Exercise 02 MDPs and the Markov Property

1. (a)

$S = \{5, 4, 3, 2, 1, S(\text{success}), F(\text{fail})\}$

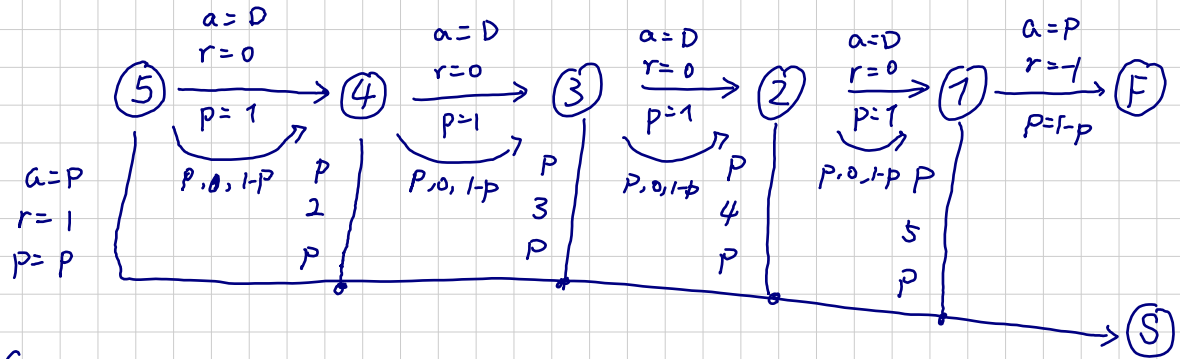
$A = \{P(\text{park}), D(\text{drive on})\}$

$R = \{0, 1, 2, 3, 4, 5, -1\}$

$p : S \times R \times S \times A \to [0,1] = \{$

$(5,0,4,D) \mapsto 1,$
$(5,1,S,P) \mapsto p,$
$(5,0,4,P) \mapsto 1-p,$
$(4,0,3,D) \mapsto 1,$
$(4,2,S,P) \mapsto p,$
$(4,0,3,P) \mapsto 1-p,$
$(3,0,2,D) \mapsto 1,$
$(3,3,S,P) \mapsto p,$
$(3,0,2,P) \mapsto 1-p,$
$(2,0,1,D) \mapsto 1,$
$(2,4,S,P) \mapsto p,$
$(2,0,1,P) \mapsto 1-p,$
$(1,5,S,P) \mapsto p,$
$(1,-1,F,P) \mapsto 1-p \}$

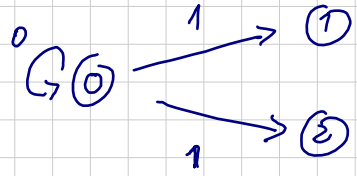
(b)



(c) no, because this is a finite MDP and expected value for all states are finite.

2. it depends on how states are modelled.
if it's a stateless model like Bandits, then no. because the known information is incompletely modelled.
if we model it as the outcome of the last 2 rounds, then yes.

3. (a) consider MDP:



with $S = \{0, 1, 2\}$
$S^+ = \{1, 2\}$

and $V_*(0) = 1$, $V_*(1) = V_*(2) = 0$.

we have $\pi_*(0) = 1$ or $\pi_*(0) = 2$. as different optimal policies.

(b).

$V_*(3) = 0$
$V_*(2) = 1$
$V_*(1) = 1$.

$V_*(2) = V_*(1)$