**Theoretical Answer**

The value of the advantage estimate reflects the contribution of the action beyond the current “average” of the state. Because the return can be change drastically between episodes, the REINFORCE algorithm gradient can suffer from high variance. This variance leads to unstable convergence. By subtracting the baseline, we can reduce the variability of the return and therefore the variance of the gradients.