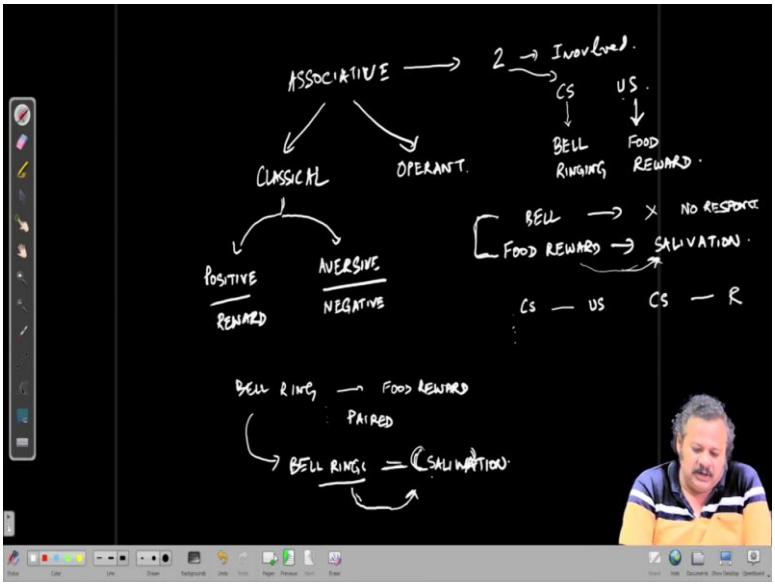
Cognition and its Computation
Prof. Rajlakshmi Guha
Prof. Sharba Bandyopadhyay
Biotechnology and Bioengineering
Indian Institute of Technology, Kharagpur

Lecture - 42
Examples

Welcome, so continuing our discussions on learning in biological networks; so, we have gone over the non-associative learning.

(Refer Slide Time: 00:42)



And now we take up the case of associative learning and here as we have mentioned we will be talking about the two broad kinds that is classical or Pavlovian and operant condition. So, when we think of the classical conditioning or the Pavlovian case of associative learning, we must be aware that there are there can be two sides to it. One that is, with positive effects or with reward and the other is aversive or with negative effects.

So, when we say associative, we of course, mean that there are two elements in the environment in our that we are interacting with that are involved. And at least two and there may be more of each kind where the two kinds we referred to as the condition stimulus in the case of classical conditioning and the unconditioned stimulus CS and US.

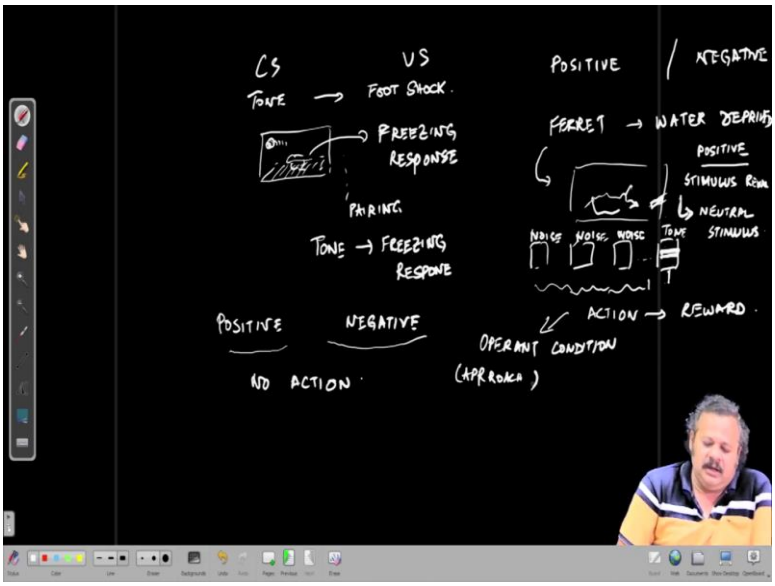
So, for example, if we go to the classical experiment of Pavlov, the condition stimulus is the bell ringing and the unconditioned stimulus is basically a food reward for the dog. So, what we know is that when the a dog does not respond to a bell; so, the experiment goes like this that when there is a bell ring this is the conditioned stimulus there is no response from the dog.

When the food reward is shown by itself, then there is salivation in the dog which is a response to the anticipation of the food that it will it is going to get; so, it is an anticipatory reaction. And so, when these two conditioned and unconditioned stimulus stimuli are paired that is the CS then followed by US. And this is repeated many times finally, when only CS is produced, we get we get the response as was observed with the unconditioned stimulus.

So, that is in the case of the dog with the bell ring every time the bell rings food reward is given. And as this pairing occurs finally, when this bell rings and there is no food reward rather; yeah, I mean when there is no food even without the food reward the response of salivation is seen that is.

Now, the bell ringing is providing the anticipation for the reward and is producing the salivation; so, this is reward based or positive aspect based classical conditioning. Similarly there are experiments with aversive or negative reinforcement and that usually has been done in animals and others with similar experiment where particular tone is associated with a foot shock.

(Refer Slide Time: 05:25)



So, a mouse is put in a cage with a floor that is basically wired in a way that small current can go through and the mouse is put in that cage. And whenever a particular tone there is a speaker and whenever a particular tone comes on, then after that immediately there is a small food shock.

And the animal learns this very fast that whenever there is a tone there will be a food shock arriving and so it freezes. So, there is a freezing response and what happens is that with this pairing going on; then finally, the tone itself in anticipation of the food shock now produces the freezing response.

So, this association of the bell with the food or the stone with the food shock this is the conditioning that is occurring and the response elicited by the unconditioned stimulus. So, in this case this is the CS and US, the response elicited by the unconditioned stimulus in. Now, in anticipation the with the condition stimulus itself we see the same response that is the condition response.

So, as we see that in both cases one that is a positive feedback and the negative feedback, in both cases the animal associates and learns the phenomena that is the conditioned stimulus is going to produce some effect. And in this case the animal does not have to perform any action animal or human, there is no action involved in on part of the organism, that is it does not have to make any choices this is happening with two different stimuli in the environment.

So, this is the big distinction between the conditioned reflex or rather conditioned learning, Pavlovian learning, and the operand learning is that. In the operand learning case the animal or the human has to make a choice of or make a perform an action that is related to the stimulus in the environment and produces an outcome. So, here like here the outcome is the reward of the food, or the freezing response which is the other outcome.

It is not the outcome of the action chosen by the animal, but it is based on just the environment and the particular stimuli. So, we will be talking about the mechanisms through which these happen, and even in the operand conditioning the basic principles are the same. So, if we take an example of operand conditioning in that case also we have the animal making a choice, and it can be either positive or negative the outcome is positive or negative.

So, and the animal has to learn to increase the positive outcomes of its action or reduce the negative outcomes of its actions and that is what the goal of the learning would be. So, for example, the example that we talked about earlier where ferret an animal that is water deprived, and usually this is how these experiments are done in order to understand the mechanisms behind the learning procedures or the structures that are involved.

That the this animal is now trained in a cage or in a box let say where the animal is head fixed or freely moving and let us say, there is a water spout here which delivers water. So, if we think of the positive reinforcement or the positive outcome-based learning we will have again a stimulus would be that is rewarding. And the other in the other case neutral stimulus, or mildly negative if required for the training purposes.

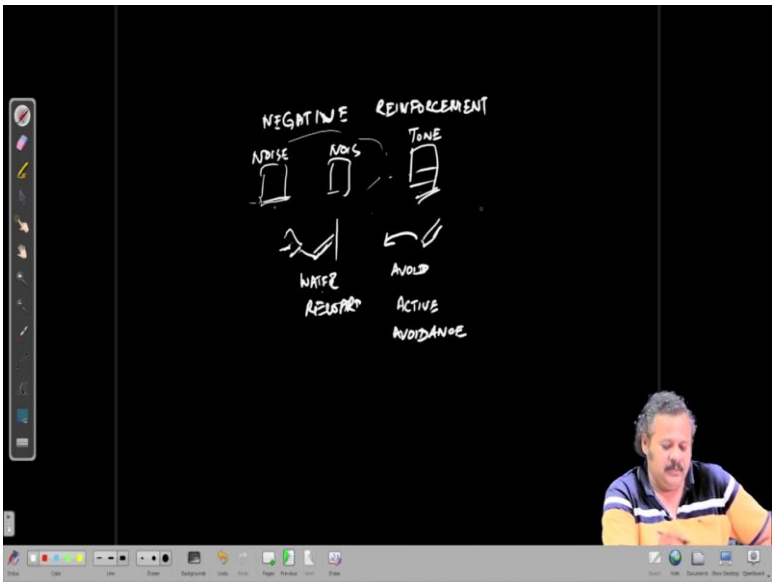
So, let us say for example, in that ferret case the water deprived animal it needs water and it takes there are sounds that are played that are noise like. And during the noise events or the noise sounds occurring the noise tokens occurring like that. And all of a sudden when the tone comes on the animal has to learn to detect the tone, and at that time only it can lick in the spout and get water.

And so, this is a positive reinforcement in the sense that the animal is learning to detect a tone emits noise sounds coming. And with its action of licking the spout choosing to lick the spout during the tone as opposed to during the noise it gets a reward.

So, the animal quickly not quickly it takes some time to learn that it cannot lick during the or it will not get a reward or water from the water spout during the noise presentations. But, during the tone it has to go and lick and get the water reward; so, it has to make a choice or its action is finally, producing the reward.

If it does not lick during the tone, it will not get the reward; so, this is the basis of it being the operand conditioning. And in this case since it is a positive reinforcement it is an approach behavior that is the animal has to go forward in the to the spout and lick and get the reward. Similarly, if the same animal can be trained in a negative reinforcement paradigm as was done by Fritz at all and the previous one was done by David at all.

(Refer Slide Time: 13:53)



And in that case the negative reinforcement is that, the during the learning procedure again we have the same thing that noise token and then noise again and so on and then all of a sudden there is a tone. So, in this case the during the noise the animal has to lick the water spout, during the noise the animal goes and has to lick the water spout and get the water reward.

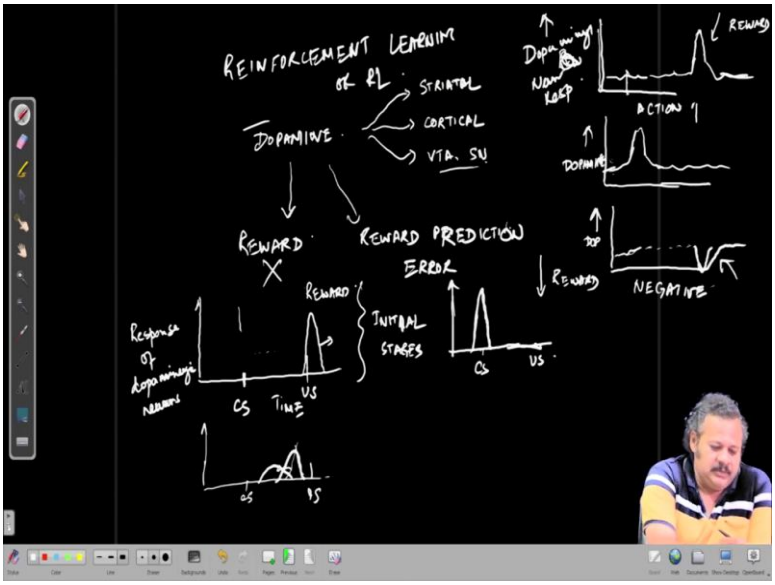
However, when the tone comes on it has to withdraw from the spout and avoid getting a tiny shock on its tongue. So, the animal in this case with the negative reinforcement the animal learns much faster than in the positive reinforcement case. And in this case this procedure would be of active avoidance active avoidance case of operand condition. So,

the animal learns to not lick during a particular sound, or particular class of sound, or particular category of sound, and not for the other category or class of sound.

So, in this case we again have associative learning, but now it is between the action that the animal has now it is mediated by the action that the animal has to do based on the difference stimuli that it is getting. And based on that an outcome is produced which can be rewarding, or based on their choices the an outcome that is a negative outcome or aversive outcome it has to it learns to reduce that kind of choice.

So, these are the two kinds of ways in which the operand conditioning works. And as we said in both the classical conditioning and the operand conditioning, we can actually think about them as reinforcement learning.

(Refer Slide Time: 16:25)



And this is or RL and so, here the main neurotransmitter system involved is dopamine. And when this and there are different structures through which this kind of learning can take place. So, in case of motor skills and so on it is through striatal circuits, and it can be through other cortical regions for other kind of tasks. And it involves either the dopaminergic neurons of VTA or the substantia nigra or the different regions in the midbrain the midbrain dopaminergic neurons.

So, although initially and even now people talk of dopamine as the reward producing neuron or is the reward neuron. The standing idea about how dopamine plays a role in

this kind of the two kinds of learning that we just talked about both classical conditioning, and reinforcement, and the operand conditioning is that dopamine actually acts as a reward, prediction, error, signal. So, it is not really a reward signal, it is what it is thought is it is a reward prediction error signal.

So, for example, what is thought is that if we have CS here in time and we are looking at response of dopaminergic neurons. Then when the US comes initially in the initial stages, the animal getting a reward here with the unconditioned stimulus is unexpected. And recordings from dopaminergic neurons show that there is a strong response or positive response from dopaminergic neurons which is signalling that unexpected reward that.

It is a prediction error that it did not predict a reward at that time initially and it is getting a reward at that time; so, it is a prediction error. And it is a positive error in the sense that it is actually getting a reward and; so, this is the firing of the dopaminergic neurons profile. As time progresses during this conditioning, the CS the dopaminergic firing reduces near the US and it gradually propagates back at different stages of the learning.

And finally, it starts responding to the CS they start responding to the CS, because this CS is not a reward, I mean the CS is not really an expected thing that will produce a reward. So, it is a reward prediction error and that is what is happening in this case that the CS now there is the dopaminergic response at the CS. And during the US when the animal has learned the association there is no response during the actual reward.

So, and gradually even in the CS it will go away once it is completely learnt there will be no responses whatsoever in the dopaminergic neurons, because it is totally predictable. So, in other words it is thought that learning occurs through surprising events and dopamine is the neurotransmitter and the dopaminergic neurons encodes this surprise or the prediction error in the reward.

The in the case of similarly in the case of operand conditioning if we see in this case that when we have a only a reward, when the this is the dopaminergic neuron response dopaminergic neuron response.

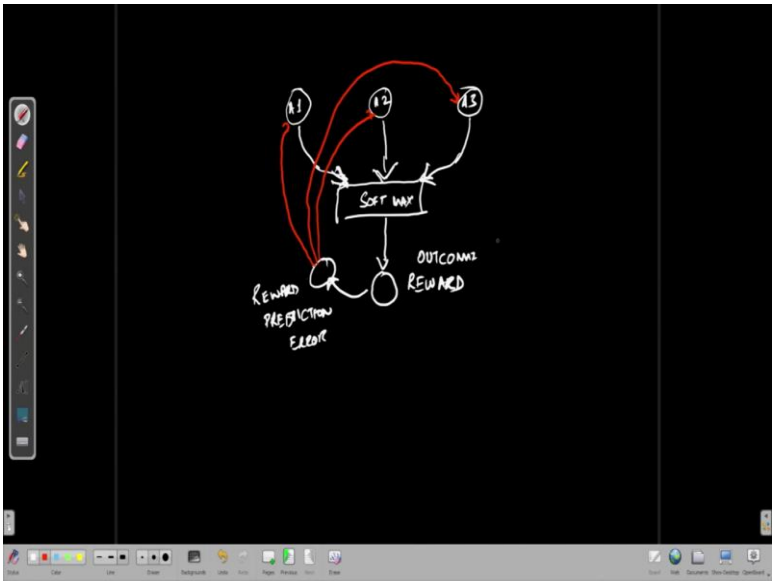
So, if you give an animal all of a sudden, a reward because of some action, it did based on a stimulus that was presented here that is a noise stimulus came it licked. And it lick

got a reward because of that action the there is firing of the neurons in the rewarding period during the reward, because this at this time the reward was unexpected.

Now, gradually when the action association with the reward is learnt, then actually the q that initiates the action that produces the dopaminergic neurons to fire. Similarly, now if so, this is we are let us say we need to put in a baseline firing here. Similarly, now once the action and the output reward is learnt; so, this is the dopaminergic neurons now. If a reward is now omitted after the q, then what happens is that there is a negative or there is a drop in the response in the dopaminergic neurons.

Because it is a negative reward prediction error unlike the positive that we saw earlier. So, in this sense although dopamine has been thought to be the reward associated element or neurotransmitter it is really over the last few decades. We know that it is really a reward prediction error and this is; and this is what actually is behind the reinforcement learning paradigm that has that we talk about.

(Refer Slide Time: 25:13)



And that is basically; so, if there are a number of actions let US say action 1, action 2 and action 3 that have to be chosen by the organism or animal at a particular time. Then given if there is uncertainty in the output produced, then what happens is there is a soft max that is applied on the value calculations of the outcome of each action. And then a reward is or some outcome is produced and that produces a reward prediction error.

And this is the dopaminergic signal that we talk about that goes and re-evaluates the value signals associated with each of these actions. And this goes on continuously in order for the learning to happen; so, that gradually this reward prediction error goes to 0. This is the principle actually applied in deep learning methodologies also. However, there are much bigger nuances that we are not talking about here in terms of reinforcement learning actually happening in the brain. And we will take up this in the open questions part in the final week of the course.

So, with this we conclude our lectures on learning and the in biological networks or learning in neural circuits. The way we have learnt is initially at the single neuron level at the synaptic level and then we have talked about those same things applied at the network or systems level. And later on in the next lectures we will talk about the different kinds of plasticity that is, here we have talked about synaptic plasticity which is more functional plasticity. And we will also talk about something that is structural plasticity more involved during development and also can occur during these kind of learning processes.

Thank you.