Chapter 8 solutions **Hanna Gábor**

1. *The nonplanning method looks particularly poor in Figure 8.3 because it is a one-step method; a method using multi-step bootstrapping would do better. Do you think one of the multi-step bootstrapping methods from Chapter 7 could do as well as the Dyna method? Explain why or why not.*

   I think it can come close. An n-step bootstrpping algorithm with big enough n (or a Monte Carlo method) would update at the end of the first episode all the action-values we encountered during the episode. I expect the policy based on these updated action-values to be quite good.

2. *Why did the Dyna agent with exploration bonus, Dyna-Q+, perform better in the first phase as well as in the second phase of the blocking and shortcut experiments?*

   At first, both algorithms might find an okay policy that is suboptimal. Due to the exploration, Dyna-Q+ will realize it sooner that it's a suboptimal policy.

3. *Careful inspection of Figure 8.5 reveals that the difference between Dyna-Q+ and Dyna-Q narrowed slightly over the first part of the experiment. What is the reason for this?*

   After finding the optimal path, Dyna-Q always uses that path, whereas Dyna-Q+ does some exploration from time to time.