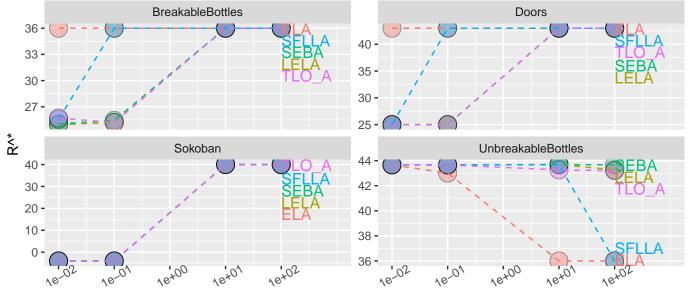Average Offline Performance across different Alignment scale levels

Across 5000 trials