

**Kvíz na policy evaluation**

-70	160	-80
-70		-80
-70		-80

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 0.7. Pro neterminální stavy spočítejte hodnoty (values)  $V(s)$  pro policy: „jed' rovně (go straight)“