Modeling Confidence using Drift Diffusion Models

Anish H S **Suraj Sahu** **Ramon Zacharias**

Introduction

What is the phenomenon you want to model? (0.5 points)

We are investigating how different modes of feedback influence the confidence in decision making. We want to check whether a person does better when given feedback on a task and also what the influence of a gamified cumulative score is on their decision behaviour. To do so we set up three experiments with different types of feedback. We assume that the drift rate parameter in drift diffusion models is correlated with the confidence of a person.

Why is that phenomenon relevant for understanding human cognition? (0.5 points)

It is relevant to understanding the decision processes in humans and how they can be influenced through the type of feedback that a person receives. Findings in this area could help shape tasks in a way that is more likely to make people succeed or influence their decision behaviour in a more controlled fashion.

We are checking what influence the different experimental conditions have on the confidence of participants. We assume that the drift is correlated with the confidence. We hypothesize that the second condition has the highest drift rate because of the positive reinforcement provided by the feedback. In the third condition we believe participants will make more random choices causing the drift rate to decrease, due to the negative score attached to not correctly as well as not answering in time.

Methods

Why is this modeling method appropriate for understanding the phenomenon? (1 point)

The Drift Diffusion Model is commonly used to describe the process of making a choice between two alternatives (Myers CE et. al, 2022). It effectively captures the fundamental aspects of the decision-making process within a simple yet robust framework.

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

We hypothesize that the second condition will the highest drift rate. We think that this will be the case because through the feedback on whether they were correct or wrong on a trial participants can correct their behaviour and have more confidence in the decisions they have made.

For the first experiment our hypothesis is that the drift rate is more random and should probably lie around 0.5 when collecting data from multiple participants.

In the third experiment we believe participants will be more cautious in order not not get a very negative score and so that the drift rate will be lower than in the other experiments. However, here it could also happen that participants make more random choices due to not answering giving them the most negative points.

Description of computational model(s)

For each experiment we are using two drift diffusion models with different prior parameter distributions, in order to represent our null hypothesis and alternate hypothesis.

The Drift Diffusion Model (DDM) is a cognitive model that describes how decisions are made between two choices through a process of evidence accumulation. It assumes that an individual accumulates evidence for each choice over time until a certain threshold is reached, leading to a decision. It simplifies the complex dynamics of decision-making into a quantifiable process, making it well suited for studying tasks that involve binary choices.

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

- **Inputs of the DDM:**
  - **Drift Rate** (v): Indicates the rate of evidence accumulation, representing the strength or quality of evidence towards a decision.
  - **Threshold** (a): The predefined level of evidence required to make a decision, balancing the trade-off between decision speed and accuracy.
  - **Starting Point** (z): Reflects initial bias or predisposition towards one decision over another at the beginning of evidence accumulation.

- **Non-Decision Time** (t_0): Accounts for time taken by processes other than decision-making itself, such as sensory processing or motor responses.

- **Outputs of the DDM:**
  - **Response Time**: Total time from the start of the trial to the decision point, combining actual decision-making time and non-decision time.
  - **Choice**: The outcome of the decision process, determined by which threshold is reached first.

- **System Properties:**
  - Incorporates **stochasticity** to represent variability.
  - Through **sequential sampling**, models the temporal aspect of accumulating evidence over time.
  - Sensitive to parameter changes, indicating how variations in drift rate, threshold, starting point, and non-decision time can influence decision outcomes and timings.
  - Can simulate **biases** in decision-making by adjusting the starting point and thresholds.

Which assumptions does each model make? (1 point)

The two drift diffusion models created for our three different experiments assume the following:

- **Null Hypothesis**: The null hypothesis assumes the initial hypothesis and the corresponding parameters to be true and fitting the experimental data. The null hypothesis assumed for the experiments are as follows:
  - **Experiment 1**: The assumption for the prior distribution for the drift rate parameter limited by a bound of (0.0, 1.0) is 0.5 with a SD of 0.2 due to the nature of the experiment without any feedback thus, giving the participants no reason to be cautious or accurate as well as giving them no sense of win/loss.
  - **Experiment 2**: The assumption for the prior distribution for the drift rate parameter limited by a bound of (0.0, 1.0) is 0.7 with a SD of 0.2 due to the nature of the experiment providing true value feedback thus, giving the participants positive reinforcement (depends on the difficulty of the experiment).
  - **Experiment 3**: The assumption for the prior distribution for the drift rate parameter limited by a bound of (0.0, 1.0) is 0.3 with a SD of 0.2 which, in this case as it provides both true value feedback and a cumulative score makes the participants more cautious and leads to lowering of confidence attributing it to lower drift rate.

- **Alternate Hypothesis**: The alternate hypothesis assumes that the data fits contrasting parameter values which don't support the initial assumptions and hypothesis. The alternate hypothesis assumed for the experiments are as follows:
  - **Experiment 1**: The assumption for the prior normal distribution for the drift rate parameter limited by a bound of (0.0, 1.0) is 0.1 with a SD of 0.1 assuming the participants to cautiously do the experiment and make quite a lot of mistakes due to the absence of feedback.
  - **Experiment 2**: The assumption for the prior normal distribution for the drift rate parameter limited by a bound of (0.0, 1.0) is 0.1 with a SD of 0.1 assuming the participants provided with the true value feedback become a lot more cautious due to making mistakes(The true value feedback work as negative reinforcement in this case).
  - **Experiment 3**: The assumption for the prior normal distribution for the drift rate parameter limited by a bound of (0.0, 1.0) is 0.7 with a SD of 0.2 assuming that the scoring and correct answers helps in improving their confidence thus, reducing reaction time. The alternate hypothesis can also fit better in the possible cases of random answering due to the high -ve scores or reduction in concentration due to the experiment length.

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

All models followed the basic definition of the drift diffusion model: $X(t) = z + v \cdot (t - t0) + \sigma \cdot W(t)$

For each experiment two models were defined. One to test the null hypothesis and one to test the alternate hypothesis. The models differed only in their choice of the prior distribution of the drift rate parameter v.

For eg: The first experiment will have models for null and alternate hypotheses which have contrasting drift rate values.

Drift rate normal distribution for null hypothesis:
$\mu = 0.5 \quad \sigma = 0.2$

Drift rate normal distribution for alternate hyppothesis:
$\mu = 0.1 \quad \sigma = 0.1$

Similarly, as defined in the last section, we take these contrasting values as the parameters to compare the models. The drift rate parameter is also modeled based on a combination of categorical factors including task transition, task type, congruence and trial number in the current experiment. This can be provided in the formula section while defining the model using HSSM.

Description of the experiment

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

We have three experiments, each one a variation of the task switching paradigm. Participants were shown moving triangles with different orientations. In each trial participants were first shown a fixation cross in blue or yellow. If it was blue they had to react to the movement of the triangles by either pressing the 'f' key, if they thought the triangles were moving to the left or the 'j' key if they thought they were moving to the right. If the fixation cross was yellow they had to react to the triangles orientation with ’f’ representing triangles oriented to the left and ’j’ representing triangles oriented to the right. From one trial to the next the task could either be switched from e.g. movement to orientation discrimination or could remain the same.

The difference between the 3 experiment lied in the type of feedback that participants received. In the first experiment participants did not get any feedback. In the second experiment they received correct/incorrect feedback and in the third experiment participants were shown a score after each trial. They get +2/-2 points for correct/incorrect answers and -4 points if they did not answer on time. We take this mathematically equivalent format of reward-punishment tradeoff from the elo algorithm seen to be used for various games like chess, football, etc (Szczecinski et. al, 2019).

The independent variables in our experiments were the type of feedback that the participant recieved, the task type(movement, orientation), the condition (congruent/incongruent) and the task transition (if the task was switched from the previous task). The dependent variables were response values and response time.

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

We collected data from three participants each of which completed all three experiments. Each experiment consisted of 48 trials. Consequently 144 trials were recorded per participant and 432 trials were recorded across all participants and experiments.

Model simulation

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

We used the hssm library to simulate data of 150 trials for all three experiments.

For the first experiment the parameters were set to $v_1 = 0.5$, $a_1 = 1.0$, $z_1 = 0.5$, and $t_{0_1} = 0.1$.

The defined parameters for the second experiment simulation are given by $v_2 = 0.7$, $a_2 = 1.0$, $z_2 = 0.5$, and $t_{0_2} = 0.1$.

The defined parameters for the third experiment simulation are given by $v_3 = 0.3$, $a_3 = 1.0$, $z_3 = 0.5$, and $t_{0_3} = 0.1$.

Model fitting

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

First models for all three experiments were initialized. They were given bound from 0 to 1 for all parameters. Afterwards the models were fit using the Markov Chain Monte Carlo (MCMC) method (van Ravenzwaaij et. al, 2018), more specifically using the No-U-Turn Sampler

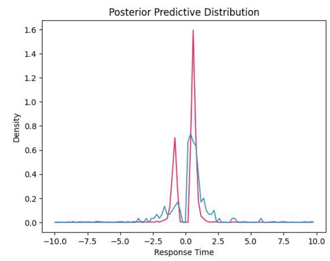


Figure 1: Posterior predictive distribution

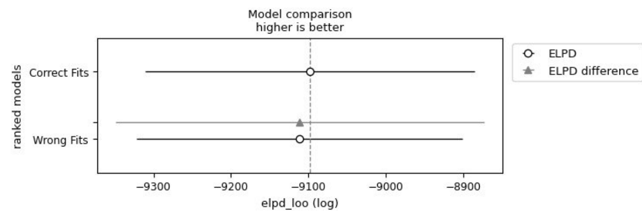


Figure 2: Drift rate recovery using bayesian inference

(NUTS). We ran 40 independent chains, with each chain drawing 200 samples from the posterior distribution. Before the actual sampling started, a tuning phase consisting of 500 steps per chain was conducted.

Parameter recovery

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

Figure 1 shows the predictive posterior distribution, which compares the response time for the true value parameters and the parameters used for the simulated model and thus provides insights on how well the model captures this using simulated data. Figure 2 shows the comparison of elpd scores of the two simulated models. Here the probability distribution of drift rate (v) is compared using these 2 models, where higher elpd suggests better fit. The recovered v value is in the normal distribution range of the assumed v value essentially capturing the recovery of the drift rate. The root mean squared error for the simulated data of the second experiment was 0.62 which shows how well the model can predict the drift rate using a linear regression while considering the effects of the lapse probability. The values for the other simulated experiments can be found in the python notebook.

Model comparison (& recovery)

Describe how you compared the models. (1 point)

We compared, for each of the experiments, two models respectively containing the values from null and alternate hypothesis as mentioned earlier.

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

Model recovery was performed by comparing the modeling of the simulated data vs experimental data by comparing the elpd loo and p loo values. The value ranges and their differences helped us in understanding how well the model could be recovered and implemented with simulated data.

Results

Disclaimer

The results observed are not optimal as after having exhausted our colab resources, we were not able to generate the same results/plots as we noticed earlier. Having tested with better parameter values, we can state the methods are reasonable. However, for the sake of completeness we still document the current results.

Simulation results

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

The change in response time and the response values which ultimately provide the drift rate value ranges can be observed given the different conditions. Based on these, assumptions about confidence/nervousness, speed-accuracy trade-offs can be made.

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

The model cannot capture personal attributes like mood, state of the participant. As these things cannot be checked or generalized through the model, supporting questionnaires can be provided prior to the experiment.

Parameter recovery

All model parameters were recovered reasonably well. For example for the model simulating data for the first experiment the actual parameters were set to $v_1 = 0.5, \quad a_1 = 1.0, \quad z_1 = 0.5, \quad and \quad t0_1 = 0.1$, while the recovered parameters rounded to two decimal points were $v_1 = 0.63, \quad a_1 = 0.97, \quad z_1 = 0.44, \quad and \quad t0_1 = 0.1$. The deviations from the true parameters were approximately the same for all three models.

Which parameters can be recovered more reliably, which less reliably? (1 point)

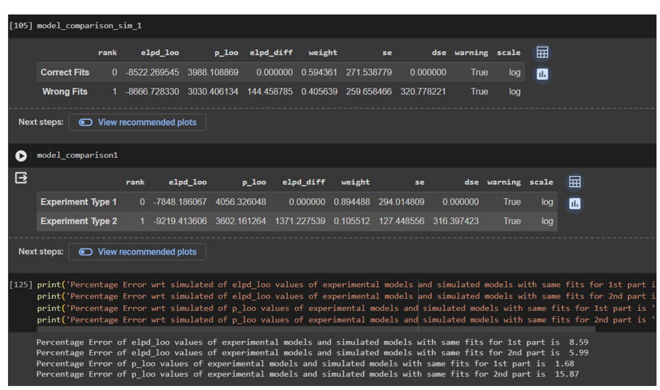
The drift parameter was recovered with less accuracy but still isn't far from the true value. When looking into the other parameters such as a, z and t, the recovery is pretty reliable for z while it is not the case for the other two parameters. These values could potentially provide valuable insights based on how much they change and the conditions.

Optional: Model recovery

Model recovery in this context would refers to the process of evaluating how well a model can recover the parameters used to generate synthetic data. The parameters obtained from model comparison elpd loo, p loo, elpd diff, weight, se, and dse play a crucial role in this process by providing quantitative measures of model fit, complexity, and uncertainty. Here's how they help in model recovery:

elpd loo (Expected Log Pointwise Predictive Density, Leave-One-Out): This metric assesses the average predictive accuracy of the model by evaluating its ability to predict individual data points. When performing model recovery, comparing and checking the closeness between the elpd loo values between the original data and the simulated data generated from the model indicates that the model is capable of accurately predicting the observed behavior.

p loo (Effective Number of Parameters): The p loo parameter quantifies the effective complexity of the model. This helps in model recovery by comparing the p loo values between the original data and the simulated data and helps in assessing whether the model accurately captures the complexity of the underlying decision-making process. If the p loo value for the simulated data closely matches that of the original data, it suggests that the model is able to capture the appropriate level of complexity. Figures 3 to 5 show the model recovery results.



```
[105] model_comparison_sim_1
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Correct Fits	0	-8522.269545	3988.108869	0.000000	0.594361	271.538779	0.000000	True	log
Wrong Fits	1	-8666.728330	3030.406134	144.458785	0.405639	259.658466	320.778221	True	log

Next steps: View recommended plots

```
model_comparison1
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Experiment Type 1	0	-7848.186067	4056.326048	0.000000	0.894488	294.014809	0.000000	True	log
Experiment Type 2	1	-9219.413606	3602.161264	1371.227539	0.105512	127.448556	316.397423	True	log

Next steps: View recommended plots

```
[125] print('Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 1st part i
      print('Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 2nd part i
      print('Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 1st part is '
      print('Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 2nd part is '

Percentage Error of elpd_loo values of experimental models and simulated models with same fits for 1st part is  8.59
Percentage Error of elpd_loo values of experimental models and simulated models with same fits for 2nd part is  5.99
Percentage Error of p_loo values of experimental models and simulated models with same fits for 1st part is  1.68
Percentage Error of p_loo values of experimental models and simulated models with same fits for 2nd part is  15.87
```

Figure 3: Model recovery results for exp. 1

```
[107] model_comparison_sim_2
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Correct Fits	0	-9098.564711	4147.465722	0.000000	0.58272	212.395479	0.000000	True	log
Wrong Fits	1	-9111.393680	3739.947940	12.828969	0.41728	210.302328	237.707163	True	log

Next steps: View recommended plots

```
[108] model_comparison2
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Experiment Type 1	0	-7968.590240	4756.796993	0.000000	0.83119	284.447637	0.000000	True	log
Experiment Type 2	1	-8760.070725	3861.901986	791.480485	0.16881	206.537062	280.387692	True	log

Next steps: View recommended plots

```
[126] print('Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 1st part is ',pe
      print('Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 2nd part is ',pe
      print('Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 1st part is ',perce
      print('Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 2nd part is ',perce

Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 1st part is  14.18
Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 2nd part is  4.01
Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 1st part is  12.81
Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 2nd part is  3.16
```

Figure 4: Model recovery results for exp. 2

```
model_comparison_sim_3
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Correct Fits	0	-8792.995925	3280.858762	0.000000	1.000000e+00	247.044691	0.000000	True	log
Wrong Fits	1	-9874.104172	3709.278880	1081.108246	1.300293e-12	4.146920	246.858462	True	log

Next steps: View recommended plots

```
[110] model_comparison3
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Experiment Type 1	0	-7103.409757	2474.202153	0.000000	0.703343	337.138959	0.000000	True	log
Experiment Type 2	1	-8398.823209	2710.352279	1295.413452	0.296657	247.361274	412.677313	True	log

Next steps: View recommended plots

```
[127] print('Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 1st part is ',per
      print('Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 2nd part is ',per
      print('Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 1st part is ',percen
      print('Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 2nd part is ',percen

Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 1st part is  23.79
Percentage Error wrt simulated of elpd_loo values of experimental models and simulated models with same fits for 2nd part is  17.57
Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 1st part is  32.6
Percentage Error wrt simulated of p_loo values of experimental models and simulated models with same fits for 2nd part is  36.86
```

Figure 5: Model recovery results for exp. 3

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

Basing upon the elpd and p loo values, the 1st and 2nd experiment models can be recovered more reliably with the current limited runs done.

Model comparison

For the first experiment our null hypothesis model had the lower rank, indicating that it has better out-of-sample predictive performance based on the $elpdloo$ values (-7848.17 vs. -9219.41). However, both models have relatively large standard errors (se) and warnings are true, indicating potential issues with the LOO estimates' reliability. Also the standard error was considerably higher for the favored null hypothesis model.

For the second experiment, our analysis revealed that the model corresponding to the null hypothesis outperformed the one for the alternate hypothesis in terms of out-of-sample predictive accuracy, as reflected by their respective $elpdloo$ values (-7968.6 vs. -8760.07). Both models are accompanied by considerable standard errors (se) and warnings indicating true, which suggests that the reliability of the Leave-One-Out Cross-Validation (LOO) estimates might be compromised.

For the third experiment again the null-hypothesis model had the lowest rank with a elpd value of -7103.41 versus -8398.8 for the alternate hypothesis model. The null-hypothesis model had lower complexity as indicated by its lower *ploo* value. However, it has a larger standard error and again warnings are set to true.

Figures 6 to 8 show the model comparison based on elpd scores for all three experiments.

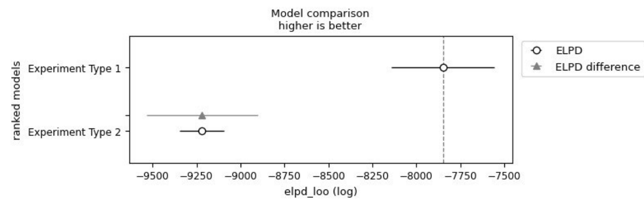


Figure 6: Elpd comparison for exp. 1

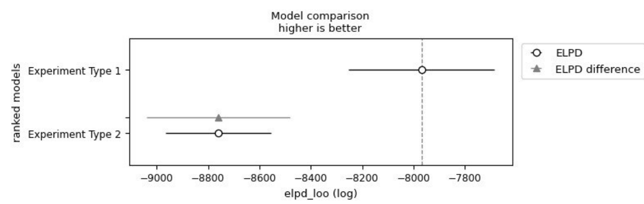


Figure 7: Elpd comparison for exp. 2

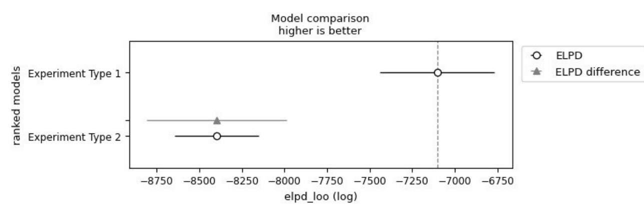


Figure 8: Elpd comparison for exp. 3

Figures 9 to 11 show more extensive model comparison results.

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Experiment Type 1	0	-7848.186067	4056.326048	0.000000	0.894488	294.014809	0.000000	True	log
Experiment Type 2	1	-9219.413606	3602.161264	1371.227539	0.105512	127.448556	316.397423	True	log

Figure 9: Model comparison for exp. 1

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Experiment Type 1	0	-7968.590240	4756.796993	0.000000	0.83119	284.447637	0.000000	True	log
Experiment Type 2	1	-8760.070725	3861.901986	791.480485	0.16881	206.537062	280.387692	True	log

Figure 10: Model comparison for exp. 2

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
Experiment Type 1	0	-7103.409757	2474.202153	0.000000	0.703343	337.138959	0.000000	True	log
Experiment Type 2	1	-8398.823209	2710.352279	1295.413452	0.296657	247.361274	412.677313	True	log

Figure 11: Model comparison for exp. 3

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

As displayed in the above paragraphs, the models built on null hypothesis fit better. That can be used a reference both for model and parameter fit.

Parameter fit

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

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

The parameters used for null hypothesis (our initial hypothesis) fit better when compared to the alternate hypothesis consisting of contrasting parameter value. This can be seen in the Figures 9, 10 and 11 which show the results for all the three experiments wherein the Experiment 1 is the null hypothesis and Experiment 2 is the alternate hypothesis.

Discussion

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

Due to the way model works, evidence collected over time (drift rate) should change based on how fast and accurately participants respond. Both of these are influenced with addition of feedback and score tracking. Hence, the hypothesis suggesting the drift rate to be influenced with the change in conditions is observed.

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

Given the model with null hypothesis fit better, we can say that the initial assumptions about the parameters reflecting confidence can be a potential insight.

This insight is supported by the experimental design to a certain extent whose design was in order bring out the 'cautiousness/confidence' of the participant with the limited time window provided to react. Although this might be true to some extent, the limitations stated above doesn't make this conclusion as robust as we would like it to be.

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

A potential weakness could be lack of data and the computational resources required to run multiple runs and parameter testing in order to optimize the results.

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

Bayesian modeling would be another possible approach for obtaining the probability distributions by defining conditional probabilities basis predefined set of rules. Another possibility might be neural networks wherein a lot of factors if, numerically defined can help in modeling the same.

Acknowledgements

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

- Anish: Wrote most of the code and ideated the theoretical basis for the experiments while providing the required inputs and analyses for the article.
- Suraj: Helped with coding, documentation, providing references and organising.
- Ramon: Helped with the interpretation of results, created the initial structure and wrote most of the article.

References

- Szczecinski, Leszek Djebbi, Aymen. (2019). Understanding and Pushing the Limits of the Elo Rating Algorithm.
- van Ravenzwaaij, D., Cassey, P. Brown, S.D. A simple introduction to Markov Chain Monte–Carlo sampling. Psychon Bull Rev 25, 143–154 (2018). https://doi.org/10.3758/s13423-016-1015-8
- Myers CE, Interian A, Moustafa AA. A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences. Front Psychol. 2022 Dec 9;13:1039172. doi: 10.3389/fpsyg.2022.1039172. PMID: 36571016; PMCID: PMC9784241.