

# Ambiente 4x3

## ■ Recompensas $R(s)$ :

- Três estados finais  $R(s) = +1$ ,  $R(s) = -1$  e  $R(s) = +0.2$  na posição (4,1)
- $R(s) = r$  (valor fixo) para todos estados não terminais

3				<div>+1</div>
2				<div>-1</div>
1	INÍCIO			<div>0.2</div>
	1	2	3	4

## ■ Implemente o algoritmo Q-Learning e retorne a política para as seguintes situações:

- $r = -0.4$
- $r = -0.04$
- Comente as diferenças entre as políticas retornadas

