Reinforcement Learning (RL)
1) RL is a frameworkunder which an agent interacts with the environment and learns the value of actions from the reward prediction error.
2) It has been shown that dopaminergic signals in the brain (rats & primates) correlate with this reward prediction prediction error. For instance, Schultz showed using SUA recordings that firing rates increase when a reward is unexpected, and decrease when an expected reward is not presented.
3) RL works well in simple problems. However, it doesn't scale to more complex problems.

Hierarchical Reinforcement Learning (HRL)
1) Problems where we need to think many steps ahead are more difficult. This is because the number of possible sequences of actions increases exponentially with the number of steps.
2) RL can be improved by including multiple levels of representation. In this example, we could represent every room as a different context. Learning how to go from one room to another is efficient because it constraints the number of possible solutions.
3) A key feature of HRL is the transition between two contexts (the hallways in the example). We call these subgoals.

Questions
1) Read question 1.
2) Read question 2.
3) We hypothesise that the dACC will be more activated during context switching because it seems to be implicated with action-outcome association and/or particularly with cognitive and response conflict.

Experiment
1) To address these questions, we have designed an fMRI task (*with humans?*) where participants navigate in a subway network. In this, stations are identified by distinctive names; lines are identified by colours.
2) The hypothesis is that two levels of representation will be used while navigating: stations and lines. The intersection between lines (exchange stations) will be treated as subgoals.
3) Participants are trained under this map before undergoing fMRI scanning, such that they become familiar with the map.

Task design
1) During the experiment, participants complete multiple journeys. they first see a cue indicating departure and goal of the journey. They can navigate freely by performing one of four actions (north/south/east/west) and they receive a reward if they reach the goal.
2) There's a fixed probability that the journey is cancelled at any single step. Participants must minimise the length of the journey to maximise reward.
3) Critically, note that there's no information about the current line during navigation. Participants must remember the identity of the line from the training.
4) These are examples of regular, exchange and elbow stations respectively.

Reaction times (RT)
1) we analyse reaction times in the first place

2) two bars on the left side is when we don’t change action/direction. two bars on the right side is when we do change.
regular stations have only 2 possible actions (same context). exchange stations have >2 actions with 2 different lines.
3) using a factorial rmANOVA, we find a double main effect:
change of action slows response
having more options also slows response.
no interaction

fMRI GLM (exchange)
1) we use a general linear model to analyse correlations between parameters of the task and BOLD signal. the first three regressors correspond to the main effects in the previous analysis (RT) and the interaction.
2) the BOLD signal in the dACC correlates with this regressor, implying that dACC is more activated in stations where you *can* change line.
3) there is no effect of changing action in the dACC. however, an interaction between the two regressors shows that dACC is specially activated when you *do* change line.

fMRI GLM (distance)
1) the other two regressors included in the analysis are the inverse values of distance to goal and distance to line change
2) BOLD signal in the vmPFC and the dmPFC correlate with distance to goal
3) BOLD signal in the dmPFC correlates with distance to subgoal

Conclusions
1) it has been shown from RT analysis that exchange points incur an additional cost
2) read point 2
3) read point 3
4) read future directions