Exercises for chapter 5 : An extended example: Tic-Tac-Toe

1. Suppose, instead of playing against a random opponent, the reinforcement learning algorithm described above played against itself, with both sides learning. What do you think would happen in this case? Would it learn a different policy for selecting moves?

   Proof. Both agents will learn to beat each other probably resulting in some equilibrium. For that they would of course learn something different than against a random opponent. □