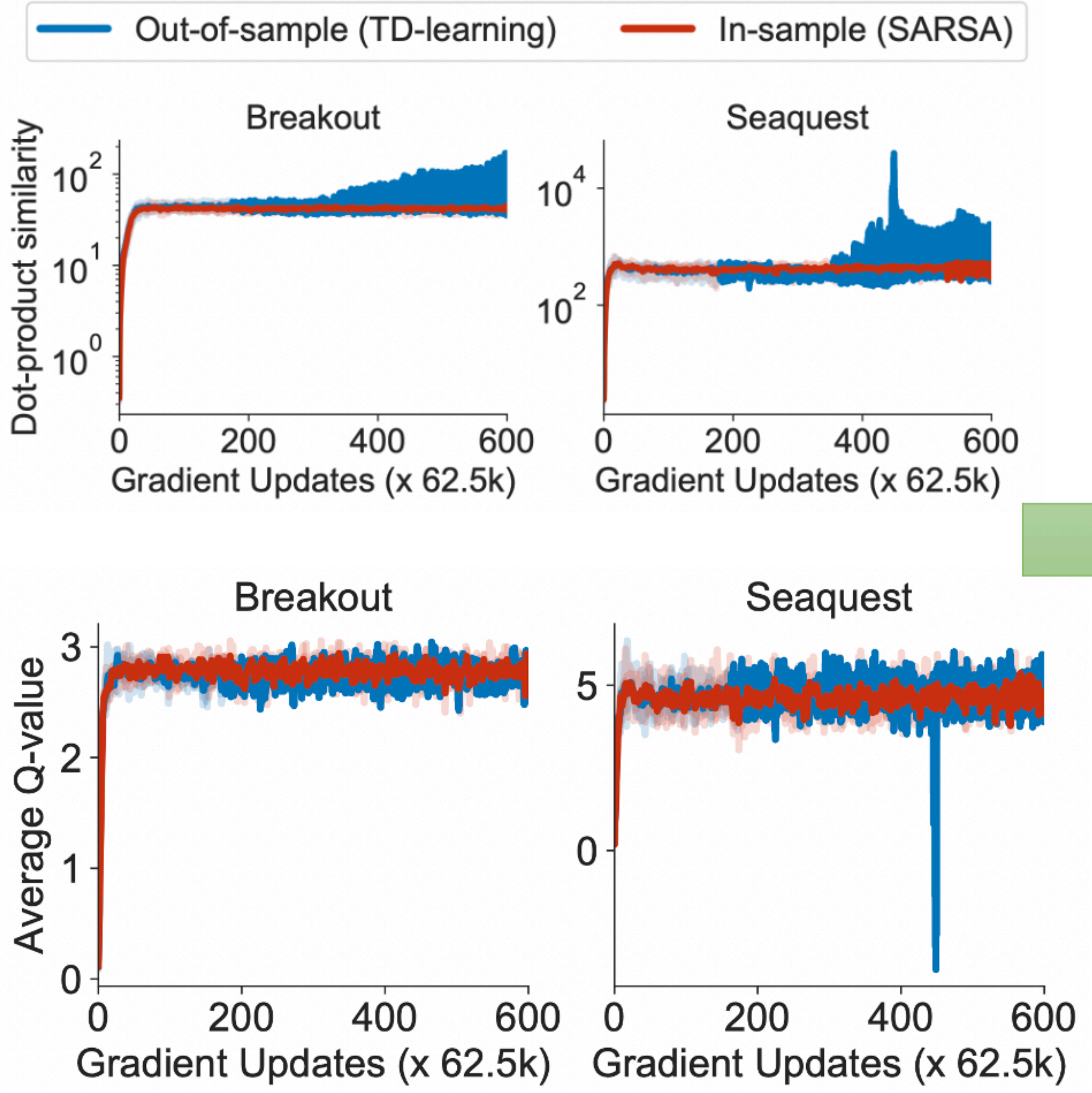
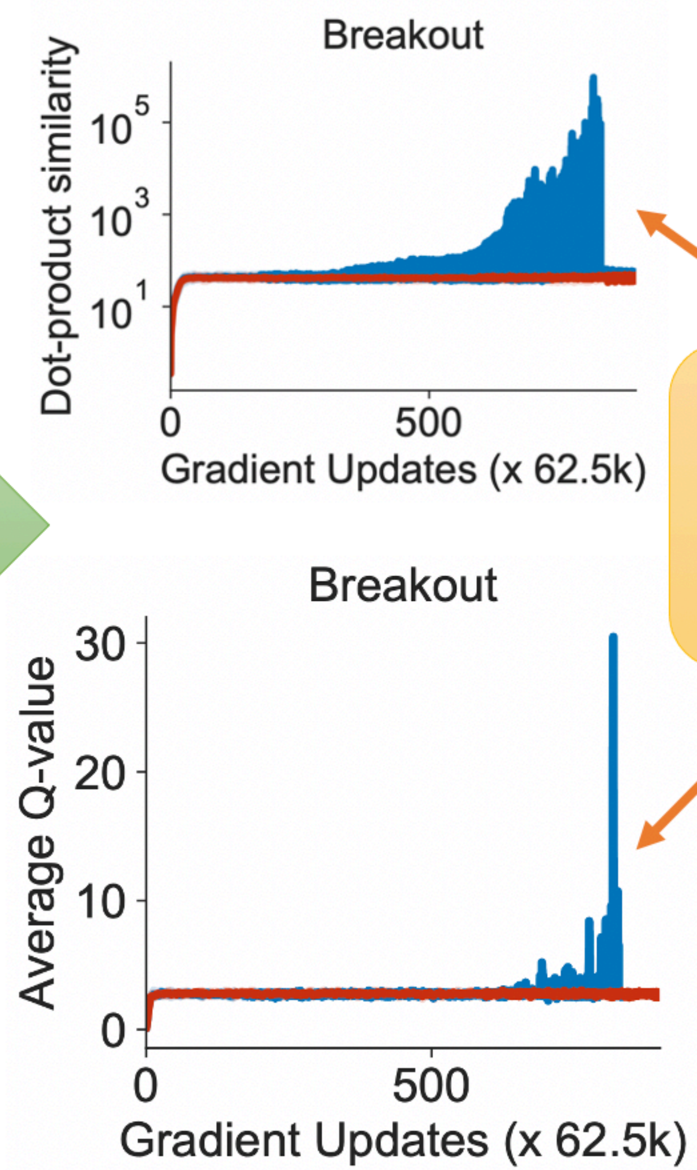
Large feature dot products arise when out-of-sample actions are used in TD-learning compared to SARSA, despite similar Q-values

Large feature dot products <u>eventually</u> imply divergent Q-values

High dot products

Similar Q-values





Incorrect Q-values found eventually; dot-products increase throughout training