Feedback Perseverance Model of the Task Switching Paradigm

Hendrik Strauss, Esma Hodzic, Sophie Terhardt, Silas Sinning, Misha Zimová, Yehor Safroniuk

Introduction

What is the phenomenon you want to model? (0.5 points)
Our primary focus is on sequential effects in the task-switching paradigm. We are mainly interested in three phenomena, (1) the performance on the switch trials decreases due to change in context, also known as switch cost. In addition we will study the effect prolonged same-task (repetition) sequences have on switch cost. Here we distinguish between two phenomena: in highly repetitive environments, the switch cost increases with repetition sequence length, while in frequently changing environments, the switch cost decreases with consecutive repeats (due to switch anticipation). Given that switches are common in our experimental setup, we investigate only the latter phenomenon (2). Lastly, we will integrate the effect of feedback on task performance, more precisely, how negative feedback leads to increased focus on the anticipation stratagem. I.e. higher accuracy on incoming switch trials, and lower on repeats (3). In practice, we explore these phenomena through accuracy measurements dependent on two key factors: (A) the specific sequence of trial types (repetition vs. switch) and (B) the provided feedback on response correctness.

Why is that phenomenon relevant for understanding human cognition? (0.5 points)
These phenomena are relevant for understanding human cognition because they reveal how expectations and the current allocation of attention could be adjusted based on prior experiences beyond the immediate task under consideration. Sequential effects and switch costs are intertwined with attentional flexibility and may be integral to understanding the mechanisms underlying task monitoring as well as to how a system integrates external feedback with internal states of attention, therefore highlighting the role of higher-order cognitive control in adapting attention in a changing environment.

Methods

Why is this modeling method appropriate for understanding the phenomenon? (1 point)
To investigate the mentioned phenomena we employ a dynamical systems modeling approach. This approach is appropriate for our goals since it, based on its own dynamics and inputs, integrates information over time, which is useful when trying to model sequential/ lasting effects. Importantly it can also describe the cognitive stability-flexibility trade-off of attention in terms of energy landscapes, by modeling states of human attention with attractors and oscillators ((Deco & Rolls, 2005)). Crucially, this method allows us to study the complex dynamics of the brain in an intuitive way by making simplifying assumptions about the dynamic interactions between different sub-networks and regions.

Which hypothesis/hypotheses do you seek to test by contrasting two (or more) models? (1 points)

We hypothesize (1) that the increase in switch cost with the length of the consecutive sequence of repeat trials (in case of a low overall number of switches) is due to repetition perseverance, while the decrease of switch cost (when switch trials are common) is via switch anticipation, both are modulated by a higher cognitive system responsible of expectation adaptation((Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2012). We anticipate that humans can reach highly attentive states while maintaining a high readiness for an attention switch if that switch is anticipated. On top of that, (2) we hypothesize that feedback on the correctness of the given response modulates the reliance on the applied strategy (i.e. task perseverance/switch anticipation) so that negative feedback leads to more reliance on the strategy, rather than Base Model attention dynamics. To test this we will be contrasting three models on their fit to human data: (A) the *Base model*, (B) the *Perseverance (P) model*, and (C) the *Feedback Perseverance (FP) model*.

Description of computational model(s)

What are the inputs, system properties, and outputs of your model(s)? (1 point)

The *Base Model* consists of two variable nodes X_1 and X_2, each representing a population of neurons responsible for integrating evidence for the two tasks. Since the tasks are mutually exclusive, the nodes inhibit each other. Furthermore, two binary (**0** or **1**) external task-relevant input

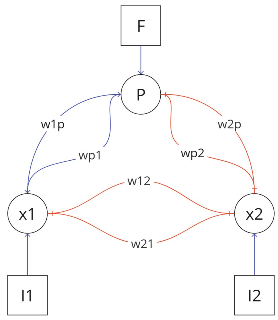


Figure 1: Feedback Perseverance Mode

parameters, I1 and I2, are included, corresponding to the cue of the assigned task. The Perseverance model (P model) extends the Base model by including an additional variable node, P, representing another population of neurons that correspond to higher cognitive control over the integration of the task-relevant evidence. P has reciprocal excitatory connections with the node for one task (X_1) and reciprocal inhibitory connections with the other(X_2)(See Figure 1). This means that P facilitates the activation of task-relevant nodes depending on its sign: a positive P increases activation of X_1, while a negative P increases activation of X_2. Lastly, for the Feedback Perseverance model (FP model), one additional binary input F is added, representing the external feedback provided in the experiment on the correctness of the response (**1** for correct and **-1** for incorrect). It directly modulates the perseverance variable **P** based on the current tendency of the system. Figure 1.
For all three models, the output is drawn from the final activation levels of X_1 and X_2, which ultimately determines the choices on the cued task.

Which assumptions does each model make? (1 point)

All models make a set of simplifying assumptions to capture essential cognitive dynamics. First, they are treated as closed systems: once given input (task cues and feedback), their internal states evolve autonomously. Neural populations are abstracted as continuous variables, allowing us to model attention and control without detailed neural-level complexity. Tasks are mutually exclusive and represented by symmetric units that inhibit each other equally, reflecting competition for cognitive resources. No inherent task bias is assumed; all weights are symmetric unless modulated by dynamics. Feedback influences behavior indirectly by adjusting a shared perseverance variable, which enhances one task's activation while proportionally suppressing the other. This ensures balance and reflects how feedback aligns with the system's current attentional tendency.

Describe the computational implementation of each model (e.g., model formulas) (1 point)

In the following is a full dynamical system model, from which, three distinct hierarchical models can be constructed. The full model consists of three interacting dynamical variables:

- X_1 and X_2:: Neural activity/attention levels for Task 1 and Task 2. higher values correspond to higher attention to task 1 or 2 respectively.
- P: Persistence variable that modulates task perseverance effects, representing higher cognitive control. $P > 0$: increased attention to task 1, $P < 0$: increased attention to task 2.

These variables are governed by the following equations:

Task Attention Dynamics:

$$\frac{dX_1}{dt} = -X_1 + S(g, \text{net}_1)$$
$$\frac{dX_2}{dt} = -X_2 + S(g, \text{net}_2)$$

where the logistic activation function S(g, ·) determines the depth of the attractor of both tasks based on the gain parameter g (> 0):

$$S(g, \text{net}) = \frac{1}{1 + e^{-g \cdot \text{net}}}$$

The net input to each task unit is defined as:

$$\text{net}_1 = -w_{12}X_2 + w_{1P}P + I_1 + \sigma\eta$$
$$\text{net}_2 = -w_{21}X_1 - w_{2P}P + I_2 + \sigma\eta$$

w_{12}, w_{21} are mutual inhibition weights, w_{1P}, w_{2P} modulate the effect of perseverance, I_1, I_2 are external task inputs, P modulates the effect of task perseverance. and $\sigma\eta$ is added Gaussian noise.

Persistence (Task Perseverance / switch anticipation) Dynamics:

$$\frac{dP}{dt} = \tau_P\left(-P + \tanh(\text{net}_P)\right)$$

$$\text{net}_P = \gamma(w_{P1}X_1 - w_{P2}X_2) + \alpha F(-\text{Tendency})$$

γ is the persistence gain parameter, where $\gamma > 0$: leads to *repeat perseverance* and $\gamma < 0$: leads to *switch anticipation* (anti-perseverance). F is external feedback signal (1: correct, -1: wrong response). α (0-1) is learning rate from the feedback. w_{P1}, w_{P2} modulate the effect of task units on persistence dynamics. tau_P time constant to regulate the speed of P in relation to X_1 X_2. Tendency is the current tendency of the system.
The Tendency term aligns the system with the current task dominance and the perseverance mechanism driven by γ:

$$\text{Tendency} = sign(\gamma(X_1 - X_2))$$

Here, $(X_1 - X_2)$ can be interpreted as task dominance.

in case of $\gamma = 0$, Tendency will align directly with task dominance.

For taking an action on a given task, the final activity levels of X_1 and X_2 are fed into a softmax function to get a correct-choice probabilities for each task. The sensitivity of the softmax to the dominant task is modulated, over time, by a parameter c, where $c > 0$ leads to more determined choices over time, $c < 0$ leads to more noisy choices over time.

This formulation captures how task-switching behavior emerges from neural competition - inhibitory tasks interactions, and feedback-modulated perseverance mechanisms.

Finally, in all our models, we fix all the weights ($w_x y$) to 1. The three models are defined as follows: *Baseline Model* (2 free parameters): $\gamma = 0$: No Perseverance effects and $F = 0$: No feedback. P will stay at 0 (or decays and stays at 0 if the initial value is not 0). As a result, the dynamics of the system are governed only by the task attention activity dynamics.

Perseverance Model (3 free parameters): $F = 0$: No feedback. P is governed by the persistence dynamics only., i.e. $\gamma > 0$: Task perseverance. $\gamma < 0$: Switch anticipation.

Feedback Perseverance Model (4 free parameters): the full model as described above.

Description of the experiment

Provide an overview of the experiment. What are the independent variables and dependent variables of the experiment? (0.5 points)

In the task-switching paradigm, participants respond to stimulus pairs consisting of a letter and a number, with the task type determined by their position on a 2×2 grid. If the stimuli appear in the top row, participants perform the **Letter Task**, pressing **B** for consonants (G, K, M, R) and **N** for vowels (A, E, I, U), while ignoring the paired number. If the stimuli appear in the bottom row, they perform the **Number Task**, pressing **B** for odd numbers (3, 5, 7, 9) and **N** for even numbers (2, 4, 6, 8), while ignoring the paired letter. After each response, feedback on its correctness is briefly displayed (a green **"correct"** or red **"wrong"**) before the next stimulus appears. A **3000 ms response time limit** is set, after which the next stimulus is automatically presented following the appearance of a red **"wrong"** text. The independent variable is the trial type, categorized as either a **repeat trial** or a **switch trial**, while the dependent variables are **accuracy** and **reaction time**.

How much data were collected (number participants and trials)? (0.5 points)

Data from 17 participants between the ages of 22 and 29 was collected, all of whom were students. All participants underwent 100 trials. However, some trials were excluded. We excluded the first trial for each participant was excluded as it had no trial type (repeat or switch); as well as the documented responses that surpassed the response time limit, as neither the RT nor the accuracy here could be determined.

Model simulation

Describe the process of simulating data from the model(s). (1 point)

To generate surrogate data, we first generate a sequence of trials matching the relevant task distribution of human experiments. Each model is then simulated on these trials using its specified free parameters and fixed hyperparameters (e.g., trial simulation time T, noise level $\sigma\eta$, and time constant tau_P). For each trial, the model runs for a duration T and produces a choice as detailed in the model implementation section. Importantly, the model's internal dynamics are continuously updated across the entire experiment. Finally, Model choices are compared to the correct responses, and the resulting data are analyzed using the same behavioral metrics applied to participant data.

Model fitting

Describe the process of fitting the model(s) to the data. Remember to describe any preprocessing steps of the data. (2 points)

The data preprocessing involved multiple cleaning steps to ensure the dataset was structured and free from inconsistencies. First, the column names were standardized, unnecessary columns were removed, and string representations of lists in the `key_choices` column were converted into actual lists. Trials unrelated to the main tasks (`letter_task` and `number_task`) were filtered out, and response times (RT) were constrained to a reasonable range (0–3000 ms) to exclude outliers. RT values were also converted from milliseconds to seconds for consistency. Additionally, boolean values were transformed into numerical representations (0,1), and a `trial_type` column was introduced to distinguish between repeat and switch trials. Finally, an accuracy column was computed as a running proportion of correct responses, and choices were adjusted to follow the same convention between surrogate and participant data.

The fitting process involved computing the log-likelihood (LL) of observed choices given the model's parameters. For each trial, input values were determined based on task type, and initial conditions were updated iteratively. The model's predicted choice probabilities were then compared to actual responses, allowing for likelihood estimation. We used a simple grid search over a predefined parameters space to find the parameter set with the highest LL estimate on the provided data, this parameter set is then returned as the best fit of the model to the data.

The free parameters were estimated within predefined ranges: $g \in [0.5, 5]$, $c \in [0.1, 10]$, $\alpha \in [0, 1]$, and $\gamma \in [0, 1]$. The hyperparameters were fixed as follows: $T = 2$, $\sigma = 0.2$, and $\tau_P = 2/T = 1$. The initial condition was set to

$x_0 = [0.5, 0.5, 0]^\top$, ensuring a balanced starting point for the state variables. τ_P was chosen to keep perseverance dynamics slower than the dynamics of X_1 and X_2 in prolonged trial simulation ($T > 2$).

Parameter recovery

Describe how you performed parameter recovery for your models. (1 points)

We performed parameter recovery by generating synthetic data using known parameter values and then fitting the model to recover these parameters. First, we simulated experimental data with predefined values of the free parameters of each model g, c, α, and γ across 25 randomized parameter sets. Each dataset was generated with 100 trials, a time interval of $T = 2$, noise level $\sigma = 0.2$

We then applied the same parameter fitting procedure as mentioned above to identify the values that maximized the log-likelihood of each set of true parameter values. The correlation between true and recovered parameters was assessed using Pearson's correlation r (See Figure 3), additionally, we analyzed cross-correlations among fitted parameters of the same model to evaluate potential inter-dependencies, seen in Figure 4.

Model comparison (& recovery)

Describe how you compared the models. (1 point)

To evaluate and compare our models, we primarily relied on the Bayesian Information Criterion (BIC) as a model fit metric, which balances model fit against complexity by penalizing the number of parameters, the choice of the BIC metric is due to prioritizing the models' interpretability over task performance.

For each candidate model, we computed the BIC using the log-likelihood obtained from the model fitting to our participant data, as well as the corresponding number of free parameters and the number of trials. The BIC values were extracted from the fitted models and ranked in terms of the best-fitted model.

In addition to BIC, we considered parameter recovery as a secondary check. After fitting, we examined whether estimated parameters aligned with expected theoretical values 6, ensuring that the model captured meaningful psychological processes rather than blindly fitting or even overfitting the data. While not a direct comparison metric, parameter recovery provided supplementary insights into model robustness.

Overall, the model with the lowest BIC was selected as the best-fitting model, with additional consideration given to parameter and model stability as well as theoretical plausibility.

Optional: Describe how you performed model recovery. (0.5 bonus points)

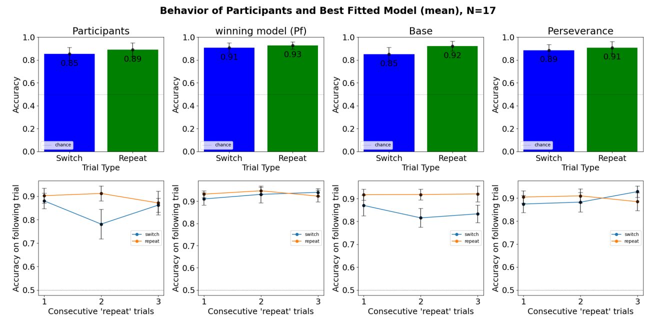


Figure 2: Mean model behavior

We performed model recovery by simulating experimental data using predefined parameter sets and fitting multiple models to this data. The true and fitted models were logged and compared using Bayesian Information Criterion (BIC) scores to identify the best-fitting model. The results were summarized in a confusion matrix and an inversion matrix, providing insights into the accuracy of model identification and predictability respectively.

Results

Simulation results

Which phenomena do the models capture and why? Make sure to support your argument with a plot. (1 point)

All models capture the classic switch cost effect, since they rely on a common architecture of two mutually inhibitory task units. In this setup, when one unit (e.g., X_1) is active, it suppresses the other (i.e., X_2), making it harder to rapidly reconfigure when the task cue changes. This mutual inhibition naturally produces a lag in task switching.

The P and FP models capture the modulation of switch cost as a function of repetition sequence length, such that, as concordant with participant data, after longer sequences of repeat trials, the switch cost decreases see Figure 2). This phenomenon emerges because the model incorporates a higher-order node, P, which integrates past trial history through reciprocal excitatory and inhibitory connections with the task units. In effect, as repetitions accumulate, P counteracts the inertia imposed by mutual inhibition, thereby enabling the system to anticipate a switch and reduce the cost when the task demands change.

Which phenomena do the models not capture and why? (1 point)

The Base model does not capture the effect of repetition sequence length on switch cost (higher switch accuracy after more repetitions) as it only has g to modulate the information integration in X_1, and X_2. While g can easily model the steepness/flatness of the attention landscape of multiple tasks, it fails to account for the ability of humans to reach a highly attentive state while simultaneously maintaining a

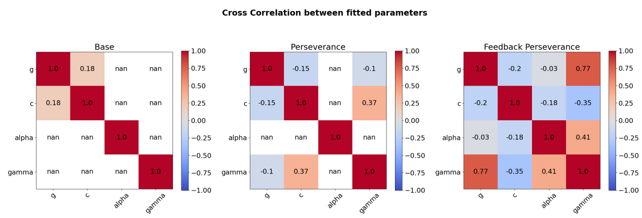
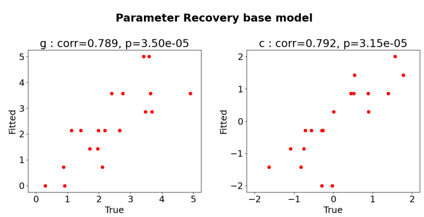


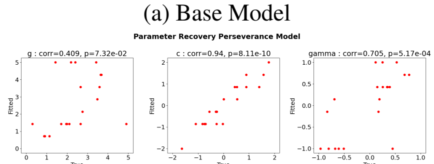
Figure 4: Cross-correlations between parameters

high readiness of attention switch it anticipated. In order to keep the focus of our study and its presentation clear, we did not fully investigate the effect of feedback on switch costs associated with repetition sequence length.

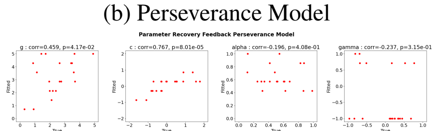
Parameter recovery



(a) Base Model



(b) Perseverance Model



(c) Feedback Perseverance Model

Figure 3: true to fitted parameters correlation

Present your parameter recovery results here. **Which parameters can be recovered more reliably, which less reliably? (1 point)**

Parameter recovery results reveal a strong difference between the recoverability of parameters in the base and P model compared to the FP model. In the FP model, α and γ exhibited weak recoverability, with Pearson correlations of -0.196 and -0.237, respectively, with g (r = 0.459) only c achieved strong recoverability (r = 0.767) (See Figure 3). On further cross-correlation analysis of the parameter (See Figure 4) we found a strong correlation between γ and g (r = 0.77) and α and γ (r = 0.41), this points to a possible reason for the poor recoverability of α and γ.

In contrast, the simpler perseverance model showed generally weaker cross-correlations between parameters (see Figure 4), except for a slight overlap between c and γ. In terms of recoverability, γ exhibited a Pearson correlation of 0.705, g had a moderate recoverability of 0.409, and c demonstrated the highest reliability at 0.94 (See Figure 3). The base model had comparable results to the P-model. These results suggest that the complexity of the FP model contributes to estimation difficulties, particularly for α and γ, while the P model allows for more stable parameter recovery.

Optional: Model recovery

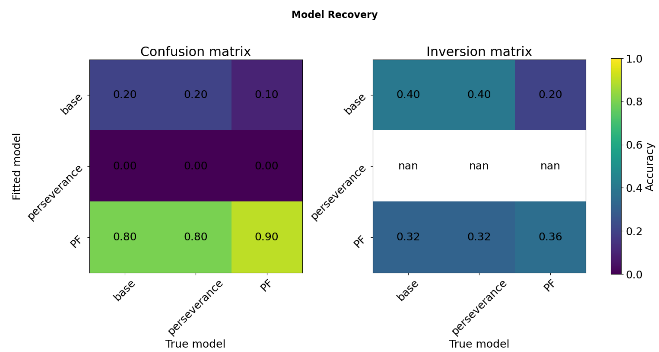


Figure 5: Confusion and inversion matrices

Which models can be recovered more reliably, which less reliably? (0.5 bonus points)

Model recovery results indicate that the P model could not be reliably recovered, as it was never identified correctly. The base model also had poor recoverability, being misclassified as the FP model 80% of the time. In contrast, the FP model exhibited the highest recoverability with a correct classification rate of 90%. The inversion matrix further highlights these discrepancies, showing weak differentiation between the base and perseverance models.

Model comparison

Table 1: Log-Likelihood (LL) Statistics

Model	Count Best	Mean	SE	95% CI
Base	0	-31.31	3.71	[-38.58, -24.03]
Perseverance	0	-30.15	3.68	[-37.36, -22.94]
PF	17	-20.47	3.47	[-27.28, -13.67]

Table 2: BIC Statistics

Model	Count Best	Mean	SE	95% CI
Base	0	71.73	7.41	[57.20, 86.26]
Perseverance	0	73.96	7.34	[59.58, 88.35]
FP	17	59.17	6.92	[45.61, 72.73]

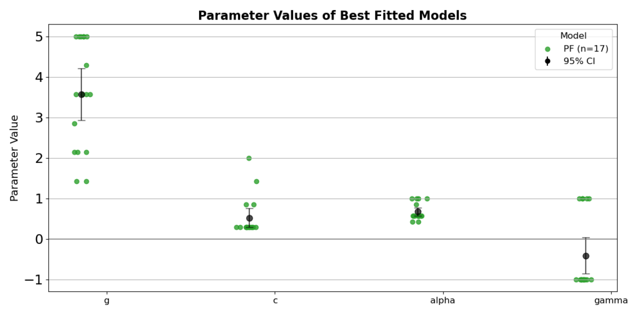


Figure 6: Distribution of best fitted parameters

Which models fit the data better and why?) (1 points)

Which models fit the data better and why? (1 point)

For all 17 participants, the FP model had the lowest BIC (μ = 59.17). One possible explanation, in line with our hypotheses, is that accurately capturing human behavior requires incorporating integral mechanisms such as switch anticipation in relation to repetition sequence length, which this model successfully does (see Figure 2. Furthermore, the feedback mechanism appears to help aid the model by offering more fine-grained control on the reliance on the anticipation/perseverance strategy based on immediate feedback. The FP model has overboard a significantly better BIC, with no overlap between its own and the other models' 95% confidence intervals. This hints at the explanatory value of the two additional parameters of the γ and α.

Parameter fit

Describe the results of fitting the winning model to the data.

Which parameter values fit the data best? (1 point)

Across all participants fitted to the (winning) FP model, the value ranges of alpha and c were the most stable. $c > 0$ with a mean of 0.52 (STD = 1.34), indicating a clear tendency to become more determined over time, and α with a mean of 0.67 (SD = 0.21) indicating considerable integration of the feedback into the attention and decision process. The parameters g and especially γ showed more notable variations across participants. g with a mean of 3.57 (SD = 1.34) indicating most participants had relatively strong current-task attention. γ with a mean of -0.41 (SD = 0.94) had the highest variation with high values in both directions $g < 0$ and $g > 0$, although collectively these results still indicate a switch anticipations tendency overall, it could also be interpreted as individual strategy-bias among participants, with some participants showing a strong preference to adhere to the repeated task rules. See Figure 6.

Discussion

Which hypothesis does your modeling support and why?. Base your answer on the model comparison (and model recovery) results. (1 point)

The P model performed worse than the Base model on several occasions, particularly when evaluating based on BIC. In contrast, the FP model provided the best fit to the data, outperforming both alternatives in terms of BIC across all conditions. Since the P model, like the FP model, is in principle capable of accounting for decreasing switch costs in higher repetition sequences it stands to reason, that the mechanism of P alone might be too "all or nothing" to fully capture the dynamic nature of human behavior in task-switching paradigms. While we did not provide a particular analysis of all the mechanisms impacted by the Feedback system, its results indicate that an anticipation strategy based upon trial history performance, rather than just trial history, more correctly emulate human behavior. Overall, it stands to reason, that the ability to modulate task-attention through higher level concepts such as feedback and stochastically driven task-anticipation plays an important role in understanding human attention in changing environments.

In addition, Though this may hint at overfitting, the FP model's ability to attune to complex dynamics allows it to generalize well to data generated by the other two models (5). Further testing with alternative model configurations that enhance parameter recoverability could disentangle the underlying particulars of the effect mechanisms.

Which other insights does your model provide? Base your answer on the parameters fits of the winning model. (1 point)

We suspect a strong individual difference in terms of appropriating task anticipation strategy. This can be seen from the fitted values of γ across participants, as most values tended to approach either one of the extremes, i.e. high task perseverance or switch anticipation (See Figure 6). To Better study these effects, we suggest an experimental design that allows for different distributions of the frequency of switch trials, this would provide an additional valuable variable for studying the interactions between expectations and task attention.

What are potential weaknesses of your modeling study? (0.5 points)

One key weakness is the high interdependence between parameters—particularly the strong correlation between γ and g in the FP model—making it difficult to isolate the contributions of individual mechanisms. This also led to poor parameter recoverability for γ and α, limiting interpretability. Additionally, the FP model, while best-fitting, may risk overfitting due to its increased expressivity. Finally, the model does not account for reaction times, which are a valuable source of insight into attentional and cognitive control processes, but remain challenging to integrate within our modeling framework.

What might be another computational modeling approach for gaining a deeper understanding of the phenomenon? (0.5 points)

Making the whole analysis more complex, by allowing α to go to both - negative and positive values - it may be able to simulate a wider range of behavior, such that also some inertia effects can be assessed. For a different approach, given the same experimental paradigm, and the goal to study repetition effects on switch cost, drift-diffusion models could provide access to reaction times, which is otherwise hard to incorporate meaningfully in our and other modeling approaches. This could possibly allow for a better understanding of the relations of the proposed two sub-networks in terms of a slow and fast system with appropriate time constants.

Acknowledgments

List which group members have been responsible for which part of the group projects.

- Hendrik Strauss: implementation, literature-research, report-writing
- Esma Hodzic: implementation, report-writing
- Sophie Terhardt: implementation/plotting, debugging, report-writing
- Silas Sinning: implementation/data-cleaning, literature research
- Misha Zimová: code-readability, report-writing
- Yehor Safroniuk: documentation, report-editing

References

Deco, G., & Rolls, E. (2005). Attention, short-term memory, and action selection: a unifying theory. *Progress in Neurobiology*, *76*(4), 236–256. doi: 10.1016/j.pneurobio.2005.08.004

Duthoo, W., Abrahamse, E., Braem, S., Boehler, C., & Notebaert, W. (2012). When predictions take control: The effect of task predictions on task switching performance. *Frontiers in Psychology*, *3*, 282. doi: 10.3389/fpsyg.2012.00282