

**Kvíz na policy evaluation**

-140	160	-100
-140		-100
-140		-100

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

Agent poslechne s pravděpodobností 0.6 a s pravděpodobností 0.2 se vydá  $\pm 90^\circ$  směrem. Koeficient zapomínání 0.8. Pro neterminální stavy spočítejte hodnoty (values)  $V(s)$  pro policy: „jed' rovně (go straight)“