A

1 Hoe willen jullie dat ik dit precies weergeef?

2 kan niet door 1?(zelfde voor 3 en 4)

3

4

B   
1 Omdat er daarna geen stappen meer zijn

2 Een voordeel van TD-learning is dat het sneller convergeert omdat het updates maakt na elke transitie. Een nadeel is dat TD-learning gevoeliger kan zijn voor de keuze van de learning rate en initialisatie.

C

1 Een Q-function is een functie die de verwachte totale beloning voorspelt voor het nemen van een specifieke actie in een bepaalde toestand.

2 omdat we geen model/kennis hebben van de omgeving.

3 'On-policy' is dat het algoritme leert en beslissingen neemt op basis van de acties die het daadwerkelijk uitvoert.

4 Ja, omdat het de agent kan beperken tot een suboptimaal pad zonder alternatieven te verkennen.