Working Memory-Supported Reinforcement Learning Related to Mental Health Phenotypes in a Representative Sample

Laura A. Bustamante, Krishn Bera, Guillaume J. Pagnier, Alexis Cruz, Alyssa Zimmerman, Deanna M. Barch, Megan A. Boudewyn, Cameron S. Carter, Molly Erickson, James Gold, Steven J. Luck, J. Daniel Ragland, Andrew P Yonelinas, Angus W. MacDonald III * & Micheal J. Frank *. Correspondence: *bustamante@wustl.edu*

Abstract

Working memory-supported reinforcement learning (RL-WM) may differ between participants with and without schizophrenia and mood disorders. It remains unexplored whether such processes are related to distinct dimensional phenotypes. This study tests whether RL-WM parameters are associated with self-reported mental health phenotypes (anxiety, depression, drug use, impulsivity, mania, motivation/pleasure, schizotypy) in a US representative online sample (N=2,300 exploratory). Participants completed a stimulus-response learning task that manipulates WM demands (set size, delay) to disentangle WM and RL contributions to performance (N=1,665 post-exclusion, ages 18-65 years). The RL-WM model estimated participants' reliance on WM (v. RL) reliance, WM capacity, WM decay (interference from intervening trials), RL learning rate, perseveration (negative feedback neglect), and undirected noise (attention lapse) parameters. We tested associations between 8 RL-WM measures and survey items using multivariate sparse partial least squares regression (m-SPLS). m-SPLS Cross-validation identified 1 component with 31 survey items predicting 7 RL-WM measures. Items reflecting schizotypy, mania, and others, were related to reduced RL and WM performance (increased perseveration, WM decay, undirected noise, decreased WM reliance). Results support RL-WM utility for computationally mechanistic community mental health research. Replicability will be tested in an independent sample (N=6,725 confirmatory).

Keywords: reinforcement learning; working memory; computational psychiatry; schizotypy; mania

Background

Reinforcement learning (RL) involves learning actions to take in an environment via experience receiving rewards/penalties. RL measures are popular in computational psychiatry, for example, comparing learning rate in individuals with and without psychiatric conditions (Halahakoon et al., 2020). However, many RL tasks and models conflate incremental RL processes (i.e., those thought to reflect dopaminergic plasticity), with fast, time-limited working memory (WM) processes (reflecting prefrontal activation-based processing) (A. G. E. Collins & Frank, 2012, 2018). This unmeasured contribution of fast-updating stimulus-action associations in WM may inflate RL learning rate estimates, such that individuals with lower WM function may inaccurately appear to have slower RL learning rate (A. G. Collins, Brown, Gold, Waltz, & Frank, 2014).

Tasks designed to disentangle reinforcement learning and working memory (RL-WM) contributions to learning and choices suggest differences between individuals with and without schizophrenia (A. G. Collins et al., 2014; A. G. Collins, Albrecht, Waltz, Gold, & Frank, 2017), and mood disorders (Cheng, Moser, Jones, & Kaiser, 2024). The present study explores whether these WM and RL processes are related to distinct dimensional phenotypes relevant to mental health in the general population. We present results from an exploratory sample, which will be used for pre-registered hypotheses to test in a confirmatory sample.

Methods

Online volunteers from Prolific (18-65 years) completed a 2-session study. Of the 2,552 submissions, 2,300 passed initial bot and attention check quality controls. The sample was reasonably comparable to 2020 US Census demographics. After providing informed consent, participants completed surveys on anxiety (Generalized Anxiety Disorder-7), depression (Patient Health Questionnaire-8), impulsivity (Abbreviated-Barrat Impulsivity Scale), disability (WHO-Disability Assessment Schedule), mania (Altman Rating Scale for Mania), motivation and pleasure (MAP-SR), schizotypy (Schizotypyal Personality Questionnaire), drug use (Drug Use Questionnaire) and a battery of 6 cognitive tasks including RL-WM.

RL-WM paradigm

The RL-WM task consists of a 'train phase' during which participants learn stimulus-response mappings in a three-alternative forced choice (10 blocks, 370 trials, Figure 1) (A. G. Collins et al., 2017), and 'test phase' involving a judgment about which of two previously encountered stimuli was associated with more reward (122 trials). The task varies WM demands by manipulating the number of unique stimuli to be learned within a block (between 2 and 5) as well as the delay between successive encounters of each stimulus.

The RL-WM model learns stimulus-response mappings using RL-style temporal-difference updates. These learning updates are based on two parallel RL and WM processes. While both processes maintain a separate state-action value representation, the learning rate of the WM system is fixed at 1 to capture the fast, one-shot updating of the contents of working memory. However, WM representations are capacity-limited and subject to decay over time. The final action policy is derived from a weighted sum of RL and WM policies. The RLWM model equations are those described in Westbrook et al. (2024) (Westbrook et al., 2024). The model was fit to data using a two-pass hierarchical maximum like-

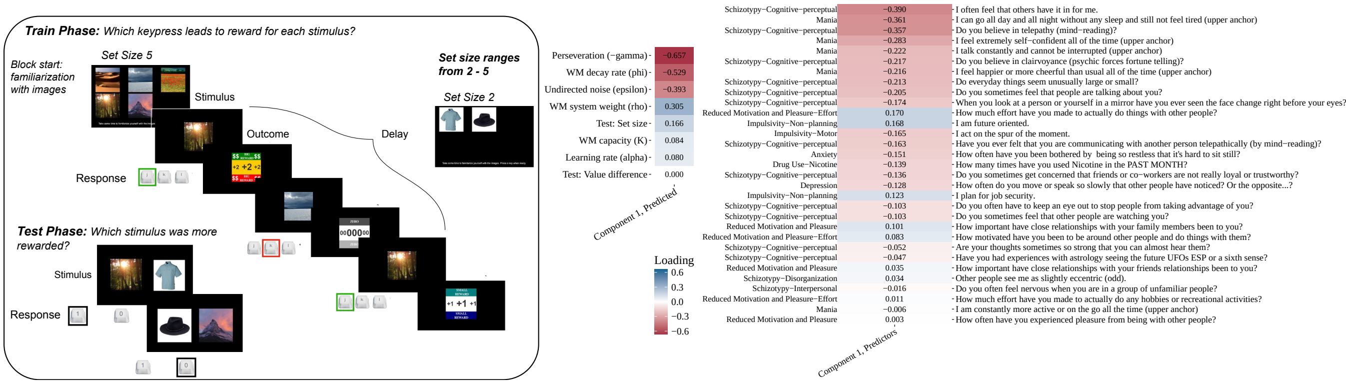

Figure 1: Left: task schematic. Righ: Multivariate sparse partial least squares regression. Top: RL-WM measure loadings. Bottom: survey item loadings, survey name is displayed on the left and item text on the right.

lihood estimation (MLE) procedure, using Nelder-Mead optimization to get the MLE estimates (`scipy.optimize`). Parameters were the WM (v. RL) reliance, WM capacity (max stimuli in WM), WM decay rate (interference from intervening trials), RL learning rate, perseveration (negative feedback neglect), and undirected noise (attention lapse). Test phase performance was measured by extracting the random effects of a logistic mixed-effects regression model predicting whether participants chose the left option by 1) the difference in set size (possibly reflecting effort avoidance), and 2) the difference in value (possibly reflecting RL performance) between the left and right option (`lme4`).

Quality control exclusions were conducted. Train phase data was excluded for participants with too few trials, or high error rate (>50%)/invalid response rate (too fast or missed deadline) (initial N=2,011, final N=1,664). Test phase data was excluded for participants that were excluded in train phase, had too few trials, a high rate of invalid responses (key other than the two valid response keys)/pressing any single key (initial N=2,004, final N=1,527). Trials with reaction times <150 ms were excluded.

Results

Consistent with previous studies, training phase accuracy was higher the greater the number of previous correct responses for a stimulus (z=72.91, p<0.042), decreased as a function of set size (z=-2.04, p<0.042), and decreased with the delay since previous correct response for a stimulus (z=-37.44, p<0.001). Test phase choices were sensitive to value (z=47.03, p<0.001) and set size differences (z=-16.6, p<0.001). This online experiment procedure yielded RL-WM behavior consistent with prior studies, and the fitted RL-WM models were able to capture key patterns in the data.

Associations to mental health phenotypes

Previous work using dimensionality reduction techniques on mental health survey items have identified latent dimensions with generalizable relationships to RL phenotypes (Gillan, Kosinski, Whelan, Phelps, & Daw, 2016; Fox et al., 2024). Here, we build on that approach by using multivariate sparse partial least squares regression (m-SPLS), a technique that simultaneously performs variable selection and dimensionality reduction on two domains of data. We predicted 8 RL-WM measures by 98 survey items using 10-fold cross-validation (`mixOmics` package). All variables were scaled residuals after controlling for demographic factors (age, sex at birth, gender, trans-gender, race, Hispanic/latine ethnicity, education, urban/rural, and income). m-SPLS selected 1 component, which included 7 of the RL-WM measures (proportion variance explained=0.072) and 31 items (proportion variance explained=0.193) (Figure 1). Overall the RL-WM loadings for this component indicated better RL and WM system performance. The strongest loadings were observed for decreased perseveration, WM decay rate, undirected noise, and increased WM system weight, and weaker loadings for increased test phase set size sensitivity, WM capacity, and learning rate. This better RL-WM performance negatively associated to survey scores, especially for items in the cognitive-perceptual schizotypy subscale and the mania scale. The top loading items were *"I often feel that others have it in for me"*, *"I can go all day and all night without any sleep and still not feel tired"*, *"Do you believe in telepathy (mind reading)?"*, and *"I feel extremely self-confident all of the time"*. These are consistent with reports of increased WM decay rate (ϕ) associated with mania scores in participants with bipolar disorder ((Cheng et al., 2024)) and between participants with and without schizophrenia ((A. G. Collins et al., 2014, 2017)). The better pattern of RL-WM performance was intriguingly positively related to *reduced* motivation and measure *"How much effort have you made to actually do things with people?"*, and increased impulsivity *"I am future oriented"* (reverse scored).

These results support the utility of the RL-WM task as a computational phenotyping tool useful for studying mechanisms of mental health (especially schizotypy, mania) in community samples, though variance explained in this large sample was low. We will test the replicability of these findings in an independent sample of 6,725 participants, drawing on the entire dataset collected to become publicly available.

References

Cheng, Z., Moser, A. D., Jones, M., & Kaiser, R. H. (2024). Reinforcement learning and working memory in mood disorders: A computational analysis in a developmental transdiagnostic sample. *Journal of Affective Disorders*, *344*, 423–431.

Collins, A. G., Albrecht, M. A., Waltz, J. A., Gold, J. M., & Frank, M. J. (2017). Interactions between working memory, reinforcement learning and effort in value-based choice: a new paradigm and selective deficits in schizophrenia. *Biological psychiatry*, *82*(6), 431.

Collins, A. G., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. *The Journal of Neuroscience*, *34*(41), 13747–13756.

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European Journal of Neuroscience*, *35*(7), 1024–1035.

Collins, A. G. E., & Frank, M. J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), 2502–2507.

Fox, C. A., Teckentrup, V., Donegan, K. R., Seow, T., Benwell, C., Tervo-Clemmens, B., & Gillan, C. (2024). Mechanistic arbitration between candidate dimensions of psychopathology. *OSF*.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife*, *5*, e11305.

Halahakoon, D. C., Kieslich, K., O'Driscoll, C., Nair, A., Lewis, G., & Roiser, J. P. (2020). Reward-Processing Behavior in Depressed Participants Relative to Healthy Volunteers: A Systematic Review and Meta-analysis. *JAMA psychiatry*, *77*(12), 1286–1295.

Westbrook, A., Bosch, R. v. d., Hofmans, L., Papadopetraki, D., Määttä, J. I., Collins, A. G., . . . Cools, R. (2024). Striatal Dopamine Can Enhance Learning, Both Fast and Slow, and Also Make it Cheaper. *bioRxiv*.