1.

$$\begin{aligned}
G_t - V_t(S_t) &= R_{t+1} + \gamma G_{t+1} - V_t(S_t) + \gamma V_t(S_{t+1}) - \gamma V_t(S_{t+1}) + \gamma V_{t+1}(S_{t+1}) - \gamma V_{t+1}(S_{t+1}) \\
&= (R_{t+1} + \gamma V_t(S_{t+1}) - V_t(S_t)) + (\gamma G_{t+1} - \gamma V_{t+1}(S_{t+1})) + (\gamma V_{t+1}(S_{t+1}) - \gamma V_t(S_{t+1})) \\
&= \delta_t + \gamma(G_{t+1} - V_{t+1}(S_{t+1})) + \gamma\alpha\delta_{t+1} \\
&= \delta_t + \gamma\alpha\delta_{t+1} + \gamma\delta_{t+1} + \gamma^2\alpha\delta_{t+2} + \gamma^2(G_{t+2} - V_{t+2}(S_{t+2})) = \dots \\
&= \sum_{k=t}^{T-1} \gamma^{k-t}(1+\alpha)\delta_k
\end{aligned}$$

2. If I get lost a few times at the beginning, then I can see why TD updates can be better. When I get lost, only the first few states learn that it takes a long time for me to go home. After I reached the highway, the predictions won't change just because I don't know the way from the new office building to the highway.

3. The first episode must have terminated on the left with a reward of -1.

   The other state values didn't change because the error in those cases was $R_{t+1} + \gamma V(S_{t+1}) - V(S_t) = 0 + 1 \cdot 0.5 - 0.5 = 0$.

   $v(A)$ changed by $\alpha(R_{t+1} + \gamma V(S_{t+1}) - V(S_t)) = 0.1 * (-1 + 0 - 0.5) = -0.15$.

4.

5. For now, let us treat the terminal points as states with values 0 and 1 respectively and all the rewards to be 0. This doesn't change the updates, but makes the notation a bit easier.

   Claim 1. *If all the state values are initialized to* 0.5*, then*

   $$0 \le V(A) \le V(B) \le V(C) \le V(D) \le V(E) \le 1$$

   holds throughout the run of the TD algorithm.

   Proof. The proof goes by induction on the number of steps. The statement clearly holds in the beginning. Suppose you are in state S and take a step to the right, arriving to S'. The updated $V(S)$ then equals to

   $$(1-\alpha)V(S) + \alpha V(S') \le (1-\alpha)V(S') + \alpha V(S') = V(S').$$

   □

   Now let S be any state in $\{A, B, D, E\}$. Let S_i and S_o denote S's neighbor closer to C and further from C, respectively.

   Claim 2. *Suppose we are in state* S *in time step* t *and the previous state was* S_i*. Then* $|\mathbb{E}(V_{t+1}(S_t)) - V_t(S_o)| < |V_t(S_t) - V_t(S_o)|$ *if and only if*

   $$(1-\alpha)|V_t(S_t) - V_{t-1}(S_i)| < |V_t(S_o)) - V_t(S_t)|$$

Proof. Using the update rule $V_t(S_i) = (1-\alpha)V_{t-1}(S_i) + \alpha V_{t-1}(S_t)$ and $V_{t-1}(S_t) = V_t(S_t)$ the following holds.

$$\begin{aligned}
\mathbb{E}(V_{t+1}(S_t)) &= V_t(S_t) + \frac{\alpha}{2}\Big(\big(V_t(S_i) - V_t(S_t)\big) + \big(V_t(S_o)) - V_t(S_t)\big)\Big) \\
&= V_t(S_t) + \frac{\alpha}{2}\Big(\big((1-\alpha)V_{t-1}(S_i) + \alpha V_{t-1}(S_t) - V_t(S_t)\big) + \big(V_t(S_o)) - V_t(S_t)\big)\Big) \\
&= V_t(S_t) + \frac{\alpha}{2}\Big(((1-\alpha)\big(V_{t-1}(S_i) - V_t(S_t)\big) + \big(V_t(S_o)) - V_t(S_t)\big)\Big)
\end{aligned}$$

$\square$