# Fitting models to behaviour

**Quentin J.M. Huys**[1,2]

**1** Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zürich and Swiss Federal Institute of Technology (ETH) Zürich, Switzerland **2** Department of Psychiatry, Psychotherapy and Psychosomatics, Hospital of Psychiatry, University of Zürich, Switzerland

**Questions:** qhuys@cantab.net

---

## 1 Background

This project will first guide you through fitting a simple reinforcement-learning model to data using maximum likelihood and then Bayesian methods. In a second part, you are invited to recreate this with your favourite task.

We will use a simple four-armed bandit task by Guitart-Masip et al. (2012) to examine how different types of learning concurrently shape inference, learning and decision-making.

Research going back to the beginning of the last century has distinguished Pavlovian conditioning from instrumental conditioning. This was initially defined in terms of experimental paradigms. In Pavlovian conditioning, a stimulus (say a buzzer) precedes and thereby predicts a reward or punishment *irrespective* of the animal's behaviour. In instrumental conditioning, the reward or punishment obtained additionally depends on what the animal does. However, it has become clear over time that they also rely on partially distinct neural circuitry, that they exist side by side, and that this co-existence and co-expression can explain a number of anomalies of learning and inference that are relevant particularly to mental health (Dayan et al., 2006; Huys et al., 2015, 2016).

In the Guitart-Masip et al. (2012) task, individuals are faced with one of four stimuli, and have to choose whether to press a button or not. The outcomes depend on both the stimulus and the choice:

- stimulus 1 (go to win): reward (with 80% probability) if go, no reward if nogo.
- stimulus 2 (nogo to win): reward (with 80% probability) if nogo, no reward if go.
- stimulus 3 (nogo to avoid loss): loss (with 80% probability) if go, no loss if nogo.
- stimulus 4 (go to avoid loss): loss (with 80% probability) if nogo, no loss if go.

Subjects do this for some 200 or so trials. Over time, they learn that some stimuli can give rewards and others losses; and that some behaviours are better for some stimuli than others. If you haven't read the paper, you might want to think about what you predict. Which stimuli will they be best at, which worst?

## 2 Fitting models

The first part of the project consists of building a generative model. This is a model that can be run, as it were, on the precise task, and that generates data as if you had run the task on a subject.

1. Assume that subjects are presented with a random sequence of the four stimuli above.
2. Assume that they learn two types of values:

$$\mathcal{V}_t(s_t) = \mathcal{V}_{t-1}(s_t) + \alpha(\beta(r_t) - \mathcal{V}_{t-1}(s_t)) \quad (1)$$

$$\mathcal{Q}_t(s_t, a_t) = \mathcal{Q}_{t-1}(s_t, a_t) + \alpha(\beta(r_t) - \mathcal{Q}_{t-1}(s_t, a_t)) \quad (2)$$

and update these values with the observed rewards ($\beta(1) \geq 0$ for reward and $\beta(2) \leq 0$ for loss, zero otherwise) on each trial $t$. Here, $a_t$ is the action on trial $t$, $s_t$ the stimulus on trial $t$. The value $\mathcal{V}$ is

the Pavlovian expected value that depends only on the stimulus presented, while the value $\mathcal{Q}$ is the instrumental value that depends on both the choice and the stimulus.

3. Next, sample choices on each trial $t$ according to:

$$p(a_t|s_t) = \frac{e^{w(a_t;s_t)}}{\sum_a e^{w(a;s_t)}} \tag{3}$$

$$w(a_t;s_t) = \begin{cases} \mathcal{Q}(a_t,s_t)+\epsilon\mathcal{V}(s_t) & \text{if } a_t \text{ is go} \\ \mathcal{Q}(a_t,s_t) & \text{else} \end{cases} \tag{4}$$

i.e. add the Pavlovian value to the go tendency. This means that positive stimuli will elicit active behaviour, and negatively value stimuli inhibit active behaviour.

4. Explore the behaviour of this model for different parameter settings $\boldsymbol{\theta} = \{\alpha, \epsilon, \boldsymbol{\beta}\}$. You can also explore the impact of changing the starting values $\mathcal{V}_0, \mathcal{Q}_0$.
5. Write a function that computes the likelihood of all choices given some parameters

$$\mathcal{L}(\boldsymbol{\theta}; A, S) = \log \prod_t p(a_t|s_t;\boldsymbol{\theta}) \tag{5}$$

6. Write a function that computes the gradients of the likelihood wrt. all the parameters. Note that the total log likelihood is the sum of the log of the individual choices, i.e. you can compute them one by one:

$$\frac{\mathcal{L}(\boldsymbol{\theta}; A, S)}{d\theta_k} = \frac{d}{d\theta_k} \log \prod_t p(a_t|s_t;\boldsymbol{\theta}) \tag{6}$$

$$= \sum_t \frac{d}{d\theta_k} \log p(a_t|s_t;\boldsymbol{\theta}) \tag{7}$$

The gradients wrt. the learning parameters can be recursively updated, for instance for the learning rate:

$$\frac{d\mathcal{V}_t(s)}{d\beta(r)} = \frac{d\mathcal{V}_{t-1}(s)}{d\beta(r)} + \alpha\left(1 - \frac{d\mathcal{V}_{t-1}(s)}{d\beta(r)}\right) \tag{8}$$

To constrain the parameters, you can reparametrise as follows:

$$\epsilon = \frac{1}{1+e^{-\epsilon'}} \qquad \text{for } 0 \le \epsilon \le 1 \tag{9}$$

$$\beta = e^{\beta'} \qquad \text{for } \beta \ge 0 \tag{10}$$

$$\tag{11}$$

and take gradients wrt. $\epsilon'$, $\beta'$ etc.

7. Check that your gradients are correct by comparing them to finite derivatives.
8. Use a gradient optimizer (in matlab you can use fminunc.m, for instance) to fit data generated in part 3 and see how well you can recover the parameters.
9. Compute the Hessian around the estimated parameters. This gives you an indication of how well-constrained your parameters are by the data. What do you observe in particular about $\beta$ and $\epsilon$? Can you derive analytically why this is so?
10. Generate parameters for a series of subjects by drawing them from normal distributions, generate data for each subject and infer the parameters for all of them. What do you observe? Is the prior variance well estimated?
11. The function that generates choices (and, paired with it, the corresponding likelihood function) is your 'hypothesis'. It incorporates what you think subjects are doing in the task. You can compare this hypothesis to other hypotheses by running different models. In this step, I suggest you modify the model in point 1 in some way and repeat all of the above.

12. Perform a simple model comparison by computing the BIC scores for each of the two models on data generated from the model, and on data generated from the other model:

$$BIC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}; A, S) + k\log(T) \tag{12}$$

where $\hat{\boldsymbol{\theta}}$ are your estimated parameters for data $A, S$, $k$ is the number of parameters and $T$ your number of choices/trials.

13. Bonus: Apply Expectation-maximisation to infer the parameters of the prior distribution (the prior mean and variance). Alternatively, ask me for some code to do this.

## 3 YOUR OWN TASK

In this part, you are invited to perform the above for your own task. Effectively, what you need to do is

1. Write a function that generates choices from the key variables in your task. You should then run it and simulate from it to see if it behaves as you expect it to.
2. Use this function to compute the likelihood of generated (or real) data.
3. Use either ML or your fancy EM fitting method to infer the parameters that generated the data.
4. Compare your hypothesis to a null hypothesis using model comparison.

## REFERENCES

Dayan, P., Niv, Y., Seymour, B., and Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw*, 19(8):1153–1160.

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., and Dolan, R. J. (2012). Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage*, 62(1):154–166.

Huys, Q. J. M., Gölzer, M., Friedel, E., Heinz, A., Cools, R., Dayan, P., and Dolan, R. J. (2016). The specificity of pavlovian regulation is associated with recovery from depression. *Psychol Med*, 46(5):1027–1035.

Huys, Q. J. M., Guitart-Masip, M., Dolan, R. J., and Dayan, P. (2015). Decision-theoretic psychiatry. *Clinical Psychological Science*, 3(3):400–421.