

Kvíz na policy evaluation

-80	130	-100
-80		-100
-80		-100

Čísla označují rewards (odměny). Červeně označené stavy jsou terminální.

Agent poslechne s pravděpodobností 0.65 a s pravděpodobností 0.175 se vydá $\pm 90^\circ$ směrem. Koeficient zapomínání 0.4. Pro neterminální stavy spočítejte hodnoty (values) $V(s)$ pro policy: „jed' vlevo (go left)“