Using Markov Decision Processes to Understand Student Thinking in Performance Tasks

Michelle M. LaMar
Educational Testing Service
mlamar@ets.org

October 1st, 2015

23 **Within a substance, atoms that collide frequently and move independently of one another are *most* likely in a**

A liquid.

B solid.

C gas.

D crystal.

CSZ20827

Traditional Assessment Task

Standard Educational Measurement Paradigm

23 **Within a substance, atoms that collide frequently and move independently of one another are *most* likely in a**

A liquid.

B solid.

C gas.

D crystal.

CSZ20827

Not very similar

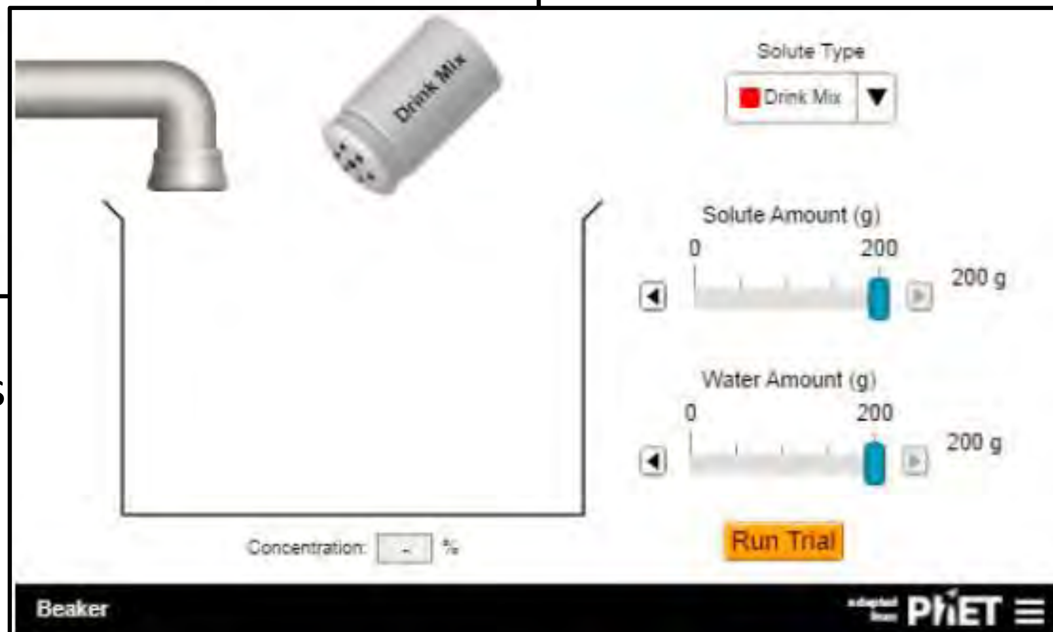
Real Science Task

Traditional Assessment Task

23 **Within a substance, atoms that collide frequently and move independently of one another are *most* likely in a**

A liquid.
B solid.
C gas.
D crystal.

Simulated Task



l Science Task

Traditional As

Assessing Science Skills

Give students some equipment and see what they do.

- What is their goal?
- How much do they care?
- Who is contributing? How much?
- Do they understand how the equipment works?
- Are they using good inquiry skills?
- Do they understand the science content?

Assessing Science Skills

Give students some equipment and see what they do.

- Goals
- Motivation
- Collaboration Skills
- Beliefs & Understanding of Task Setup
- Science Process Skills
- Science Content Knowledge

Assessing Science Skills

Give students a standard assessment item.

- ~~Goals~~
- ~~Motivation~~
- ~~Collaboration~~
- ~~Beliefs & Understanding~~
- ~~Science Process~~
- Science Knowledge

23 **Within a substance, atoms that collide frequently and move independently of one another are *most* likely in a**

A liquid.

B solid.

C gas.

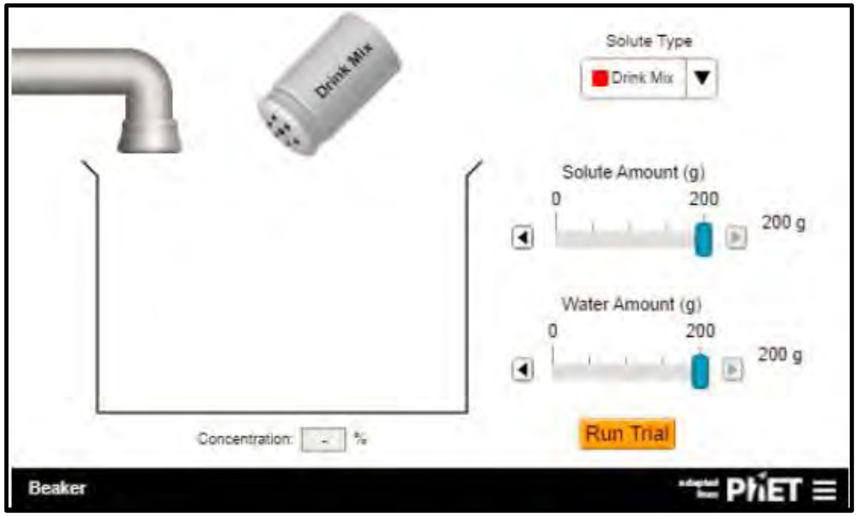
D crystal.

CSZ20827

Assessing Science Skills

Give students some equipment and see what they do.

- Goals
- Motivation
- Collaboration
- Beliefs & Understanding
- Science Process
- Science Knowledge

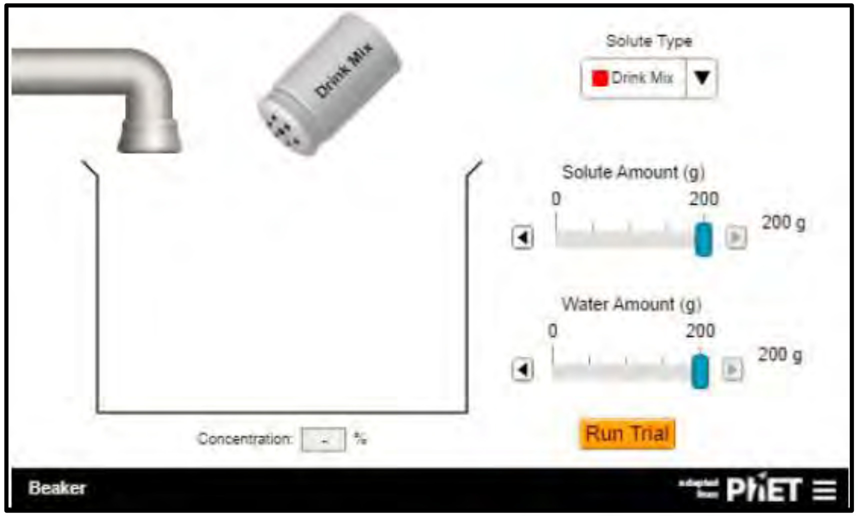


Assessing Science Skills

Give students some equipment and see what they do.

- Goals
- Motivation
- Collaboration
- Beliefs & Understanding
- Science Process
- Science Knowledge

Model
All Factors
of Interest



Latent-trait Models

Mislevy: The conditional probability model-fragments:

$$p\left(X_{ijk} \middle| \theta_i, \beta_j, \zeta_k\right)$$

X_{ijk} is the "observable" variable from the action(s) of "Person *i*" in "Situation *j*" given other relevant contextual variables *k*;

θ_i is the "proficiency" variable for "Person *i*" (might also subscript for time *t*);

β_j is the effect of "Situation *j*"; and

ζ_k is the effect of other relevant contextual variables *k*.

Cognitive Process Models

Action choice based on human and environment:

$$p(a_{ijk} | \theta_i, \beta_j, \zeta_k)$$

a_{ijk}	is the "observable" actions of "Person *i*" in "Situation *j*" given other relevant contextual variables *k*;
θ_i	is the "proficiency" variable for "Person *i*" (might also subscript for time *t*);
β_j	is the effect of "Situation *j*"; and
ζ_k	is the effect of other relevant contextual variables *k*.

Outline

- Peg Solitaire Example
- Markov Decision Process Measurement Model
  - The MDP
  - The MDP for Measurement
- MDP-MM in Action
  - Peg Solitaire
  - Microbes
  - SimCityEDU Pollution Challenge
- Conclusions

Example: Peg Solitaire Game

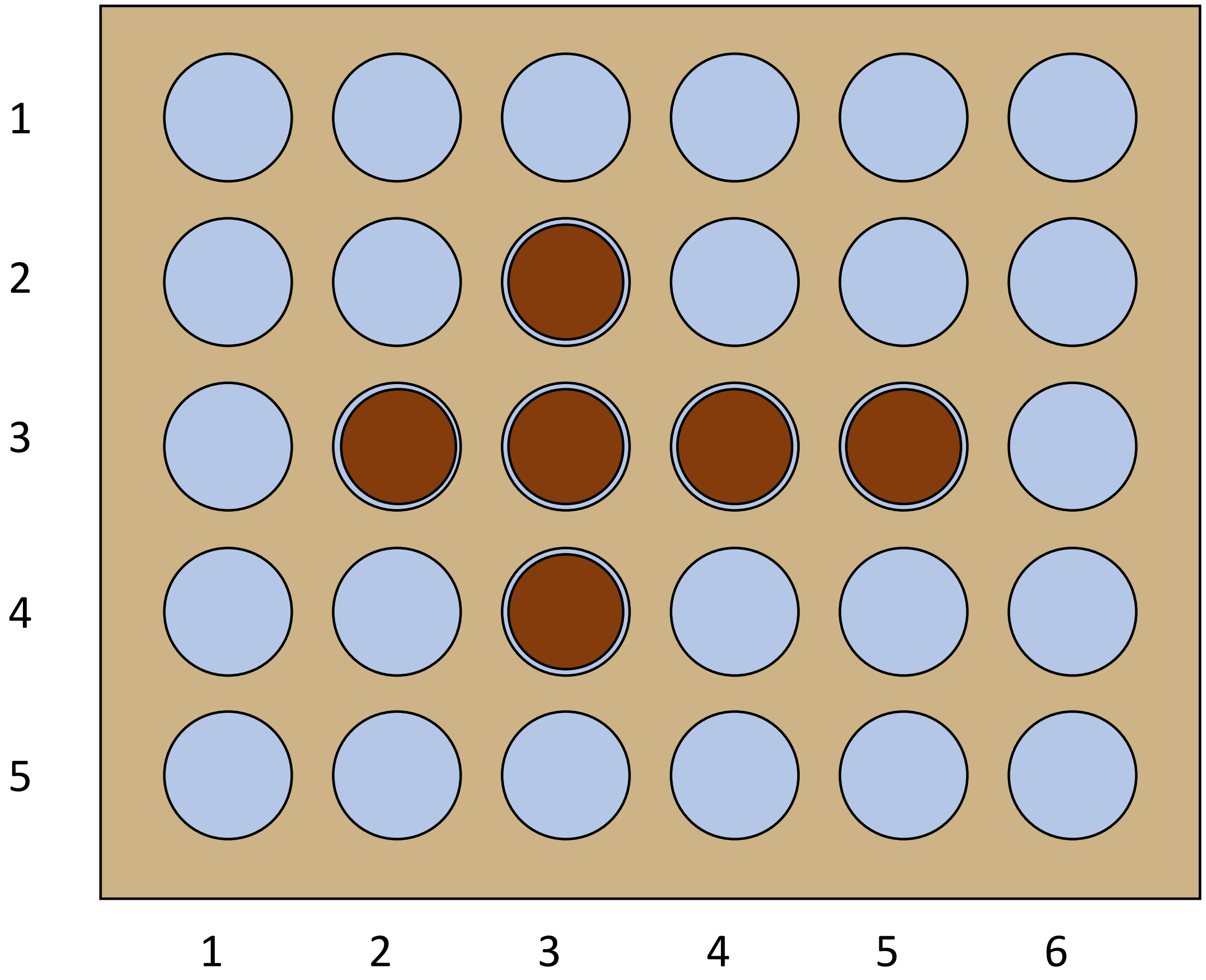
- Goal: leave as few pegs on the board as possible
- Jump pegs to remove them

Example: Peg Solitaire Game

- Goal: leave as few pegs on the board as possible
- Jump pegs to remove them

Can we estimate student strategic ability from a single game play record?

Example: Peg Solitaire Game



Example: Peg Solitaire Game

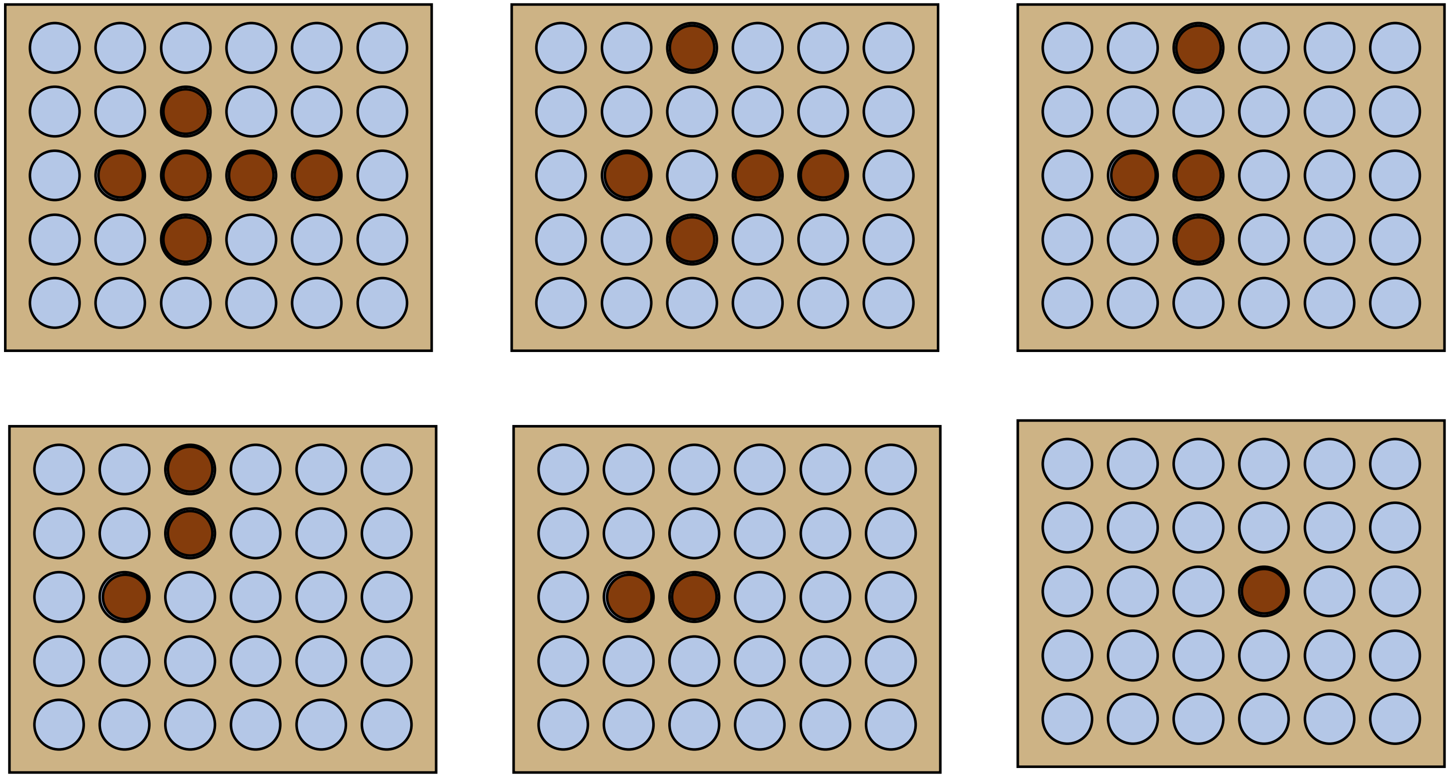
Process Data, Action Sequence:

$$(3,3) \longrightarrow (1,3)$$
$$(3,5) \longrightarrow (3,3)$$
$$(4,3) \longrightarrow (2,3)$$
$$(1,3) \longrightarrow (3,3)$$
$$(3,2) \longrightarrow (3,4)$$

Score

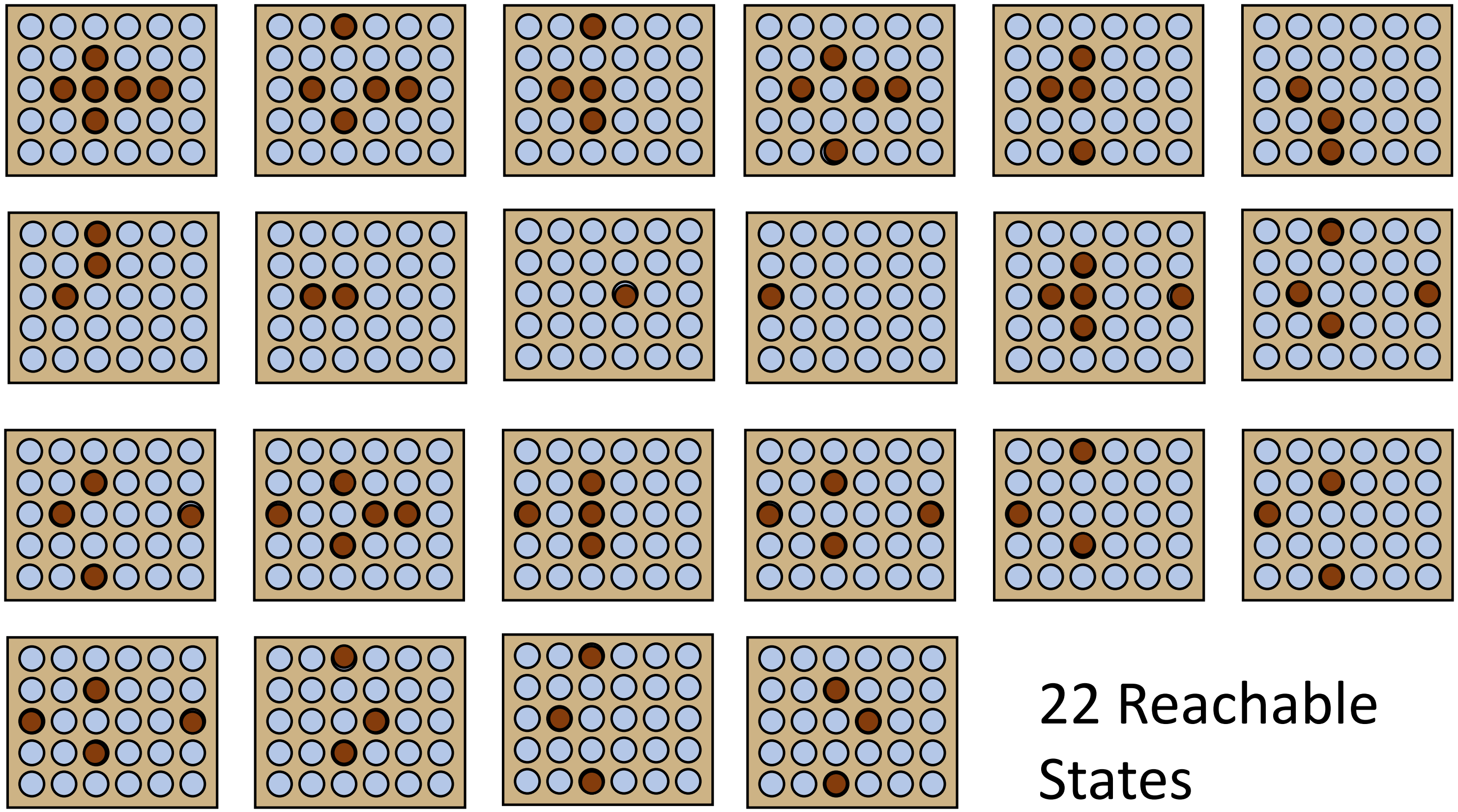
Example: Peg Solitaire Game

State Sequence



Example: Peg Solitaire Game

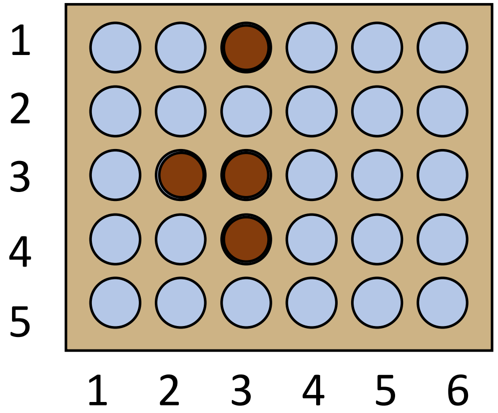
State Space



22 Reachable States

Example: Peg Solitaire Game

Each state presents a choice:



Available Actions A_s

$(3,3) \longrightarrow (1,3)$

$(3,3) \longrightarrow (5,3)$

$(3,2) \longrightarrow (3,4)$

$(4,3) \longrightarrow (2,3)$

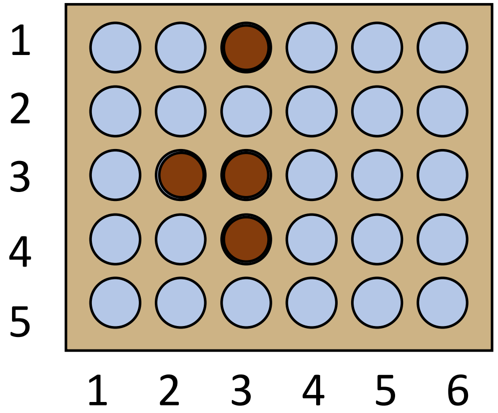
Score

Reset

Want $\quad p(a|s,\theta_j) = f(\boldsymbol{\theta_j}, \xi_s)$

Example: Peg Solitaire Game

Each state presents a choice:



Available Actions A_s

$(3,3) \longrightarrow (1,3)$

$(3,3) \longrightarrow (5,3)$

$(3,2) \longrightarrow (3,4)$

$(4,3) \longrightarrow (2,3)$

Score

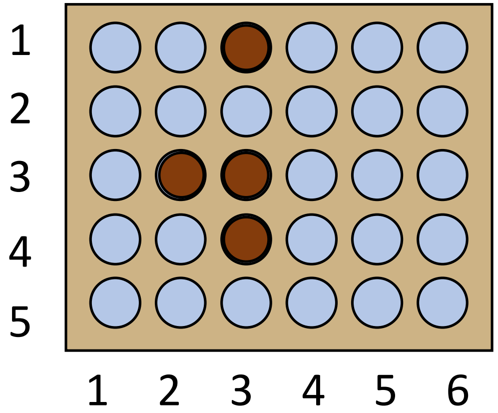
Reset

Want $p(a|s,\theta_j) = f(\boldsymbol{\theta_j}, \xi_s)$

Cognitive Process Model

Example: Peg Solitaire Game

Each state presents a choice:



Available Actions A_s

$(3,3) \longrightarrow (1,3)$

$(3,3) \longrightarrow (5,3)$

$(3,2) \longrightarrow (3,4)$

$(4,3) \longrightarrow (2,3)$

Score

Reset

Want $p(a|s,\theta_j) = f(\boldsymbol{\theta_j}, \xi_s)$ ⟸ **Markov Decision Process**

Markov Decision Process

- Model for sequential planning in the presence of uncertainty.
- Developed in the 1950s for process optimization in robotics (Bellman 1957).
- Recently used in cognitive science to model how we infer another person's motivations and beliefs (Baker, Saxe, Tennenbaum, 2009)

Markov Decision Process

State Space	$S = \{s_1, s_2, \dots s_{\bar{S}}\}$
Action Set	$A = \{a_1, a_2, \dots a_{\bar{A}}\}$
Transition Function	$T(s, a, s') = p(s' \mid s, a)$
Reward Structure	$R(s, a, s')$

Policy: $p(a \mid s, \xi)$

Markov Decision Process

$$p(a|s,\xi) = f\,(\textbf{The value of action a})$$

Markov Decision Process

$p(a|s,\xi) = f($**The value of action a**$)$

Expected Total Rewards

Markov Decision Process

$p(a|s,\xi) = f($**The value of action a**$)$

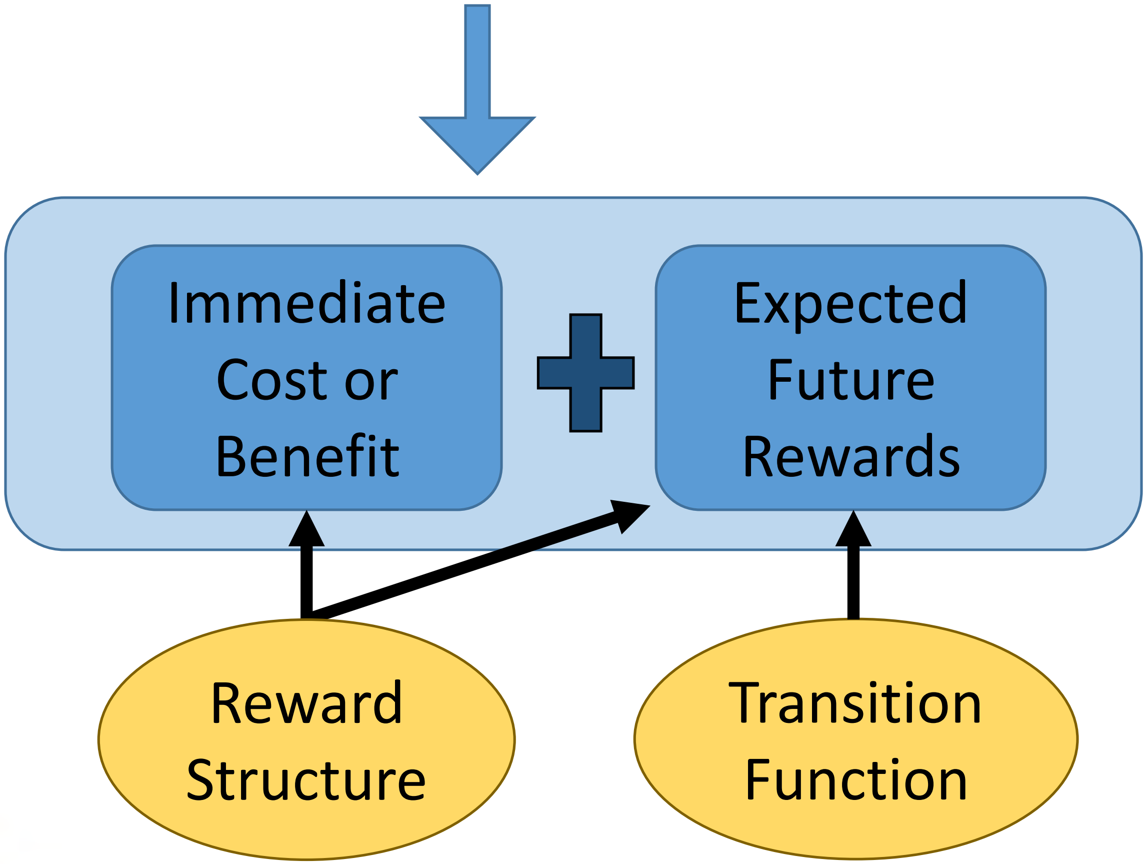
Immediate Cost or Benefit

+

Expected Future Rewards

Markov Decision Process

$p(a|s,\xi) = f($**The value of action a**$)$



Markov Decision Process

The expected rewards for taking action a in state s is expressed by the Q-function (Bellman, 1957):

$$Q(s,a) = \sum_{s' \in S} p(s'|s,a)\left(R(s,a,s') + \gamma \sum_{a' \in A} p(a'|s')Q(s',a')\right)$$

Markov Decision Process

The expected rewards for taking action a in state s is expressed by the Q-function (Bellman, 1957):

$$\underbrace{Q(s,a)}_{\substack{\text{Value of} \\ \text{choosing action} \\ a \text{ in state } s}} = \sum_{s' \in S} \underbrace{p(s'|s,a)}_{\substack{\text{Transition} \\ \text{probability}}} \left(\underbrace{R(s,a,s')}_{\substack{\text{Immediate} \\ \text{Reward}}} + \underbrace{\gamma \sum_{a' \in A} p(a'|s')Q(s',a')}_{\substack{\text{Discounted Expected} \\ \text{Future Reward}}} \right)$$

Decision Process

In robotics, solve for the optimal policy:

$$\pi(s) \equiv \underset{a \in A}{argmax}\big(Q^*(s,a)\big), \qquad p(a \in \pi(s)|s) = 1$$

In psychology, the Boltzmann policy is used

$$p\big(x_{sj} = a\big|s\big) \propto e^{\beta Q(s,a)}$$
$$\beta \in [0, \infty)$$

Consider β_j as a person-specific "capability"

$$p\big(x_{sj} = a\big|\beta_j, s\big) \propto e^{\beta_j Q\left(s,a|\beta_j\right)}$$

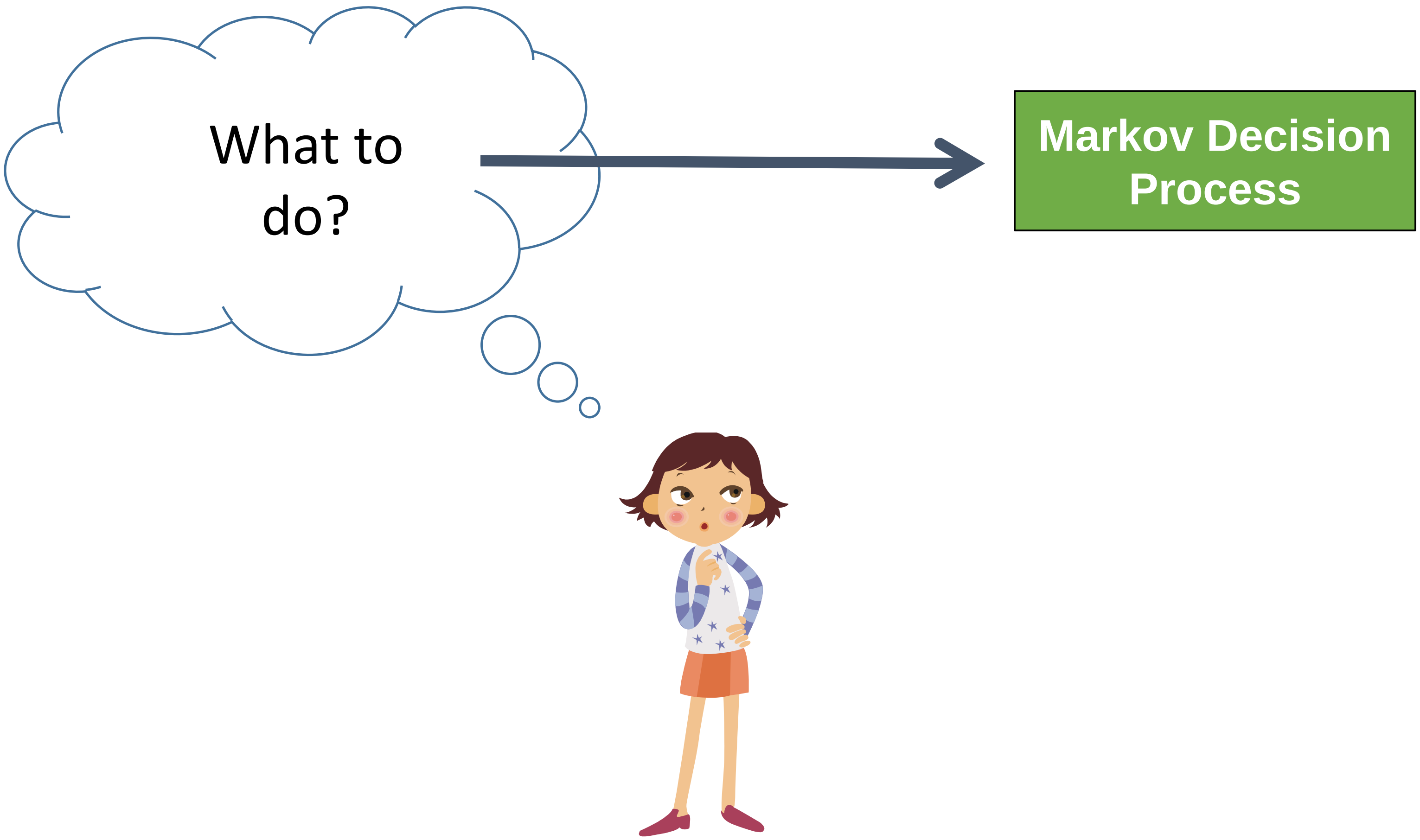
MDP as a Measurement Model

Full MDP Measurement model:

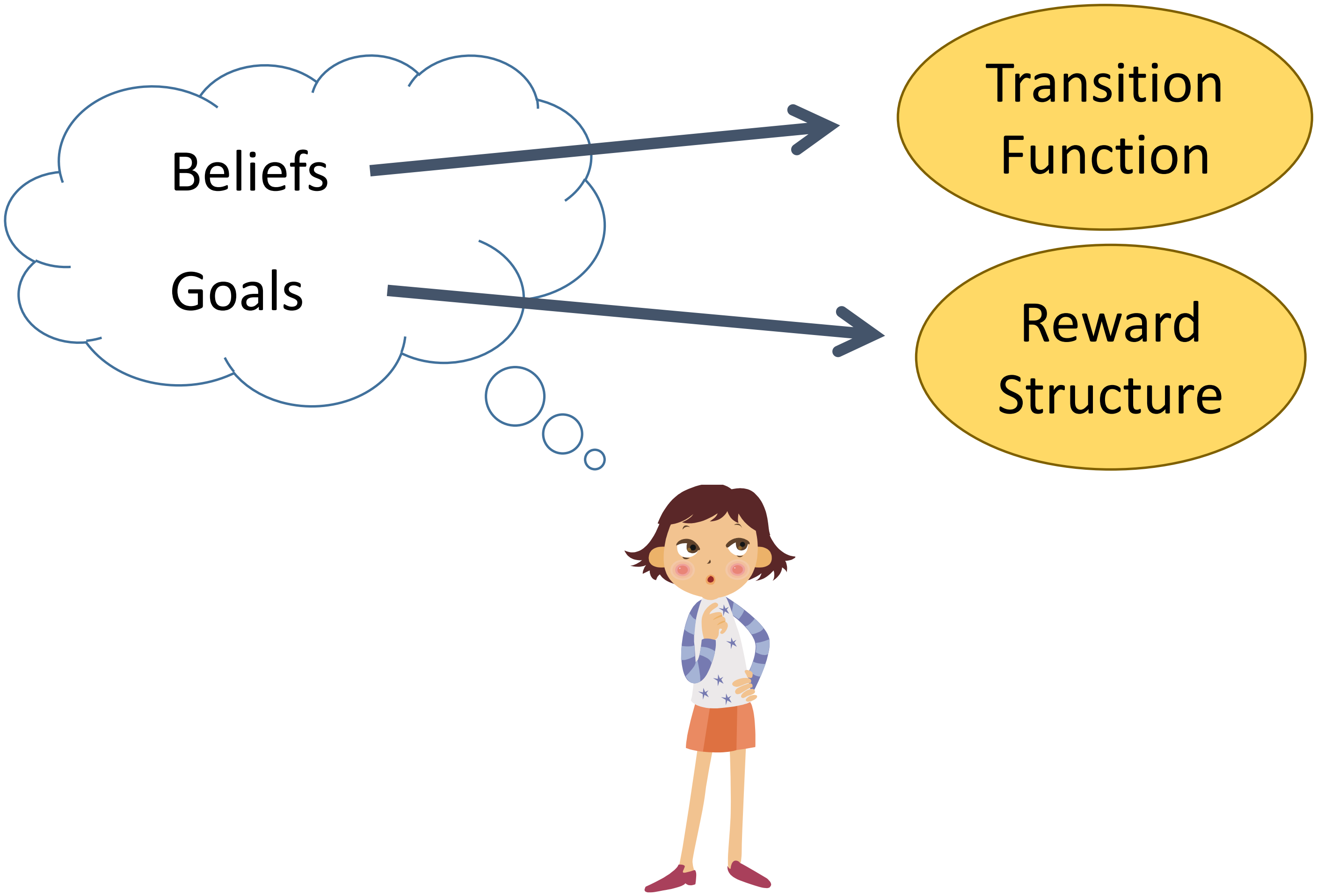
$$p\left(x_{sj} = a \middle| s, \beta_j\right) = \frac{\exp\left(Q(s, a|\beta_j)\beta_j\right)}{\sum_{a' \in A_s} \exp\left(Q\left(s, a'|\beta_j\right)\beta_j\right)}$$

$$\beta_j \sim ln\mathrm{N}(\mu, \sigma)$$

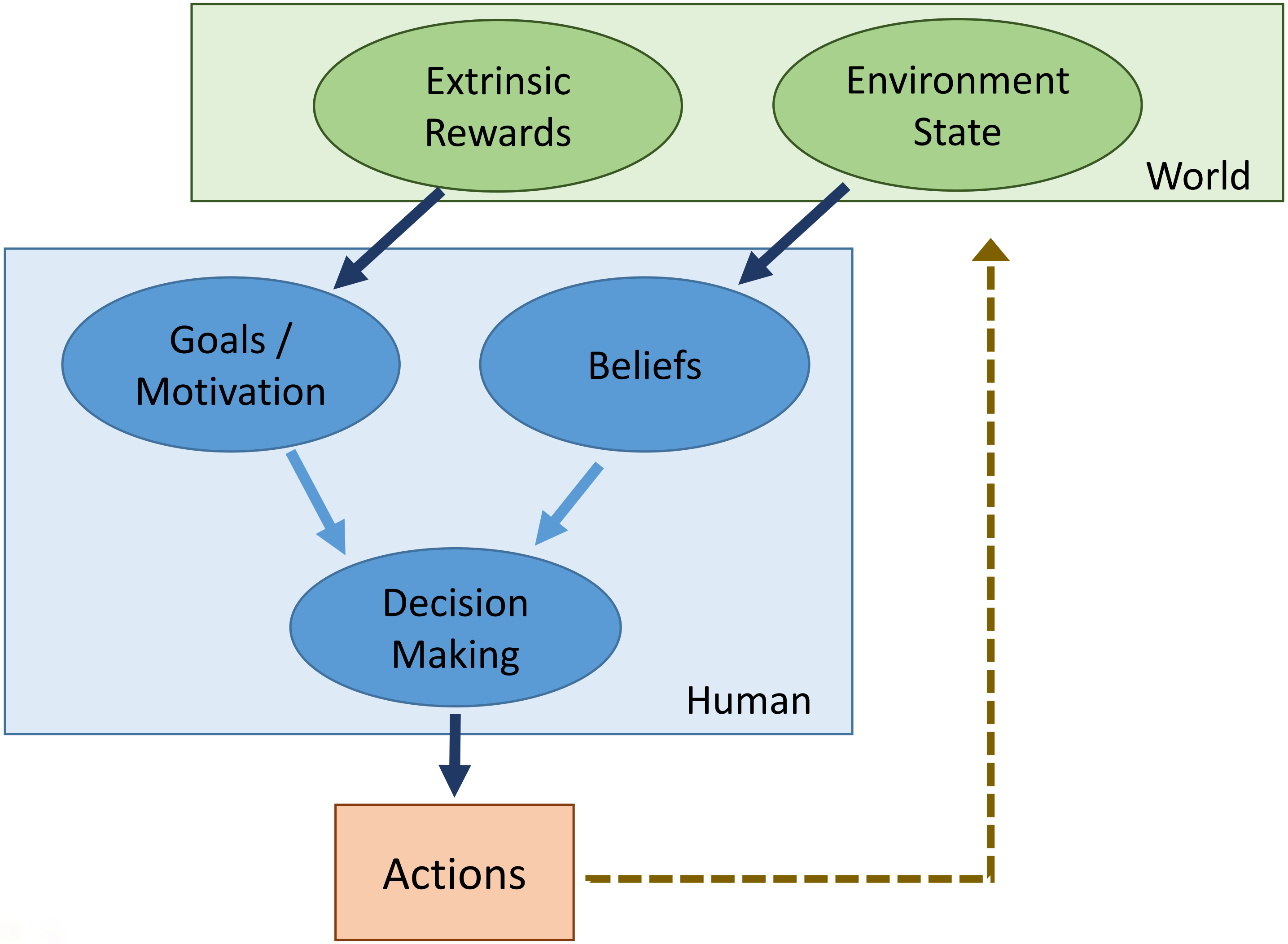
MDP as a Cognitive Model



MDP as a Cognitive Model

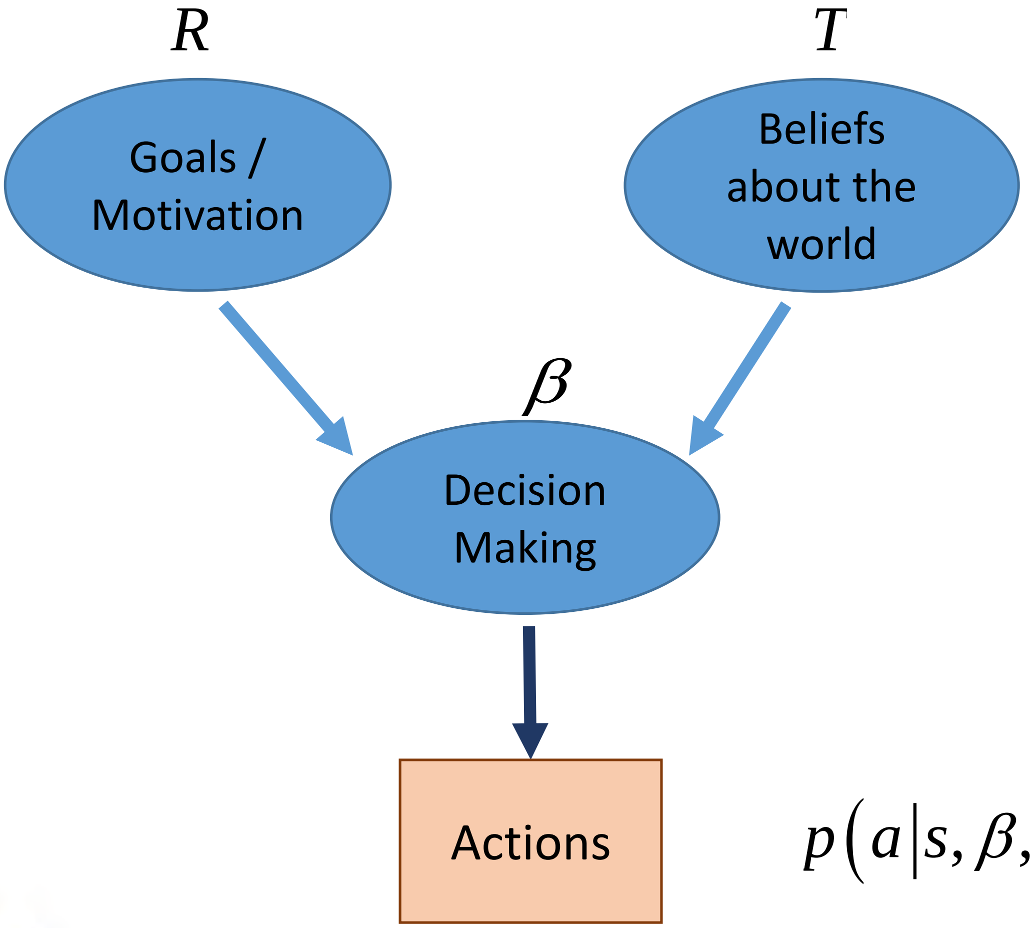


MDP as a Cognitive Model



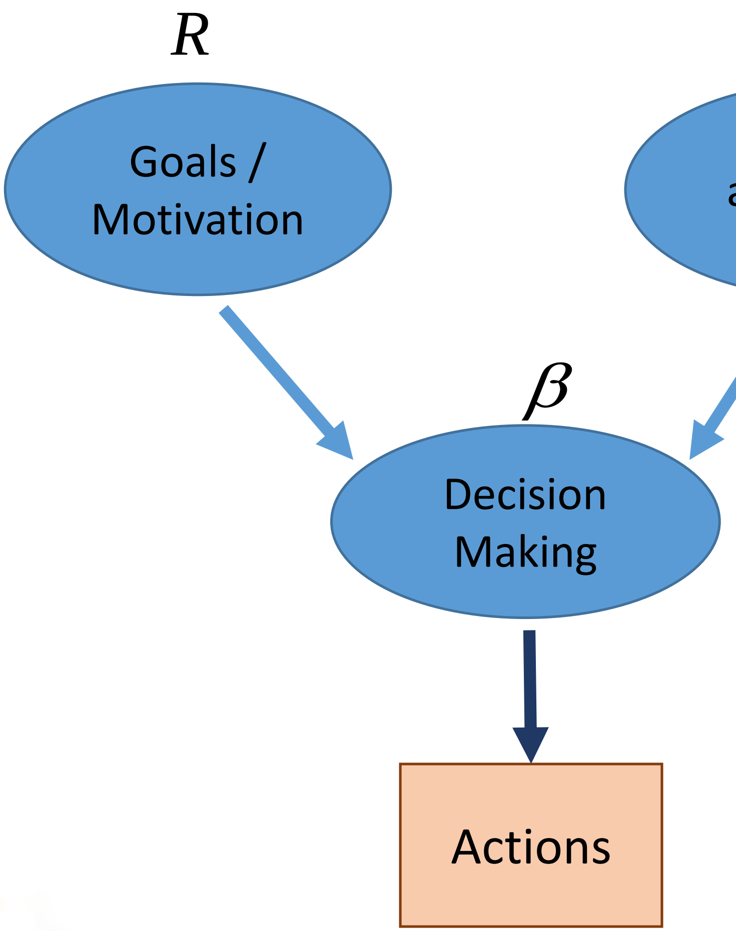
Adapted from: Baker, C., Saxe, R., & Tenenbaum, J. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the thirty-third annual conference of the cognitive science society* (p. 2469{2474).

MDP-MM Parameter Space



$$p\left(a \middle| s, \beta, R, T\right)$$

MDP-MM Parameter Space



Parameter Treatment	
Fixed	$\{5,1,\dots\}$
Population	R
Group	R_g
Person	R_j

$$p\left(a \middle| s, \beta, R, T\right)$$

MDP-MM Parameter Space

Transition Parameters

Totally Free

A1	S1	S2	S3	S4	S5	S6	S7	S8	…
S1	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	
S2	λ_{21}	λ_{22}	λ_{23}	λ_{24}	λ_{25}	λ_{26}	λ_{27}	λ_{28}	
S3	λ_{31}	λ_{32}	λ_{33}	λ_{34}	λ_{35}	λ_{36}	λ_{37}	λ_{38}	
S4	λ_{41}	λ_{42}	λ_{43}	λ_{44}	λ_{45}	λ_{46}	λ_{47}	λ_{48}	
S5	λ_{51}	λ_{52}	λ_{53}	λ_{54}	λ_{55}	λ_{56}	λ_{57}	λ_{58}	
S6	λ_{61}	λ_{62}	λ_{63}	λ_{64}	λ_{65}	λ_{66}	λ_{67}	λ_{68}	
S7	λ_{71}	λ_{72}	λ_{73}	λ_{74}	λ_{75}	λ_{76}	λ_{77}	λ_{78}	

MDP-MM Parameter Space

Transition Parameters

Fixed by Objective Reality

A1	S1	S2	S3	S4	S5	S6	S7	S8	…
S1	1	0	0	0	0	0	0	0	
S2	0	1	0	0	0	0	0	0	
S3	0	0	1	0	0	0	0	0	
S4	0	0	0	**0.1**	0	0	0	**0.9**	
S5	0	0	0	0	1	0	0	0	
S6	0	0	0	0	0	1	0	0	
S7	0	0	0	0	0	0	1	0	

MDP-MM Parameter Space

Transition Parameters

Targeted

A1	S1	S2	S3	S4	S5	S6	S7	S8	…
S1	1	0	0	0	0	0	0	0	
S2	0	1	0	0	0	0	0	0	
S3	0	0	1	0	0	0	0	0	
S4	0	0	0	λ_1	0	0	0	**1-λ_1**	
S5	0	0	0	0	1	0	0	0	
S6	0	0	0	0	0	1	0	0	
S7	0	0	0	0	0	0	1	0	

MDP-MM Parameter Space

Transition Parameters

Fixed by Misconception

A1	S1	S2	S3	S4	S5	S6	S7	S8	…
S1	1	0	0	0	0	0	0	0	
S2	0	1	0	0	0	0	0	0	
S3	0	0	1	0	0	0	0	0	
S4	0	0	0	**1**	0	0	0	0	
S5	0	0	0	0	**0.1**	0	0	**0.9**	
S6	0	0	0	0	0	1	0	0	
S7	0	0	0	0	0	0	1	0	

MDP-MM Parameter Space

Transition Parameters

Categorical by Belief:

$$T = \{H_1, H_2\}$$

H_1 → A1 may work in S4
H_2 → A1 may work in S5

MDP-MM Estimation

We use marginal maximum likelihood (MML) to estimate the population and group level parameters

$$L(\xi|O) = \prod_{j=1}^{N}\prod_{t=1}^{T_j} p\left(a_{jt}\middle|s_{jt},\xi\right)$$

$$L(\mu,\sigma|O) = \int \prod_{t=1}^{T_j} \frac{\exp\left(Q\left(s_t,a_t|\beta_j\right)\beta_j\right)}{\sum_{a'\in A}\exp\left(Q\left(s_t,a'|\beta_j\right)\beta_j\right)} P\left(\beta_j|\mu,\sigma^2\right)d\beta_j\ ,$$

$$\beta_j \sim ln\mathrm{N}(\mu,\sigma^2)$$

And MLE to estimate the person level parameters.

MDP-MM Estimation

Q-Function is recursive – must be solved using dynamic programming.

$$Q(s,a) = \sum_{s' \in S} p(s'|s,a)\left(R(s,a,s') + \gamma \sum_{a' \in A} p(a'|s')Q(s',a')\right)$$

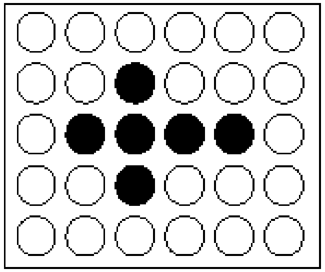
MDP-MM in Action

Peg Solitare

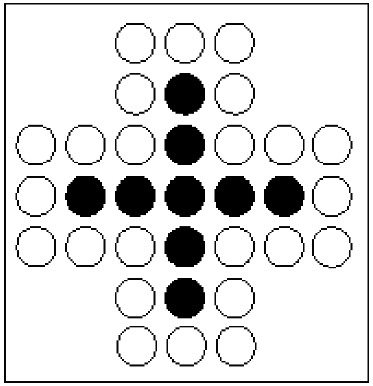
Peg Solitaire Simulation Studies

Game boards with varying complexity

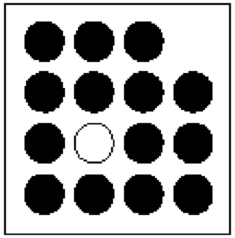
Board	Path Length	Move Actions	Reachable States
Tiny Cross	5	12	22
Big Cross	8	22	153
Big-L	13	30	807
Diamond	11	70	5923



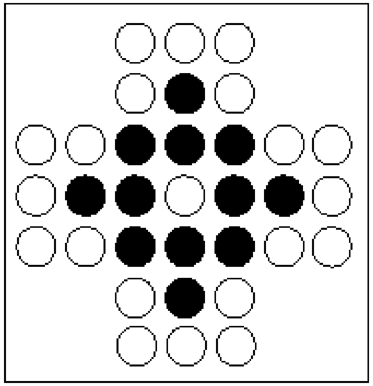
Tiny Cross



Big Cross



Big L



Diamond

Peg Solitaire Parameters

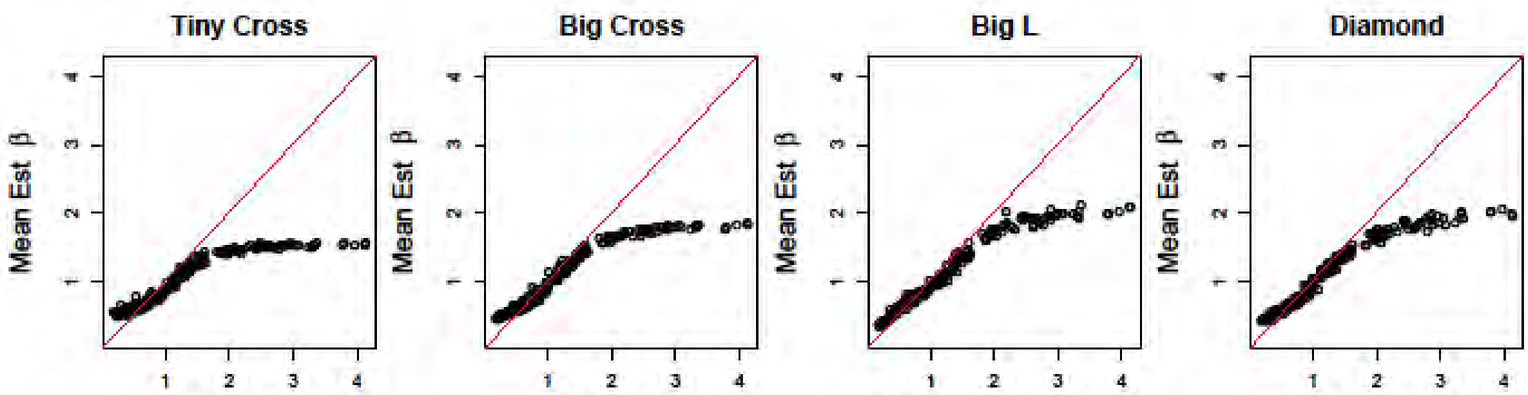
No Transition Parameters.

Capability parameters: β_j, μ, σ

Rewards:

Parameter	Function	Example Value	
R_{win}	Reward for scoring with one peg left	5.0	Fixed
R_{peg}	Add to reward for each extra peg	-1.0	Fixed
R_{move}	Cost of a move	-0.1	Est.
R_{reset}	Cost of reset	-1.0	Est.

Estimating Capability



Board	Ceiling Thresh.	Students Remaining	β_j Bias	β_j RMSE
Tiny Cross	2.03	0.80	-0.064	0.395
Big Cross	2.33	0.84	-0.036	0.362
Big-L	2.62	0.88	-0.072	0.365
Diamond	2.28	0.84	-0.045	0.327

Estimating Capability & Motivation

At the population level.

200 students/group. 25 games/student/board.

Sample	Capability	Motivation	μ	σ	R_{move}
1	High	High	0.5	0.75	-0.05
2	High	Low	0.5	0.75	-0.75
3	Low	High	-0.5	0.75	-0.05
4	Low	Low	-0.5	0.75	-0.75

Estimating Capability & Motivation

At the population level.

200 students/group. 25 games/student/board.

Sample	Capability	Motivation	μ	σ	R_{move}
1	High	High	0.5	0.75	-0.05
2	High	Low	0.5	0.75	-0.75
3	Low	High	-0.5	0.75	-0.05
4	Low	Low	-0.5	0.75	-0.75

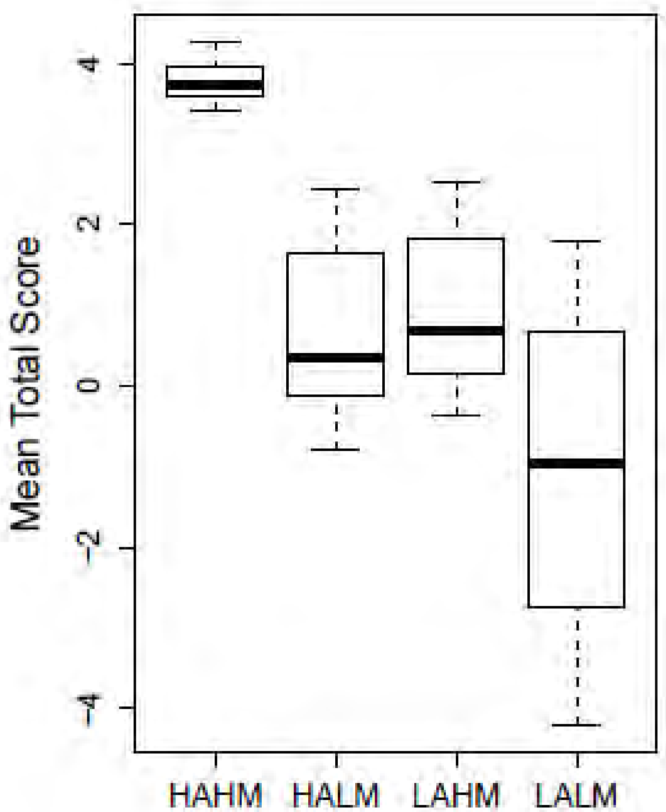
Estimating Capability & Motivation

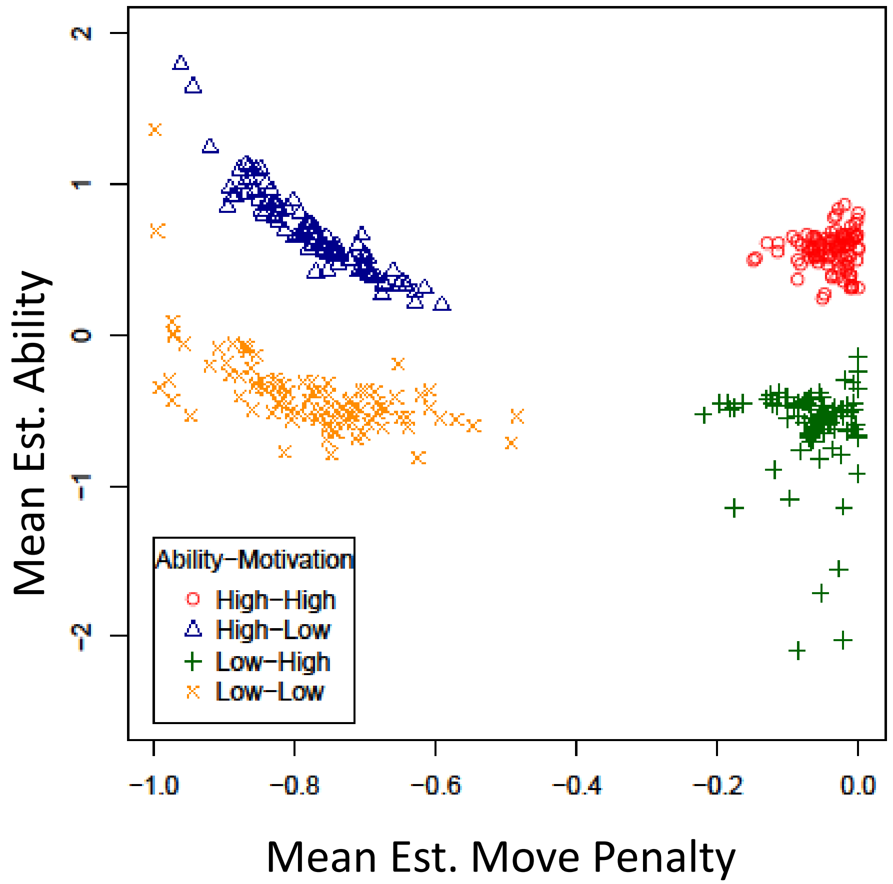
At the population level.

200 students/group. 25 games/student/board.

Sample	Capability	Motivation	μ	σ	R_{move}
1	High	High	0.5	0.75	-0.05
2	High	Low	0.5	0.75	-0.75
3	Low	High	-0.5	0.75	-0.05
4	Low	Low	-0.5	0.75	-0.75

Estimating Capability & Motivation





MDP-MM in Action

PBS-Kids Microbes

Application: Microbes

N = 238 (148)

Application: Microbes

N = 238 (148)

MDP Model for Microbes

6 Game Levels. Each modeled as a separate MDP

State Space	State Variables: • Microbe Config = 484 States • Win History
Action Set	Buy Mito, Buy Chloro Play Level, Stop
Rewards	Win, Lose, Buy
Transitions	Play => $\begin{cases} win & p(win \mid s, a = play) \\ lose & 1 - p(win \mid s, a = play) \end{cases}$

Estimating Capability

- Transition parameters are fixed.
- Rewards either fixed or estimated at the population level.

	Post-test Correlations	AIC
MDP-MM Fixed R	0.507	15465
MDP-MM Est R	0.516	11243
IRT First Try	0.317	
IRT Multi-try PC	0.379	

The estimates for β_j from the MDP models correlated better with the posttest than the IRT estimates for θ_j.

Microbes Transition Parameters

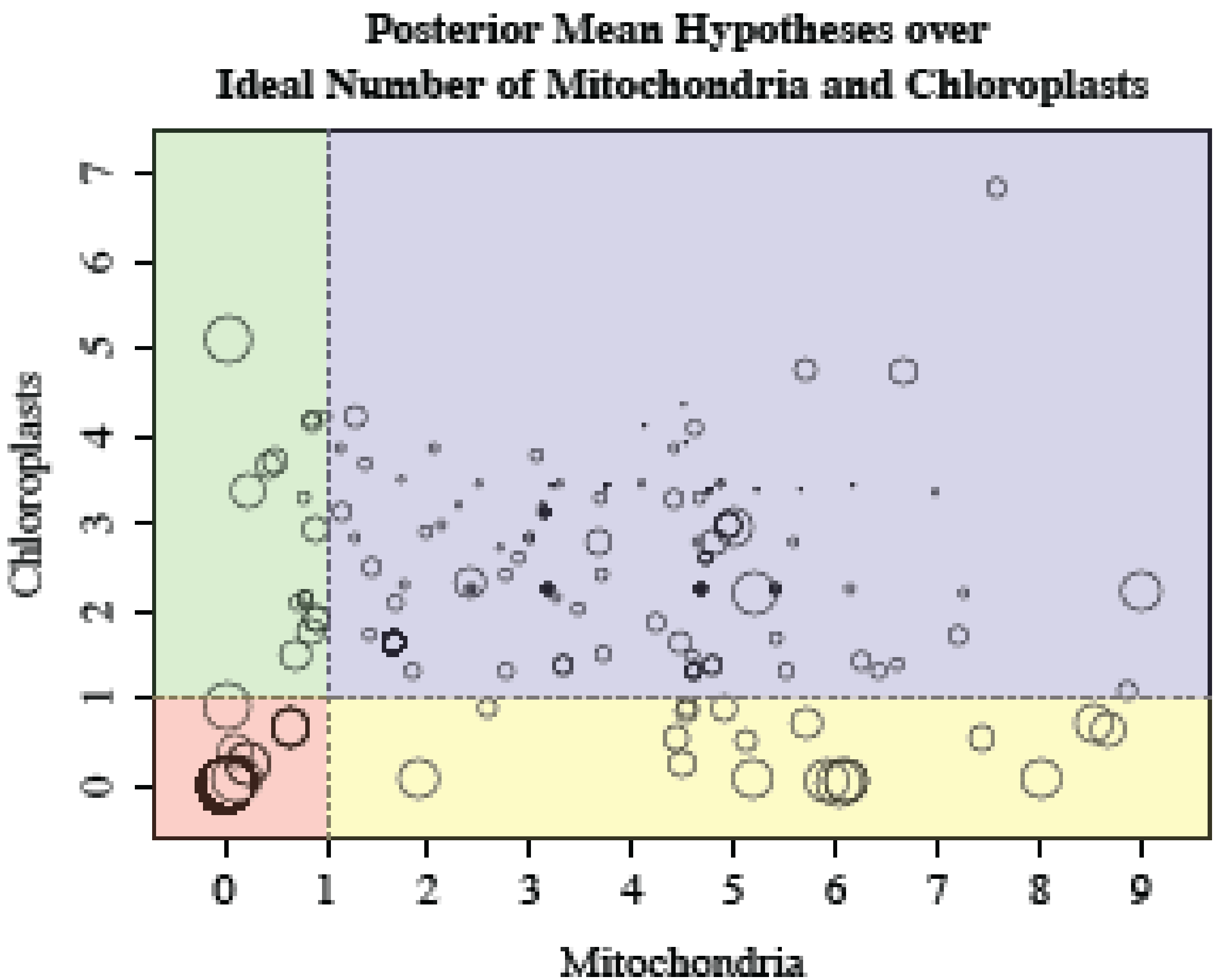
Play $\rightarrow \begin{cases} win & p(win|s, a = play) \\ lose & 1 - p(win|s, a = play) \end{cases}$

To get at student beliefs, assume each student has an ideal microbe configuration.

c_j = student j's ideal # of chloroplasts
m_j = student j's ideal # of mitochondria

$$\max(p(win|s, a = play)) = p(win|s = \{c_j, m_j\}, a = play)$$

Estimating Beliefs/Understanding



Median posttest ability estimates:

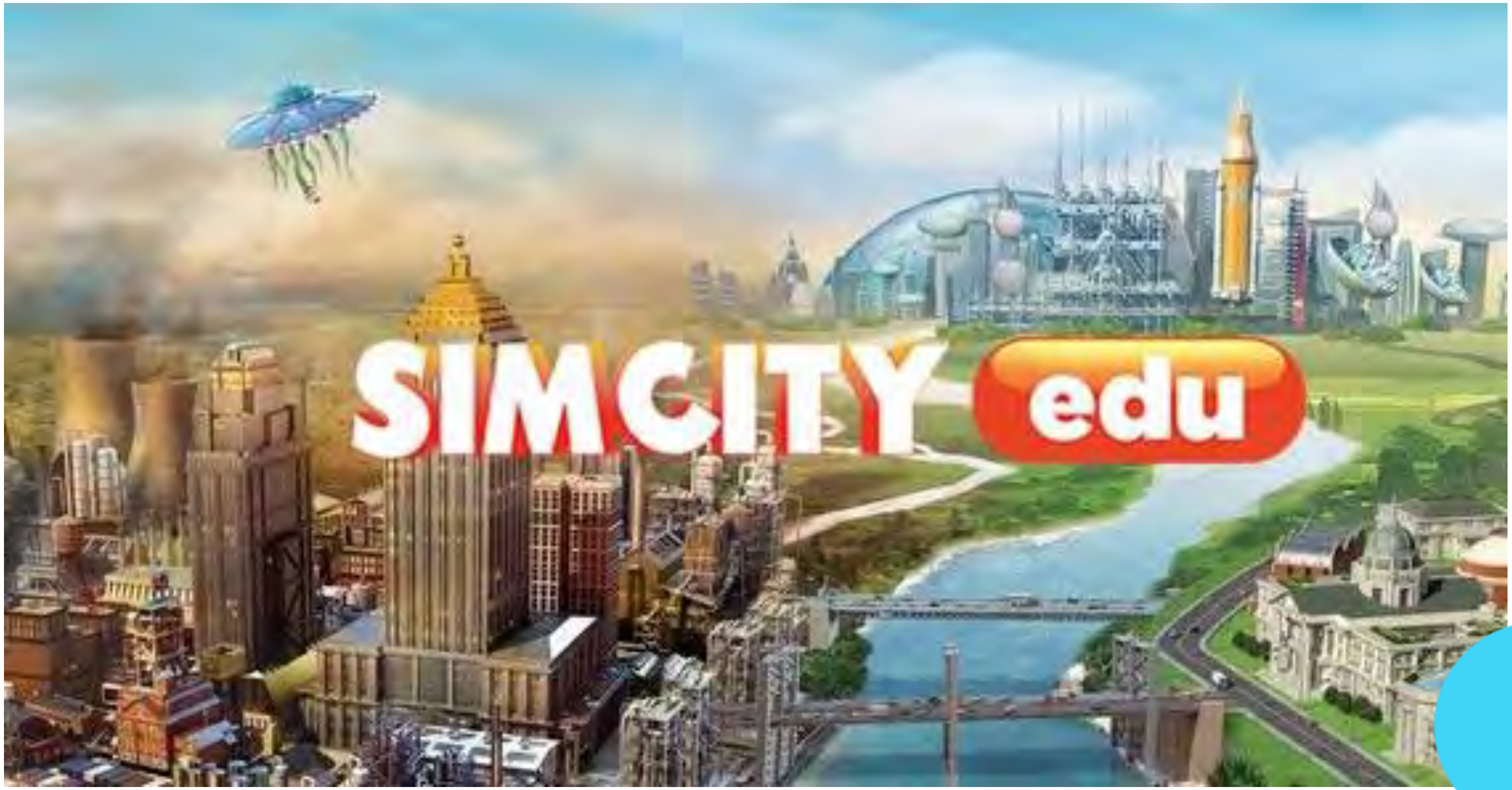
-0.18	-0.36	-0.54	0.33

(Rafferty et al., 2015)

MDP-MM in Action

SimCityEDU Pollution Challenge

Application: SimCityEDU

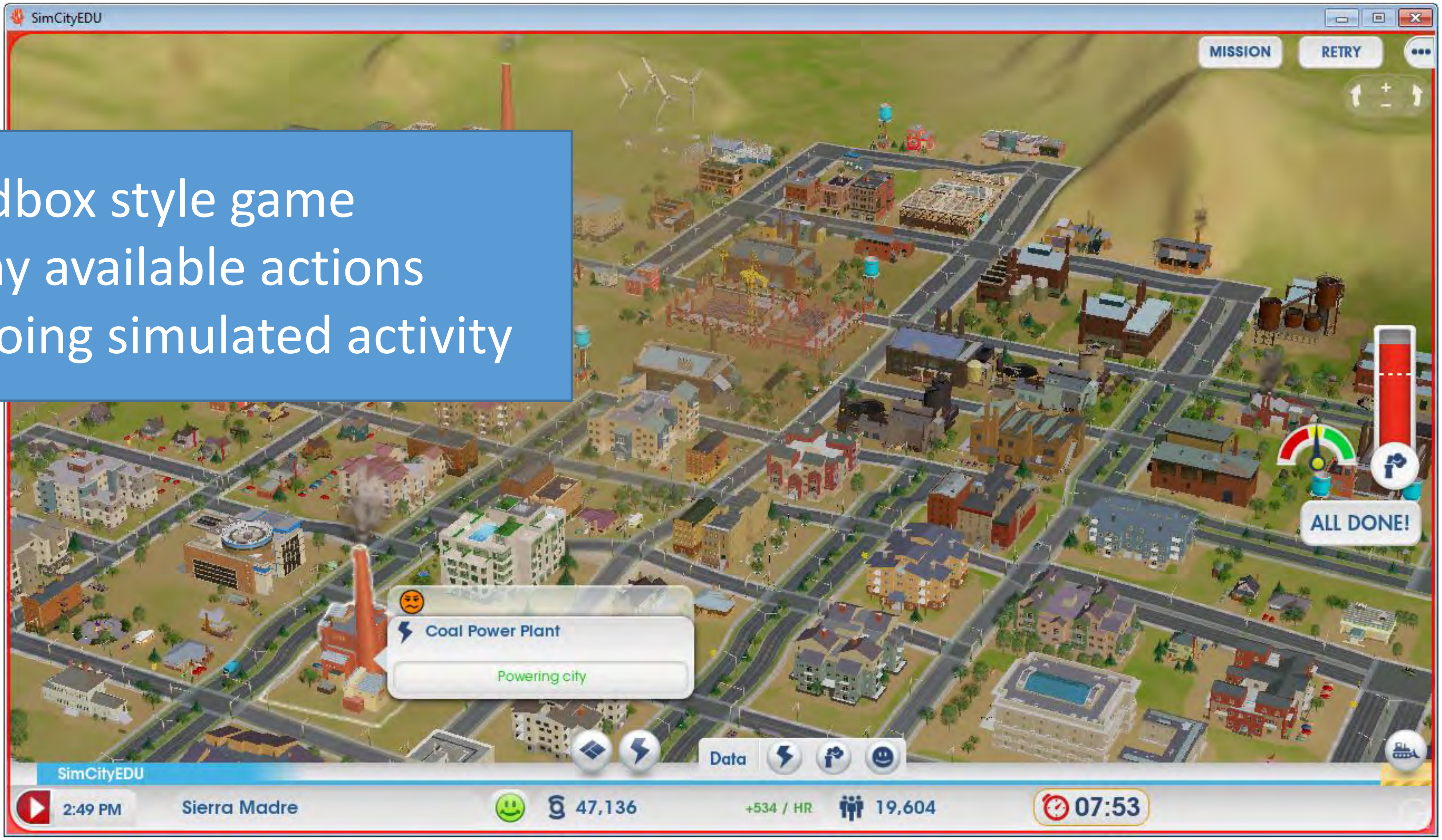


N = 224

GlassLab

SimCity EDU

- Sandbox style game
- Many available actions
- Ongoing simulated activity

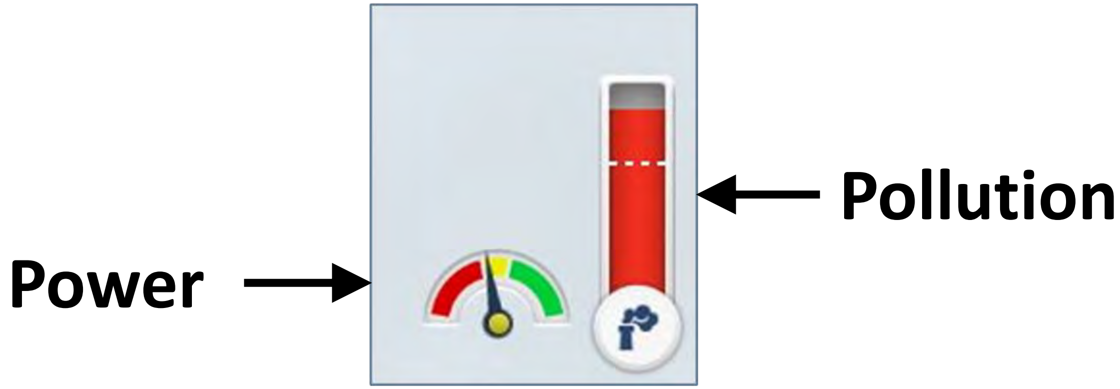


SimCity EDU


Mission 4: Pollution Problems
Welcome to Sierra Madre!
This city has a problem with high air pollution. Find out how to lower pollution and keep the power level optimal.
You have 10 minutes. Can you avoid power blackouts?
NAVIGATION
GO!

SimCity Assessment

- Designed to assess Systems Thinking
- Students must optimize two variables simultaneously



SimCity MDP

Action set is huge

- Follow Sim named Joe Smith
- View Apartment Building Status
- Upgrade Garbage Dump
- Build Large Solar Power Plant
- Build Small Solar Power Plant
- Dezone Commercial (23,45)
- Zone as Residential (302,82)
- Bulldoze the Smith's House
- Turn off Coal Plant
- Expand School
- Build Statue at City Hall …

State space is huge

Game State includes

- Location and status of
  - Every Sim
  - Every Building
- Time of day
- City funds ($$)
- Severity and location of Pollution
- Wind direction and speed …

Need to trim down to important subset!

SimCity Actions

	Wind	Solar	Coal
Build			
Turn On			
Turn Off			
Upgrade			
Bulldoze			

+ Wait
End Mission
= 17 actions

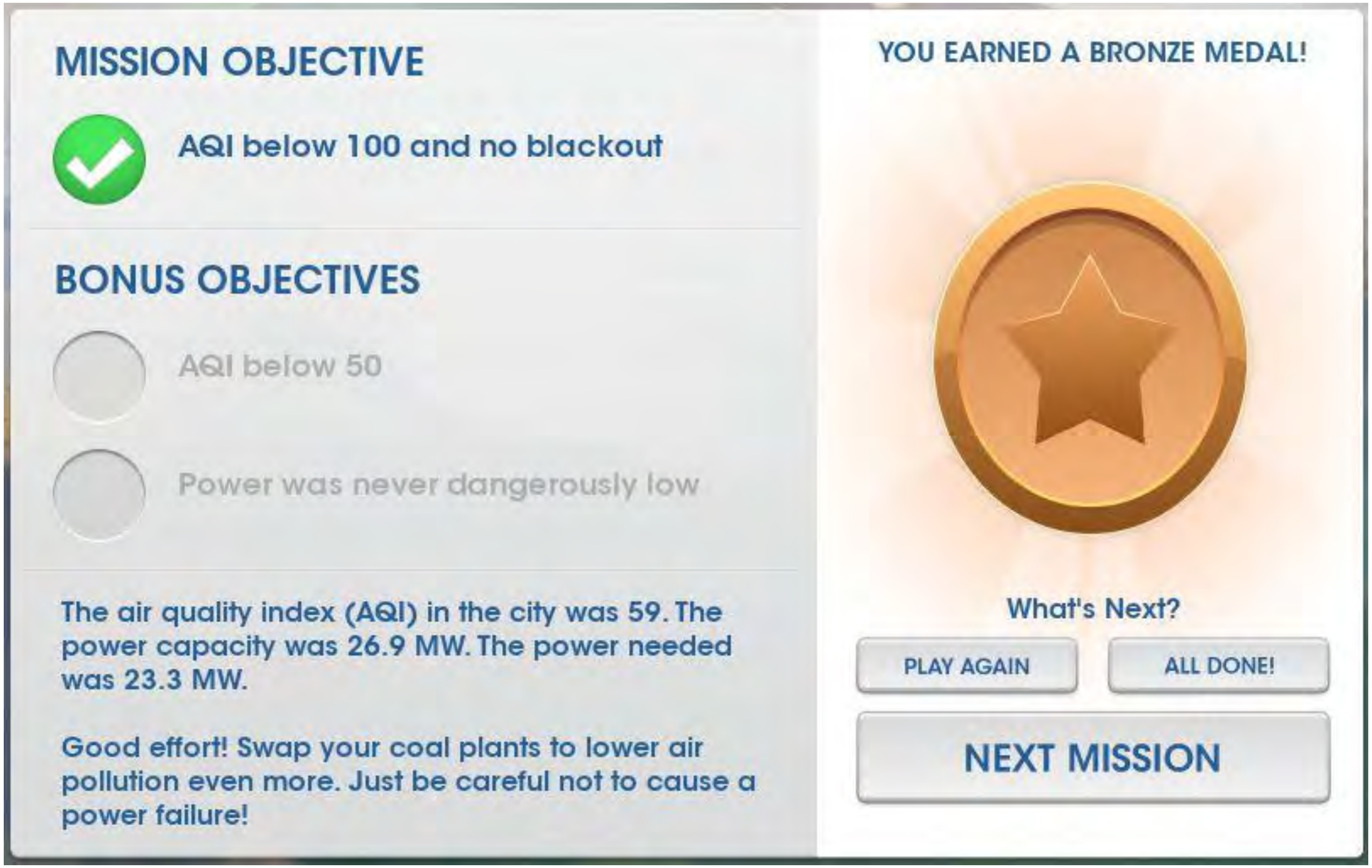
SimCity State Space

	Min	Max	# Values
# Coal Generators On	0	3	4
# Coal Generators Off	0	3	4
# Wind Turbines	0	10	11
# Solar Panels	0	2	3
Power Balance	-8	7	16
Pollution	0	3	4
Cash	0	30	31

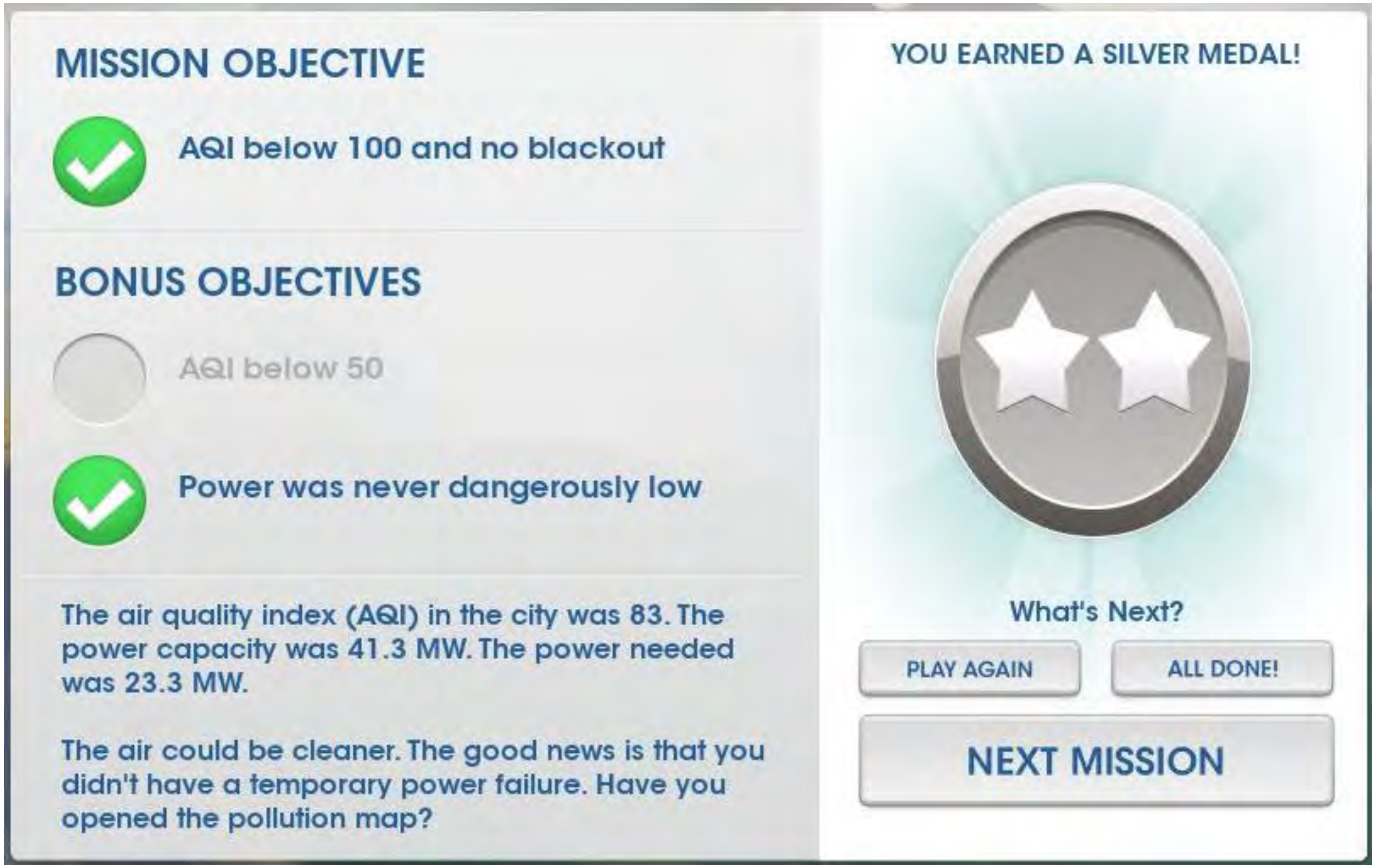
Total # of States: 2,856,960

But only 25,420 reachable states

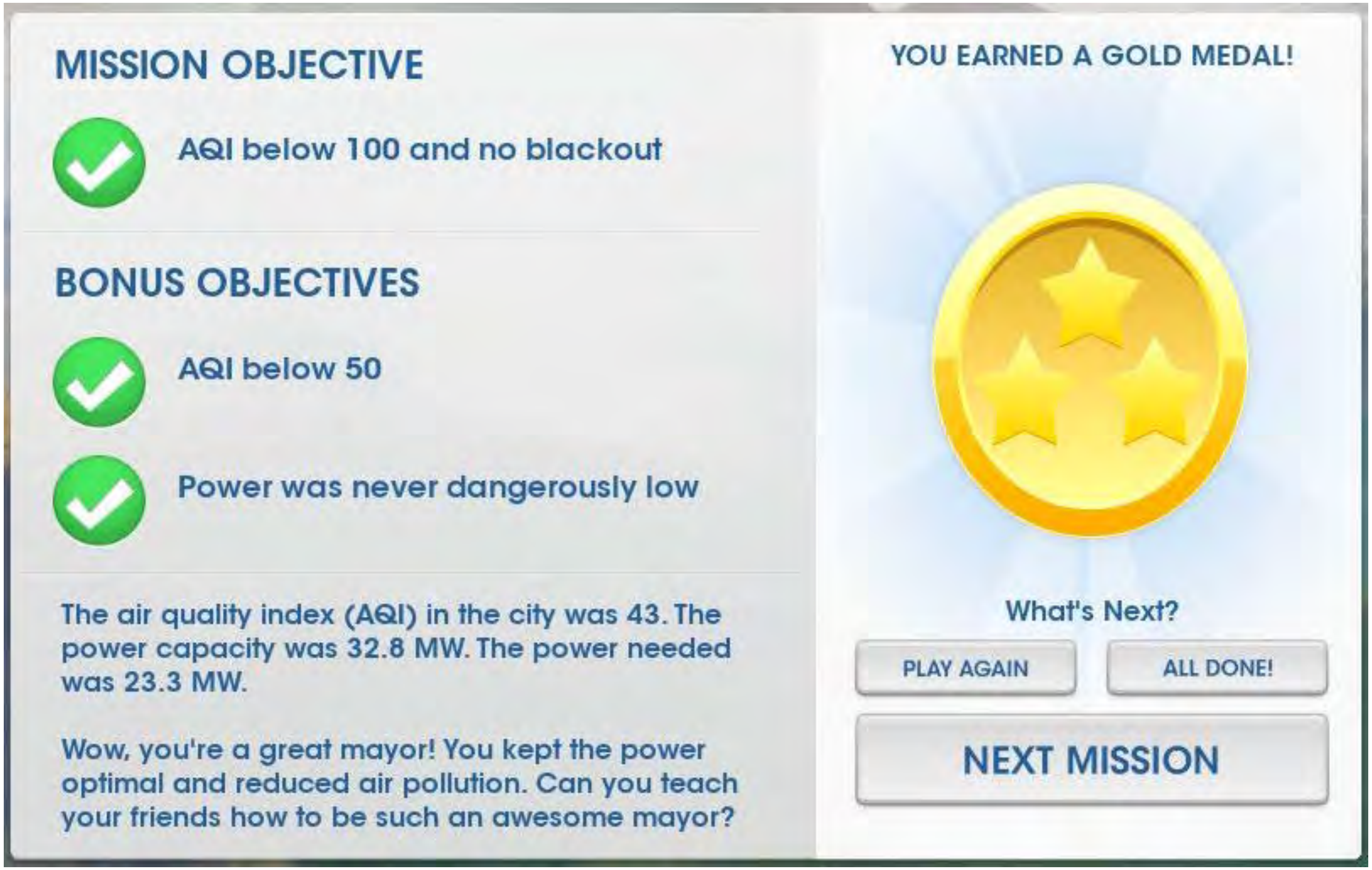
SimCity Rewards



SimCity Rewards



SimCity Rewards



SimCity Rewards

Goal Hypotheses

	Bronze Medal	Pollution Silver	Power Silver	Gold Medal
Medals	+5	+5	+5	+10
Just Win	+10	0	0	0
Pollution	+5	+5	0	0
Power	+5	0	+5	0

Estimating Goals

	Log-likelihood	Num Students Classified
Medals	-17565.1	38
Just Win	-17974.5	28
Pollution	-17974.6	38
Power	-17915.0	24

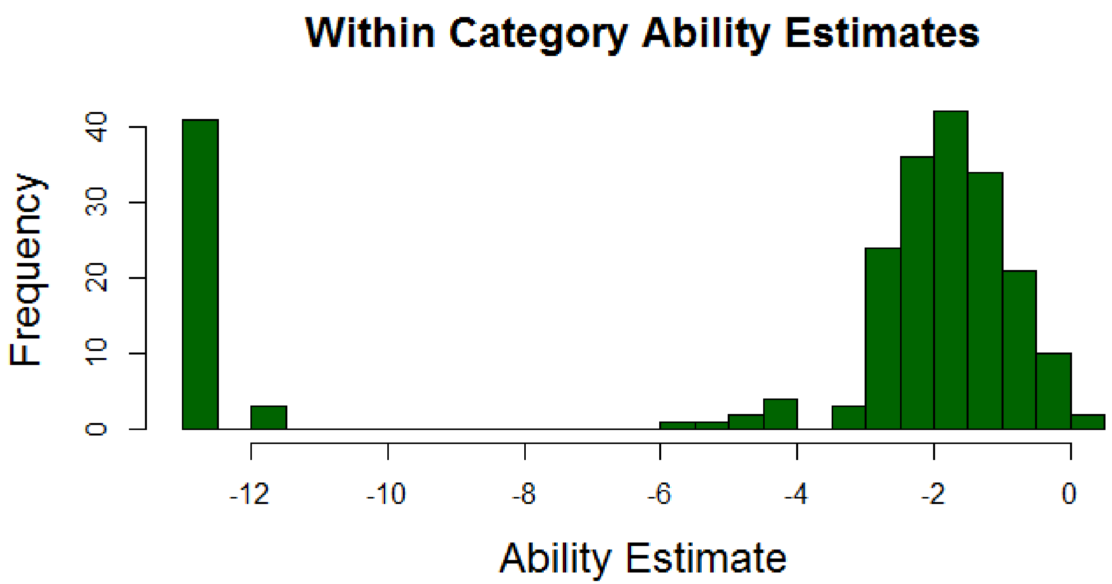
Estimating Goals

	Log-likelihood	Num Students Classified
Medals	-17565.1	38
Just Win	-17974.5	28
Pollution	-17974.6	38
Power	-17915.0	24

53 Students who fit none – Posteriors were flat

Overall Sufficiency of SC MDP

Strong implication that our model is over simplified for many of the students.



Consider expanding model:

- Zoning actions?
- Sim Happiness as a goal?

Conclusions

Markov Decision Process Measurement Model

- Potential as a flexible framework for assessment
  - Estimate general ability from task process data
  - Separate student motivation, system understanding and strategic ability

- Sensitive to specification of cognitive processes

Conclusions

- Early work; much yet to do
  - Improve algorithms & estimation
  - Gather more validity evidence
  - Partially Observable MDP (POMDP)

Just one example of Cognitive Process Models for assessment

Center for Research on Computational Psychometrics

Other work:

- Multi-modal analytics: evidence from stream data
  - Emotion detection
  - Gestures, posture, and actions
  - Voice tone and fluency
- Assessing Collaboration
  - Collaborative Assessment Frame
  - Collaborative Dialog Analysis
  - Social Network Models
  - Hawkes Process Models

This work was made possible by …

Colleagues

Anna Rafferty
Tom Griffiths
Sophia Rabe-Hesketh
Mark Wilson
Matt Silverglitt
Malcolm Bauer
Alina von Davier
Bob Mislevy

Data Provided By

WestEd

RED HILL STUDIOS

GlassLab

Using Markov Decision Processes to Understand Student Thinking in Performance Tasks

Michelle M. LaMar
Educational Testing Service
mlamar@ets.org

October 1st, 2015

References

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 113(3), 329–349.

Bellman, Richard 1957. *Dynamic Programming*. Princeton University Press.

Rafferty, A., Brunskill, E., Griffiths, T., & Shafto, P. (2011). Faster teaching by POMDP planning. *Artificial Intelligence in Education* (pp. 280–287).

Rafferty, A., LaMar, M., & Griffiths, T. (2015). Inferring learners' knowledge from their actions. *Cognitive Science,* 39 (3), 584-618.