Curricula for Learning Agents

Matthew E. Taylor

http://eecs.wsu.edu/~taylorm

WASHINGTON STATE UNIVERSITY

This Session's Focus

- Idea-focused
- Participation Encouraged

2004: Real Life Reinforcement Learning

Where's My Adaptive & Learning Agent?

Our goal (for the sake of discussion)

We want to have a deployed, consumer learning agent in the next 5 years.

- Assume
  - If an embodied agent, assume someone else is building the (right) hardware
  - An MBA can make it profitable

- What task(s) should it achieve?
- What are we missing to make this technically feasible?

- I’m ignoring it (spam filter, amazon): cloud/ server side
- Siri
- Preference, recommender
- Don’t trust agents: sensible exploration
- Serve my interests
- NEST
- Multi-agent?
- Trading agents: bidding for adwords, etc.
- HCI
- Trust in “physical type things”

Learning

Real life agent learning

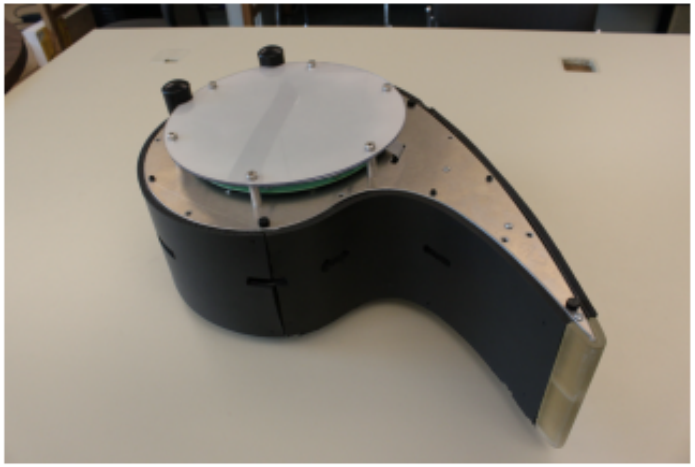
- Reliability
- Speed

How do agents get a better prior?

- Better learning algorithms
- Leverage past experience
- Human knowledge

Leveraging Past Experience

- Lifelong Learning
  - Sebastian Thrun
  - Mark Ring
  - Sutton+: Critterbot, Horde Architecture



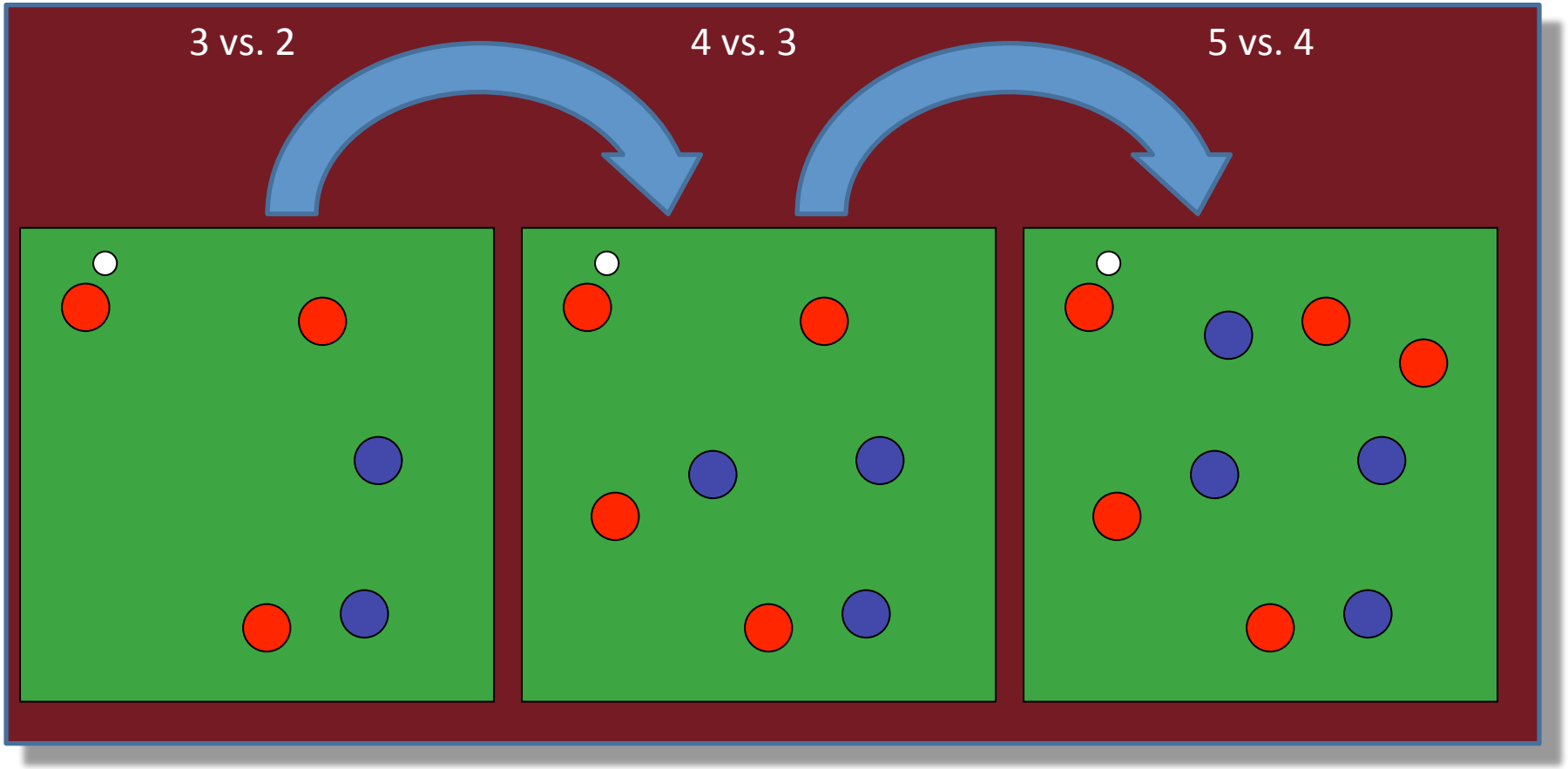
- Multi-task Learning
  - Fernandez, Lazaric

- Transfer Learning
  - Re-use past knowledge
  - Use to set bias
  - Automatically learn how tasks are similar

Leveraging Past Experience

- Determining how tasks are related
- How past info is used
  - Q-values
  - Policy
  - Model
  - Options
  - Reward functions
  - High-level rules/advice
  - Features

Multi-step Transfer



Sequential learning can outperform direct learning: can take ½ the time!

Programmer Knowledge

- Algorithm, parameters, function approximator, etc.
- Smart feature selection
- Clever actions
  - Macro actions, options, etc.
- Reward shaping
  - Andrew Ng
  - Sam Devlin+

Goals for Human Interaction?

- How to get knowledge from human
  - Limited effort
  - Non-optimal
  - Non-technical
- HCI/HRI

Human Interaction

- Learning from Demonstration
  - Brenna Argall+: Survey
- Imitation learning
  - Price & Boutilier: Implicit Imitation
- Learning from Feedback
  - Brad Knox: TAMER
- Giving NLP Advice
  - Rich Maclin: RATLE

- Demonstration is critical: grandma!
- Good for Human to Robot
  - Robot having model of people? Quirky way people will respond
- GIANT problem
  - Psych / Econ: agent shrinks
  - Why doing what they're doing?
  - Don't need to look at your perspective: sit in robot's shoes
- BUT human shouldn't have more involvement than wanted

Curriculum Learning

- Assisting Transfer-Enabled Machine Learning Algorithms: Leveraging Human Knowledge for Curriculum Design (2009)

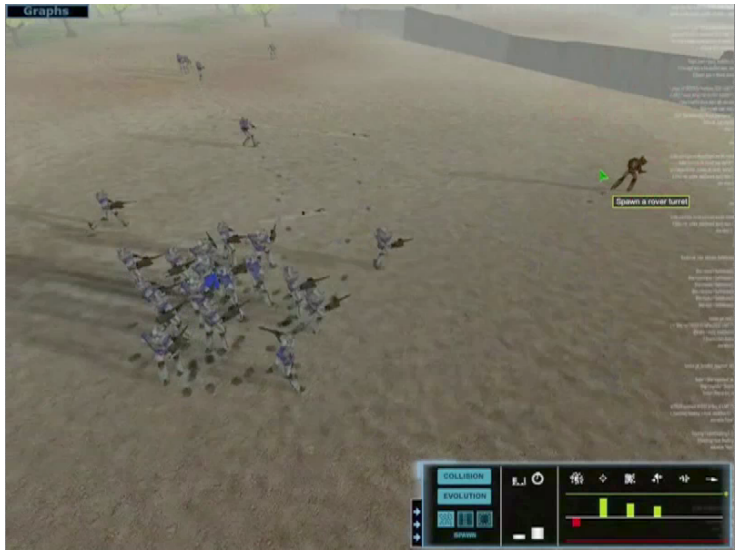
- Multiple possible goals
  - General knowledge
  - Specific final task

Curriculum Learning

- ML, HCI/HRI, education?
  - Thomaz & Breazeal: Teachable Robots: Understanding Human Teaching Behavior to Build More Effective Robot Learners
  - Roberts & Littman+: Human Experiments
  - Knox+: Understanding Human Teaching Modalities in Reinforcement Learning Environments: A Preliminary Report

Curriculum Learning

- Isbell & Thomaz+ (2010)
  - Training regimens
- Bengio+ (2009)
  - Supervised Learning
- Stanley+
  - Nero Video Game
- Consumer level?
  - How do humans *want* to pick tasks?
  - Can normal people do this well?



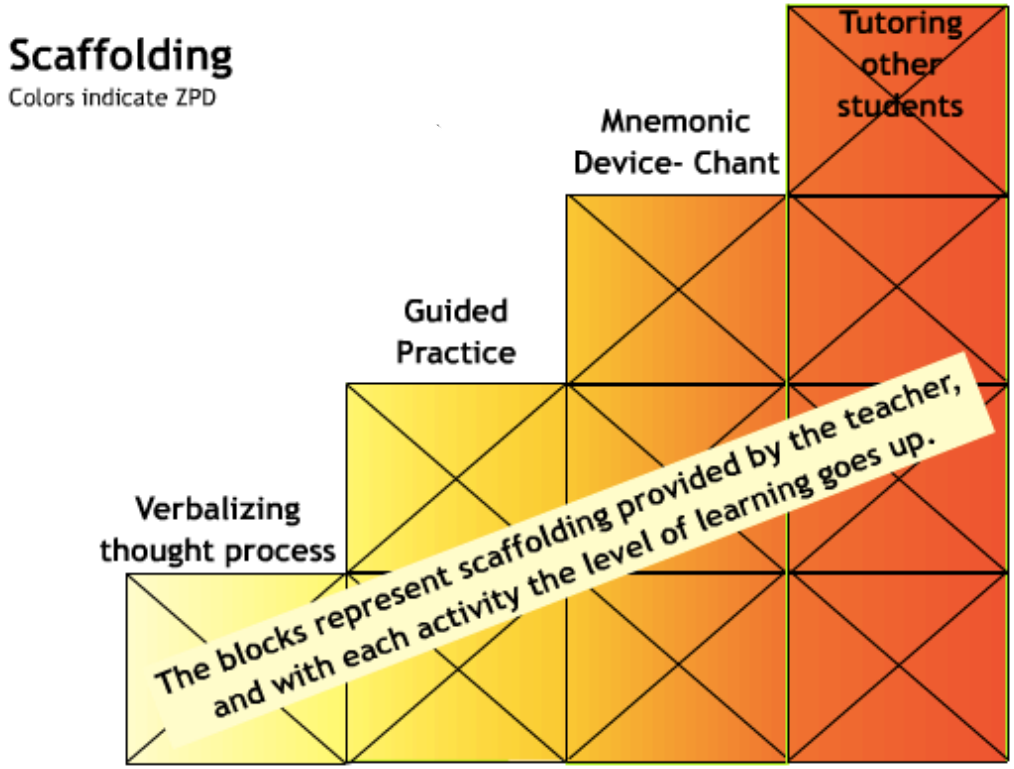
Instructional Scaffolding (1950s)

- Soft scaffolding
  - circulating around room and answering questions / providing feedback
- Hard scaffolding:
  - identify hints/cues before assigning problem

Picking Tasks

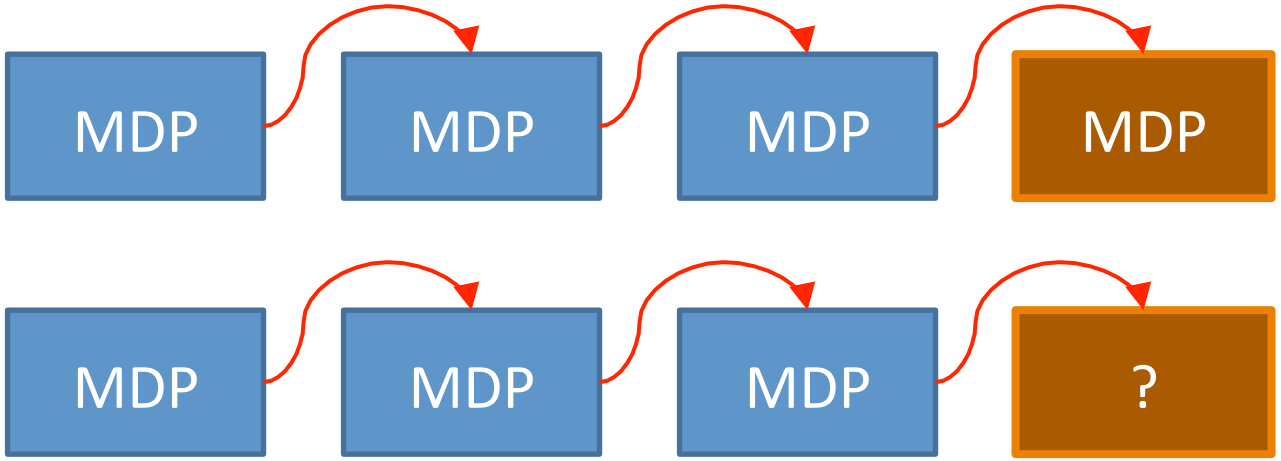
Zone of Proximal Development

- Expert state: What learner can do on own
- Pedagogical State: Can be achieved with the support of a instructor



Automatic Curriculum Design?

- Meta-planning problem for agent learning

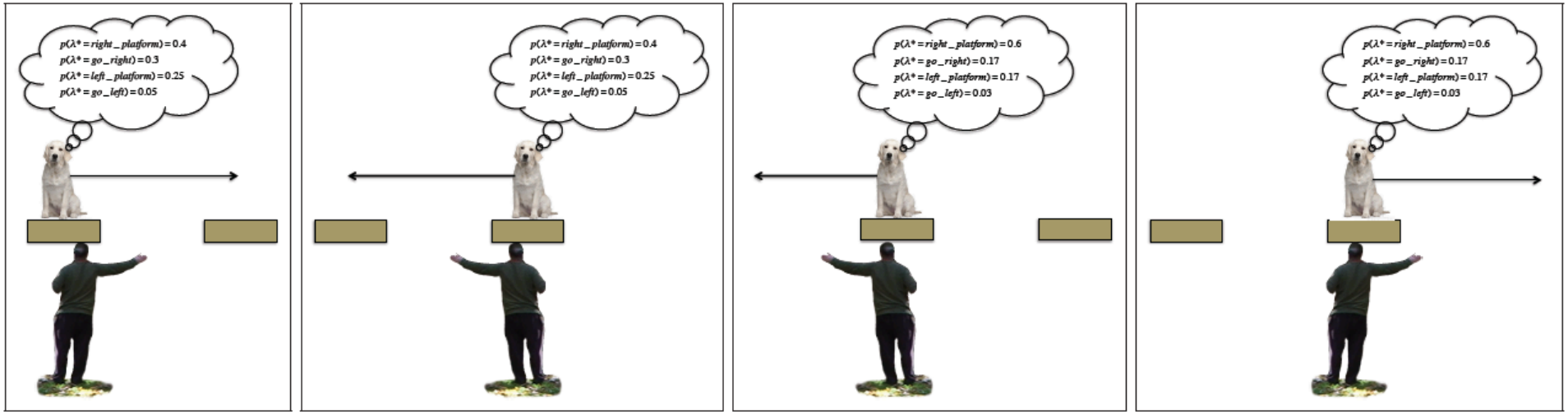


- Post-hoc analysis: determine “optimal curriculum”?
- Model the student essential?

Other Ideas

- Task relatedness measures
  - Bou Ammar+: MDP similarity for TL usefulness

- Intentionality of task sequence:
  - Roberts, Littman+: Dog learning



(a) To Platform (right) (b) To Platform (left) (c) In Direction (left) (d) In Direction (right)

Helpful?

- Learn the right state features
- How should the state be represented (function approximation)
- Learn a prior over reward functions, policies, etc.
- Build up a library of policies
- Bias action selection
- Set a decent learning rate / tune learning params
- …

Summary

- Leveraging past knowledge sets biases
- Easy way for humans to help agents learn
- Automate curricula creation

- Non-RL applicability?
- Other challenges?
- Where to start?
- Weaknesses in ideas/approach?

- MDP: needs bias. But MDP is a limitation? How do we go beyond by rethinking s/a/timing
- Some easy cases fail: negative transfer
  - What are the steps to go from trivial to complex
  - How can we make this more reliable?
- Disagree that learning is the problem
  - Not good at mobile/manipulate
  - Blame Willow Garage
- Hard to get enough data though from humans….
- Giving a good context for learning (even hard for humans to learn with proper context)
- Training happening in home or in factory/lab?
  - Home: Might require human involvement human isn't prepared
  - Factory: Could be OK if better performance than engineering….

Agents Teaching Agents

- Nick Carboni
  - A few minutes from now
- Lisa Torrey
  - Thursday, F4 – Learning 1
  - Teaching on a Budget: Agents Advising Agents in Reinforcement Learning