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*], 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: 41

Number of Figures & Tables: 5/1

Number of words in the abstract: 271

Number of words in the introduction: 791

Number of words in the discussion: 1960

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

Abstract

Humans and animals learn optimal behaviors by interacting with the environment. Research suggests that a fast, capacity-limited working memory (WM) system and a slow, incremental reinforcement learning (RL) system jointly contribute to instrumental learning. Situations that strain WM resources alter several decision-making processes and the balance between multiple decision-making systems: under WM loads, learning becomes slow and incremental, while reward prediction error (RPE) signals become stronger; the reliance on computationally efficient learning increases as WM demands are balanced against computationally costly strategies; and action selection becomes more random. Meanwhile, instrumental learning is known to interact with Pavlovian learning, a hard-wired system that motivates approach to reward and avoidance of punishment. However, the neurocognitive role of WM load during instrumental learning under Pavlovian influence remains unknown, while conflict between the two systems sometimes leads to suboptimal behavior. 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); WM load was manipulated with dual-task conditions. Behavioral and computational modeling analyses revealed that WM load compromised learning by reducing the learning rate and increasing 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 decelerates instrumental 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 makes learning slower and decisions noisier in an instrumental learning task with Pavlovian components. FMRI analysis revealed that the WM load weakens the functional connectivity between the prefrontal cortex and the striatum when people anticipate outcomes 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, the deterioration of either system can alter the learning and 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

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

While reductions of WM resources substantially alter instrumental learning, another important factor known to shape instrumental learning is the Pavlovian system. Through the motivation of hard-wired responses, such as active responses to appetitive cues and inhibitory responses toward aversive cues [37–40], the Pavlovian system may facilitate certain instrumental behaviors and impede others. This bias in instrumental learning is known as Pavlovian bias [41–44]. Pavlovian bias is generally presumed to be associated with maladaptive behaviors such as substance use disorder and compulsivity-related disorders [45–49].

Although it is well known that the enhancement of WM load alters instrumental learning in several ways, it remains unclear how WM load changes instrumental learning when it is under Pavlovian influence. To investigate this relationship, we conducted a functional magnetic resonance imaging (fMRI) study in which participants completed an instrumental learning task that involved Pavlovian–instrumental conflicts [50], with and without additional 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, as 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 [50–54]. 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 decelerated and action selection became increasingly random; these findings supported hypotheses 1 and 3 but not 2. Increased striatal RPE signaling suggests that the increased contribution of RL and decreased contribution of WM may explain slower learning. 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 [50], a model-free learning task that contained Pavlovian–instrumental conflicts. In the WM load condition, a 2-back task was added to the GNG task; the modified task was named the working memory go/no-go (WMGNG) task (see Materials and Methods for more detail).

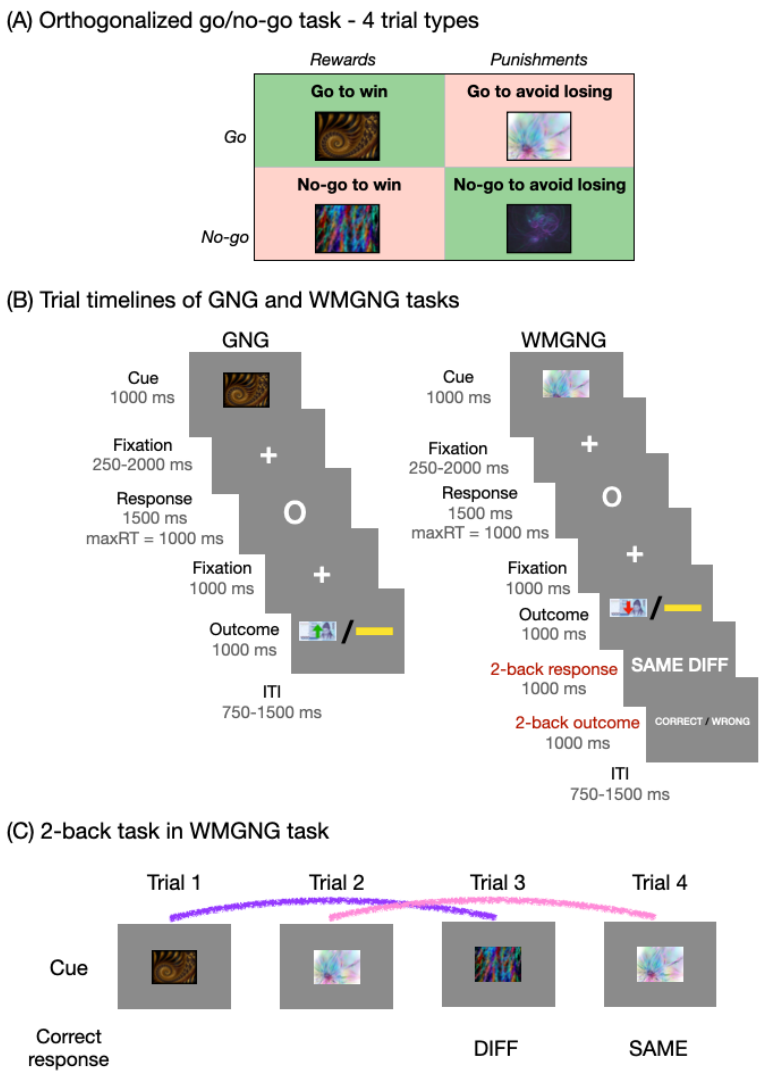


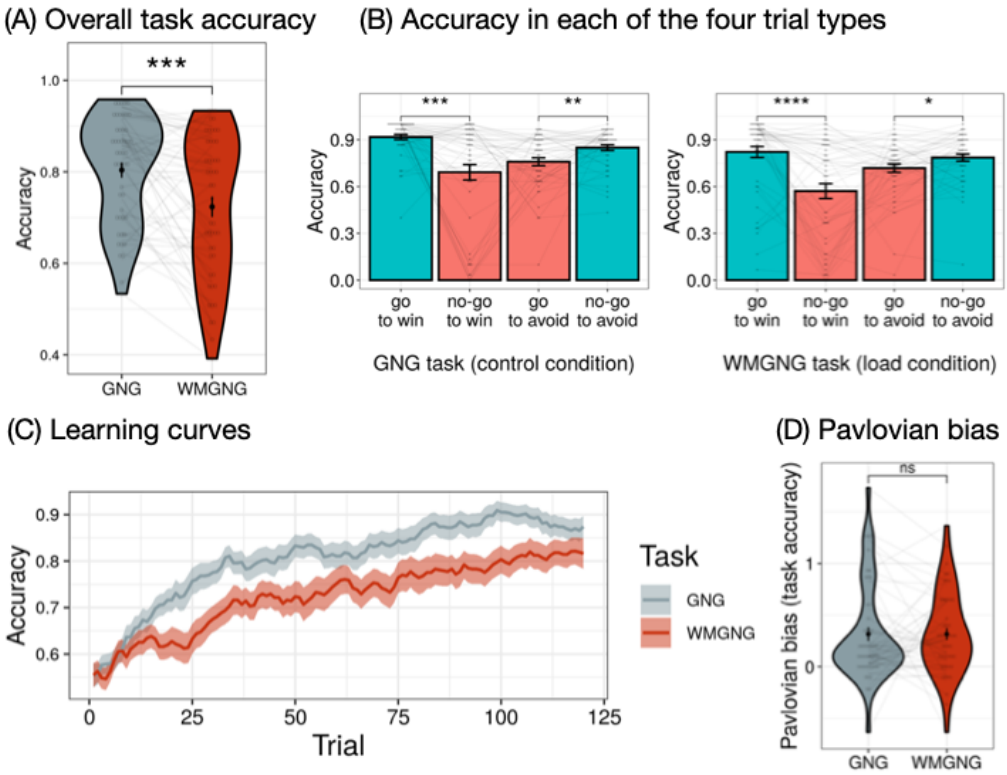
Fig 1. The GNG and WMGNG tasks. (A) In both tasks, four fractal cues indicated the combination of action (go/no-go) and valence at the outcome (win/loss). (B) In each trial, a fractal cue was presented, followed by a variable delay. After the delay, actions were required in response to a circle, and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

Comparison of overall task accuracy between the two tasks confirmed that our dual-task manipulation with a 2-back task successfully imposed WM load. 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, $t(48)=3.86$, $p<0.001$, $d=0.55$). We also confirmed that participants exhibited go bias and Pavlovian bias in both tasks, thus replicating the findings of earlier studies [50,51,56–62]. 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 effect of WM load on learning speed (hypothesis 1, **Fig 2C**). While the learning curves indicated that participants learned during both tasks, the learning curve was slower in the WMGNG task than in the GNG task (i.e., WM load reduced learning speed and overall accuracy). To test the effect of WM load on Pavlovian bias (hypothesis 2, **Fig 2D**),

184 we quantified Pavlovian bias by subtracting the accuracy in Pavlovian-incongruent conditions
185 ("no-go to win" and "go to avoid losing") from accuracy in Pavlovian-congruent conditions ("go
186 to win" and "no-go to avoid losing"), then compared it between the two tasks. No significant
187 difference in Pavlovian bias was observed between the GNG and WMGNG tasks, confirming
188 that WM load did not affect Pavlovian bias.
189



190

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

Computational modeling: WM load influences learning rate and irreducible noise

We used a computational modeling approach to test the three hypotheses. For this purpose, we developed eight nested models that assumed different learning rate, Pavlovian bias, or irreducible noise parameters under WM load. These models were fitted to the data using hierarchical Bayesian analysis, then compared using the leave-one-out information criterion (LOOIC), where a lower LOOIC value indicated better out-of-sample predictive accuracy (i.e., better fit) [63]. Importantly, the use of computational modeling allowed us to test hypothesis 3 regarding whether WM load would increase random choices; this would have not been possible if we had performed behavioral analysis alone.

Based on earlier studies [50,64], we constructed a baseline model (model 1) that used a Rescorla-Wagner updating rule and contained learning rate (ε), Pavlovian bias, irreducible noise, go bias, and separate parameters for sensitivity to rewards and punishments (Materials and Methods). In the model, state-action values are updated with the prediction error; learning rate (ε) modulates the impact of the prediction error. Reward/punishment sensitivity (ρ) scales the effective size of outcome values. Go bias (b) and cue values weighted by Pavlovian bias (π) are added to the value of go choices. Here, as the Pavlovian bias parameter increases, the go tendency increases under the reward condition whereas the go tendency is reduced under the punishment condition; this results in an increased no-go tendency. Computed action weights are used to estimate action probabilities, and irreducible noise (ξ) determines the extent to which information about action weights is utilized to make decisions. As irreducible noise increases, action probabilities will be less reflective of action weights, indicating that action selection will become increasingly random.

In models 2, 3, and 4, we assumed that WM load affects only one parameter. For example, in model 2, a separate Pavlovian bias parameter (π_{wm}) was assumed for the WM load condition. Models 3 and 4 assumed different learning rates (ε_{wm}) and irreducible noise (ξ_{wm}) parameters in their respective WM load conditions. In models 5, 6, and 7, we assumed 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, it demonstrated that 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 the lack of support for hypothesis 2 (Pavlovian bias) in the behavioral results. Conversely, the learning rate was credibly lower, while irreducible noise was credibly greater in the WMGNG than in the GNG task. These results support hypothesis 1 (i.e., WM load will reduce learning rate) and hypothesis 3 (i.e., WM load will 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 S3**), 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 Supplementary Material for the posterior distributions of individual parameters; **Fig S2-S5**).

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.

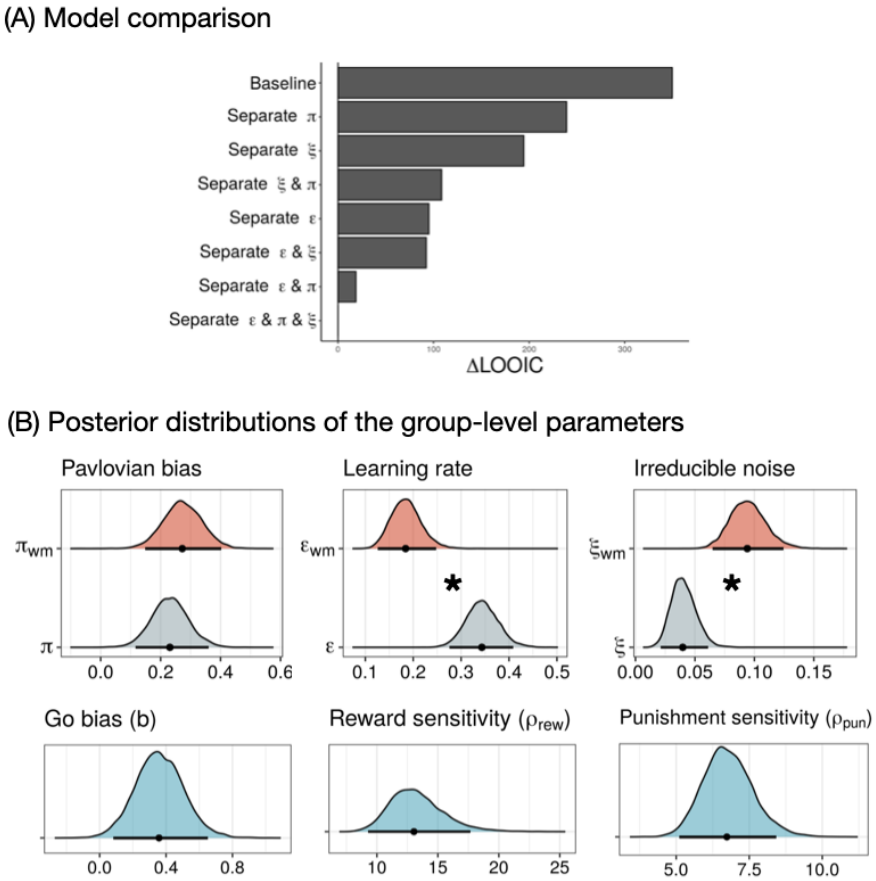


Fig 3. Model comparison results and posterior distribution of the group-level parameters of the best-fitting model (N=49). (A) Relative LOOIC difference indicates the difference in LOOIC between the best-fitting model and each of the other models. The best-fitting model was the full model, which assumed 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).

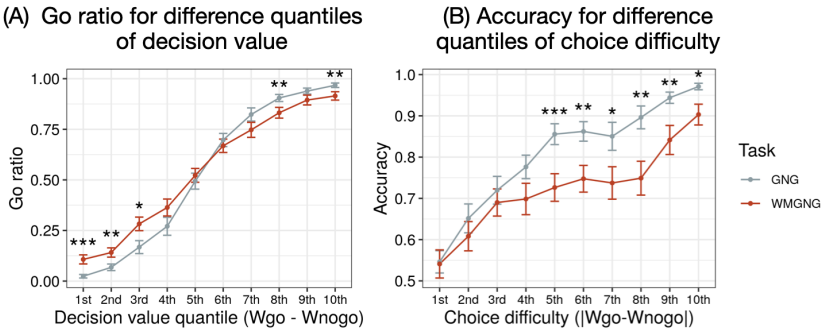


Fig 4. Choice consistency (N=49). (A) Mean percentage of go choices for different quantiles of action weight differences (W_{go} - W_{nogo}) between "go" and "no-go" choices, where higher quantiles indicate higher decision values for "go" choices. Under WM load, the increase in go ratio according to 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 ± standard errors of the mean. 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 hypothesis 1, which predicts an increased

317 contribution of the RL system and decreased contribution of the WM system, to learning under
318 WM load. We also tested hypothesis 2 regarding Pavlovian bias, but we found no main effect
319 of Pavlovian bias between the GNG and WMGNG tasks (WMGNG > GNG [Pavlovian-
320 congruent > Pavlovian-incongruent]) within the striatum or SN/VTA ($p < 0.05$ SVC, a priori
321 ROIs). Note that previous studies showed no significant result for the same contrast
322 (Pavlovian-congruent > Pavlovian-incongruent) within the same regions [50]. With regard to
323 hypothesis 3 concerning random choices, we observed no main effect of WM load on random
324 choice (WMGNG > GNG [W_{chosen} - $W_{unchosen}$]) within the ventromedial prefrontal cortex (vmPFC;
325 $p < 0.05$ SVC, a priori ROI). See Supplementary Material for further details regarding these
326 findings (**Table S5**).

327 Increased RPE signals under WM load may indicate reduced WM contribution and
328 increased RL contribution to learning because of the load, suggesting diminished cooperation
329 between the two systems for learning. Therefore, we conducted a psychophysiological
330 interaction (PPI) analysis [66] using the gPPI toolbox [67] to test whether functional
331 connectivity between areas associated with RL and WM systems would weaken under WM
332 load. Specifically, we explored differences between the two tasks in terms of functional
333 coupling between the striatum, which showed increased RPE signaling under WM load during
334 the feedback phase, and other regions when computing reward expectations. The striatum
335 showed significantly decreased connectivity with the vmPFC (MNI space coordinates $x = 13$,
336 $y = 56$, $z = 0$, $Z = -4.90$, $p < 0.05$ whole-brain cluster-level familywise error rate (FWE)) and
337 dlPFC (MNI space coordinates $x = -20$, $y = 63$, $z = 23$, $Z = -4.24$, $p < 0.05$ whole-brain cluster-
338 level FWE, **Fig 5B**, **Table S6**) in the WMGNG task, compared with the GNG task.

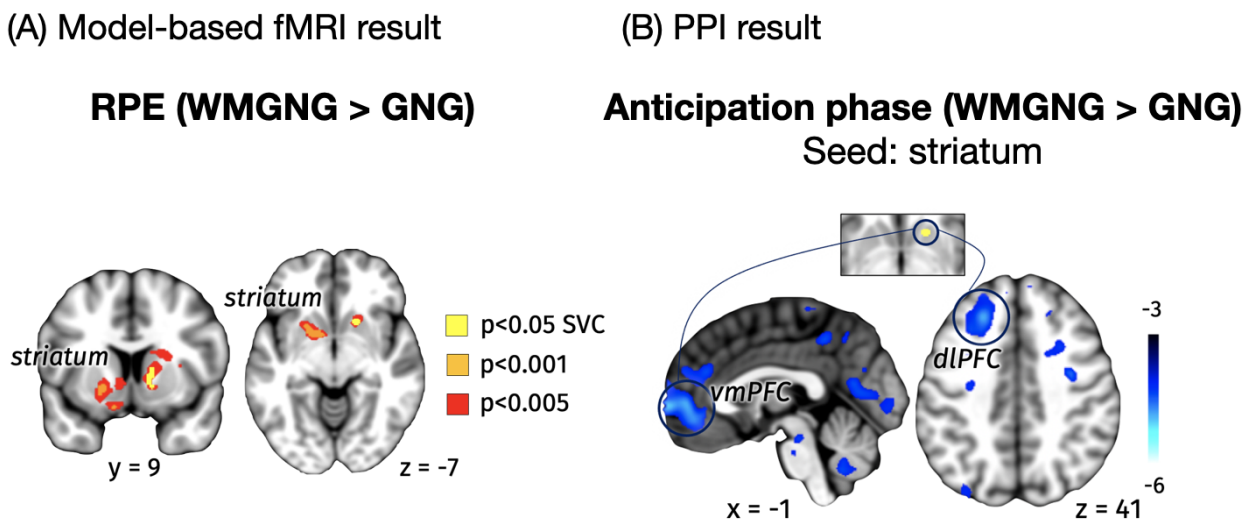


Fig 5. fMRI results (N=44). (A) In the predefined ROI-based analysis, RPE signaling in the striatum was stronger in the WMGNG task than in the GNG task. Effects that were significant at $p < 0.05$ (SVC) are shown in yellow. (B) Functional connectivity between the striatum (seed region, top) and prefrontal 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-based cluster-level FWE). 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 increasingly random. First, we found that the limitation of WM resources according to 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.

359 **Decreased contribution of the WM system and increased contribution of the RL system**
360 **under WM load**

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

364 These findings can be attributed to the reduced contribution of the WM system and
365 increased contribution of the striatal RL system, consistent with previous findings that WM
366 improves learning efficiency (in parallel with RL), as well as reward prediction precision in RL
367 processes. First, as a rapid and immediate learning system, WM learns in parallel with the
368 slow and incremental RL system by directly storing associations between states and actions
369 [1,15–17,68]. Specifically, WM and RL systems compete with each other based on their
370 reliability in a given situation. Under WM load, the fast and capacity-limited WM system
371 becomes less reliable; thus, the slow and incremental RL system supersedes the WM system,
372 causing learning to occur more slowly and incrementally [1,16,30]. Second, RL computations
373 themselves are intertwined with WM; WM can represent feed reward expectations to the RL
374 system [69–72] and improve reward prediction precision, which can reduce RPE and improve
375 learning efficiency [15,18]. In our study, WM load, which limited the contribution of WM, may
376 have increased the striatal RL contribution while reducing the accuracy of RL reward
377 computation. Consistent with this interpretation, we found that RPE signaling in the striatum –
378 a marker of RL – was strengthened under WM load (**Fig 5A**). This is consistent with previous
379 findings that RPE-associated neural signals were increased under higher WM load [15,18].

380 Overall, these findings suggest that WM load led to reduced cooperation between RL
381 and WM by interrupting and reducing the contribution of WM. This notion is further supported
382 by the finding that the striatum showed weakened functional connectivity with the dlPFC during
383 reward prediction under WM load (**Fig 5A**). Taken together, these findings suggest that WM
384 load may have weakened the interplay between WM in the dlPFC and RL in the striatum during
385 the value estimation process, which subsequently led to stronger RPE signals. However,

386 further research is necessary to demonstrate the directionality of functional connectivity
387 between the two systems during reward prediction; frontostriatal connectivity is reportedly
388 bidirectional, such that the striatum may also provide prefrontal regions with inputs that relate
389 to reward information [73,74].

390 Notably, we observed weakened connectivity between the vmPFC and the striatum.
391 The vmPFC has been identified as a critical neural correlate of value-based decision-making;
392 it integrates reward predictions [70], represents value signals or decision value [75–78], and
393 affects reward anticipation/processing in the striatum [79,80]. Our findings suggest that value
394 integration through the cortico-striatal loop was also weakened under WM load.

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

396 Contrary to our hypothesis, WM load did not influence Pavlovian bias. Behavioral and
397 modeling results showed that Pavlovian bias did not significantly differ between the GNG and
398 WMGNG tasks (**Fig 2D; Fig 3**), while fMRI analysis revealed that neural signaling associated
399 with Pavlovian bias did not significantly differ between the two tasks (**Table S6**). These
400 findings indicate that the brain did not exhibit greater reliance on the computationally efficient
401 system under WM load, in contrast to the results of previous studies [32]. We identified two
402 possible explanations for this discrepancy. First, instrumental and Pavlovian learning require
403 similar amounts of WM resources; second, the WM system may not be involved in modulating
404 the balance between Pavlovian and instrumental systems.

405 In the first potential explanation, the amounts of WM resources may be similar for
406 Pavlovian and instrumental learning (especially model-free learning), in contrast to model-
407 based and model-free learning. Model-based learning system constructs an internal model to
408 compute the values of actions; thus, it requires greater WM resources to compute and retain
409 the model online [20,81–84]. However, the model-free system simply uses the action-reward
410 association history to compute action values (i.e., “cached values”), and does not require the
411 internal model [81,82,85]. Pavlovian learning is similar to model-free learning but differs in
412 terms of the dimensions for value learning–the Pavlovian system learns state-outcome

413 associations, while the instrumental system learns state-action-outcome associations [42,86].
414 Therefore, the difference in WM demands between model-based and model-free learning
415 could be significantly greater than the difference between model-free instrumental and
416 Pavlovian learning. In our task, in particular, the instrumental learning was model-free; both
417 instrumental and Pavlovian systems were required to learn the associations without prior
418 information. Thus, the difference in WM demands may not have been sufficient to trigger a
419 trade-off between the two learning systems. Rather than depending more on Pavlovian
420 learning which has little computational benefit in our task, the participants may simply have
421 compromised overall learning.

422 In the second potential explanation, WM resources may be unimportant with respect
423 to modulating the Pavlovian–instrumental interaction, despite earlier studies' suggestions to
424 the contrary. Several studies have proposed that prefrontal WM control systems are crucial
425 for controlling Pavlovian bias. Electroencephalogram studies demonstrated that midfrontal
426 theta oscillations are important for controlling Pavlovian bias [64,87], suggesting top-down
427 prefrontal control over Pavlovian bias [64]. Furthermore, recruitment of the inferior frontal
428 gyrus (IFG) is involved in appropriate response inhibition, helping to overcome Pavlovian bias
429 [50]. Finally, there is indirect evidence that administration of levodopa, which increases
430 dopamine levels, reduced Pavlovian influences on instrumental learning; such a reduction was
431 speculated to result from increased dopamine levels in the PFC, which may have facilitated
432 the operation of prefrontal WM functions [59]. A related finding suggested that genetic
433 determinants of prefrontal dopamine function may be important in overcoming Pavlovian bias
434 [62].

435 While the results of the present study appear to contradict these findings, several
436 complex possibilities exist. In particular, although previous findings implied the involvement of
437 prefrontal mechanisms (e.g., model-based prefrontal control [64] and WM [54,59] in controlling
438 the Pavlovian system, they did not directly suggest active recruitment of the prefrontal WM
439 system. First, while Cavanagh et al. speculated that midfrontal theta power could be indicative
440 of "model-based top-down prefrontal control" [64], a subsequent study by Swart et al.

441 suggested that midfrontal theta signals could only be involved in the detection of conflict by
442 signaling “the need for control” [87,88], rather than being a source of direct control. Next, the
443 IFG showed an increased BOLD response only in the “no-go” condition [50], implying that the
444 IFG is important for “inhibitory” motor control (i.e., as a brake [89]); it does not participate in
445 active maintenance or representation of goal-directed behaviors including both “go” and “no-
446 go,” which would be more closely associated with WM [29,90,91]. Finally, elevated dopamine
447 levels should be cautiously interpreted as improvements in prefrontal WM function [59]. While
448 dopamine has been shown to enable successful cognitive control in the prefrontal cortex, it
449 may have three roles: gating behaviorally relevant sensory signals; maintaining and
450 manipulating information in WM to guide goal-directed behavior; and relaying motor
451 information to premotor areas for action preparation [92]. Moreover, distinct mechanisms have
452 been known to modulate the influence of dopamine on WM in the PFC through distinct types
453 of dopamine receptors [92]. Thus, there may be several ways to interpret the observation that
454 dopamine level [59] or function [62] was associated with the modulation of Pavlovian
455 influences. Considerable research is needed to fully understand the mechanisms by which
456 dopamine levels affect Pavlovian bias. Alternatively, the role of prefrontal WM in controlling
457 Pavlovian bias may not require vast resources. It may only be responsible for signaling a need
458 for control [87], promoting response inhibition [50], or influencing subcortical areas (e.g., the
459 striatum and subthalamic nucleus [93,94]).

460 **Increased random choices under WM load**

461 Another notable finding was that random choice increased under WM load. Our
462 modeling analysis revealed that irreducible noise parameter estimates were greater in the
463 WMGNG task than in the GNG task (**Fig 3**), suggesting increased random action selection
464 under WM load. Further analysis using the modeling outputs revealed that participants’
465 choices were less affected by the relative value difference between the “go” and “no-go”
466 actions under WM load (**Fig 4A**). Moreover, analysis using the absolute difference between
467 the two options (**Fig 4B**) revealed that the increase in accuracy became smaller as the

468 absolute difference increased (i.e., the choice became easier). Both findings suggest that WM
469 involvement led to an increase in random choices, regardless of value comparison and choice
470 difficulty.

471 Our findings are broadly consistent with the results of previous studies concerning the
472 role of WM and prefrontal regions in action selection and execution [95–102]. In particular,
473 several studies have demonstrated that the interruption of WM function via WM load could
474 increase the frequency of random choices in value-based decision-making tasks [34–36].
475 Additionally, transcranial direct current stimulation, a brain stimulation method, over the left
476 PFC led to increased random action selection during an RL task, suggesting that the prefrontal
477 WM component influenced action selection [103]. Furthermore, the importance of WM in
478 action selection during learning tasks is supported by the indirect evidence that individual
479 differences in WM capacity were correlated with appropriate exploratory action selection in
480 multi-armed bandit tasks [104]. Overall, the reduced availability of WM resources because of
481 WM load in our study may have compromised the participants' abilities to actively represent
482 their current goals and actions, leading to reduced WM control over consistent choice based
483 on value computation.

484 No significant neural correlates were identified with respect to the increased random
485 choices. We assumed that random action selection would be associated with the reduced
486 sensitivity to value difference or value comparison between the two options ("go" and "no-go")
487 [55]; thus, we hypothesized that value comparison signals would decrease under WM load.
488 Contrary to our hypothesis, no significant differences in value comparison signaling in ROIs
489 were observed between GNG and WMGNG tasks. There are several possible explanations
490 for this null finding. Our assumption of value sensitivity may not be the source of the random
491 choice observed here. Alternatively, subsequent attentional lapse [105,106] or value-
492 independent noise [107] may have led to inconsistent action selection despite the presence of
493 value comparison signals. Further research is necessary to distinguish these possibilities.

494 In summary, the present study has shown that WM load compromises overall learning
495 by reducing learning speed via weakened cooperation between RL and WM; it also increases

random action selection without affecting the balance between Pavlovian and instrumental systems. To our knowledge, this is the first study to investigate the neurocognitive effect of WM load during interactions between Pavlovian and instrumental systems. By investigating how learning and decision-making using different systems are altered in the presence of WM load and by linking such behaviors to their underlying neural mechanisms, this study contributes to our understanding of how distinct cognitive components interact with each other and synergistically contribute to learning. Because impairments in learning, balance among multiple systems, and action selection have been reported in various neurological and psychiatric disorders [2,108], our findings represent an important step toward improved understanding of various symptoms.

Materials and Methods

Participants

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

The behavioral analysis included 49 participants (29 women; 24.3±3.3 y.o); the fMRI analysis included 44 participants (27 women; 24.2±3.3 y.o). Four participants were excluded because of technical issues; one participant was excluded because they slept during the task. Two participants were excluded because of poor performance in the 2-back task since the results in the dual-task paradigm could only be valid and interpretable when participants actually performed both tasks. The accuracy cutoff was 0.575, a value that rejects the null hypothesis that participants would randomly choose one of two options. After assessment of preprocessed image quality, five participants were excluded from the fMRI analysis because of head movements in the scanner, which can systematically alter brain signals; four out of

these five were excluded because the mean framewise displacement exceeded 0.2 mm [109], while the remaining one was excluded after visual assessment of carpet plots [110].

Experimental design and task

The experiment was performed in two blocks: one contained the original GNG task [50] and one contained the GNG task paired with the 2-back task as a secondary task. The order of task completion was counterbalanced among participants. The GNG and WMGNG tasks consisted of two blocks (four blocks in total); each block consisted of 60 trials. Therefore, each task contained 120 trials (240 trials in total). Participants underwent fMRI while performing the tasks for approximately 50 min, with a short (~60 s) break after each set of 60 trials. Before scanning, participants performed 20 practice trials each of GNG task and WMGNG task to help them become accustomed to the task structure and response timing. Participants received additional compensation based on their accuracy in the two tasks, along with the standard participation fee at the end of the experiment.

Orthogonalized go/no-go (GNG) task

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

Each trial consisted of three phases: fractal cue presentation, response, and 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

549 monetary reward, a red arrow pointing downwards on a ₩1000 bill indicated monetary
550 punishment, and a yellow horizontal bar indicated no reward or punishment.

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

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

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

567 **Computational modeling**

568 *Baseline RL model with Pavlovian bias*

569 We adopted a previously implemented version of an RL model [50] that can model
570 Pavlovian bias and choice randomness as well as learning rate. In our baseline model, we
571 assumed no difference in parameters between the control and load conditions.

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

575 $$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \epsilon(\rho r_t - Q_{t-1}(a_t, s_t))$$

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.

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.

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 [65,111]. 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 [112].

620 Moreover, bounded parameters such as learning rates and irreducible noise (∈ [0, 1]) were
621 estimated within an unconstrained space; they were then probit-transformed to the
622 constrained space, thus maximizing MCMC efficiency within the parameter space [111,113].
623 We ran four independent chains with 4000 samples each, including 2000 warm-up
624 samples (i.e., burn-in) to ensure that the parameters converged to the target distributions.
625 Four chains were run to ensure that the posterior distributions were not dependent on initial
626 starting points [114]. We visually checked convergence to target distributions by observing
627 trace plots (**Fig S1**) and computing the R statistics - a measure of convergence across chains
628 [115]. R statistics were < 1.1 for all models, indicating that the estimated parameter values
629 converged to their target posterior distributions (**Table S1**).
630 Models were compared using the LOOIC, which is an information criterion calculated
631 from the leave-one-out cross-validation [63]. This method is used to estimate the out-of-
632 sample predictive accuracy of a fitted Bayesian model for model comparison and selection.
633 The LOOIC is computed using the log-likelihood evaluated from posterior distributions or
634 simulations of the parameters. The R package loo [63], which provides an interface for the
635 approximation of leave-one-out cross-validated log-likelihood, was used to estimate the
636 LOOIC for each model. Lower LOOIC values indicated better fit.

637 **fMRI scans: acquisition and protocol**

638 fMRI was performed on the same scanner (Simens Tim Trio 3 Tesla) using a 32-
639 channel head coil across all participants. A high-resolution T1-weighted anatomical scan of
640 the whole brain resolution was also acquired for each participant (TR = 2300ms, TE = 2.36ms,
641 FOV = 256mm,1mm×1mm×1mm) to enable spatial localization and normalization. The
642 participant's head was positioned with foam pads to limit head movement during acquisition.
643 Functional data was acquired using echo-planar imaging (EPI) in four scanning sessions
644 containing 64 slices (TR = 1500ms, TE = 30ms, FOV = 256mm, 2.3mm × 2.3mm × 2.3mm).
645 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) [116,117], which is based on Nipype 1.5.1 (RRID:SCR_002502) [118,119]. Details of preprocessing with fMRIPrep are provided in Supplementary Material. 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 to obtain participant-specific estimates for each effect. These estimates were entered into group-level analyses, with participants regarded as random effects, using a one-sample t-test against a contrast value of 0 at each voxel. The group-level model included covariates for gender, age, and the task order. For all GLM analyses, we conducted ROI analysis; the results were corrected for multiple comparisons using small volume correction (SVC) within ROIs.

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

673 outcome was the product of feedback multiplied by reward/punishment sensitivity. The a priori
674 ROI was the striatum, which is widely known to process RPE [12,14,120]. The striatum ROI
675 was constructed anatomically by combining the AAL3 definitions of bilateral caudate, putamen,
676 olfactory bulb, and nucleus accumbens.

677 *GLM2 (Hypothesis 2):* GLM2 was used to test hypothesis 2: neural responses
678 associated with Pavlovian bias would be increased under WM load. Specifically, the GLM
679 examined whether the difference between the anticipatory response to fractal cues in
680 Pavlovian-congruent trials and Pavlovian-incongruent trials was greater in the WMGNG task
681 than in the GNG task in regions associated with Pavlovian bias. Therefore, GLM included the
682 following regressors: (1) cue onset of “go to win” trials, (2) cue onset of “no-go to win” trials,
683 (3) cue onset of “go to avoid losing” trials, (4) cue onset of “no-go to avoid losing” trials, (5)
684 target onset of “go” trials, (6) target onset of “no-go” trials, (7) outcome onset of win trials, (8)
685 outcome onset of neutral trials, (9) outcome onset of loss trials, (10) wait onset (i.e., inter-trial
686 interval). We compared the main effect of Pavlovian bias (Pavlovian-congruent trials -
687 Pavlovian-incongruent trials) between two tasks ($[(1) + (4) - ((2) + (3))]_{WMGNG} - [(1) + (4) ((2) +$
688 $(3))]_{GNG}$). The a priori ROIs were the striatum and SN/VTA, which are considered important in
689 Pavlovian bias [50–54]. The striatum ROI was constructed anatomically by combining the
690 AAL3 definitions of bilateral caudate, putamen, olfactory bulb, and nucleus accumbens.
691 Furthermore, the SN/VTA was constructed by combining the AAL3 definitions of bilateral SN
692 and VTA.

693 *GLM3 (Hypothesis 3):* GLM3 was used to test hypothesis 3: value comparison signals
694 would decrease under WM load. GLM3 was also implemented with a model-based fMRI
695 approach: (1) cue onset of all trials, (2) cue onset parametrically modulated by the trial-by-trial
696 decision values ($W_{chosen} – W_{unchosen}$), (3) target onset of “go” trials, (4) target onset of “no-go”
697 trials, (5) outcome onset, and (6) wait onset (i.e., inter-trial interval). Decision value regressors
698 were calculated by subtracting the action weights of the unchosen option ($W_{unchosen}$) from the
699 action weights of the chosen option (W_{chosen}). We then compared the main effect of decision
700 value between two blocks ($(2)_{WMGNG}$-$(2)_{GNG}$). The a priori ROI for GLM3 was the vmPFC, which

701 was suggested as a region that encodes the relative chosen value ($W_{chosen} - W_{unchosen}$)
702 [121,122]. Here, ROI masks were created by drawing a sphere with a diameter of 10 mm
703 around the peak voxel reported in the previous studies ([-6,48,-8] for vmPFC [121]).
704 *PPI analysis:* In addition to GLMs, we used PPI analysis to test whether WM load led
705 to reduced cooperation between WM and RL systems for learning [15,18] by using PPI
706 analysis. Here, to examine differences between the two blocks in terms of functional coupling
707 between the prefrontal areas and the area computing RPE after choices, we performed PPI
708 analysis using the gPPI toolbox [67]; the physiological variable was the time course of the
709 striatum in the anticipation phase, and the psychological variable was the effect of WM load
710 during the same phase. As a seed region (i.e., a physiological variable), the cluster striatum
711 ROI (peak MNI space coordinates x = 13, y = 14, z = -3) was derived from the results of GLM1,
712 which revealed stronger RPE signaling in the WMGNG task than in the GNG task during the
713 feedback phase. The entire time series throughout the experiment was extracted from each
714 participant in the striatum ROI. To create the PPI regressor, these normalized time series were
715 multiplied by task condition vectors for the anticipation phase, which consisted of the cue
716 representation and fixation phases as in GLM1. A GLM with PPI regressors of the seed region
717 was thus generated together with movement regressors. The effects of PPI for each
718 participant were estimated in the individual-level GLM; the parameter estimates represented
719 the extent to which activity in each voxel was correlated with activity in the striatum during the
720 anticipation phase. The contrast was constructed by subtracting activity during the anticipation
721 phase in the GNG task from activity in the WMGNG task (WMGNG vs. GNG in the anticipation
722 phase). Individual contrast images for functional connectivity were then computed and entered
723 into one-sample t-tests in a group-level GLM together with nuisance covariates (i.e., gender,
724 age, and task order). Whole-brain cluster correction was applied for PPI analysis.
725

726 **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.

Supporting Information

Text S1. fMRIPrep details.

Text S2. fMRI results details.

Fig S1. 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 S2. Posterior distributions of individual go bias, reward sensitivity, punishment sensitivity parameters. Dots indicate medians and thick bars indicate 95% HDIs.

Fig S3. Posterior distributions of individual Pavlovian bias and Pavlovian bias under WM load parameters. Dots indicate medians and thick bars indicate 95% HDIs.

Fig S4. Posterior distributions of individual learning rate and learning rate under WM load parameters. Dots indicate medians and thick bars indicate 95% HDIs.

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

Fig S6. Results of fMRI replication analyses. (A) The main effect of WM load during the anticipation phase. As reported in previous meta-analysis studies [29,123], 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

754 had cognitive loads in the brain level. (B) The main effect of reward outcome. The reward was
755 significantly associated with the signal in the striatum (MNI space coordinates x = 18, y = 5, z
756 = -12, Z = 4.47, p < 0.05 whole-brain cluster-level FWE) and ventromedial PFC (MNI space
757 coordinates x = -3, y = 68, z = 4, Z = 4.92, p < 0.05 whole-brain cluster-level FWE). (C) The
758 main effect of loss outcome. Loss-related regions such as the insula (MNI space coordinates
759 x = -36, y = 19, z = -10, Z = 6.77, p < 0.05 whole-brain cluster-level FWE) showed increased
760 BOLD response. (B) and (C) are consistent with the previous findings suggesting reward- and
761 loss-related regions [124,125]. Overlays are shown with a threshold of p < 0.001 (uncorrected).
762 Color scale indicates t-values.

763 **Table S1. Statistics of posterior distributions of group parameters for the best-fitting**
764 **model.** $\hat{R}$ values [115] for all parameters were close to 1.00 (< 1.1), which indicates that the
765 estimated parameter values converged to their target posterior distributions.

766 **Table S2. LOOIC for each model.**

767 **Table S3. Whole-brain results from replication analyses.** Whole-brain cluster-level family-
768 wise error (FWE) for multiple comparison with a cluster-forming threshold of p < 0.001. * The
769 insula (peak in insula: x = -36, y = 19, z = -10, *Z* = 6.77) reported in the result (**Fig S6C**) was
770 included in this cluster of which the peak coordinate was located in orbital part of IFG.

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

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

776 **Table S6. Whole-brain results from PPI analysis.** * The dlPFC (peak in dlPFC: x = -20 y =
777 63, z = 23, Z =-4.24) reported in the result (**Fig 5B**) was included in this cluster of which peak
778 coordinate was located in medial SFG.

779

780 ## Author Contributions

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

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

Formal analysis: Heesun Park, Hoyoung Doh.

Funding acquisition: Woo-Young Ahn.

Investigation: Heesun Park, Hoyoung Doh, Harhim Park.

Methodology: Heesun Park, Hoyoung Doh, Harhim Park.

Project administration: Heesun Park, Woo-Young Ahn.

Resources: Woo-Young Ahn.

Software: Heesun Park, Hoyoung Doh, Harhim Park.

Supervision: Woo-Young Ahn.

Validation: Heesun Park, Hoyoung Doh, Harhim Park.

Visualization: Heesun Park, Hoyoung Doh.

Writing – original draft: Heesun Park.

Writing – review & editing: Hoyoung Doh, 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.

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

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

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

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

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

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

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

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

44. Williams DR, Williams H. AUTO-MAINTENANCE IN THE PIGEON: SUSTAINED

PECKING DESPITE CONTINGENT NON-REINFORCEMENT 2. J Exp Anal Behav. 1969;12: 511–520.

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

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

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

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

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

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

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

52. 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.

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

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

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

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

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

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

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

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

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

76. Lim S-L, O'Doherty JP, Rangel A. The Decision Value Computations in the vmPFC and

Striatum Use a Relative Value Code That is Guided by Visual Attention. J Neurosci. 2011;31: 13214–13223.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1015 93. Albrecht MA, Waltz JA, Cavanagh JF, Frank MJ, Gold JM. Reduction of Pavlovian Bias
1016 in Schizophrenia: Enhanced Effects in Clozapine-Administered Patients. PLoS One.
1017 2016;11: e0152781.

1018 94. Cools R. The costs and benefits of brain dopamine for cognitive control. Wiley
1019 Interdiscip Rev Cogn Sci. 2016;7: 317–329.

1020 95. Barrouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and Cognitive Load
1021 in Working Memory. J Exp Psychol Learn Mem Cogn. 2007;33: 570–585.

1022 96. Dalley JW, Cardinal RN, Robbins TW. Prefrontal executive and cognitive functions in
1023 rodents: neural and neurochemical substrates. Neurosci Biobehav Rev. 2004;28: 771–
1024 784.

1025 97. Granon S, Vidal C, Thinus-Blanc C, Changeux J-P, Poucet B. Working memory,
1026 response selection, and effortful processing in rats with medial prefrontal lesions. Behav
1027 Neurosci. 1994;108: 883.

1028 98. Oberauer K. Working Memory and Attention – A Conceptual Analysis and Review.
1029 Journal of Cognition. 2019;2: 36.

1030 99. Ridderinkhof KR, Van Den Wildenberg WPM, Segalowitz SJ, Carter CS. Neurocognitive
1031 mechanisms of cognitive control: the role of prefrontal cortex in action selection,
1032 response inhibition, performance monitoring, and reward-based learning. Brain Cogn.
1033 2004;56: 129–140.

1034 100. Ripke S, Hübner T, Mennigen E, Müller KU, Rodehacke S, Schmidt D, et al. Reward
1035 processing and intertemporal decision making in adults and adolescents: The role of
1036 impulsivity and decision consistency. Brain Res. 2012;1478: 36–47.

1037 101. Seo M, Lee E, Averbeck BB. Action selection and action value in frontal-striatal
1038 circuits. Neuron. 2012;74: 947–960.

1039 102. Szmalec A, Vandierendonck A, Kemps E. Response selection involves executive
1040 control: Evidence from the selective interference paradigm. Mem Cognit. 2005;33: 531–
1041 541.

1042 103. Turi Z, Mittner M, Opitz A, Popkes M, Paulus W, Antal A. Transcranial direct current
1043 stimulation over the left prefrontal cortex increases randomness of choice in
1044 instrumental learning. Cortex. 2015;63: 145–154.

1045 104. Laureiro-Martinez D, Brusoni S, Tata A, Zollo M. The Manager's Notepad: Working
1046 Memory, Exploration, and Performance. Journal of Management Studies. 2019;56:
1047 1655–1682.

1048 105. Master SL, Eckstein MK, Gotlieb N, Dahl R, Wilbrecht L, Collins AGE. Disentangling
1049 the systems contributing to changes in learning during adolescence. Dev Cogn
1050 Neurosci. 2020;41: 100732.

1051 106. Nassar MR, Frank MJ. Taming the beast: extracting generalizable knowledge from
1052 computational models of cognition. Current opinion in behavioral sciences. 2016;11: 49–
1053 54.

1054 107. Talmi D, Dayan P, Kiebel SJ, Frith CD, Dolan RJ. How humans integrate the
1055 prospects of pain and reward during choice. Journal of Neuroscience. 2009;29: 14617–
1056 14626.

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

109. 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.

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

111. 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.

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

113. 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.

114. 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.

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

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

117. 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.

118. 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.

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

120. 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.

121. 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.

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

123. 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.

124. 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.

125. 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.