CS395T
Reinforcement Learning: Theory and Practice
Fall 2004

Peter Stone

Department or Computer Sciences
The University of Texas at Austin

Week2a: Tuesday, September 7th
Good Afternoon Colleagues
Good Afternoon Colleagues

● Are there any questions?
Logistics

- Nice responses!
Logistics

- Nice responses!
  - Length and content good
Logistics

- Nice responses!
  - Length and content good
  - Look for programming assignment opportunities
Logistics

- Nice responses!
  - Length and content good
  - Look for programming assignment opportunities
  - Send responses to the discussion leader?
Logistics

- Nice responses!
  - Length and content good
  - Look for programming assignment opportunities
  - Send responses to the discussion leader?
    * You? Me? Just selected ones?
Logistics

- Nice responses!
  - Length and content good
  - Look for programming assignment opportunities
  - Send responses to the discussion leader?
    * You? Me? Just selected ones?

- Discussion leader week 8?
Logistics

- Nice responses!
  - Length and content good
  - Look for programming assignment opportunities
  - Send responses to the discussion leader?
    * You? Me? Just selected ones?

- Discussion leader week 8?

- Stars on the sections?
Let’s Play!
Let’s Play!

- I’m a 2-armed bandit
Let’s Play!

- I’m a 2-armed bandit
- As a class, you choose which arm: 3 times around.
Let’s Play!

- I’m a 2-armed bandit
- As a class, you choose which arm: 3 times around.
- Maximize your payoff.
Let’s Play!

- I’m a 2-armed bandit
- As a class, you choose which arm: 3 times around.
- Maximize your payoff.
- The answer:
Let’s Play!

- I’m a 2-armed bandit
- As a class, you choose which arm: 3 times around.
- Maximize your payoff.
- The answer:

```lisp
(defun l () (+ 7 (random 7)))
```

```lisp
(defun r ()
    (let ((x (random 3)))
        (case x
            (0 20)
            (1 0)
            (2 (+ 6 (random 11)))
        )))
```
N-armed bandit in practice?

- Choosing mechanics
- Choosing a barber/hairdresser
What’s all this about RL vs. Supervised?

- What exactly are they comparing?
What’s all this about RL vs. Supervised?

- What exactly are they comparing?
- Why is instructional feedback better? (p.32, bottom)
What’s all this about RL vs. Supervised?

- What exactly are they comparing?
- Why is instructional feedback better? (p.32, bottom)
- Why $L_{R-I}$ better on bandit A than B?
What’s all this about RL vs. Supervised?

- What exactly are they comparing?
- Why is instructional feedback better? (p.32, bottom)
- Why $L_{R-I}$ better on bandit A than B?
  - Why better than action values?
Exercises

• 2.1
Exercises

- 2.1
- 2.6
Exercises

- 2.1
- 2.6
- 2.8
Exercises

• 2.1

• 2.6

• 2.8