Problem 1

Q-learning is an off-policy algorithm and SARSA-learning is an on-policy algorithm. Assuming the same starting state for each learning episode, if the policy used by an agent is the "greedy policy", then the ensuing states and actions that are taken in these states would be the same for both Q and SARSA learning. Thus, the "greedy policy" would make on-policy and off-policy learning do the same thing, making them equivalent.

Problem 2

A positive reward for taking the opponent's pieces can distract the agent into learning a policy that may favor the agent taking the opponent's pieces at the expense of winning the game. Thus, the agent might learn a policy that only helps it take the opponent's pieces, but not one that is a good strategy for a win. It is true that taking the opponent's pieces helps towards a possible win in the game (i.e. a positive reward for taking pieces may indicate the agent how it could win), but taking pieces is not sufficient for a win. In general, the reward structure should tell the agent what to do, not how to do it.

Problem 3

When the state/action space is large, one can use a neural network to represent the value function. The neural network would thus take the state-action pair as its input, and give the value of this pair as the output. As such, value updates happen indirectly via updating the weights of the neural network.