20th November

Reinforcement learning

Classical conditioning

Pavlov's dog

- Learning to expect things to happen, not learning to act.
- Conditioned stimulus, e.g. bell, light
- Unconditioned stimulus (reinforcement, r), e.g. food, shock
- Conditioned response, e.g. anticipatory behaviour (salivation, freezing)
- Pair stimulus with significant event and measure anticipatory behaviour

Rescorla-Wagner rule (1972)

- Came up with up a rule to try and understand how a stimulus, S, leads to a reward.
- In the first phase, the stimulus always lead to reward- acquisition. Then test for the response.
- Extinction learning- in the second phase, having learnt the association, present stimulus without reward. There will be no response. This could be that the animal learnt a new response.
- Partial reinforcement- present stimulus before reward sometimes and without reward. There will be a weak response.

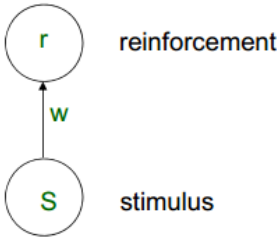
Experimental terms

	Phase 1:	Phase 2:	Test:
Acquisition:	S → r		S? response
Extinction:	S → r	S → -	S? –
Partial Reinf.:	S → r or -		S? weak resp

The model consists of a stimulus and a reinforcement neuron and a weighted connection between them. In the brain there will be lots of different neurons representing reinforcement and lots of neurons representing stimulus. This model is simple.

If the stimulus is present, the connection weight should change.

$w \rightarrow w + \varepsilon S\delta;\ S = r - wS$



If the reward is present the connection should increase in value. So the connection weight becomes its previous value plus a small positive constant multiplied by the size of the reward.

$w \rightarrow (1-\varepsilon)\, w + \varepsilon r$

If the reward is not present, the connection weight should decrease. So it becomes its previous value multiplied by a value that is less than one, so the connection weight decays.

We can think this as a simple 'delta rule' model.

Delta is the actual reward delivered minus the expectation of the reward, which is the net input strength, S x w. If stimulus is present, S=1. If stimulus not present, S=0.

$w \rightarrow w + \varepsilon S\delta;\ \delta = r - wS$

This could be viewed as error-driven learning.