ALEC PORTELLI

①

The value iteration backup is:

$$V_{k+1}(s) = \max_a \sum_{s',r} p(s',r \mid s,a)\left[r + \gamma V_k(s')\right]$$

If we switch to action values, the equation becomes the following:

$$q_{k+1}(s,a) = \sum_{s',r} p(s',r \mid s,a)\left[r + \gamma \max_{a'} q_k(s',a')\right]$$