Chapter One Solution

1.1 Self-Play

Answer

According to game theory, it is a finite game, thus it must have a Nash equilbrium, either mixed policy or pure policy.
With a more clever opponent, it may learn a different policy.

1.2 Symmetries

Answer

Using a appropriate "state" and "action" to take advantage of symmetry.
Our strategy on whether use the symmetry should be the same with our opponent's.

1.3 Greedy Play

Answer

As mention in the book, a challenge of RL is the trade-off between exploration and exploitation. But greedy strategy only exploits and thus this strategy will always have a worse result.

1.4 Learning from Exploration

Answer

Without exploration, the learned probabilities can't express the complete distribution of the states and actions.

1.5 Other Improvements

Answer

Ways to improve:

- Replay, in order to approach the real distribution of states and actions.
- Asynchronous training, speeding up the training.

As for tic-tac-toe game, the state space is so small that it can be solve be traditional search method such as recursive search and adversarial search.