Computational Cognitive Neuroscience Coursework

Reinforcement Learning for Computational Psychiatry. March 2025

Lecturer: Peggy Seriès
Teaching Assistant: Lars Werne

Practicalities

Deadline is 24 March 2025 at Noon (standard late policies apply). Please submit your report as a **single PDF** on **Learn**. Marking will be done through **Gradescope**; please match the pages of your report to the questions on Gradescope (2 marks will be deducted if you forget).

- Please keep your submission anonymous. The submission should be no longer than **6 pages**. References, appendices and code do not count toward this limit.
- You **do** have to **include code in your report**. Please attach it to the end of your PDF as an appendix.
- Plots should always include axis labels and units. Figures should always have a caption and be referenced in the main text. The presentation and format will count for 10% of the final mark. Note that using default settings for plots is probably not the best idea (e.g. the line-widths are often too thin).
- Show your work. If there are more than one (sensible) ways to calculate something be sure to describe how you did it. Be concise and precise in how you report your results. Don't include lots and lots of separate graphs: you can superimpose different graphs in the same plot.
- Copying results is not allowed. It is okay to ask for help from your friends. However, this help must not extend to copying code or written text that your friend has written, or that you and your friend have written together. You are assessed on the basis of what you are able to do by yourself. Similarly, we ask you NOT to use Generative AI (e.g., ChatGPT) to write the code and descriptions that you will submit. Please adhere to the University guidelines on the limited, acceptable uses of GenAI in academia. Failure to do so, where we are able to detect it, will be penalized and reported to the University's Academic Misconduct Officers.
- Your programming style will not be assessed but your code should be readable/decipherable in a reasonable amount of time. If your code is incorrect, you may still be able to get points by realising that the results are not correct and describing how you think they should look like.
- It is recommended that you try to complete the tasks in the order they are given. Please clearly indicate which question you are answering in your report.
- If something is not clear please create a new post on **Piazza** so that the answer can be shared with all the other students. If you are unsure whether your question can be shared with all the students, you can create a private Piazza question which, in case, can be made public by the course staff.

Marks for Style of Presentation (10 marks)

You should present your findings in the style of a scientific report. By this we mean that the document should include text broken into appropriate sections (e.g., by question number and/or topic), equations (if relevant), figures (with axes clearly labeled, font size similar to the font in the main text, with figure captions to explain their content, and with numbering/lettering of figures and subpanels in order to refer to them from the main text), and references (if relevant). You should consider how to best communicate your findings in a clear and concise way to the reader. This includes designing figures appropriately, for example by making use of subplots and plotting multiple things on the same graph to save space or for aid of visual comparison, setting the axis limits appropriately, using figure legends and colour schemes to delineate different things plotted on the same graph, etc. There is no need to aim for a 6 page report if you can convey your findings more succinctly, and marks may be deducted if you provide excessively long-winded, incoherent, or unstructured answers. To take these factors into account, 10 marks have been allocated for the quality of presentation with regards to the overall layout and style, the quality of figures, and the scientific clarity and precision of written answers.

Assignment (90 marks)

Hyperarousal – which entails feeling 'on edge' and reacting disproportionally to surprising stimuli – is a key symptom of posttraumatic stress disorder (PTSD). In this assignment we are interested in understanding whether participants with PTSD differentially adapt their behaviour following surprising (harmful) events, in the context of a reinforcement learning task. We will focus on modelling behavioural impairments during a gain-loss learning (two-arm bandit) task, inspired by a study conducted in 2018 by Vanessa M Brown et al. (1). The participants of the study were US military veterans that had been exposed to combat trauma, satisfying criterion A1 of PTSD according to the DSM-IV (2). At the start of the experiment, structured interviews were held, to identify a subset of participants meeting the criteria for a PTSD diagnosis. In our (hypothetical) version of the experiment, participants completed a PTSD self-report measure (PCL-5 (3)) as part of this procedure. The PCL-5 questionnaire consists of 20 items, corresponding to 20 PTSD symptoms, which are answered using scores from 0 ('not at all') to 4 ('extremely'). A total symptom severity score (range $0-80$) can be obtained by summing all individual scores. Summing the scores within certain groups of items yields *cluster severity scores* (e.g., items $1-5$ relate to memory intrusions).

During the task, participants were asked to choose between two actions on each trial, both of which had a chance to lead to a 'good' or to a 'bad' outcome. The experiment consisted of blocks of 'gain' and 'loss' trials. On a 'gain' trial, good/bad outcomes respectively corresponded to a large/small gain of (virtual) money ($+£1.00$ / $+£0.20$) – and to a small/large loss of money on a 'loss' trial ($-£0.20$ / $-£1.00$). Subjects started the experiment with a balance of $£10.00$. If a subject ever reached a negative balance, the experiment continued as normal.

Introduction

You are given, on the CCN opencourse page, three csv files containing simulated data for 50 participants. The experiment consists of 200 trials per participant. Blocks of 'gain' and 'loss' trials alternated as follows: The first 25 trials are 'gain' trials, the next 25 trials are 'loss' trials, and so on. In each block, one of the subjects' two available actions ('A' or 'B') had a 75%, the other a 25% chance of yielding the 'good' outcome, as defined above. Action 'A' was the 'better' action in blocks $2, 5, 7$ and 8, action 'B' was linked with the higher success rate in blocks $1, 3, 4$ and 6. Participants were aware that there was a good and a bad action to choose within each block of trials, but did not know which was which and were not aware of the assigned probabilities or their changes – i.e., they had to infer them from the observed outcomes of their actions, with the aim of maximizing their monetary gain. Brown and colleagues modelled the participants' behaviour using a selection of reinforcement learning models similar to the ones you have seen in the CCN lectures. In this coursework, you will implement versions of these models, starting with the one defined in Box 1 below:

Box 1: Temporal Difference Reinforcement Learning (Q-learning)

Following standard Q-learning (Sutton and Barto, 1987 (4)), the expected value (Q) of the *chosen* action on the next trial ($t+1$) is updated as follows:

$$Q_C(t+1) = Q_C(t) + \alpha \cdot \delta(t) \tag{1}$$

where $C \in \{A, B\}$ denotes the *chosen* option. The value of the *unchosen* action remains unchanged. The trial-by-trial prediction error $\delta(t)$ is computed as the difference between the actual outcome (R) and the expected value (Q_C):

$$\delta(t) = R(t) - Q_C(t) \tag{2}$$

The probability of choosing action A on the *next* trial is modeled with a softmax function incorporating inverse temperature ($\beta > 0$):

$$P_A(t+1) = \frac{e^{\beta Q_A(t+1)}}{e^{\beta Q_A(t+1)} + e^{\beta Q_B(t+1)}} \tag{3}$$

Importantly, participants are aware that action values may change at the start of every block of 25 trials. Thus, *before* the first trial in a gain (or loss) block, the expected values of both actions are initialized as follows:

$$Q_A = Q_B = 0.6 \quad \text{(gain block)}, \quad Q_A = Q_B = -0.6 \quad \text{(loss block)}.$$

The hyperparameters α and β are assumed to remain *constant* across trials.

Question 1. Exploring the data (8 marks)

Load and explore the provided data. For each participant, compute their total PCL-5 symptom severity score, as well as *cluster severity scores* for the following blocks of questionnaire items:

- Item $1-5$: Intrusive thoughts or memories
- Item $6-7$: Avoidance
- Item $8-14$: Negative thought patterns and mood
- Item $15-20$: Altered physical and emotional reactions

High PCL-5 scores are intended to support a PTSD diagnosis. In the structured interviews, the first 25 participants were found to have PTSD, while the last 25 were trauma-exposed, but healthy, controls. We will assume that these diagnoses are correct.

Task 1a): Compute the mean, standard deviation and median of the total score, as well as each of the four cluster scores. Your answers should be rounded to 3 significant figures. Do this for all participants, then for the 25 PTSD subjects and for the 25 controls separately. Would you say that the PCL-5 scores align with the given diagnoses?

Task 1b): It is suggested that a PCL-5 score of at least 32 is indicative of probable PTSD (3). If this rule had been used as the only diagnostic criterion, how many healthy controls would have been diagnosed as having PTSD? How many subjects with PTSD would have received no diagnosis?

Task 1c): Turning to the experimental part of the study, compute the 'total money gained', 'total money lost' and 'net profit' achieved by each participant. Report the mean value of each of these quantities across individuals in the trauma/control groups. What would be the *expected value* of these quantities at the end of the experiment, for an agent choosing a random action (with $50-50$ probabilities) on each trial? Did our participants perform the task well?

Question 2. Simulations (7)

Since we are working with a generative model, we are able to simulate data, which can be extremely useful. Use equations (1) to (3) to write a function that lets you generate data from known parameters (the parameter values should be an input to the function). Generate the outcomes with probabilities which are changing every 25 trials as described in the introduction.

Hint: You can take inspiration from the code from the lectures to help you get started.

Simulate 200 choices with parameter settings $\alpha = 0.3, \beta = 8$ a number of times. (Choose a reasonable number so you can average the simulations). Illustrate the average evolution of values Q_A and Q_B. Illustrate the average evolution of the difference in Q values of the two stimuli (i.e. show how $Q_A - Q_B$ changes, on average, over the course of the simulated experiments). Very briefly explain what is observed and why the shape of this evolution makes sense.

Question 3. Exploring parameter settings (6)

Simulate 200 choices several times for a number of different parameter settings. Systematically vary settings of α and β to explore how different values affect the average net profit at the end of the simulation. Plot the average net profit as a function of the parameter settings for α and β, which

means there will be three dimensions. Choose sensible ranges for the parameters (e.g. $0 < \alpha < 1$ and $0 < \beta < 15$).

Briefly describe your precise approach and comment on how the expected performance during the experiment is related to different settings of the parameters.

Question 4. Likelihood function (6)

To find the parameter values that best capture each participant's behavior, we need to define the likelihood of the parameters. Write a function that takes as input the data (choices and outcomes) for an individual and a vector of parameters (learning rate and inverse temperature). The function should return the negative log likelihood (NLL) of these parameters, where θ is the parameter vector containing α and β.

$$NLL = -\sum_{c \in \text{Choices}} \log p(c|Q, \theta).$$

Hint: You can use the code from the lectures to help you get started.

Report the NLL for the first and the 10th participant using parameter setting $\alpha = 0.3, \beta = 8$.

Question 5. Model fitting (8)

Task 5a): Find the parameters that minimize the NLL for each individual: Pass your NLL function and a set of starting parameters (use $\alpha = 0.3, \beta = 8$) to a optimization function performing unconstrained minimization.

Hint: For Python, you could have a look at `scipy.optimize` and particularly the Nelder-Mead algorithm.

Task 5b): Calculate and report mean and variance of the fitted parameter values for learning rate and inverse temperature. Illustrate the results: Plot the participant index on the x-axis, parameter values on the y-axis. Can any difference between the PTSD and control groups be seen from the plot? Comment briefly on what you observe. Do all the values make sense? If not, what can you do about it?

Task 5c): Calculate and report the Pearson's correlation coefficient between estimated parameters (i.e. between α and β across all participants). Consider the first 25 participants as your PTSD group. Also, calculate and report the Pearson's correlation coefficient between estimated parameters separately for participants within each group.

Question 6. Parameter recovery (8)

We now want to check the reliability and identifiability of our parameter estimates.

Task 6a): Sample 50 sets of parameter values of learning rate and inverse temperature from a multivariate normal distribution. Choose sensible numbers for the mean of this distribution and describe how you chose them; choose small numbers for the variance; set the covariance to zero. Illustrate the sampled values and highlight and exclude (resample) nonsensical values.

Task 6b): Use the sampled parameter values to simulate 50 sets of data (as in Question 2). Fit new parameter values to these simulated data sets (as in Question 5). Calculate, report and

illustrate the Pearson correlation between the parameter values you used to simulate the data and the parameter values that you obtained from fitting the model to the simulated data. Repeat this process 5 times and report the Pearson correlation each time (there is no need for you to plot the sampled values all 5 times, the correlation coefficients are enough).

Task 6c): Comment briefly. Does this parameter recovery simulation meet your expectations. Explore and describe how the number of trials and the number of simulated data sets affect the performance of the parameter recovery.

Question 7. Parameter splitting (9)

Reward and punishment have been argued to have differential influences on behaviour, in experiments similar to the one considered here.

Task 7a): Hence, fit another model to the data of each participant, as in Question 5, but assume that the model uses a different set of (α, β) parameters on gain- than on loss trials. Each model will thus have four trainable parameters.

Task 7b): For each participant you should now have two negative log likelihood values: One for a model with parameters combined across loss and gain trials (Question 5), and one for a model separating all parameters by condition (Question 7). Compare these negative log likelihood values between models. Explain what you observe. Does it make sense?

Task 7c): For each participant, compute Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) scores for each model. For what percentage of participants does the BIC score improve when parameters are separated by condition? And the AIC score? Briefly discuss what this indicates.

Box 2: Model selection criteria

For your reference, these are the formulas used to compute AIC and BIC:

$$\text{AIC} = 2k - 2\ln(L)$$

$$\text{BIC} = k\ln(n) - 2\ln(L)$$

where k is the number of parameters in the model, L is the maximized value of the model's likelihood function (i.e., $\ln(L)$ is the negative of the minimized NLL), and n is the number of samples used to train the model. Generally, lower AIC/BIC scores indicate a better fit.

Going forward, we will consider the model with split parameters, as the processing of monetary gains and losses have commonly been argued to be affected differentially in PTSD (e.g. (5)).

Question 8. Group comparison (4)

Use a two sample t-test (describe which exact test you are using) to test whether the estimated parameter values (Question 7) are significantly different across groups (consider the first 25 participants as your PTSD group). Report the t statistic, degrees of freedom and p value if applicable. Explain briefly how we can interpret the results and how this relates to our hypotheses.

Question 9. Alternative model (9)

Now, we include another model which introduces an *associability* parameter. *Associability* aims to measure the amount of attention paid to a particular stimulus or to the outcomes of a particular action. The associability of a stimulus is argued to increase when observing it was recently linked with *surprise*. The model is defined in Box 3 below.

Box 3: Associability

For the associability RL model, the learning rate (α) is modulated on a trial-by-trial basis by an associability value (κ) for the chosen stimulus.

An associability weight parameter (η; range 0 to 1) controls the extent to which the magnitude of previous prediction errors updates the trial-by-trial associability value for each stimulus. Associability values are initialized at 1 and updated separately for each stimulus.

The associability value of the chosen stimulus ($\kappa_C; C \in \{A, B\}$) on the next trial ($t+1$) is updated as follows:

$$\kappa_C(t+1) = (1-\eta) \cdot \kappa_C(t) + \eta \cdot |\delta(t)| \tag{4}$$

where $\delta(t)$ is the trial-by-trial prediction error, as defined in the standard RL model.

The expected value of the chosen action (Q_C) on the next trial ($t+1$) is updated using the associability-modulated learning rate:

$$Q_C(t+1) = Q_C(t) + \alpha \cdot \kappa_A(t) \cdot \delta(t) \tag{5}$$

As for the previous model, associability and expected values are reset at the start of each block of 25 trials.

Implement simulation and negative log likelihood functions for this new model. Similar to Question 7, assume that gain and loss trials use *separate* parameters α, β and η. Fit the model to the data, using $\alpha_{\text{gain}} = \alpha_{\text{loss}} = 0.3, \eta_{\text{gain}} = \eta_{\text{loss}} = 0.6, \beta_{\text{gain}} = \beta_{\text{loss}} = 8$ as initial values.

As in Question 5, illustrate the fitted parameter values.

Question 10. Model comparison (7)

We would now like to compare the two models fitted using separate parameters for gain and loss trials (Questions 7 and 9).

Task 10a): For each participant, and for both models, compute AIC and BIC scores. Sum up the participant scores for each model (i.e. for each model you will have a single BIC and a single AIC score). Report the results. Comment briefly. Which model would you choose as the 'best' model?

Task 10b): For the parameters of the new model, compute group comparisons, as in Question 8. Report t statistic, degrees of freedom and p value if applicable. Do you find statistically significant differences for a parameter between the PTSD and control groups? How could these be interpreted? What might we conclude if the data was real?

Question 11. Model recovery and confusion matrix (10)

We now want to check the reliability of our model comparison procedure using model recovery simulations. This will give us some indication about how much we can trust our results from Question 10. For the models from questions 7 and 9, simulate data (using the functions you already implemented) multiple times. For each of these simulated data sets, fit each model and use model comparison (i.e., AIC and BIC scores) to choose the best model. Display your results in two confusion matrices – one based on AIC, one on BIC. Comment briefly. (Make sure to explain your approach and results in appropriate detail.)

Question 12. Discussion (8)

Summarize and discuss the findings of your report. Did your model fitting procedures demonstrate notable differences between the PTSD and control groups? What implications might the differences you found, if any, have on trauma treatments (e.g., exposure therapy)? Point out any limitations of your experiments that may affect the interpretability of your results. Write a conclusion for your report, as you would expect to find in a scientific paper.

References

[1] Brown VM, Zhu L, Wang JM, Frueh BC, King-Casas B, Chiu PH. Associability-modulated loss learning is increased in posttraumatic stress disorder. Elife. 2018;7:e30150.

[2] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.

[3] Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5); 2013. Accessed: 2025-02-08. https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp.

[4] Sutton RS, Barto AG. A temporal-difference model of classical conditioning. In: Proceedings of the ninth annual conference of the cognitive science society. Seattle, WA; 1987. p. 355-78.

[5] Boukezzi S, Baunez C, Rousseau PF, Warrot D, Silva C, Guyon V, et al. Posttraumatic stress disorder is associated with altered reward mechanisms during the anticipation and the outcome of monetary incentive cues. NeuroImage: Clinical. 2020;25:102073.