Neurobiological and computational mechanisms of complex planning and risk in navigation

Slide 1 – Introduction

My name is
Project proposed for my DPhil about "Planning and risk during navigation"
Already started in Chris' lab
I will talk about the work already done

Slide 2 – Background

Decision-making involves learning about the value of our actions.

Reinforcement learning (RL – cornerstone of dm) explains how we update the value of actions on the basis of outcome prediction error.

Neurobiology has provided evidence from fMRI and single-cell recording studies for RL implemented in the brain.

[Point to the picture]

Slide 3

Model-free RL starts from the idea that we receive a direct reward for every action we make.

In the example, for pacman getting the power-pill, he may not be reinforced during the first 3 steps before getting the power-pill.

When several steps are required to harvest a reward, decisions can be improved by planning.

Slide 4

Model-based RL descibres how we can plan based on a model of the environment.

Without planning, in this example, we just learn the value of the last action before reaching the goal.

Planning allows us to update more action values by explorating our model of the world. This is done by mental simulation of our actions.

And, thus, we learn faster.

Slide 5 – Models

Up to here, nothing is new.

But humans are limited-capacity agents.

And so, we need to deal with an accuracy/complexity trade-off in our strategies, the same as happen in artificial intelligence. We do this thanks to heuristic algorithms.

We already know that for an optimal solution to complex navigation we need to take into account uncertainty for planning: for example, about finding shortcuts in an unknown village, the other day just by exploring I discovered that I could go from Iffley to Cowley without going through the Magdalen Bridge! Without taking uncertainty into account I wouldn't explore and increase the accuracy of my representations.

Planning has a high computational complexity cost.

We need heuristics to solve this.

Slide 6 – Research Questions

We can formulate some first questions:

MSDP = multi-step decision problems

1) Are humans optimal when making plans to solve MSDP?
→ This is interesting because 40 years of economical behavioural research says 'No'.

2) How do humans deal with exploration/exploitation trade-off in MSDP?
→ Show new insights about heuristics in navigation for Artificial Intelligence

3) What are the neural substrates underlying those strategies?
→ There are just a few publications about model-based fMRI.
→ This will allow us to validate between different models.

Slide 7 – Experiment 1

Subjects have to plan journeys between stations in an unfamiliar subway network.
The start position is here.
And... the goal destination is here.
They are asked plan by minimizing the travel time.
Each line has a different speed, and for a number of journeys the map doesn't change, but the start and end positions do. The speed of all trains in a same line has a fixed mean with some variability.
While planning journeys in the same map, participants should learn the speed of the lines, probably exploring at the early trials.

Slide 8

Let's say that a participant has already found that the red line is quite fast and the green one is deadly slow.
In this example it may be better to take the red line than the green one.
So the quickest route is not always the shortest.

Slide 9 – Pilot data

We have runned a pilot of the experiment with 10 subjects.
For each map, they had 10 journeys.

Slide 10 – Comparison

This is a bit hard to explain.
This is a comparison between human data and data from 3 computational models.
The red model is the 'omniscient' one. We call it like that because it knows from the first trial the real value of the speed of each line. So it is optimal without needing to explore.
The plot shows average time spent in a journey across journeys (so trials). In fact it's the relative time, proportional to the time spent by this omniscient model.
The cyan model doesn't take into account the speed (so it chooses the sortest way).
The green one is an ideal observer.

The blue one plots the human results.

We see that humans learn the speed of the lines because they have a better performance than the distance model.

It's interesting because they seem to explore (having a worse performance than the distance model) at the first trials.

We can ask about how humans improve their total time per journey: by decreasing the distance, or increasing the average speed of the lines they choose.

Slide 11 – Stop

As we can see, humans increase the total distance of their journeys.

Slide 12 – Speed

In fact, what it seems to happen, is that humans tend to increase the speed of the lines they choose. Even more than what the other models do.

Slide 13 – Future plans and conclusions

This experiment is just a proof of concept. It seems that
Second aim → second experiment (vary variance)
Third aim → after starting my PhD