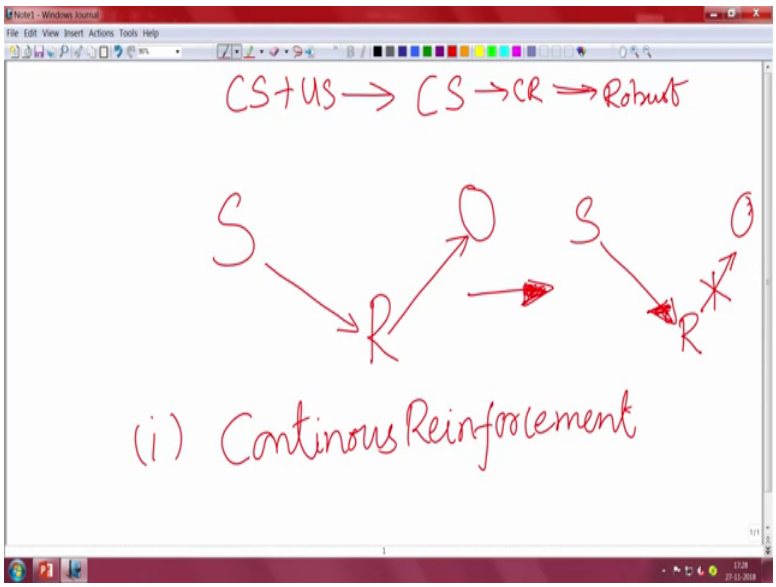
Learning about Learning A Course on Neurobiology of Learning and Memory
Prof. Balaji Jayaprakash
Centre for Neuroscience
Indian Institute of Science, Bangalore

Lecture - 13
Sign Tracking vs Goal Tracking, Linking complex behaviors to simple molecules

Hello and welcome to the lecture 13 of the learning about learning lecture series. we are in to studying different aspects of reinforcement learning. specifically we are into understanding the scheduling of the reinforcement or the outcomes in response to the expression of the stimulus induce response by the animals. There we said there are two major classes of presentation of these outcomes or the reinforcements one of them we said is a contiguous presentation or continuous reinforcement scheduling.

Versus the second one being the partial reinforcement scheduling. Why would you so apart from the fact that skinner ran out of the footpad pellets and then he is trying to be see conserving the food. why would you even get into the notion of partial partial reinforcement. It turns out when you when we look into the behaviour of this reinforcement learning.

(Refer Slide Time: 01:57)



Ah we started out by saying that you have a stimulus we call it as S. And you want the animal to actually develop a response in order to develop that what you are doing is that you are re reinforcing that response by presenting a favourable outcome for that animal a

likeable out outcome for that animal. When you repeatedly keep doing this all right the notion is eventually when you just present the stimuli the animal should be able to get this response.

But what we would like to ask is how good is this retention? How good is the retention of the stimulus eliciting a response right. Remember this is equivalent to we talking about a CS paired with a US acquiring an ability to elicit a response all by itself when presented later right, all this all goes in time right. So, this we know it is pretty robust you can go ahead and keep testing it right. You can keep on testing it it is you can test it after variant many number of times it is pretty robust.

Of course, if you test it very many times such that there is no us being presented mean the test does not involve a presentation of the us right. And we know that extinction kicks in same thing here. When you have the stimulus response outcome and then when you are testing the rob robustness of the stimulus eliciting response you are not necessarily presenting any outcome ok. Of course, you can do that, but you can decouple that.

So, when you do not present it and then train this animals right you say that we have decoupled the response to the outcome. This is an important phenomena for the animal to capture because we said; the school of learning in the living beings exist so that they can modify their behaviour to suitably respond to the stimuli right. If no longer you are coupling the response to the outcome indirectly you are saying the stimulus is no longer good predictor of an outcome. So, you should be able to modify your response there is no point keep on responding to a stimuli.

How well do they do it? It turns out if you do it in a if you train them in the stimulus response outcome unlike the CS US association the answer is little bit bit more complex in the CS US association is very simple you keep on doing that you will cause it is a robust. But it you will cause an extinction you can modify it. Here whether you will cause an extinction whether the animal will extinguish the stimulus response depends on how you actually train them. What do you mean by that; remember towards the end of the last lecture we were talking about scheduling the responses right.

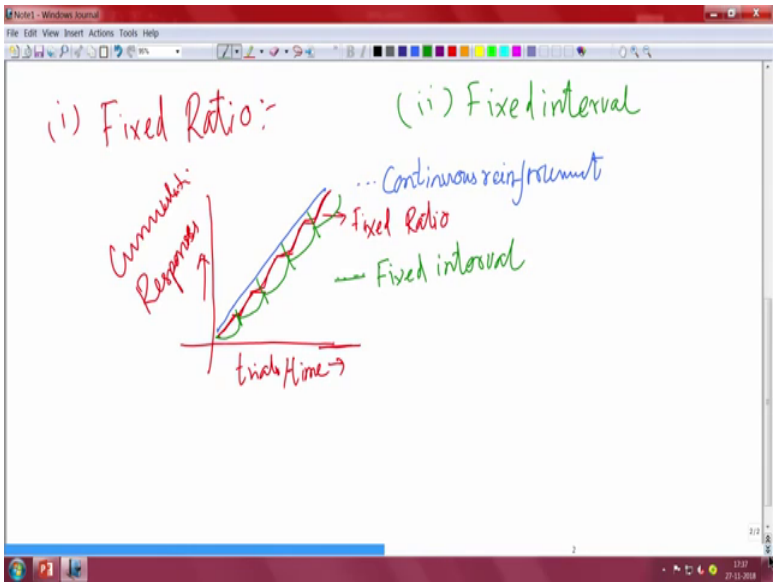
It turns out if you were to schedule such that the stimulus response outcome is acquired through contiguous manner right which is your reinforcement is continuous reinforcement. That is to say every single response of the animal is rewarded is

reinforced. Now you decouple them the extinction is pretty quick neat and clean. In fact, that is a problem the whole paraphernalia of the stimulate response outcome be study that you are doing is with the motivation.

That we would be able to develop some habits develop some responses we can bias these responses by presenting this reinforcement such that in future I can have this response for this stimuli without having to worry about this reinforcement that is the motive right. And if that motive is completely gone here because I cannot keep on giving this reinforcement all the time. At some point I need to get rid of this such that you are having a stable behaviour and it turns out that is not the case at all when you are doing continuous reinforcement they will extinguish pretty quickly.

However, when you do partial reinforcement does not matter which of the scheduling you are using. But if you are using a partial reinforcement that is good enough and the extinction in the partial reinforcement is extremely hard to combine. It is not that you cannot but it is just hard. And clearly that is one of the reasons why you really would like to study the way of press, how the way of presentation it is called scheduling affects the strength of the stimulus response relationships. Now what we are going to do is that, we are going to take a little bit more detailed look at how these responses develop in the various different. The four different scheduling paradigms that we talked about number 1 being the fixed ratio right.

(Refer Slide Time: 08:15)



So, what were going to do is that we are going to plot the responses of these animals as a function of the as a function of time or as a function of presentations you can think of right. So, in a contiguous presentation you would expect at the earlier part of the learning, we are not talking about the saturation where we have reached the v max and stuff. At the earlier part of the learning you would expect the responses here so in this axis and trials or time in this axis.

So, you would expect for the contiguous presentation it should be a very simple straight line and every instance for every response you are actually reinforcing that. So, I am not plotting that out, but you will see a pretty simple straight line right. That is contiguous continuous reinforcement; however, when you are doing this on a fixed ratio alright. What is going to happen is that the animal is going to understand that for until unless it is actually reaching a particular response number right you are actually looking at this is a cumulative response here that is when you are cumulative response.

So, unless until you reach that fixed number you are not going to get the food. So, what it is going to do is that, it is going to keep pressing it keep pressing or keep pressing the lever or keep responding just like the contiguous thing no no doubt about it. Right after it gets the reinforcement it understands for some time or some process, it is not going to get the food at all. So, you will see a little pass here it is not necessarily going to get have any responses there, but then it is soon quickly understands ok, I am not going to get the food unless I keep pressing it.

So, it keeps doing this these are straight lines with small passes here. So, that is what you would get for a fixed ratio. As against a fixed interval in time so these are straight lines flat parallel lines to the x axis. And what you see here is that they wait there is a little bit of lag then pick it up little bit of a lag pick it up. However, when you do this fixed interval and let us mark it through green dark green. What you see is that the animal clearly understands until that certain time is elapsed there is no point I am going to get the response.

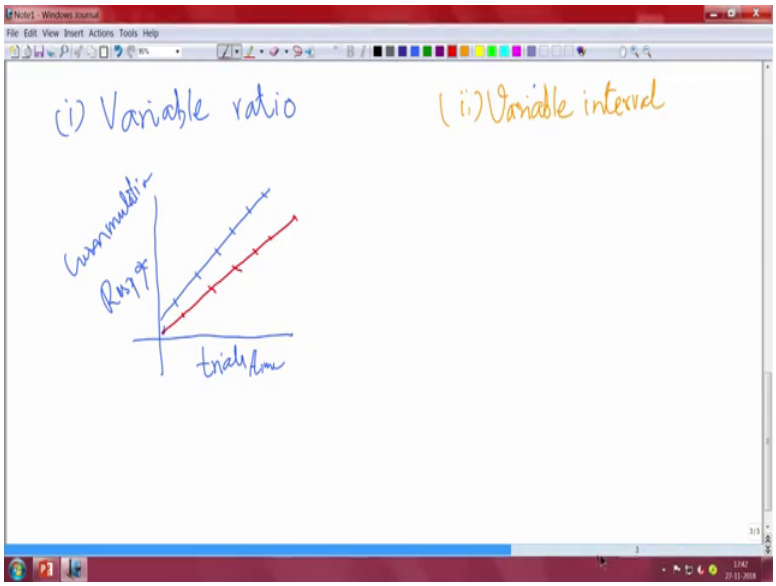
So, they alter the rate at which they are doing this lever process according to that. So, you here you can see it is a rate of lever process are same right. That is why the slope is same the slope represents the rate and that is why you have straight lines in bits and pieces in here that is not going to be a straight line. But in fact it is going to be a curve

that is doing that. So, what is happening here? The rate at which they are doing the lever process is slow lower and then it quickly picks up because they know as the time for the next response comes in I mean next reinforcement comes in they know that they are going to get reinforce.

So, they quickly pick up and they get the reinforcement right here all right. And then again a lull period fix it up right here and so on so forth. So, clearly if you look at the way that the response their behaviour is changing with respect to the scheduling. You can see there is a wide difference between the fixed ratio and fixed interval learning as against that continuous reinforcement to.

So, the green here would be fixed interval. So, now, what is happening here oh we need to talk about to one more thing to number 2 is, when two other things fixed interval all right. So, we talked about the fixed ratio fixed interval right. Two more things which are variable ratio and variable interval. Let us quickly take a look at that is going to look and then let us discuss what is happening here ok.

(Refer Slide Time: 14:23)



In this case I will go back to the blue, where I am going to say it is going to be a variable ratio variable interval. So, now, what do we expect here? In a variable ratio in this scenario since the responses are distributed about a mean ratio being number of responses elapsed responses before it can be reinforced. The animal has no way of

actually estimating that right they can estimate roughly that there is going to be a reward, but they do not know when they really cannot predict when the reward is going to come.

So, what they do is that, they start to respond continuously at a constant rate. So, what do we see is if you time or the trials. Again the same axis here and then a cumulative responses we are plotting out here just like the last graph. You would see them learn pretty much like a continuous thing if I did not tell you whether it is a variable ratio or a continuous different thing just by looking at the cumulative responses you will not tell them they are both different. However, the slope the rate at which they are that line is going up could be different all right.

But the interesting point being that, they are traveling in a straight line without any brakes like these are the places where you are actually having a the rewards are the reinforcement. How about the variable interval? The variable interval again is very non predictable right. You the animal understands that the reinforcement will be ready for them to take at a given interval of time, but they do not know when that interval is going to be. And they also know that if they do not respond they are not going to get the food.

So, they constantly keep responding though the rate at which they will be responding will be much much much lower because it is the interval and not the number of responses they do understand that. So, what you will see is that, a curve that sometimes parallel or that is still a straight line, but with a lesser slope indicating that they are responding lesser and this is these are the places where you are probably giving the reinforcement. Now let us put these two things together the two graphs together and let us look at what is what exactly is happening.

So, now clearly you are developing some kind of an expectation it is not the expectation that in the sense of what we talked about in earlier. But some kind of an expectation you can say in terms of when the reward is going to come and so on and so forth. And many of the experiments would go on and prove that such kind of expectation do have do happen. And there is also a counter argument to say that it is I mean the debate is not about whether there is an expectation or not. But it is about at what level do the animal understand that, we are the it is expecting or is just a reflexive on I mean it is not going through this conscious cognitive system.

But it just goes through it is a reflexive as a reflexive response either way what is what we do not have to debate about is this notion that we actually developed. In the very first and the very beginning which is during this stimulus response outcome you are the animal is developing this association good. So, now, when you are actually taking away this outcome it is in direct conflict with what it is known what it has learned. So, far when it is a continuous reinforcement right.

So, it knows for sure it should have been there it is not there. So, I should change my response well that signal goes down in magnitude in various amounts depending on which of the scheduling which of the partial reinforcement scheduling that you take clearly in a partial reinforcement scheduling, you are taking away that liberty right is just because you do not get an outcome does not mean you will never get an outcome. So, the animal does not know our animal cannot change it is response to zero right.

It is just that it is very unpredictable how unpredictable you can make you can make it I mean and still be contingent right. That is the key here if you if it is not contingent you are not going to make an association you have to be contingent. But still there is this notion of at some point I would get a reward right that is what is keeps driving there keeps driving them. And in fact, we use that for a good effect by reinforcing or keeping this stimulus response association longer mean even after you remove that coupling of the response to the outcome you can actually make them have that stimulus response.

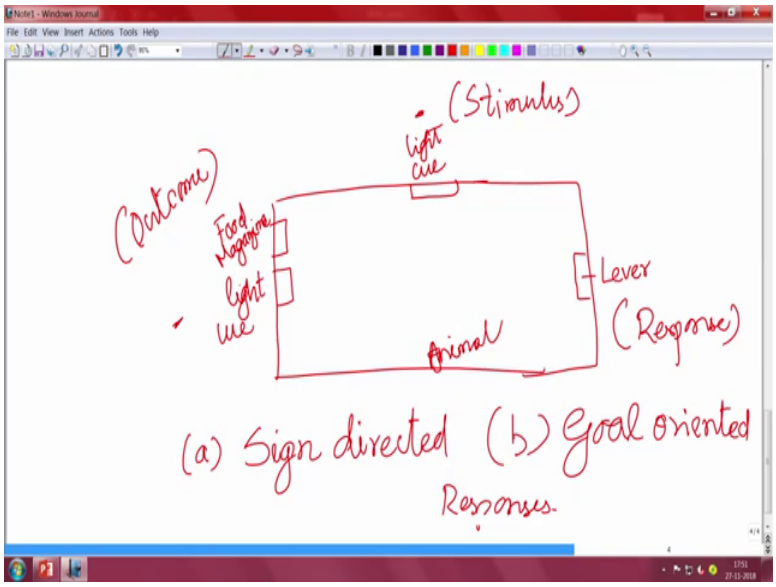
Um behaviour for a longer period of time just by changing the nature of presentation of the outcomes at which instance you are presenting them good. So, we have learned couple of quite a few things about the reinforcement process or the operant conditioning process. Now, let us revisit our original idea of what is it the animal is actually learning is it learning. And do all the animals learn in a same manner right revisit the idea of is it reflexive or cognitive and is it a complete phenomena or a universal phenomena, that is common through outs all the animals clearly the answer is no right it is not common for all the animals.

Different animals have different propensity to learn this stimulus response outcome in different ways. So, number 1 we say that both these processes can happen that is cognitive as well as reflexive things can happen and number 2 we say that different individuals have different propensity. And of course, the value of the outcome again

changes as a function of time function of context come function of so, many different things for an individual itself.

So, it is no surprise across an individual this same phenomena can induce learning in a different manner. So, it all in 2011 or I think two thousand nine 2011 decided to address this issue using this stimulus response learning paradigm in rats we will get into that in detail in the next lecture. However, let me give a little bit of a prelude of what exactly the behaviour is so that we can connect these phenomena's that we have learnt to that experiment. And then in the next lecture we can take it forward from where we are left there.

(Refer Slide Time: 22:57)



So, in short we are going to be using a skinner box kind of a instrument. Where there is a light cue that being the stimulus that being one of the stimulus there are many other stimuli that are present there right. So, the that being as soon as the light cue comes on the rat need to go and press a lever that is complete that is on the I mean that is on the diametrically opposite side of the box.

When it presses the lever it is going to get the food that is again it is in the food magazine it is in the other side. People pick this kind of an arrangement I mean the light cue can be here or here where it does not matter in one of those places for a particular reason the reason. Here being the lever when the animal presses the lever the food gets

released on the food magazine. But it stays there not for infinite amount of time, but for a very finite amount of time.

You can tighter the time such that it is just about sufficient for the animal to run and get it there. So, clearly it will be very very advantageous if the animal starting from here they need not have to travel it from across this box if it directly can go and get this food that will save it is time. And for sure ensure the availability of the food. So, it is that is the advantage; however, the contingency here is unless and until it presses the lever the food is not going to be there. So, that is light cue you can either this or this we can we can call it as stimulus.

That is the response and that is our outcome all right. So, now what are they going to do, they are actually going to utilize this geometry to probe two kinds of behaviour that are natural to these rats. A behaviour a we are going to call it as sign directed and behaviour two I am going to call it as goal oriented responses. And then go on to probe what determines a given animal will be either responding in a sign directed manner or a goal oriented manner. And can we bias them? And if you have to bias them, how do you bias them. With that I will end this lecture.

Thank you and see you in the next one.