**Lei, W. et al.: Imbalanced goal-directed and habitual control in individuals with Internet Gaming Disorder**

https://doi.org/10.1556/2006.2025.00037

**Supplementary materials**

## Supplementary Methods

### *Computational model of choice behaviours*

For the computational modelling of the choice behaviours, we closely followed the methodology employed in the previous study (Daw, Gershman, Seymour, Dayan, & Dolan, 2011). The model script was adopted based on the *ts_par7.stan* from the *hBayesDM* package in R (Ahn, Haines, & Zhang, 2017). Our code is nearly identical to the *ts_par7.stan*, except we did not include hyperparameters in our implementation. The model was fitted using *rstan* and its default *No-U-Turn Sampler* in R (Carpenter et al., 2017). For the sake of completeness, we report the algorithm used in our implementation here.

The task involves two states, $s_i$, that are visited during each trial $t$, where the index $i$ indicates the state. The 1st stage state, $s_{1,t}$, is always the same. The 2nd stage states are denoted as $s_{21,t}$ and $s_{22,t}$, with only one visited per trial. Actions $a_{i,t} = \{1, 2\}$ lead to transitions between the successive states, which are noted as $a_{1,t}$ at the 1st stage and $a_{2,t}$ at the second stage. The 1st stage rewards, $r_{1,t}$, which is always zero, while the 2nd stage rewards, $r_{2,t}$, can be either zero or one, depending on the reward probability.

### Learning algorithm of the habitual system

We used the SARSA(λ) temporal difference learning algorithm to model habitual choices. Here, the action values $Q_{TD}(s_{i,t}, a_{i,t})$ are used to compute the reward prediction error (RPE), $\delta_{i,t}$, which is the difference between the obtained reward and expected reward based on the current action value:

$$(1)\quad \delta_{i,t} = r_{i,t} + Q_{TD}(s_{i+1,t}, a_{i+1,t}) - Q_{TD}(s_{i,t}, a_{i,t})$$

$Q_{TD}$ values are then updated based on the RPE using the SARSA(λ) update rule:

$$(2)\quad Q_{TD}(s_{i,t}, a_{i,t}) = Q_{TD}(s_{i,t}, a_{i,t}) + \alpha_i \delta_{i,t}$$

Where $\alpha_i$ represents the free learning parameters. Separate learning rates $\alpha_1$ and $\alpha_2$ were used for 1st and 2nd stages, respectively. Moreover, the eligibility trace parameter $\lambda$ was used to update the 1st stage $Q$ values based on 2nd stage RPE:

(3) $Q_{TD}(s_{1,t}, a_{1,t}) = Q_{TD}(s_{1,t}, a_{1,t}) + \alpha_1 \lambda \delta_{2,t}$

**Learning algorithm of the goal-directed system**

Unlike the habitual system, which does not consider the structure of the transitions between states, the goal-directed system computes the 1st stage action values $Q_{MB}(s_{1,t}, a_{1,t})$ at each trial by constructing a model-based representation of the task. This involves building a tree-like structure of possible future states, actions, and transitions, and then weighting each action's expected outcome by the probability of its occurrence:

(4) $Q_{MB}(s_{1,t}, a_{1,t}) = P(s_{21,t}|s_{1,t}, a_{1,t}) \max_a Q(s_{21,t}, a_{21,t}) +$
$P(s_{22,t}|s_{1,t}, a_{1,t}) \max_a Q(s_{22,t}, a_{22,t})$

Subjects were instructed that transitions between the 1st and 2nd stages are probabilistic, with one transition being more probable than the other (common transition vs. rare transition). In the computational model, the probability of the common transition was set to 0.7, while the rare transition probability was set to 0.3.

The Goal-directed and habitual action values were then combined via weighting parameter $\omega$:

(5) $Q_{net}(s_{1,t}, a_{1,t}) = (1 - \omega) Q_{TD}(s_{1,t}, a_{1,t}) + \omega Q_{MB}(s_{1,t}, a_{1,t})$

Here, the relative contribution of model-free and model-based action values is determined by the free weighting parameter $\omega$. A parameter value of $\omega = 0$ indicates fully model-free choice and $\omega = 1$ indicates fully model-based choice. At the 2nd stage, $Q$ values are assumed to be identical for both goal-directed and habitual systems: $Q_{net}(s_{2,t}, a_{2,t}) = Q_{TD}(s_{2,t}, a_{2,t}) = Q_{MB}(s_{2,t}, a_{2,t})$. The integrated $Q$ values were passed through a softmax function to compute action probabilities:

(6) $P(a_{i,t} = a|s_{i,t}) = \frac{\exp(\beta_i [Q_{net}(s_{i,t}, a_{i,t}) + \pi\, rep(a)])}{\sum_A \exp(\beta_i [Q_{net}(s_{i,t}, A_{i,t}) + \pi\, rep(A)])}$

This is governed by free inverse-temperature parameters $\beta$, which control the noisiness of the choices. When $\beta = 0$, the choices are purely random, regardless of the action values. When $\beta = \infty$, the choices deterministically prefer the actions with the highest expected value. In this

model, separate inverse-temperature parameters were used for the 1st and 2nd stages of the task, denoted as $\beta_1$ and $\beta_2$ respectively.

Additionally, the model included a free parameter $\pi$ that captures the effect of first-order perseveration or switching in the first-stage choices. For second-stage actions, $\pi$ was set to zero. The effects of 1st stage choice repetition were implemented via the indicator variable *rep(a)*, which was set to 1 if the current 1st stage action was the same as the previous trial's action, and 0 otherwise (i.e., $a_{1,t} = a_{1,t-1} \rightarrow rep(a) = 1$ and for $a_{1,t} \neq a_{1,t-1} \rightarrow rep(a) = 0$).

In total, the model contains seven free parameters $\theta = (\beta_1, \beta_2, \alpha_1, \alpha_2, \lambda, \omega, \pi)$.

**First-level General Linear Model (GLM) for fMRI data analysis**

We also followed the approach described by Daw et al. (2011) to set up the first-level GLM for fMRI data analysis. Specifically, we computed RPEs with respect to the integrated action values $Q_{net}$, which captures a weighted mixture of both model-based and model-free action values:

$$(7)\quad \delta_{net;i,t} = Q_{net}(s_{i+1,t}, a_{i+1,t}) + r_{i,t} - Q_{net}(s_{i,t}, a_{i,t})$$

This effectively defines a generalized version of Equation (1). Following Daw et al. (2011), we calculated the difference between $\delta_{net;i,t}$ computed for purely model-based ($\omega = 1$) versus purely model-free ($\omega = 0$) choice.

The $RPE_{MF}$ and $RPE_{\Delta MB}$ (denoted as $RPE_{MB}$ in this study) were parametric modulators locked onto at the second-stage onset and reward receipt. The time series of these two modulators are our contrasts of interest. Furthermore, we included several nuisances and regressors of no interest. This included an onset regressor at the outcome presentation to control for general differences in mean activation between choice and outcome events. An onset regressor of the 1st stage stimulus presentation with two additional parametric modulators of no interest: $P(a_{1,t}|s_A)$ from Equation 6 "as a normalized measure of the 1st stage action value" and "its partial derivative with respect to ω" (see Daw et al., 2011 in Supplementary). In addition, we included the six motion parameters from the realignment of the fMRI data preprocessing to control movement artifacts.

**Supplementary Results**

**Table S1.** Random effects in mixed-effects logistic regression.

| | Variance | SD | $\chi^2$(df=1) | p |
|---|---|---|---|---|
| **HC** | | | | |
| **(Intercept)** | 1.98 | 1.41 | | |
| **Reward** | 0.05 | 0.22 | 6.19 | **0.013** |
| **Transition** | 0.09 | 0.30 | 0.44 | 0.508 |
| **Reward×Transition** | 0.55 | 0.74 | 4.10 | **0.043** |
| **IGD** | | | | |
| **(Intercept)** | 1.92 | 1.39 | | |
| **Reward** | 0.10 | 0.31 | 22.60 | **<0.001** |
| **Transition** | 0.53 | 0.73 | 0.62 | 0.431 |
| **Reward×Transition** | 0.25 | 0.50 | 4.60 | **0.032** |

Note: Significant values were highlighted in bold fonts.

**Table S2. Model Comparisons Between Hybrid Model and Its Special Cases**

| | AIC | BIC | LL | No. subjs favoring Hybrid | Aggregate LRT favouring Hybrid | p-value for LRT |
|---|---|---|---|---|---|---|
| **Hybrid** | 27083.21 | 32014.19 | -12046.61 | - | - | - |
| **TD only** | 27206.49 | 32137.48 | -12108.25 | 10 | -123.28 | <0.001 |
| **MB only** | 27788.77 | 32719.76 | -12399.39 | 42 | -705.56 | <0.001 |
| **Hybrid λ = 0** | 27238.21 | 32169.20 | -12124.11 | 16 | -155.00 | <0.001 |
| **Hybrid λ = 1** | 27095.81 | 32026.80 | -12052.91 | 2 | -12.60 | <0.001 |
| **Null** | 53350.52 | 54422.47 | -26350.26 | 65 | -28607.31 | 0 |

**Note:** TD: temporal-difference (TD) algorithm SARSA(λ); MB: model-based; Hybrid λ = 0: hybrid model with λ restricted to 0; Hybrid λ = 1: hybrid model with λ restricted to 1; Null: null model where choices were assumed made by chance; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LL: raw log-likelihood; No. subjs favoring Hybrid: the number of subjects favoring the hybrid model on a likelihood ratio test ($p < 0.05$); Aggregate LRT favoring Hybrid: test statistic and p-value for a likelihood ratio test against the hybrid model, aggregated across subjects.

## References


Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry (Cambridge, Mass.), 1*, 24.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A Probabilistic Programming Language. *J Stat Softw, 76*.

Daw, Nathaniel D., Gershman, Samuel J., Seymour, B., Dayan, P., & Dolan, Raymond J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204-1215.