

# POMDP applications

Jason Williams  
AT&T Labs - Research

# What POMDPs solve

- Problems with...
  - Hidden state
  - Sequential decision making
  - Uncertain action effects
  - Objective can be cast as a reward function
- Especially useful at valuing actions which gather information vs. "do" something
- POMDPs are good solutions to applications with these problems

# POMDP Lineage

*EJ Sondik. **1978**. The Optimal Control of Partially Observable Markov Decision Processes over the Infinite Horizon: Discounted Costs. Operations Research, Vol 26, No. 2.*

From abstract: "The paper closes with a detailed example illustrating the application of the algorithm to **the two-state partially observable Markov process**."

Although POMDPs have a long tradition of theoretical robustness, applications have long been "toy" in nature.

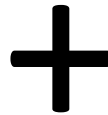
# What is an application?

## Domain

Example domains:

- Assistive care
- Navigation
- Network management
- Avoiding being eaten by a tiger while trying to find some gold

Reasonably clear  
to me



## Realism

"Levels" of realism:

- Imagined
- Sort-of-real
- End-to-end system

Mostly vague to me;  
need to clarify

# What is realism?

How is problem presented?

Model of dynamics (T, O, S) given

Environment given

Invented

Measured

Simulated

Real

How is evaluation done?

Average reward,  
as reported by  
solver

—

—

X

X

Average reward,  
in generative  
simulation

R1

X

X

Average reward,  
interacting with  
environment

X

X

R2

R3

Measuring an  
extrinsic quantity  
(not reward)

—

—

Average reward, as reported by solver	—	—	X	X
Average reward, in generative simulation	R1		X	X
Average reward, interacting with environment	X	X	R2	R3
Measuring an extrinsic quantity (not reward)	—	—		

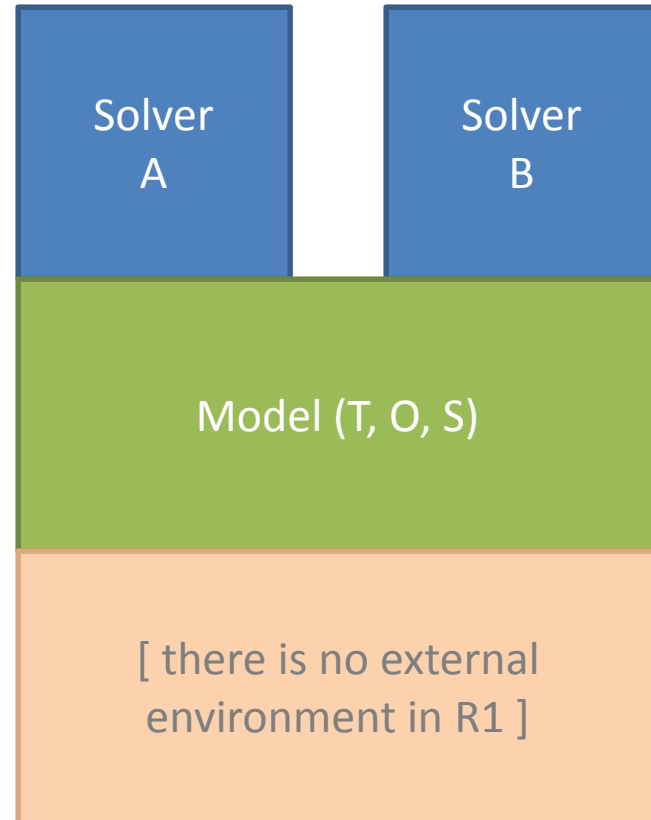
# R1

## Pros

- Facilitates comparing algorithms
- Easy to share problems (e.g., Cassandra format)
- Helps clarify which solvers are suited to which problems
- Tradition which assumes models are correct ("Markov" is in our name)

## Cons

- Doesn't test mismatch between model and world ("All models are wrong")
- Reward is a proxy for something (Got what I asked for vs. got what I wanted)



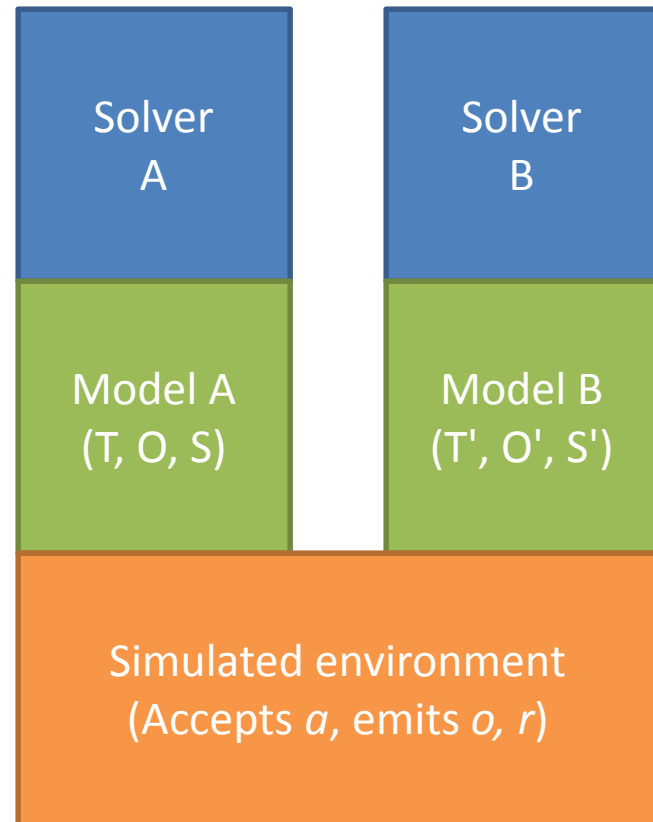
# R2

## Pros

- Still easy to share problems
- Facilitates comparisons to other RL techniques

## Cons

- Coupling between Model and Solver makes comparisons between solvers harder. E.g., changing "S" may re-order effectiveness of algorithms
- Is the simulated environment convincing? Not easy!



# Environment simulations

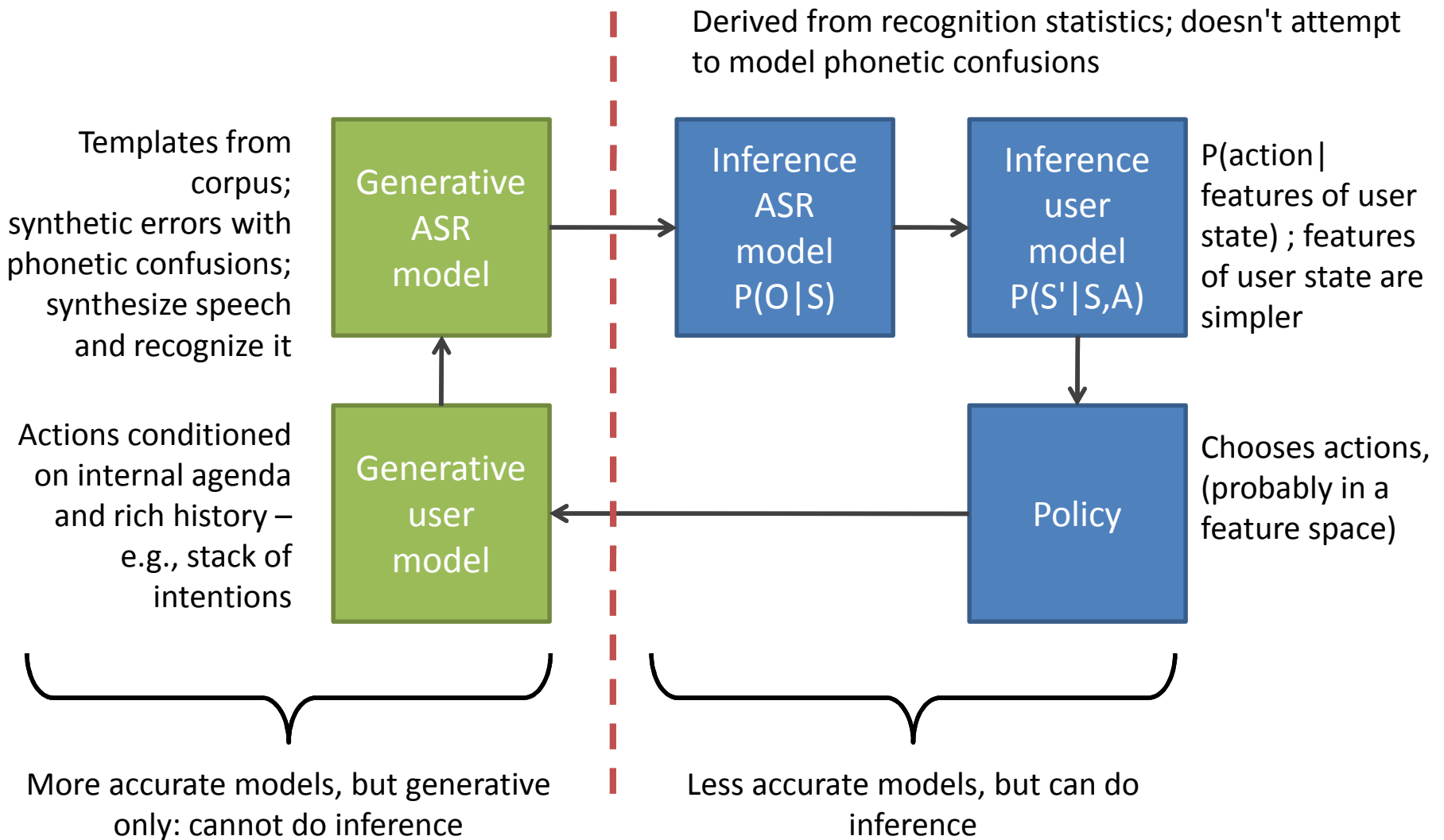
"We derive stochastic policies from simulations on a simple planar model. We then run the policy for the stochastic model in a high-fidelity dynamics simulation and measure average total reward per episode. **Note that the stochastic model and the high-fidelity model differ in some substantial details: the dimensions of the block and the geometry of the fingers are different and the actual sensor and detailed control behavior are different. Therefore, some of the trajectories that are most common in the high-fidelity simulation have relatively low probability in the stochastic model.** These simulations gives us a measure of how much the mis-estimation of the probabilities in the stochastic model decreases performance."

*Kaijen Hsiao, Leslie Pack Kaelbling and Tomas Lozano-Perez, "Grasping POMDPs," IEEE Conference on Robotics and Automation, 2007.*

# Dialog system example

## Simulated environment

## POMDP model



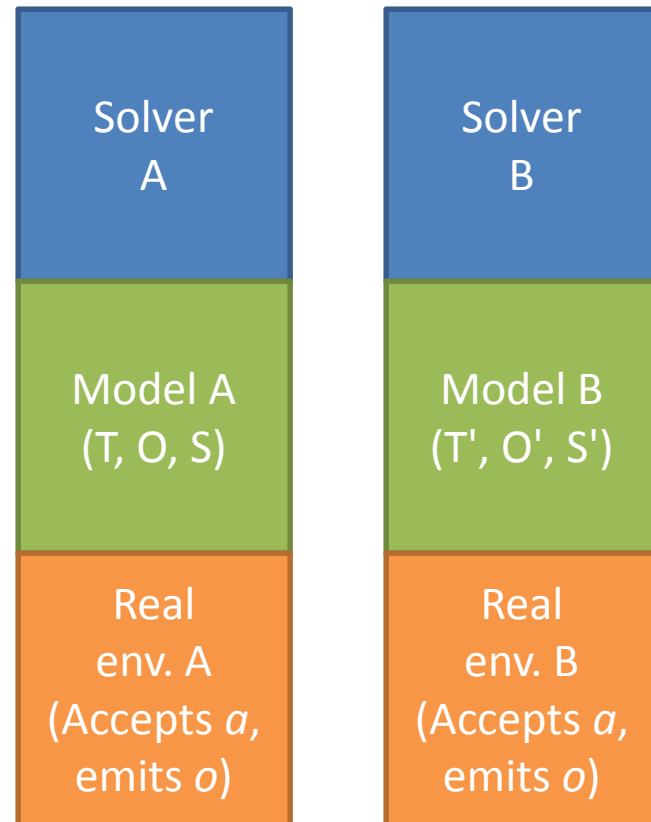
# R3

## Pros

- Faithful use of what we think POMDPs are good for

## Cons

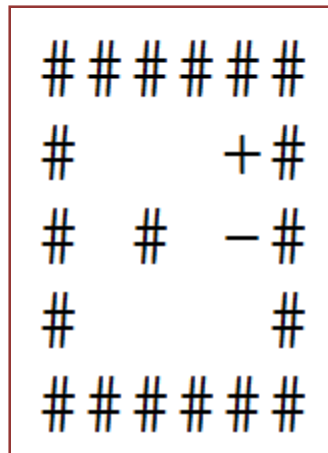
- In general, the real environment can't be gzipped (e.g., patient with dementia).
- Tons of free variables – very hard to do "identical" experiments
- Even in a controlled setting, need to do statistical analysis to compare algorithms properly
- Expensive; system-building



# R1 Applications: Domains

"Benchmark problems" a la Tony Cassandra's POMDP page

- Navigation, spatial goal-finding, mazes, avoiding death/plundering treasure
  - Tiger, Hallway, Hallway2
- Information gathering with a noisy channel
  - RockSample, TinyTravel (Spoken Dialog System)
- Network operation
  - Slotted aloha
- Manufacturing



```
states:          none-bad one-bad two-bad  
actions:         Manufacture Examine Inspect Replace  
observations:    good defective
```

# Robotics (R2)

State	Noisy sensors; local readings BUT reliable internal map
Action	Motors are imprecise; trajectories are stochastic
Temporal	Shortest path may involve stopping for sensor readings
Reward	Goal state; get there fast but don't spill my coffee

## Grasping with a robotic arm

Evaluated in a high-fidelity dynamics simulation

*Kaijen Hsiao, Leslie Pack Kaelbling and Tomas Lozano-Perez, "Grasping POMDPs," IEEE Conference on Robotics and Automation, 2007.*

## Navigating an office

Simulation with continuously-valued positions & real camera images

*M. T. J. Spaan and N. Vlassis. Perseus: Randomized point-based value iteration for POMDPs. Journal of Artificial Intelligence Research, 24:195-220, 2005.*

# Assistive care/home care (R3)

State	Elderly abilities; Image processing; speech recognition + robotics
Action	People don't/can't understand/comply
Temporal	Long-term goals; sequence-dependent steps
Reward	Task accomplishment vs. speed

## Handwashing

Evaluated with actors mimicking dementia patients.

*Jesse Hoey, Axel von Bertoldi, Pascal Poupart, and Alex Mihailidis. Assisting Persons with Dementia during Handwashing Using a Partially Observable Markov Decision Process. Proc ICVS, Biefeld, Germany, 2007.*

## Pearl – retirement home assistant

Trial in a retirement home.

*J. Pineau, M. Montemerlo, M. Pollack, N. Roy, & S. Thrun "Towards robotic assistants in nursing homes: Challenges and results". Special issue on Socially Interactive Robots, Robotics and Autonomous Systems 42 (3-4). 2003.*

# Spoken dialog systems (R2-R3)

State	User's intentions corrupted by speech recognition errors
Action	User behavior is non-deterministic
Temporal	When to confirm vs. when to commit (e.g., bill user for ticket)
Reward	Task accomplishment vs. speed; user frustration; call center cost

## Comparing effects of user simulations on POMDP dialog policies (R2)

Interesting non-symmetric effects for training vs. evaluation

*Dongho Kim, Hyeong Seop Sim, Kee-Eung Kim, Jin Hyung Kim, Hyunjeong Kim, Joo Won Sung, Effects of User Modeling on POMDP-based Dialogue Systems, Proceedings of Interspeech, 2008.*

## Comparing POMDP vs. MDP dialog strategies (R3)

Tested on recruited subjects in a working system.

*B. Thomson, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, K. Yu, S. Young(2008). "User study of the Bayesian Update of Dialogue State approach to dialogue management." Interspeech 2008, Brisbane, Australia.*

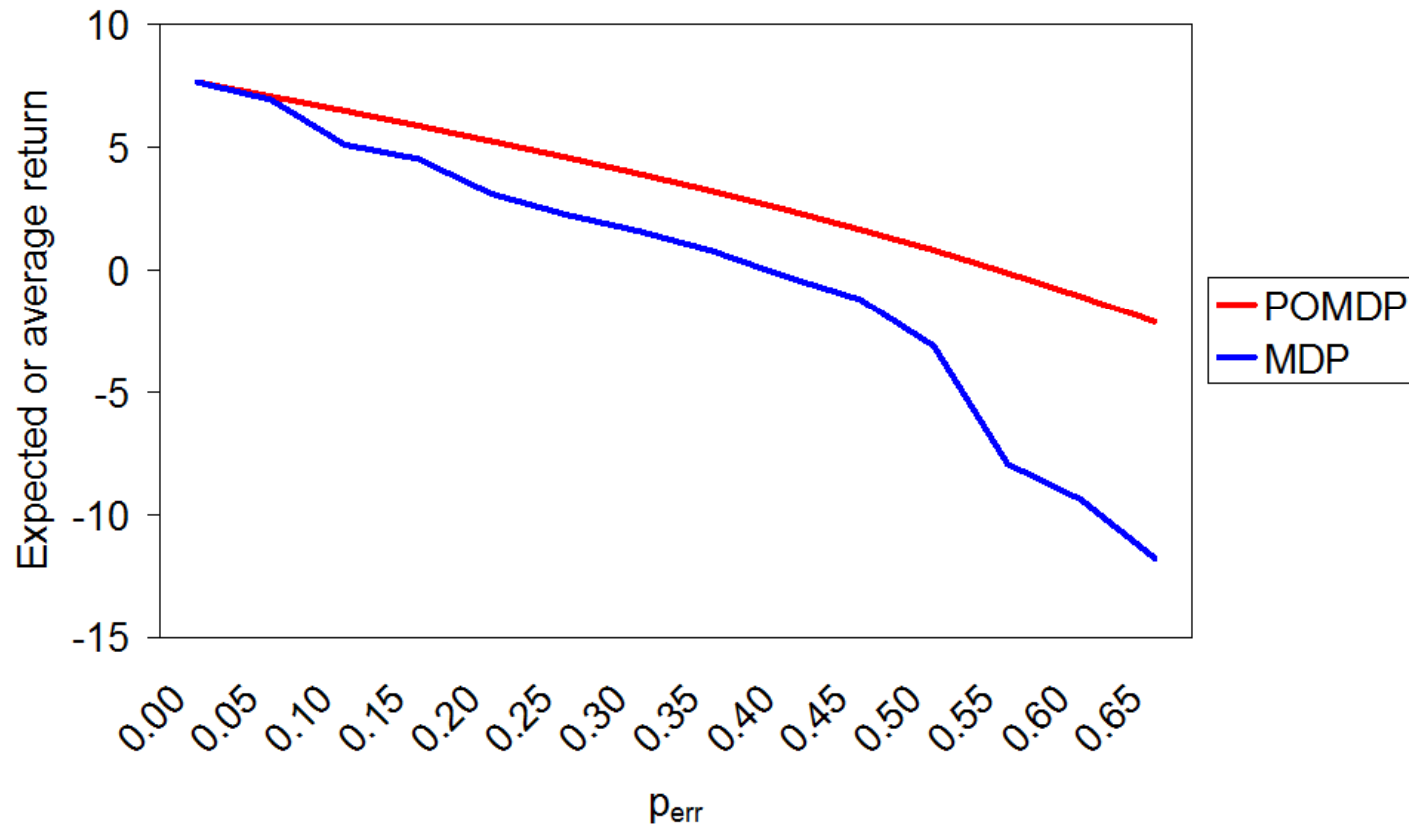
# One customer's story (mine)

## Chapter 1: Researcher meets POMDP, researcher falls in love

1. 2003: Realized POMDPs were appealing for dialog control
2. 2003: Constructed a toy (R1) problem
  - 1945 states, ~20 observations & actions
3. 2004: Figured out how to optimize it
4. 2005: Cautious but positive reception

# 2005

## Results: POMDP vs. MDP



## Chapter 2: Try to apply to real world; things get complicated

1. 2005: Realized planning needs to be done in a compressed belief space
2. 2007: Saw that exploration may never find a good policy, suggesting the benefit of merging expert knowledge and optimization
3. 2008: Discrete, observable features points to planning in a general "feature space" rather than a belief space

# 2008

Mozilla Firefox  
http://attda.research.att.com/pomdpDialer/cgi-bin/showSession.pl?tn=9733608138

## POMDP Dialer : call from 9733608138

Previous system action	Belief State	State Features
First and last name?	<b>Remaining mass</b> [33 partition(s)]	Best name Best phone type Phones available: one Name confirmed?: no Name is ambiguous?: no
<b>Recognition result</b> 35 james mullins james meyers jason williams james lentz jason raines kay simmons gene simmons james watts james lynch jason yates james mullins usa jason wint james meyers usa jason henson jason williams usa [ 16 more N-Best entries not shown]	<b>Remaining mass</b> jason wing redditch, eng (igbr) james mullins auburn, ca (usa) jay choi florham_park, nj (usa) jason wint middletown, nj (usa) jason li columbia, md (usa) jason lee silver_spring, md (usa) jason lee denver, co (usa) patricia renz morristown, nj (usa) patricia moore dallas, tx (usa) jason yates alpharetta, ga (usa)	Allowed Actions AskName Sorry, first and last name? ConfirmName jason wing. $\{(\hat{a}_m, a_m)\}$ Action Search Values at point 297 (distance 0.012) 16.853 AskName 15.985 ConfirmName $\hat{Q}(\hat{x}, \hat{a}_m)$ Output system action Sorry, first and last name? $a_m^*$

$O$   $b(s)$   $\hat{x}$   $\hat{Q}(\hat{x}, \hat{a}_m)$   $a_m^*$

You can hang up and make another call from the same phone at any time  
Choose another phone number

# Out there in the wild

POMDPs grew up with...	But in the wild we'll need...
Discrete observations and actions	Vectors of features with continuous and discrete elements (possibly with unseen elements)
Discrete states	Complex states with continuous and discrete elements; networks/manifolds
Deterministic elements just copied into the belief space	Rich state feature sets with some observable and some uncertain features
Planning in the full belief space	Planning in a compressed space; automatically extracting features from actions and belief states
Planning from <i>tabula rasa</i>	Templates; business rules; policy constraints
Comparisons to other POMDP techniques	Comparisons to state-of-the-art RL techniques
Given models	Synthesizing initial models with learning from experience
Stationary, known environments	Non-stationary environments; different environments; transfer learning

# Applications: Food for thought

- What would more evaluations on simulated environments (rather than given models) teach us?
  - Would "standard environment simulations" be more compelling benchmark problems than "standard models" (of T, O, S)?
- Is there a real domain which *is* a computer program? (Simulation = Reality)
- Is there a real-world problem with a steady flow of real interactions? (2nd Life?)

# Where is our killer app?

- A common view: "POMDPs are a great approach; the (temporary) problem is scalability."
- Suppose scalability were not a concern. Then, which real problems would we tackle?
  - What is the competition for solving those problems?
  - What would POMDPs hope to do better than the competition?

# Thanks!

Jason Williams  
AT&T Labs - Research