The goal of this slide is to find an optimal solution for foraging using MVT for some very simplistic setting.
We are taking the most simplistic case of the environment. The reward we get for each harvest is given by r=E[7-0.5n-Ɲ(0,0.025)]. Since the random term is very low we can assume that to be 0 for each harvest in this simplistic setting. Also, another assumption is that the harvest time is 1 sec.
Now the reward function in this scenario is r=E[7-0.5n]=7-0.5n (because both terms are constant for a given state).

Variables	Meaning
T (this is the variable we are trying to optimize)	No of times we decide to harvest a patch. (Alternatively total time (in secs) we spend on a patch since harvest time =1 second)
t (constant for this env)	Travel time from one patch to the next

Also to be noted that the time is discrete.
On an average the time spent in a patch is sum of time of harvesting a particular patch + time of travelling to another patch.

$$T_u = T + t$$

Total energy gained by harvesting a patch for T time is given by

$$h(T) = \sum_{k=0}^{T-1}(7 - 0.5k)$$

$$\Rightarrow h(T) = \frac{29T - T^2}{4}$$

average energy per time we receive from a patch is given by

$$E_a = h(T)/T_u$$

$$\Rightarrow E_a = \frac{29T - T^2}{4(t+T)}$$

The goal is to maximise E_a w.r.t T so as to optimise the total reward.

Optimising E_a yields $\frac{dE_a}{dT} = \frac{29-2T}{4(t+T)} - \frac{29T-T^2}{4(t+T)^2}$

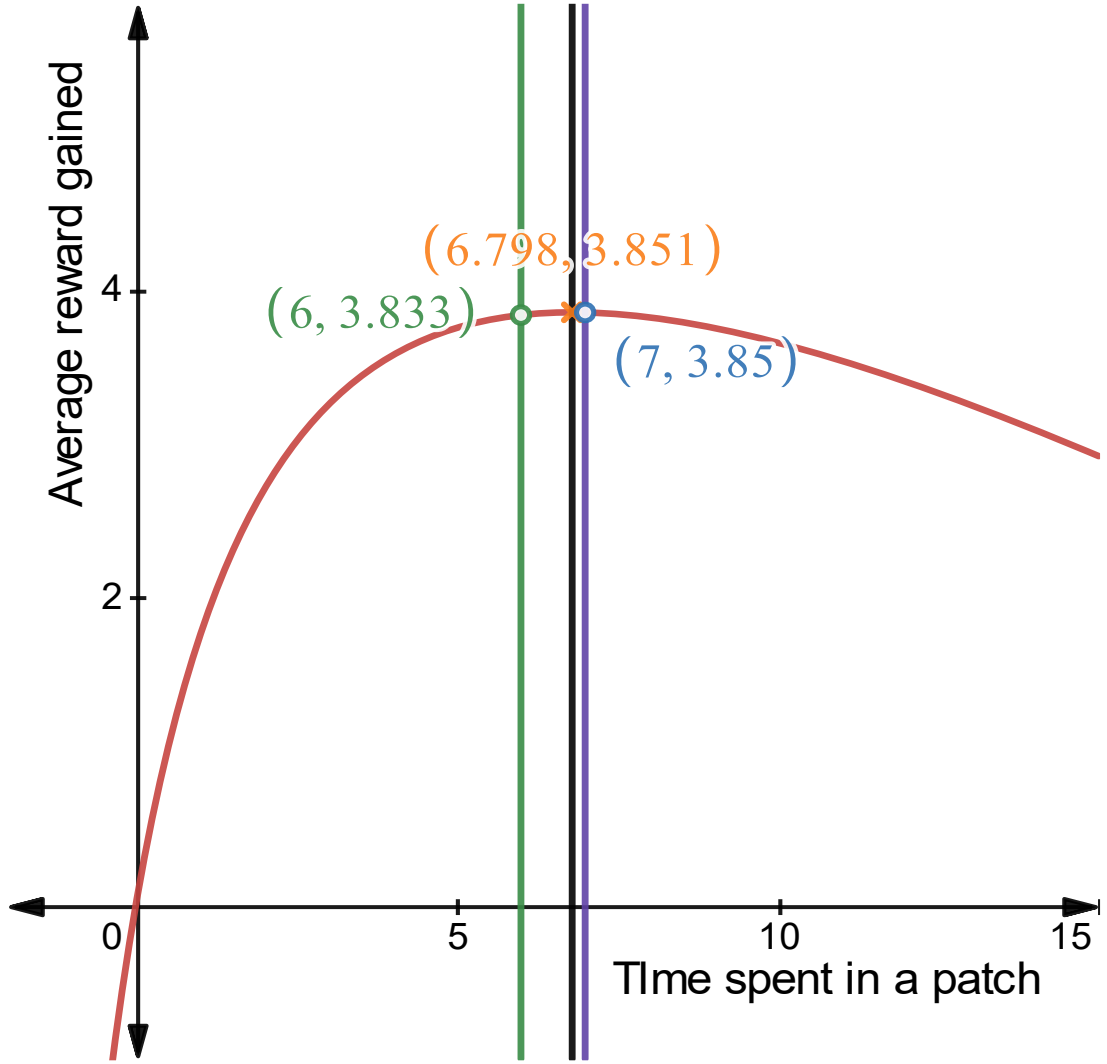
Setting the derivative to 0. $\frac{dE_a}{dT} = 0 \Rightarrow T = \sqrt{t^2 + 29t} - t$

Now for t=3, T=6.798. Since T can be discrete, so there are 2 cases.

But from the graph we can see that setting T=7 gives a higher

Reward on average. So T=7 for t=3 is the most optimal solution.

Similarly if we set t=10, then optimal solution is given by T=10.



<u>Advantage of Using RL agents over the analytical solution.</u>

MVT is known to suffer from issues due to boundary conditions.

For e.g. let's say the total time remaining in our environment is 10

Seconds. Let's say that travel time from one patch to another is 7

Seconds. So MVT predicts that we should leave the patch after

harvesting 7 times (i.e. for 7 seconds). But, our common sense

would say "why not harvest for the remaining 3 seconds instead of travelling to another patch which basically would give a 0 reward". To overcome this we can use RL agents which can plan their decisions and take decisions which are predicted by MVT in intermediate times, and still can act intelligently near boundary times to get even better rewards than that achieved by following a policy generated by following MVT.