selection [105]. Furthermore, the importance of WM in
533 action selection during learning tasks is supported by the indirect evidence that individual
534 differences in WM capacity were correlated with appropriate exploratory action selection in
535 multi-armed bandit tasks [106]. Overall, the reduced availability of WM resources because of
536 WM load in our study may have compromised the participants’ abilities to actively represent
537 their current goals and actions, leading to reduced WM control over consistent choice based
538 on value computation.

539 No significant neural correlates were identified with respect to the increased random
540 choices. We assumed that random action selection would be associated with the reduced
541 sensitivity to value difference or value comparison between the two options (“go” and “no-go”)
542 [55]; thus, we hypothesized that value comparison signals would decrease under WM load.
543 Contrary to our hypothesis, no significant differences in value comparison signaling in ROIs
544 were observed between GNG and WMGNG tasks. There are several possible explanations
545 for this null finding. Our assumption of value sensitivity may not be the source of the random
546 choice observed here. Alternatively, subsequent attentional lapse [107,108] or value-

547 independent noise [109] may have led to inconsistent action selection despite the presence of
548 value comparison signals. Further research is necessary to distinguish these possibilities.
549 In summary, the present study has shown that WM load compromises overall learning
550 by reducing learning speed via weakened cooperation between RL and WM; it also increases
551 random action selection without affecting the balance between Pavlovian and instrumental
552 systems. To our knowledge, this is the first study to investigate the neurocognitive effect of
553 WM load during interactions between Pavlovian and instrumental systems. By investigating
554 how learning and decision-making using different systems are altered in the presence of WM
555 load and by linking such behaviors to their underlying neural mechanisms, this study
556 contributes to our understanding of how distinct cognitive components interact with each other
557 and synergistically contribute to learning. Because impairments in learning, balance among
558 multiple systems, and action selection have been reported in various neurological and
559 psychiatric disorders [2,110], our findings represent an important step toward improved
560 understanding of various symptoms.
561

562 ## Materials and Methods

563 ### Participants

564 Fifty-six adults participated in this study (34 women; 24.5±3.6 years old). All
565 participants were healthy, right-handed; they had normal or corrected-to-normal visual acuity.
566 They were screened prior to the experiment to exclude individuals with a history of
567 neurological, or psychiatric illness. All participants provided written informed consent, and the
568 study protocol was approved by the Institutional Review Board of Seoul National University.
569 The behavioral analysis included 49 participants (29 women; 24.3±3.3 y.o); the fMRI
570 analysis included 44 participants (27 women; 24.2±3.3 y.o). Four participants were excluded
571 because of technical issues; one participant was excluded because they slept during the task.
572 Two participants were excluded because of poor performance in the 2-back task since the
573 results in the dual-task paradigm could only be valid and interpretable when participants

574 actually performed both tasks. The accuracy cutoff was 0.575, a value that rejects the null
575 hypothesis that participants would randomly choose one of two options. After assessment of
576 preprocessed image quality, five participants were excluded from the fMRI analysis because
577 of head movements in the scanner, which can systematically alter brain signals; four out of
578 these five were excluded because the mean framewise displacement exceeded 0.2 mm [111],
579 while the remaining one was excluded after visual assessment of carpet plots [112].

580 **Experimental design and task**

581 The experiment included two tasks: the original GNG task [45] and the GNG task
582 paired with the 2-back task as a secondary task. The order of task completion was
583 counterbalanced among participants; some participants performed the GNG task first while
584 others began with the WMGNG task. The GNG and WMGNG tasks consisted of two blocks
585 (four blocks in total); each block consisted of 60 trials. Therefore, each task contained 120
586 trials (240 trials in total). Participants underwent fMRI while performing the tasks for
587 approximately 50 min, with a short (~60 s) break after each set of 60 trials. Before scanning,
588 participants performed 10 practice trials for each of the GNG and WMGNG tasks to help them
589 become accustomed to the task structure and response timing. Participants received
590 additional compensation based on their accuracy in the two tasks, along with the standard
591 participation fee at the end of the experiment. We used different sets of fractal stimuli for the
592 practice, first, and second tasks.

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

594 Four trial types were implemented depending on the nature of the fractal cue (**Fig 1A**):
595 press a button to gain a reward (go to win); press a button to avoid punishment (go to avoid
596 losing); do not press a button to earn a reward (no-go to win); and do not press a button to
597 avoid punishment (no-go to avoid losing). The meanings of fractal images were randomized
598 among participants.

599 Each trial consisted of three phases: fractal cue presentation, response, and
600 probabilistic outcome. **Fig 1B** illustrates the trial timeline. In each trial, participants were

presented with one of four abstract fractal cues for 1000 ms. After a variable interval drawn from a uniform probability distribution within the range of 250-2000 ms, a white circle was displayed on the center of the screen for 1000 ms. When the circle appeared, participants were required to respond by pressing a button or not pressing a button. Next, the outcome was presented for 1000 ms: a green arrow pointing upwards on a ₩1000 bill indicated monetary reward, a red arrow pointing downwards on a ₩1000 bill indicated monetary punishment, and a yellow horizontal bar indicated no reward or punishment.

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

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

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

Computational modeling

Baseline RL model with Pavlovian bias

We adopted a previously implemented version of an RL model [45] that can model Pavlovian bias and choice randomness as well as learning rate. In our baseline model, we assumed no difference in parameters between the control and load conditions.

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

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

where ε is the learning rate. The learning rate (ε) is a step size of learning [4] that modulates how much of the prediction error, a teaching signal, is incorporated into the value update.

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

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

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

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

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

If the Pavlovian bias parameter (π) is positive, it increases the action weight of "go" in the reward conditions because $V_t(s_t)$ is positive. In the punishment conditions, positive π decreases the action weight of "go" because $V_t(s_t)$ is negative. A value of π close to 0 means that the agent relies mostly on the instrumental system (and the go bias), while a larger value of π means that the Pavlovian system also has a great impact on the agent's action selection.

Action probabilities were dependent on these action weights $W(a_t,s_t)$, which were passed through a squashed softmax [4]:

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

where ξ was the irreducible noise in the decision rule; it was free to vary between 0 and 1 for all models. The irreducible noise parameter explains the extent to which information about action weights is utilized in making a choice. As the irreducible noise increases, the influence of the difference between the action weights is reduced, indicating that action selection becomes random. We did not include the widely used inverse temperature parameter because ρ_{rew} and ρ_{pun} serve the same role as inverse temperature, deciding the level of determinism when generating actions based on action weights [56]. On the other hand, ξ represents the constant randomness regardless of action weights, and has been widely used by researchers using similar models to ours [45,56,64,113,114].

Additional models

To test our hypotheses regarding the effects of WM load on parameters, we constructed seven additional nested models assuming different Pavlovian biases (π), learning rate (ε), and irreducible noise (ξ) under WM load (**Table 1**). Model 1 is the baseline model. Model 2 assumed a separate Pavlovian bias parameter (π) for the WM load condition. Similarly, models 3 and 4 assumed different learning rates (ε) and irreducible noises (ξ) in the WMGNG block, respectively. To address the possibility that two of the three parameters would be affected by the WM load, we constructed three additional models with eight free parameters: model 5 with different Pavlovian bias (π) and learning rate (ε); model 6 with different Pavlovian bias (π) and irreducible noise (ξ); and model 7 with different learning rate (ε) and irreducible noise (ξ). Finally, we constructed the full model, which assumed that all of these three parameters would be affected by WM load, leading to nine free parameters.

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

Procedures for model fitting and model selection

Model parameters were estimated using hierarchical Bayesian analysis (HBA). Group-level distributions were assumed to be normally distributed, with mean and standard deviation parameters set as two free hyperparameters. We employed weakly informative priors to minimize the influences of those priors on the posterior distributions [65,115]. Additionally, for parameter estimation, the Matt trick was used to minimize the dependence between group-level mean and standard deviation parameters; it also facilitated the sampling process [116]. Moreover, bounded parameters such as learning rates and irreducible noise ($\in$ [0, 1]) were estimated within an unconstrained space; they were then probit-transformed to the constrained space, thus maximizing MCMC efficiency within the parameter space [115,117].

We ran four independent chains with 4000 samples each, including 2000 warm-up samples (i.e., burn-in) to ensure that the parameters converged to the target distributions. Four chains were run to ensure that the posterior distributions were not dependent on initial starting points [118]. We visually checked convergence to target distributions by observing trace plots (**Fig S3**) and computing the R statistics - a measure of convergence across chains [119]. R statistics were < 1.1 for all models, indicating that the estimated parameter values converged to their target posterior distributions (**Table S1**).

Models were compared using the LOOIC, which is an information criterion calculated from the leave-one-out cross-validation [63]. This method is used to estimate the out-of-sample predictive accuracy of a fitted Bayesian model for model comparison and selection.

The LOOIC is computed using the log-likelihood evaluated from posterior distributions or simulations of the parameters. The R package loo [63], which provides an interface for the approximation of leave-one-out cross-validated log-likelihood, was used to estimate the LOOIC for each model. Lower LOOIC values indicate better fit.

fMRI scans: acquisition and protocol

fMRI was performed on the same scanner (Simens Tim Trio 3 Tesla) using a 32-channel head coil across all participants. A high-resolution T1-weighted anatomical scan of the whole brain resolution was also acquired for each participant (TR = 2300ms, TE = 2.36ms, FOV = 256mm,1mm×1mm×1mm) to enable spatial localization and normalization. The participant's head was positioned with foam pads to limit head movement during acquisition. Functional data was acquired using echo-planar imaging (EPI) in four scanning sessions containing 64 slices (TR = 1500ms, TE = 30ms, FOV = 256mm, 2.3mm × 2.3mm × 2.3mm, multiband acceleration factor = 4). For the GNG task, functional imaging data were acquired in two separate 277-volume runs, each lasting about 7.5 min. For the WMGNG task, data were acquired in two separate 357-volume runs, each lasting about 9.5 min.

fMRI scans: general linear models

Preprocessing was performed using fMRIPrep 20.2.0 (RRID:SCR_016216) [120,121], which is based on Nipype 1.5.1 (RRID:SCR_002502) [122,123]. Details of preprocessing with fMRIPrep are provided in Supporting Information. Subsequently, images were smoothed using a 3D Gaussian kernel (8mm FWHM) to adjust for anatomical differences among participants. BOLD-signal image analysis was then performed using SPM12 [http://www.fil.ion.ucl.ac.uk/spm/] running on MATLAB v9.5.0.1067069(R2018b).

We built participant-specific GLMs, including all runs – two runs for the GNG block and two runs for the WMGNG block – and calculated contrasts to compare the two blocks at the individual level. The first-level model included six movement regressors to control the movement-related artifacts as nuisance regressors. Linear contrasts at each voxel were used

722 to obtain participant-specific estimates for each effect. These estimates were entered into
723 group-level analyses, with participants regarded as random effects, using a one-sample t-test
724 against a contrast value of 0 at each voxel. The group-level model included covariates for
725 gender, age, and the task order. For all GLM analyses, we conducted ROI analysis; the results
726 were corrected for multiple comparisons using small volume correction (SVC) within ROIs.

727 *GLM1:* GLM1 was used to test the hypothesis that WM load would increase the
728 contribution of the slow, incremental RL system because the resources for the fast WM system
729 are constrained. We expected an stronger RPE signaling in the striatum in the WMGNG than
730 the GNG task. GLM was implemented by the model-based fMRI approach and included the
731 following regressors: (1) cue onset of “go to win” trials, (2) cue onset of “no-go to win” trials,
732 (3) cue onset of “go to avoid losing” trials, (4) cue onset of “no-go to avoid losing” trials, (5)
733 target onset of “go” trials, (6) target onset of “no-go” trials, (7) outcome onset, (8) outcome
734 onset parametrically modulated by the trial-by-trial RPEs, and (9) wait onset (i.e., inter-trial
735 interval). The regressor of interest was “RPE”; we compared the main effect of RPE between
736 two tasks ($RPE(8)_{WMGNG}$ - $RPE(8)_{GNG}$). RPE regressors were calculated by subtracting the
737 expected values (Q) from the outcome for each trial. Here, the outcome was the product of
738 feedback multiplied by reward/punishment sensitivity. The a priori ROI was the striatum, which
739 is widely known to process RPE [12,14,124]. The striatum ROI was constructed anatomically
740 by combining the AAL3 definitions of bilateral caudate, putamen, olfactory cortex, and nucleus
741 accumbens.

742 *GLM2:* GLM2 was used to test the hypothesis that neural responses associated with
743 Pavlovian bias would increase under WM load. Specifically, the GLM examined whether the
744 difference between the anticipatory response to fractal cues in Pavlovian-congruent trials and
745 Pavlovian-incongruent trials was greater in the WMGNG task than in the GNG task in regions
746 associated with Pavlovian bias. Therefore, GLM included the following regressors: (1) cue
747 onset of “go to win” trials, (2) cue onset of “no-go to win” trials, (3) cue onset of “go to avoid
748 losing” trials, (4) cue onset of “no-go to avoid losing” trials, (5) target onset of “go” trials, (6)
749 target onset of “no-go” trials, (7) outcome onset of win trials, (8) outcome onset of neutral trials,

750 (9) outcome onset of loss trials, (10) wait onset (i.e., inter-trial interval). We compared the main
751 effect of Pavlovian bias (Pavlovian-congruent trials - Pavlovian-incongruent trials) between
752 two tasks ($[(1) + (4) - ((2) + (3))]_{WMGNG} - [(1) + (4) ((2) + (3))]_{GNG}$). The a priori ROIs were the
753 striatum and SN/VTA, which are considered important in Pavlovian bias [45,46,52–54]. The
754 striatum ROI was constructed anatomically by combining the AAL3 definitions of bilateral
755 caudate, putamen, olfactory cortex, and nucleus accumbens. Furthermore, the SN/VTA was
756 constructed by combining the AAL3 definitions of bilateral SN and VTA.

757 *GLM3:* GLM3 was used to the hypothesis that there would be less neural responses
758 associated with value comparison signals under WM load. GLM3 was also implemented with
759 a model-based fMRI approach: (1) cue onset of all trials, (2) cue onset parametrically
760 modulated by the trial-by-trial decision values ($W_{chosen} – W_{unchosen}$), (3) target onset of "go" trials,
761 (4) target onset of "no-go" trials, (5) outcome onset, and (6) wait onset (i.e., inter-trial interval).
762 Decision value regressors were calculated by subtracting the action weights of the unchosen
763 option ($W_{unchosen}$) from the action weights of the chosen option (W_{chosen}). We then compared
764 the main effect of decision value between two blocks ($(2)_{WMGNG}$-$(2)_{GNG}$). The a priori ROI for
765 GLM3 was the vmPFC, which was suggested as a region that encodes the relative chosen
766 value ($W_{chosen} – W_{unchosen}$) [125,126]. Here, ROI masks were created by drawing a sphere with
767 a diameter of 10 mm around the peak voxel reported in the previous studies ([-6,48,-8] for
768 vmPFC [125]).

769 *PPI analysis:* In addition to GLMs, we used PPI analysis to test whether WM load led
770 to reduced cooperation between WM and RL systems for learning [15,18] by using PPI
771 analysis. Here, to examine differences between the two blocks in terms of functional coupling
772 between the prefrontal areas and the area computing RPE after choices, we performed PPI
773 analysis using the gPPI toolbox [67]; the physiological variable was the time course of the
774 striatum in the anticipation phase, and the psychological variable was the effect of WM load
775 during the same phase. As a seed region (i.e., a physiological variable), the cluster striatum
776 ROI (peak MNI space coordinates x = 13, y = 14, z = -3) was derived from the results of GLM1,
777 which revealed stronger RPE signaling in the WMGNG task than in the GNG task during the

feedback phase. The entire time series throughout the experiment was extracted from each participant in the striatum ROI. To create the PPI regressor, these normalized time series were multiplied by task condition vectors for the anticipation phase, which consisted of the cue representation and fixation phases as in GLM1. A GLM with PPI regressors of the seed region was thus generated together with movement regressors. The effects of PPI for each participant were estimated in the individual-level GLM; the parameter estimates represented the extent to which activity in each voxel was correlated with activity in the striatum during the anticipation phase. The contrast was constructed by subtracting activity during the anticipation phase in the GNG task from activity in the WMGNG task (WMGNG vs. GNG in the anticipation phase). Individual contrast images for functional connectivity were then computed and entered into one-sample t-tests in a group-level GLM together with nuisance covariates (i.e., gender, age, and task order). Whole-brain cluster-level FWE correction was applied for PPI analysis.

Corrections for multiple comparisons

For each GLM, we performed whole-brain FWE corrections at the cluster level (corrected $p < 0.05$, with the height threshold of $p < 0.001$) for multiple comparisons [127,128] using SPM12. For a priori ROIs, we used SVC using SPM12.

Data availability

Behavioral data and codes for behavioral analysis and modeling are available at https://github.com/CCS-Lab/project_WMDM_public. Unthresholded t maps are available on Neurovault: https://identifiers.org/neurovault.collection:12860. We have shared raw neuroimaging data in OpenNeuro: https://openneuro.org/datasets/ds004647/versions/1.0.2.

Supporting Information

Text S1. fMRIPrep details.

Text S2. fMRI results details.

Fig S1. Performance in the Pavlovian-congruent and Pavlovian-incongruent conditions (N=49). (A) Participants' accuracy in the Pavlovian-congruent and Pavlovian-incongruent conditions, separately for the GNG and WMGNG tasks. Participants showed lower accuracy in the WMGNG compared to the GNG task both in the Pavlovian-congruent and Pavlovian-incongruent conditions. (B) Prediction from the best model regarding the accuracy in the Pavlovian-congruent and Pavlovian-incongruent conditions. We used a one-step ahead prediction for generating the model predictions. In line with our participants' data, our model predictions showed lower accuracy in the WMGNG than the GNG task in both types of conditions. In all figures, error bars indicate means ± standard errors of the means.

Fig S2. Temporal development of Pavlovian bias (N=49). The behavioral measure of Pavlovian bias, which is the difference in accuracy between the Pavlovian-congruent and Pavlovian-incongruent conditions, plotted as a function of trial separately for the GNG and WMGNG tasks. While the peak in the WMGNG task appeared later than the GNG task, Pavlovian bias was similar between the tasks after the peaks. Moving average smoothing was applied with filter size 3. Lines indicate group means and ribbons indicate means ± standard errors of the means. Squares indicate the peak values of Pavlovian bias during the two tasks.

Fig S3. Trace plots of group parameters for the best-fitting model. The trace plots show that MCMC samples were well mixed and converged. Note that the plots excluded burn-in samples.

Fig S4. Posterior distributions of individual go bias, reward sensitivity, punishment sensitivity parameters. Dots indicate medians and thick bars indicate 95% HDIs.

Fig S5. Posterior distributions of individual Pavlovian bias and Pavlovian bias under WM load parameters. Dots indicate medians and thick bars indicate 95% HDIs.

Fig S6. Posterior distributions of individual learning rate and learning rate under WM load parameters. Dots indicate medians and thick bars indicate 95% HDIs.

Fig S7. Posterior distributions of individual irreducible noise and irreducible noise under WM load parameters. Dots indicate medians and thick bars indicate 95% HDIs.

Fig S8. Results of fMRI replication analyses. (A) The main effect of WM load during the anticipation phase. As reported in previous meta-analysis studies [29,129], under WM load, regions of the lateral prefrontal cortex (PFC), including bilateral superior gyrus (MNI space coordinates x = -22, y = 0, z = 50, Z = 6.03, $p < 0.05$ whole-brain cluster-level family-wise error (FWE)) and middle frontal gyrus (MNI space coordinates x = 29, y = 2, z = 57, Z = 5.44, $p < 0.05$ whole-brain cluster-level FWE), and left precentral gyrus (MNI space coordinates x = -45, y = 5, z = 30, Z = 5.79, $p < 0.05$ whole-brain cluster-level FWE) and left inferior parietal cortex (MNI space coordinates x = -38, y = -51, z = 41, Z = 5.42, $p < 0.05$ whole-brain cluster-level FWE) showed increased BOLD signal. These results indicate that participants indeed had cognitive loads in the brain level. (B) The main effect of reward outcome. The reward was significantly associated with the signal in the striatum (MNI space coordinates x = 18, y = 5, z = -12, Z = 4.47, $p < 0.05$ whole-brain cluster-level FWE) and ventromedial PFC (MNI space coordinates x = -3, y = 68, z = 4, Z = 4.92, $p < 0.05$ whole-brain cluster-level FWE). (C) The main effect of loss outcome. Loss-related regions such as the insula (MNI space coordinates x = -36, y = 19, z = -10, Z = 6.77, $p < 0.05$ whole-brain cluster-level FWE) showed increased BOLD response. (B) and (C) are consistent with the previous findings suggesting reward- and loss-related regions [130,131]. Overlays are shown with a threshold of $p < 0.001$ (uncorrected). Color scale indicates t-values.

Table S1. Statistics of posterior distributions of group parameters for the best-fitting model. $\hat{R}$ values [119] for all parameters were close to 1.00 (< 1.1), which indicates that the estimated parameter values converged to their target posterior distributions.

Table S2. LOOIC for each model.

Table S3. Whole-brain results from replication analyses. Whole-brain cluster-level family-wise error (FWE) for multiple comparison with a cluster-forming threshold of $p < 0.001$. * The

insula (peak in insula: x = -36, y = 19, z = -10, *Z* = 6.77) reported in the result (**Fig S6C**) was included in this cluster of which the peak coordinate was located in orbital part of IFG.

Table S4. ROI and whole-brain results of the main effect of WM load on RPE. * p: small-volume corrected FWE within an anatomical striatum ROI defined from aal3 atlas with a cluster-forming threshold of $p < 0.001$. ** p: whole-brain cluster-level FWE for multiple comparison with a cluster-forming threshold of $p < 0.001$.

Table S5. Neural statistics of the effect of WM load on Pavlovian bias.

Table S6. Whole-brain results from PPI analysis. * The dlPFC (peak in dlPFC: x = -22 y = 63, z = 23, Z =-4.24) reported in the result (**Fig 5B**) was included in this cluster of which peak coordinate was located in medial SFG.

Author Contributions

Conceptualization: Heesun Park, Hoyoung Doh, Harhim Park, Woo-Young Ahn.

Data curation: Heesun Park, Hoyoung Doh, Eunhwi Lee, Harhim Park.

Formal analysis: Heesun Park, Hoyoung Doh, Eunhwi Lee.

Funding acquisition: Woo-Young Ahn.

Investigation: Heesun Park, Hoyoung Doh, Eunhwi Lee, Harhim Park.

Methodology: Heesun Park, Hoyoung Doh, Eunhwi Lee, Harhim Park, Woo-Young Ahn.

Project administration: Heesun Park, Woo-Young Ahn.

Resources: Woo-Young Ahn.

Software: Heesun Park, Hoyoung Doh, Harhim Park, Woo-Young Ahn.

Supervision: Woo-Young Ahn.

Validation: Heesun Park, Hoyoung Doh, Eunhwi Lee, Harhim Park.

Visualization: Heesun Park, Hoyoung Doh, Eunhwi Lee.

Writing – original draft: Heesun Park.

Writing – review & editing: Heesun Park, Hoyoung Doh, Eunhwi Lee, Woo-Young Ahn.

References

1. Collins AGE, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci. 2012;35: 1024–1035.

2. Huys QJM, Browning M, Paulus MP, Frank MJ. Advances in the computational understanding of mental illness. Neuropsychopharmacology. 2021;46: 3–19.

3. Rmus M, McDougle SD, Collins AGE. The role of executive function in shaping reinforcement learning. Current Opinion in Behavioral Sciences. Elsevier Ltd; 2021. pp. 66–73. doi:10.1016/j.cobeha.2020.10.003

4. Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press; 2018.

5. Bornstein AM, Daw ND. Multiplicity of control in the basal ganglia: computational roles of striatal subregions. Curr Opin Neurobiol. 2011;21: 374–380.

6. Khamassi M, Lachèze L, Girard B, Berthoz A, Guillot A. Actor--Critic models of reinforcement learning in the basal ganglia: from natural to artificial rats. Adapt Behav. 2005;13: 131–148.

7. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. Journal of neuroscience. 1996;16: 1936–1947.

8. Niv Y. Reinforcement learning in the brain. J Math Psychol. 2009;53: 139–154.

9. Schultz W. Dopamine neurons and their role in reward mechanisms. Curr Opin Neurobiol. 1997;7: 191–197.

10. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol. 1998;80: 1–27.

11. Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. Science. 1997;275: 1593–1599.

12. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev. 2013;37: 1297–1310.

13. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science. 2004;304: 452–454.

14. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal Difference Models and Reward-Related Learning in the Human Brain. Neuron. 2003;38: 329–337.

15. Collins AGE, Ciullo B, Frank MJ, Badre D. Working Memory Load Strengthens Reward Prediction Errors. J Neurosci. 2017;37: 4332–4342.

16. Collins AGE. The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. J Cogn Neurosci. 2018;30: 1422–1432.

17. Yoo AH, Collins AGE. How Working Memory and Reinforcement Learning Are Intertwined: A Cognitive, Neural, and Computational Perspective. J Cogn Neurosci. 2022;34: 551–568.

18. Collins AGE, Frank MJ. Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. Proceedings of the National Academy of Sciences. 2018;115: 201720963.

19. Dayan P. Goal-directed control and its antipodes. Neural Netw. 2009;22: 213–219.

20. Dolan RJ, Dayan P. Goals and habits in the brain. Neuron. Cell Press; 2013. pp. 312–325. doi:10.1016/j.neuron.2013.09.007

21. Tanaka SC, Balleine BW, O'Doherty JP. Calculating consequences: brain systems that encode the causal effects of actions. Journal of Neuroscience. 2008;28: 6750–6755.

22. Valentin VV, Dickinson A, O'Doherty JP. Determining the neural substrates of goal-directed learning in the human brain. Journal of Neuroscience. 2007;27: 4019–4026.

23. Collins AGE, Shenhav A. Advances in modeling learning and decision-making in neuroscience. Neuropsychopharmacology. 2021; 1–15.

24. Baddeley AD, Hitch G. Working Memory. Psychol Learn Motiv. 1974;8: 47–89.

25. Barbey AK, Koenigs M, Grafman J. Dorsolateral prefrontal contributions to human working memory. Cortex. 2013;49: 1195–1205.

26. Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci. 2003;7: 415–423.

27. Funahashi S. Prefrontal cortex and working memory processes. Neuroscience. 2006;139: 251–261.

28. Funahashi S, Kubota K. Working memory and prefrontal cortex. Neurosci Res. 1994;21: 1–11.

29. Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, et al. Modelling neural correlates of working memory: A coordinate-based meta-analysis. Neuroimage. 2012;60: 830–846.

30. Collins AGE, Albrecht MA, Waltz JA, Gold JM, Frank MJ. Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. Biol Psychiatry. 2017;82: 431–439.

31. McDougle SD, Collins AGE. Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. Psychon Bull Rev. 2020; 1–20.

32. Otto AR, Gershman SJ, Markman AB, Daw ND. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci. 2013;24: 751–761.

33. Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci. 2008;9: 545–556.

34. Franco-Watkins AM, Pashler H, Rickard TC. Does Working Memory Load Lead to Greater Impulsivity? Commentary on Hinson, Jameson, and Whitney (2003). J Exp Psychol Learn Mem Cogn. 2006;32: 443–447.

35. Franco-Watkins AM, Rickard TC, Pashler H. Taxing Executive Processes Does Not Necessarily Increase Impulsive Decision Making. Exp Psychol. 2010;57: 193–201.

964 36. Olschewski S, Rieskamp J, Scheibehenne B. Taxing Cognitive Capacities Reduces
965 Choice Consistency Rather Than Preference: A Model-Based Test. J Exp Psychol Gen.
966 2018;147: 462–484.

967 37. Dickinson A, Balleine B. The role of learning in the operation of motivational systems.
968 2002.

969 38. Mackintosh NJ. Conditioning and associative learning. Clarendon Press Oxford; 1983.

970 39. Wasserman EA, Franklin SR, Hearst E. Pavlovian appetitive contingencies and
971 approach versus withdrawal to conditioned stimuli in pigeons. J Comp Physiol Psychol.
972 1974;86: 616–627.

973 40. Wasserman EA, Miller RR. What's elementary about associative learning? Annu Rev
974 Psychol. 1997;48: 573–607.

975 41. Breland K, Breland M. The misbehavior of organisms. Am Psychol. 1961;16: 681–684.

976 42. Dayan P, Niv Y, Seymour B, Daw ND. The misbehavior of value and the discipline of
977 the will. Neural Netw. 2006;19: 1153–1160.

978 43. Hershberger WA. An approach through the looking-glass. Anim Learn Behav. 1986;14:
979 443–451.

980 44. Williams DR, Williams H. AUTO-MAINTENANCE IN THE PIGEON: SUSTAINED
981 PECKING DESPITE CONTINGENT NON-REINFORCEMENT 2. J Exp Anal Behav.
982 1969;12: 511–520.

983 45. Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go
984 learning in reward and punishment: Interactions between affect and effect. Neuroimage.
985 2012;62: 154–166.

986 46. Guitart-Masip M, Duzel E, Dolan R, Dayan P. Action versus valence in decision making.
987 Trends Cogn Sci. 2014;18: 194–202.

988 47. Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from
989 actions to habits to compulsion. Nat Neurosci. 2005;8: 1481–1489.

990 48. Garbusow M, Schad DJ, Sommer C, Jünger E, Sebold M, Friedel E, et al. Pavlovian-to-
991 Instrumental Transfer in Alcohol Dependence: A Pilot Study. Neuropsychobiology.
992 2014;70: 111–121.

993 49. Garbusow M, Schad DJ, Sebold M, Friedel E, Bernhardt N, Koch SP, et al. Pavlovian-
994 to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol
995 dependence. Addict Biol. 2016;21: 719–731.

996 50. Glasner SV, Overmier JB, Balleine BW. The role of Pavlovian cues in alcohol seeking in
997 dependent and nondependent rats. J Stud Alcohol. 2005;66: 53–61.

998 51. Lüscher C, Robbins TW, Everitt BJ. The transition to compulsion in addiction. Nat Rev
999 Neurosci. 2020;21: 247–263.

1000 52. Boer L de, Axelsson J, Chowdhury R, Riklund K, Dolan RJ, Nyberg L, et al. Dorsal
1001 striatal dopamine D1 receptor availability predicts an instrumental bias in action
1002 learning. Proceedings of the National Academy of Sciences. 2018;116: 201816704.

1003 53. Chowdhury R, Guitart-Masip M, Lambert C, Dolan RJ, Düzel E. Structural integrity of

the substantia nigra and subthalamic nucleus predicts flexibility of instrumental learning in older-age individuals. Neurobiol Aging. 2013;34: 2261–2270.

54. Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science. 2004;306: 1940–1943.

55. Gläscher JP, O'Doherty JP. Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. Wiley Interdiscip Rev Cogn Sci. 2010;1: 501–510.

56. Adams RA, Moutoussis M, Nour MM, Dahoun T, Lewis D, Illingworth B, et al. Variability in Action Selection Relates to Striatal Dopamine 2/3 Receptor Availability in Humans: A PET Neuroimaging Study Using Reinforcement Learning and Active Inference Models. Cereb Cortex. 2020;30: 3573–3589.

57. Betts MJ, Richter A, Boer L de, Tegelbeckers J, Perosa V, Baumann V, et al. Learning in anticipation of reward and punishment: perspectives across the human lifespan. Neurobiol Aging. 2020;96: 49–57.

58. Ereira S, Pujol M, Guitart-Masip M, Dolan RJ, Kurth-Nelson Z. Overcoming Pavlovian bias in semantic space. Sci Rep. 2021;11: 3416.

59. Guitart-Masip M, Economides M, Huys QJM, Frank MJ, Chowdhury R, Duzel E, et al. Differential, but not opponent, effects of l-DOPA and citalopram on action learning with reward and punishment. Psychopharmacology . 2014;231: 955–966.

60. Perosa V, Boer L de, Ziegler G, Apostolova I, Buchert R, Metzger C, et al. The Role of the Striatum in Learning to Orthogonalize Action and Valence: A Combined PET and 7 T MRI Aging Study. Cereb Cortex. 2020;30: 3340–3351.

61. Richter A, Guitart-Masip M, Barman A, Libeau C, Behnisch G, Czerney S, et al. Valenced action/inhibition learning in humans is modulated by a genetic variant linked to dopamine D2 receptor expression. Front Syst Neurosci. 2014;8: 140.

62. Richter A, Boer L de, Guitart-Masip M, Behnisch G, Seidenbecher CI, Schott BH. Motivational learning biases are differentially modulated by genetic determinants of striatal and prefrontal dopamine function. J Neural Transm. 2021; 1–16.

63. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017;27: 1413–1432.

64. Cavanagh JF, Eisenberg I, Guitart-Masip M, Huys Q, Frank MJ. Frontal Theta Overrides Pavlovian Learning Biases. J Neurosci. 2013;33: 8541–8548.

65. Kruschke J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. 2014.

66. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage. 1997;6: 218–229.

67. McLaren DG, Ries ML, Xu G, Johnson SC. A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. Neuroimage. 2012;61: 1277–1286.

68. Tsujimoto S, Sawaguchi T. Neuronal representation of response--outcome in the primate prefrontal cortex. Cereb Cortex. 2004;14: 47–55.

69. Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of

1045 information in an uncertain world. Nat Neurosci. 2007;10: 1214–1221.

1046 70. Ballard IC, Murty VP, Carter RM, MacInnes JJ, Huettel SA, Adcock RA. Dorsolateral
1047 Prefrontal Cortex Drives Mesolimbic Dopaminergic Regions to Initiate Motivated
1048 Behavior. Journal of Neuroscience. 2011;31: 10340–10346.

1049 71. Kahnt T, Heinzle J, Park SQ, Haynes J-D. Decoding different roles for vmPFC and
1050 dlPFC in multi-attribute decision making. Neuroimage. 2011;56: 709–715.

1051 72. Lee D, Seo H. Mechanisms of reinforcement learning and decision making in the
1052 primate dorsolateral prefrontal cortex. Ann N Y Acad Sci. 2007;1104: 108–122.

1053 73. Wallis JD, Miller EK. Neuronal activity in primate dorsolateral and orbital prefrontal
1054 cortex during performance of a reward preference task. Eur J Neurosci. 2003;18: 2069–
1055 2081.

1056 74. Park SQ, Kahnt T, Beck A, Cohen MX, Dolan RJ, Wrase J, et al. Prefrontal cortex fails
1057 to learn from reward prediction errors in alcohol dependence. Journal of Neuroscience.
1058 2010;30: 7749–7753.

1059 75. Pasupathy A, Miller EK. Different time courses of learning-related activity in the
1060 prefrontal cortex and striatum. Nature. 2005;433: 873–876.

1061 76. Economides M, Guitart-Masip M, Kurth-Nelson Z, Dolan RJ. Anterior Cingulate Cortex
1062 Instigates Adaptive Switches in Choice by Integrating Immediate and Delayed
1063 Components of Value in Ventromedial Prefrontal Cortex. J Neurosci. 2014;34: 3340–
1064 3349.

1065 77. Lim S-L, O'Doherty JP, Rangel A. The Decision Value Computations in the vmPFC and
1066 Striatum Use a Relative Value Code That is Guided by Visual Attention. J Neurosci.
1067 2011;31: 13214–13223.

1068 78. O'Doherty JP. Contributions of the ventromedial prefrontal cortex to goal-directed action
1069 selection. Ann N Y Acad Sci. 2011;1239: 118–129.

1070 79. Smith DV, Hayden BY, Truong T-K, Song AW, Platt ML, Huettel SA. Distinct Value
1071 Signals in Anterior and Posterior Ventromedial Prefrontal Cortex. Journal of
1072 Neuroscience. 2010;30: 2490–2495.

1073 80. Hiser J, Koenigs M. The Multifaceted Role of the Ventromedial Prefrontal Cortex in
1074 Emotion, Decision Making, Social Cognition, and Psychopathology. Biol Psychiatry.
1075 2018;83: 638–647.

1076 81. Pujara MS, Philippi CL, Motzkin JC, Baskaya MK, Koenigs M. Ventromedial Prefrontal
1077 Cortex Damage Is Associated with Decreased Ventral Striatum Volume and Response
1078 to Reward. J Neurosci. 2016;36: 5047–5054.

1079 82. Balleine BW, O'doherty JP. Human and rodent homologies in action control:
1080 corticostriatal determinants of goal-directed and habitual action.
1081 Neuropsychopharmacology. 2010;35: 48–69.

1082 83. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and
1083 dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005;8: 1704–1711.

1084 84. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on
1085 humans' choices and striatal prediction errors. Neuron. 2011;69: 1204–1215.

1086 85. Keramati M, Dezfouli A, Piray P. Speed/accuracy trade-off between the habitual and the
1087 goal-directed processes. PLoS Comput Biol. 2011;7: e1002055.

1088 86. Dickinson A. Actions and habits: the development of behavioural autonomy. Philos
1089 Trans R Soc Lond B Biol Sci. 1985;308: 67–78.

1090 87. Dorfman HM, Gershman SJ. Controllability governs the balance between Pavlovian and
1091 instrumental action selection. Nat Commun. 2019;10: 5826.

1092 88. Swart JC, Frank MJ, Määttä JI, Jensen O, Cools R, Ouden HEM den. Frontal network
1093 dynamics reflect neurocomputational mechanisms for reducing maladaptive biases in
1094 motivated action. PLoS Biol. 2018;16: e2005979.

1095 89. Cavanagh JF, Frank MJ. Frontal theta as a mechanism for cognitive control. Trends
1096 Cogn Sci. 2014;18: 414–421.

1097 90. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one
1098 decade on. Trends Cogn Sci. 2014;18: 177–185.

1099 91. Levy R, Goldman-Rakic PS. Segregation of working memory functions within the
1100 dorsolateral prefrontal cortex. Executive control and the frontal lobe: Current issues.
1101 2000; 23–32.

1102 92. Petrides M. The role of the mid-dorsolateral prefrontal cortex in working memory. Exp
1103 Brain Res. 2000;133: 44–54.

1104 93. Ott T, Nieder A. Dopamine and Cognitive Control in Prefrontal Cortex. Trends Cogn Sci.
1105 2019;23: 213–234.

1106 94. Albrecht MA, Waltz JA, Cavanagh JF, Frank MJ, Gold JM. Reduction of Pavlovian Bias
1107 in Schizophrenia: Enhanced Effects in Clozapine-Administered Patients. PLoS One.
1108 2016;11: e0152781.

1109 95. Cools R. The costs and benefits of brain dopamine for cognitive control. Wiley
1110 Interdiscip Rev Cogn Sci. 2016;7: 317–329.

1111 96. Nord CL, Lawson RP, Huys QJM, Pilling S, Roiser JP. Depression is associated with
1112 enhanced aversive Pavlovian control over instrumental behaviour. Sci Rep. 2018;8: 1–
1113 10.

1114 97. Barrouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and Cognitive Load
1115 in Working Memory. J Exp Psychol Learn Mem Cogn. 2007;33: 570–585.

1116 98. Dalley JW, Cardinal RN, Robbins TW. Prefrontal executive and cognitive functions in
1117 rodents: neural and neurochemical substrates. Neurosci Biobehav Rev. 2004;28: 771–
1118 784.

1119 99. Granon S, Vidal C, Thinus-Blanc C, Changeux J-P, Poucet B. Working memory,
1120 response selection, and effortful processing in rats with medial prefrontal lesions. Behav
1121 Neurosci. 1994;108: 883.

1122 100. Oberauer K. Working Memory and Attention – A Conceptual Analysis and Review.
1123 Journal of Cognition. 2019;2: 36.

1124 101. Ridderinkhof KR, Van Den Wildenberg WPM, Segalowitz SJ, Carter CS.
1125 Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action
1126 selection, response inhibition, performance monitoring, and reward-based learning.

Brain Cogn. 2004;56: 129–140.

102. Ripke S, Hübner T, Mennigen E, Müller KU, Rodehacke S, Schmidt D, et al. Reward processing and intertemporal decision making in adults and adolescents: The role of impulsivity and decision consistency. Brain Res. 2012;1478: 36–47.

103. Seo M, Lee E, Averbeck BB. Action selection and action value in frontal-striatal circuits. Neuron. 2012;74: 947–960.

104. Szmalec A, Vandierendonck A, Kemps E. Response selection involves executive control: Evidence from the selective interference paradigm. Mem Cognit. 2005;33: 531–541.

105. Turi Z, Mittner M, Opitz A, Popkes M, Paulus W, Antal A. Transcranial direct current stimulation over the left prefrontal cortex increases randomness of choice in instrumental learning. Cortex. 2015;63: 145–154.

106. Laureiro-Martinez D, Brusoni S, Tata A, Zollo M. The Manager's Notepad: Working Memory, Exploration, and Performance. Journal of Management Studies. 2019;56: 1655–1682.

107. Master SL, Eckstein MK, Gotlieb N, Dahl R, Wilbrecht L, Collins AGE. Disentangling the systems contributing to changes in learning during adolescence. Dev Cogn Neurosci. 2020;41: 100732.

108. Nassar MR, Frank MJ. Taming the beast: extracting generalizable knowledge from computational models of cognition. Current opinion in behavioral sciences. 2016;11: 49–54.

109. Talmi D, Dayan P, Kiebel SJ, Frith CD, Dolan RJ. How humans integrate the prospects of pain and reward during choice. Journal of Neuroscience. 2009;29: 14617–14626.

110. Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016;19: 404–413.

111. Gu S, Satterthwaite TD, Medaglia JD, Yang M, Gur RE, Gur RC, et al. Emergence of system roles in normative neurodevelopment. Proceedings of the National Academy of Sciences. 2015;112: 13681–13686.

112. Power JD. A simple but useful way to assess fMRI scan qualities. Neuroimage. 2017;154: 150–158.

113. Ousdal OT, Huys QJ, Milde AM, Craven AR, Ersland L, Endestad T, et al. The impact of traumatic stress on Pavlovian biases. Psychol Med. 2018;48: 327–336.

114. Moutoussis M, Bullmore ET, Goodyer IM, Fonagy P, Jones PB, Dolan RJ, et al. Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. PLoS Comput Biol. 2018;14: e1006679.

115. Ahn W-Y, Haines N, Zhang L. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. Computational Psychiatry. 2017;1: 24–57.

116. Papaspiliopoulos O, Roberts GO, Sköld M. A general framework for the parametrization of hierarchical models. Stat Sci. 2007; 59–73.

117. Wetzels R, Vandekerckhove J, Tuerlinckx F, Wagenmakers E-J. Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. J Math Psychol. 2010;54: 14–27.

118. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-normalization, folding, and localization: An improved $\widehatR$ for assessing convergence of MCMC. arXiv preprint arXiv:190308008. 2019.

119. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci. 1992;7. doi:10.1214/ss/1177011136

120. Esteban O, Blair R, Markiewicz CJ, Berleant SL, Moodie C, Ma F, et al. fMRIPrep. Softw Pract Exp. 2018. doi:10.5281/zenodo.852659

121. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods. 2019;16: 111–116.

122. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. Front Neuroinform. 2011;5: 13.

123. Gorgolewski KJ, Esteban O, Markiewicz CJ, Ziegler E, Ellis DG, Notter MP, et al. Nipype. Softw Pract Exp. 2018. doi:10.5281/zenodo.596855

124. Chase HW, Kumar P, Eickhoff SB, Dombrovski AY. Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. Cogn Affect Behav Neurosci. 2015;15: 435–459.

125. Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS. How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron. 2009;62: 733–743.

126. Lee SW, Shimojo S, O'Doherty JP. Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. Neuron. 2014;81: 687–699.

127. Nichols TE. Multiple testing corrections, nonparametric methods, and random field theory. Neuroimage. 2012;62: 811–815.

128. Woo C-W, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. Neuroimage. 2014;91: 412–419.

129. Wang H, He W, Wu J, Zhang J, Jin Z, Li L. A coordinate-based meta-analysis of the n-back working memory paradigm using activation likelihood estimation. Brain Cogn. 2019;132: 1–12.

130. Bartra O, McGuire JT, Kable JW. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage. 2013;76: 412–427.

131. Liu X, Hairston J, Schrier M, Fan J. Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. Neurosci Biobehav Rev. 2011;35: 1219–1236.