

# ExTra: Transfer-guided Exploration Supplementary Material

Paper #1073

## 1 EXTRA VARIANTS OF TRADITIONAL EXPLORATION STRATEGIES

---

---

### Algorithm 1: ExTra + vanilla $\epsilon$ -greedy Q-learning

```

1: step = 0
2: while step < MAXSTEPS do
3:   with probability  $\epsilon$ 
4:     with probability  $\epsilon_{bisim}$ 
5:        $a_2 \sim \pi_{ExTra}(\cdot | s_2, \mathcal{M}_1, \pi_1^*)$ 
6:     with probability  $1 - \epsilon_{bisim}$ 
7:        $a_2 \sim \text{uniform}(A_2)$ 
8:   with probability  $1 - \epsilon$ 
9:      $a_2 \leftarrow \arg \max_{a'_2} Q_2(s_2, a'_2)$ 
10:   $r = \text{take\_step}(a_2)$ 
11:   $\text{update\_}Q(Q_2(s_2, a_2), r)$ 
12:  step = step + 1
13: end while

```

---



---

---

### Algorithm 2: ExTra + Softmax

```

1: step = 0
2: while step < MAXSTEPS do
3:   with probability  $\epsilon$ 
4:      $a_2 \sim \pi_{ExTra}(\cdot | s_2, \mathcal{M}_1, \pi_1^*)$ 
5:   with probability  $1 - \epsilon$ 
6:      $a_2 \sim \text{softmax}(Q_2(s_2, \cdot))$ 
7:    $r = \text{take\_step}(a_2)$ 
8:    $\text{update\_}Q(Q_2(s_2, a_2), r)$ 
9:   step = step + 1
10: end while

```

---



---

---

### Algorithm 3: ExTra + Pursuit

```

1: step = 0
2:  $\pi_{pursuit} = \text{Uniform}(A_2)$ 
3: while step < MAXSTEPS do
4:   with probability  $\epsilon$ 
5:      $a_2 \sim \pi_{ExTra}(\cdot | s_2, \mathcal{M}_1, \pi_1^*)$ 
6:   with probability  $1 - \epsilon$ 
7:      $a \leftarrow \arg \max_{a'_2} Q_2(s_2, a'_2)$ 
8:      $\text{update\_}\pi_{pursuit}(a)$ 
9:      $a_2 \leftarrow \text{Sample}(\pi_{pursuit})$ 
10:   $r = \text{take\_step}(a_2)$ 
11:   $\text{update\_}Q(Q_2(s_2, a_2), r)$ 
12:  step = step + 1
13: end while

```

---



---

---

### Algorithm 4: ExTra + MBIE-EB

```

1: step = 0
2: while step < MAXSTEPS do
3:   with probability  $\epsilon$ 
4:      $a_2 \sim \pi_{ExTra}(\cdot | s_2, \mathcal{M}_1, \pi_1^*)$ 
5:   with probability  $1 - \epsilon$ 
6:      $a_2 \leftarrow \arg \max_{b'} Q_2(s_2, a'_2)$ 
7:    $r = \text{take\_step}(a_2) + \frac{\beta}{\sqrt{n(s_2, a_2)}}$ 
8:    $\text{update\_}Q(Q_2(s_2, a_2), r)$ 
9:   step = step + 1
10: end while

```

---

## 2 HYPERPARAMETERS FOR OPTIMISTIC BISIMULATION TRANSFER

Transfer SixLarge	Parameters Rooms	FourLarge NineLarge	Rooms Rooms
$c_R$	0.1	0.2	0.1
$c_T$	0.9	0.9	0.9
Threshold	0.01	0.01	0.01

## 3 HYPERPARAMETERS FOR BASELINE EXPLORATION STRATEGIES

$\epsilon$ -greedy	Q Learning Rate	0.2
	$\epsilon$	0.5
Softmax	Q Learning Rate	0.2
	$\tau$	8.1
MBIE-EB	Q Learning Rate	0.2
	cb- $\beta$	0.005
	$\epsilon$	0.2
Pursuit	Q Learning Rate	0.2
	$\beta$	0.007
ExTra	Q Learning Rate	0.5
	$\epsilon$	0.2
	$\alpha$	1e-6