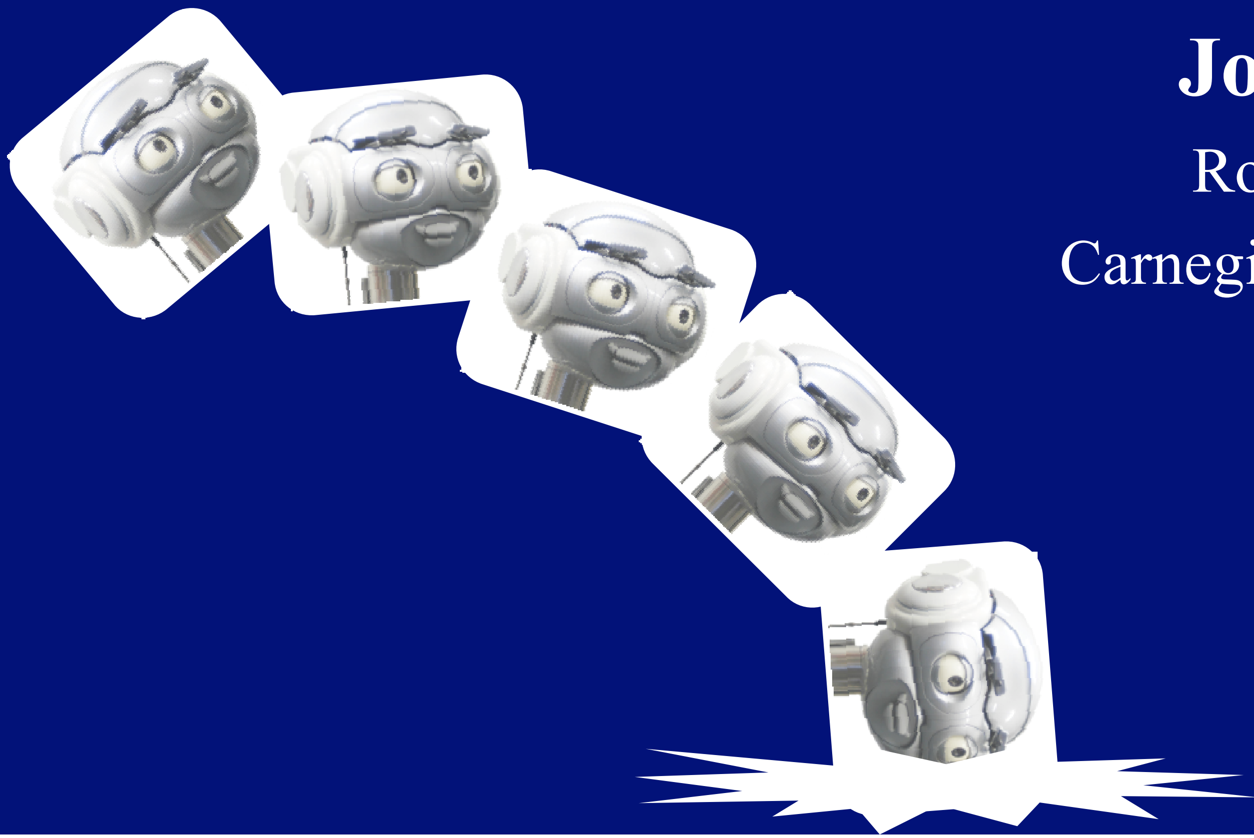
Tractable Planning Under Uncertainty:

Exploiting Structure



Joelle Pineau

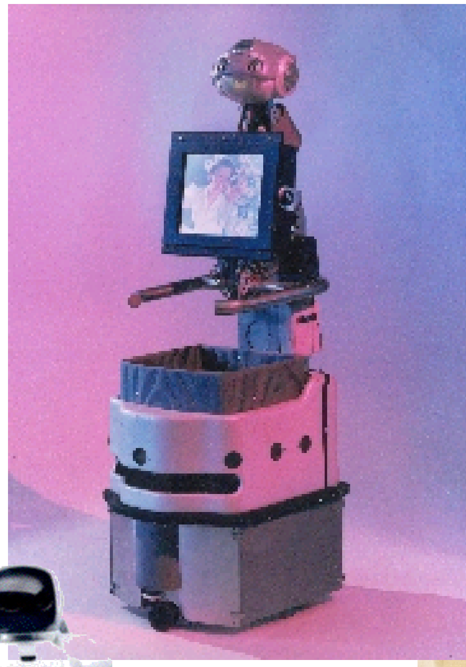
Robotics Institute

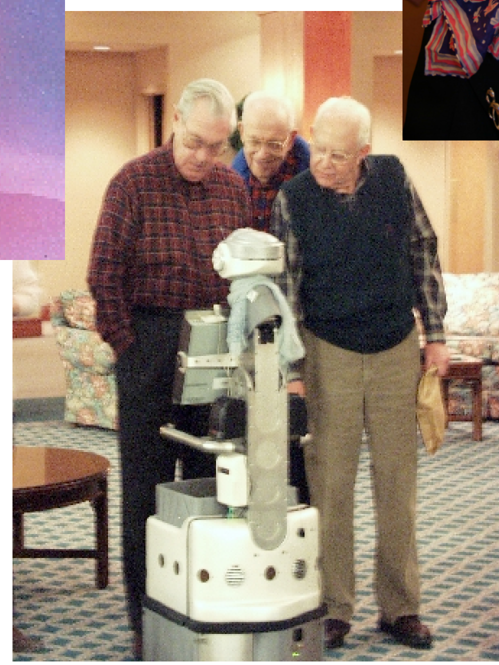
Carnegie Mellon University

Thesis Oral

June 14, 2004

Bringing robots into unstructured environments

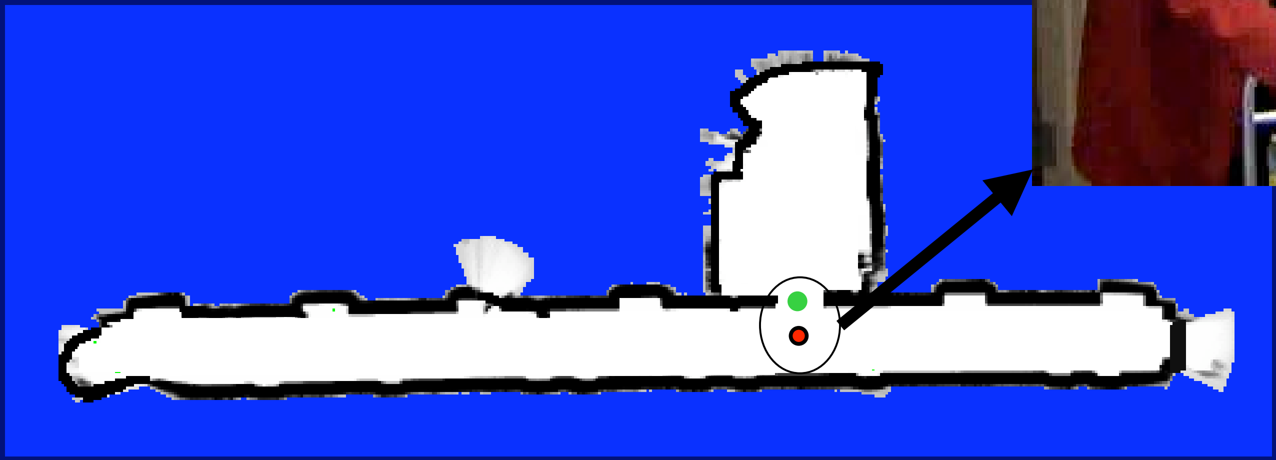




Dealing with state uncertainty

Objective:

To optimize plans which are robust to incomplete, ambiguous, outdated, incorrect sensor information.



The technical challenge

Tracking state uncertainty is hard:

- Kalman filter, particle filter.

Conventional planning is hard:

- Path planning, MDP.

Planning with state uncertainty is harder!

- POMDP framework is well-established *[Sondik, 1970]*.
- Tractability is the major obstacle.

Thesis statement

Planning under uncertainty can be made tractable for complex problems by exploiting structure in the problem domain.

- **Geometric structure: PBVI**
- **Hierarchical control structure: PolCA+**

Talk outline

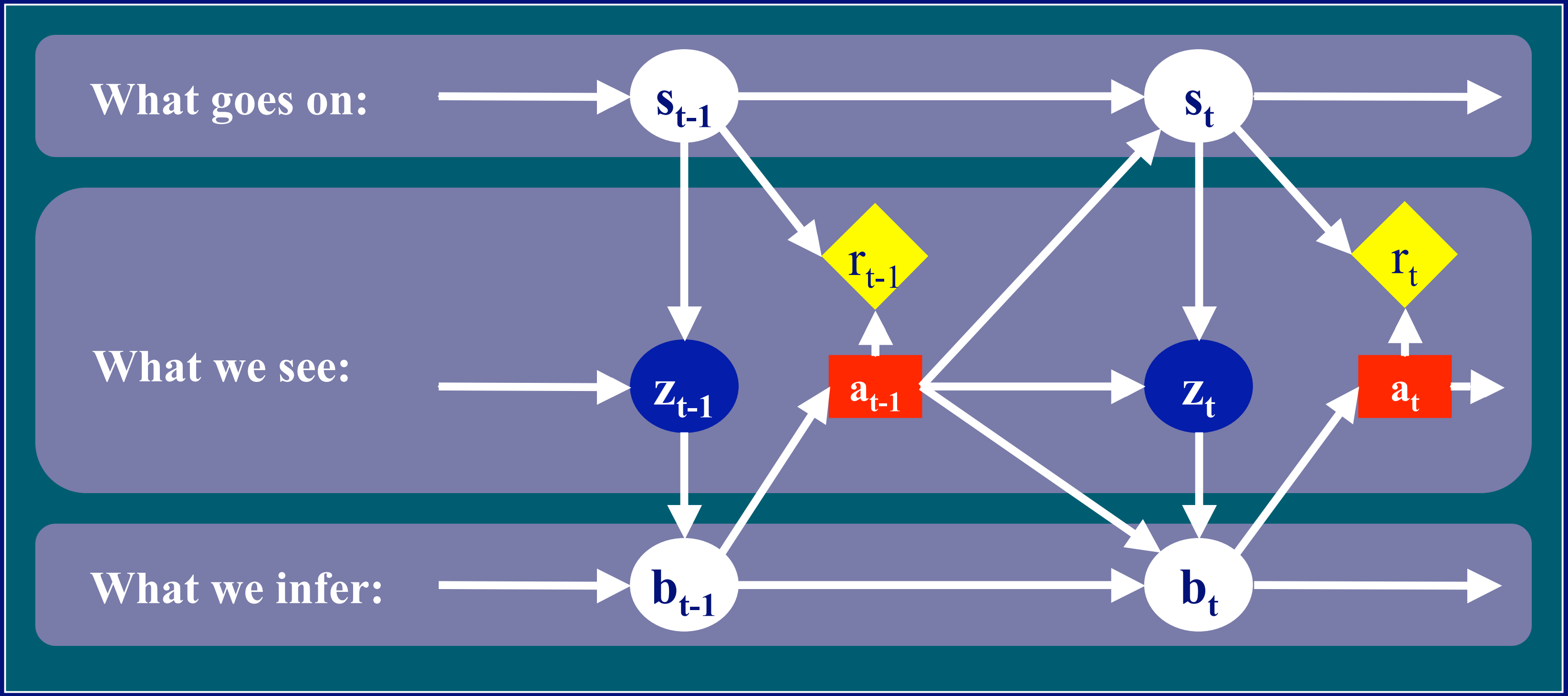
- Uncertainty in plan-based robotics
- Partially Observable Markov Decision Processes (POMDPs)
- Exploiting geometric structure
  - » Point-based value iteration (PBVI)
- Exploiting hierarchical control structure
  - » Policy-contingent abstraction (PolCA+)

POMDP model

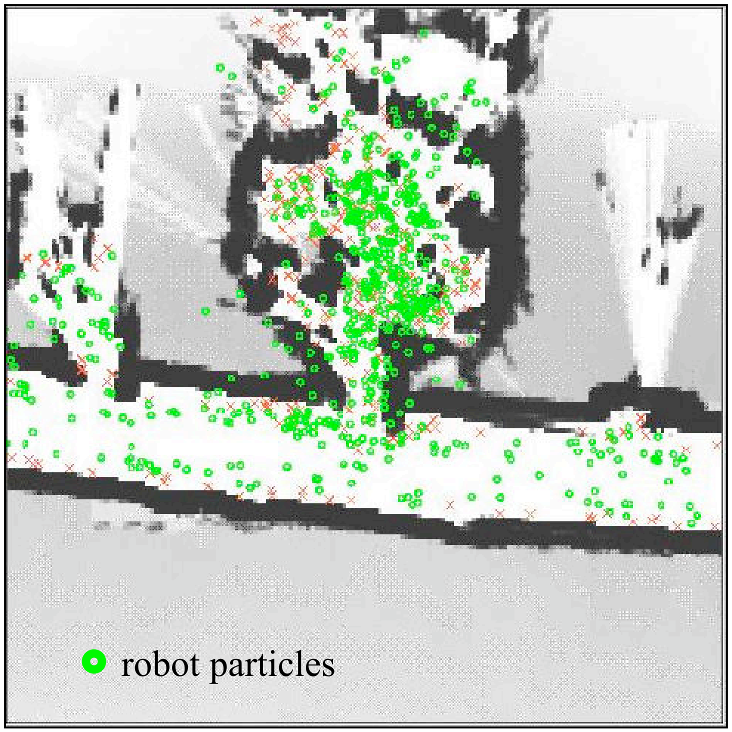
POMDP is n-tuple { *S, A, Z, T, O, R* }:

S = state set
A = action set
Z = observation set

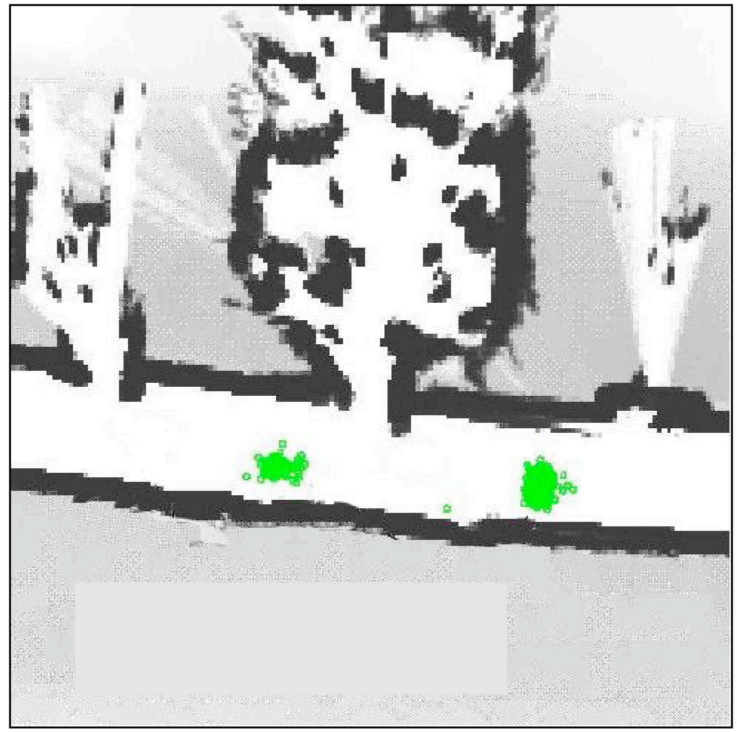
T: $Pr(s'|s,a)$ = state-to-state transition probabilities
O: $Pr(z|s,a)$ = observation generation probabilities
$R(s,a)$ = reward function



Examples of robot beliefs



Uniform belief



Bi-modal belief

Pictures courtesy of Nicholas Roy.

POMDP solving

Objective: Find the sequence of actions that maximizes the expected sum of rewards.

$$V(b) = \max_{a \in A} \left[R(b,a) + \gamma \sum_{b' \in B} T(b,a,b') V(b') \right]$$

Value function

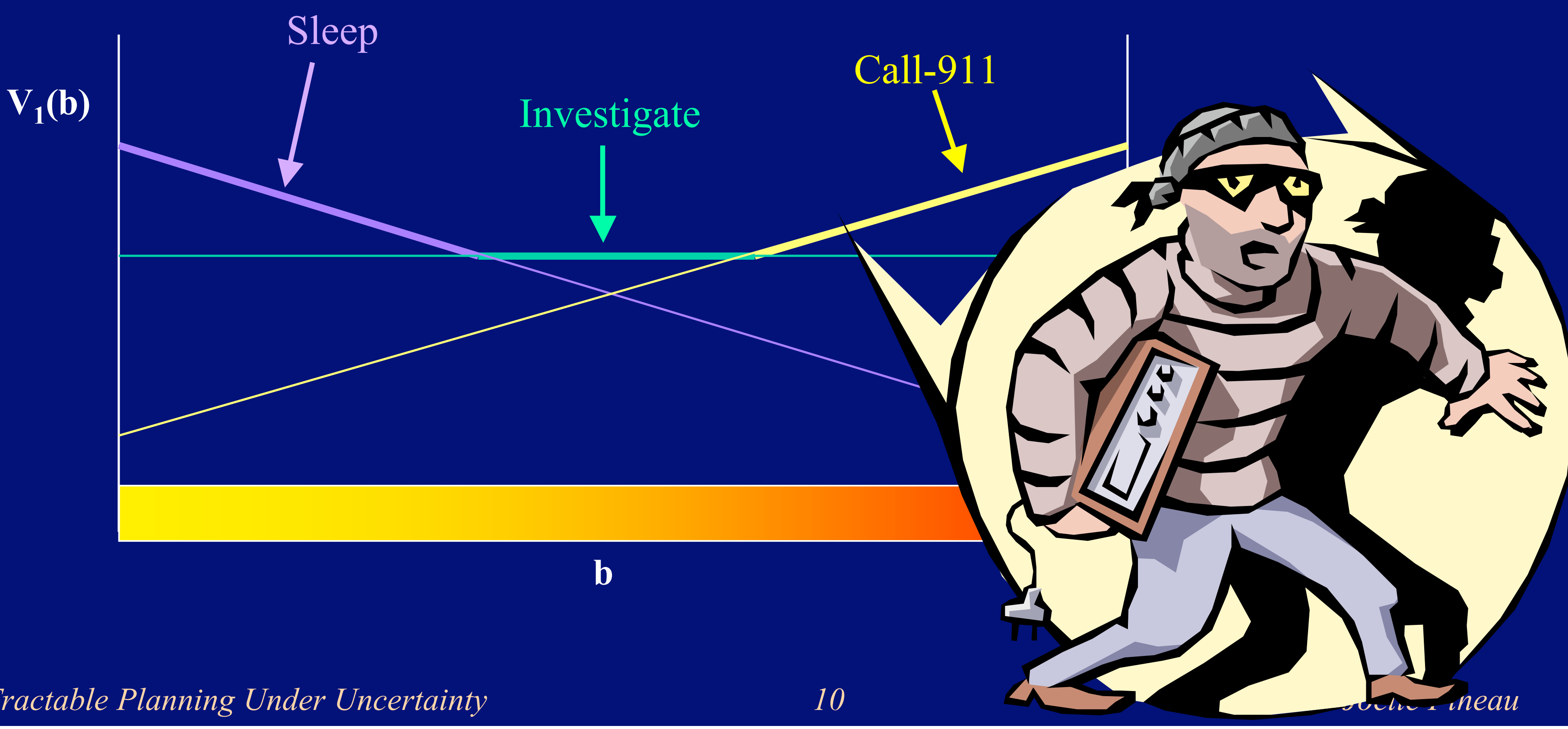
Immediate reward

Future reward

Optimal POMDP solving

- Simple problem: *2 states, 3 actions, 3 observations*

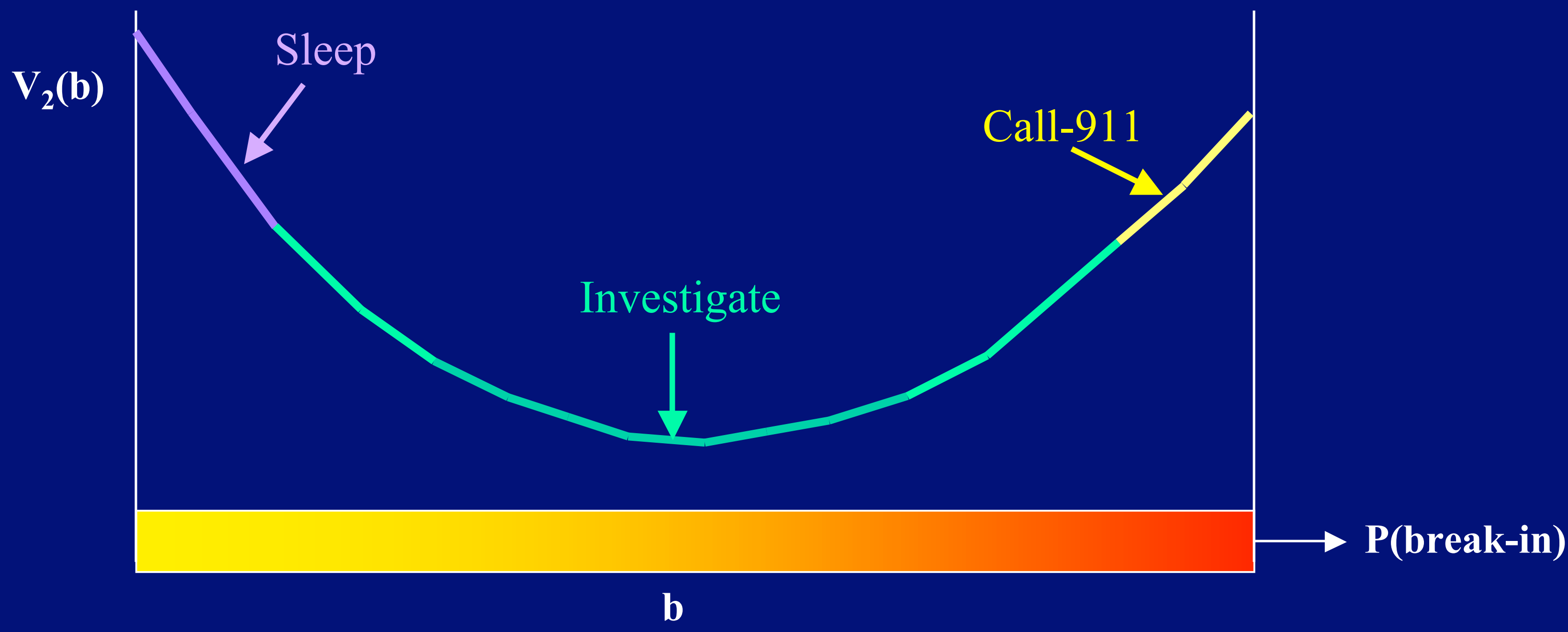
Plan length	# vectors
1	3



Optimal POMDP solving

- Simple problem: *2 states, 3 actions, 3 observations*

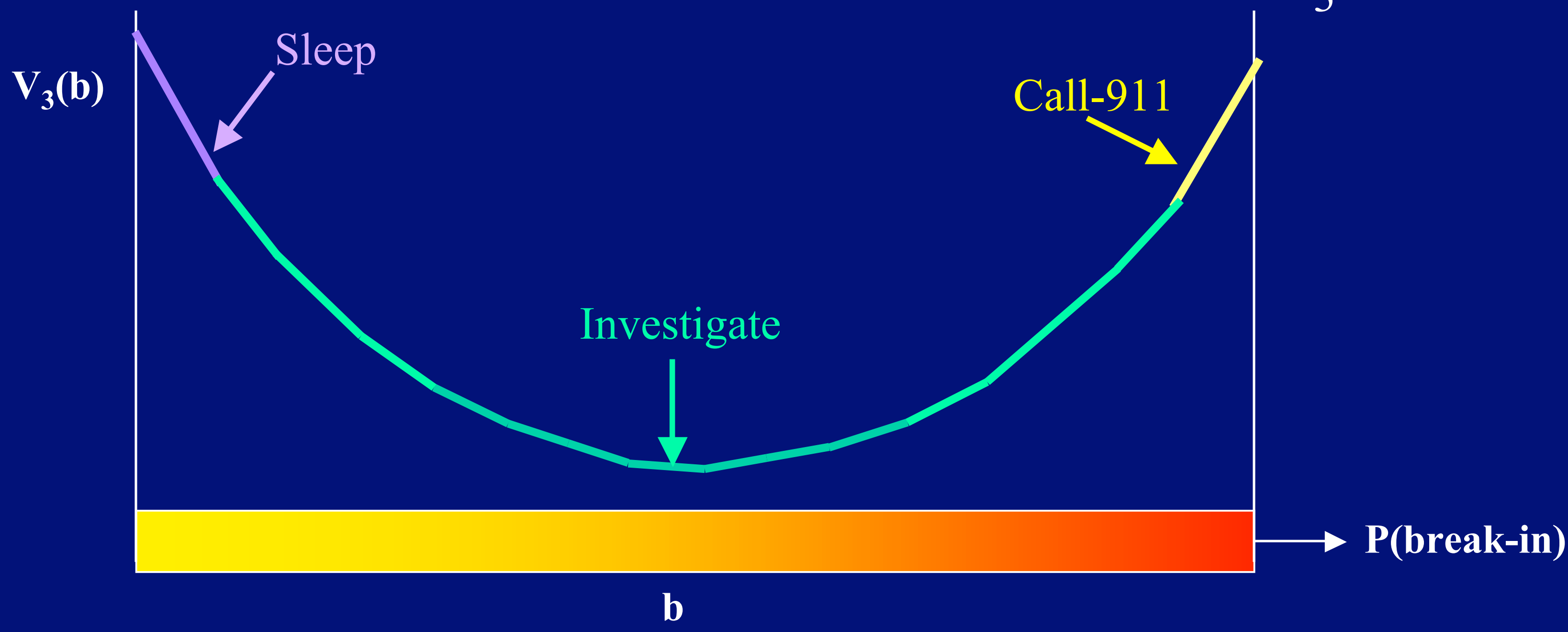
Plan length	# vectors
1	3
2	27



Optimal POMDP solving

- Simple problem: *2 states, 3 actions, 3 observations*

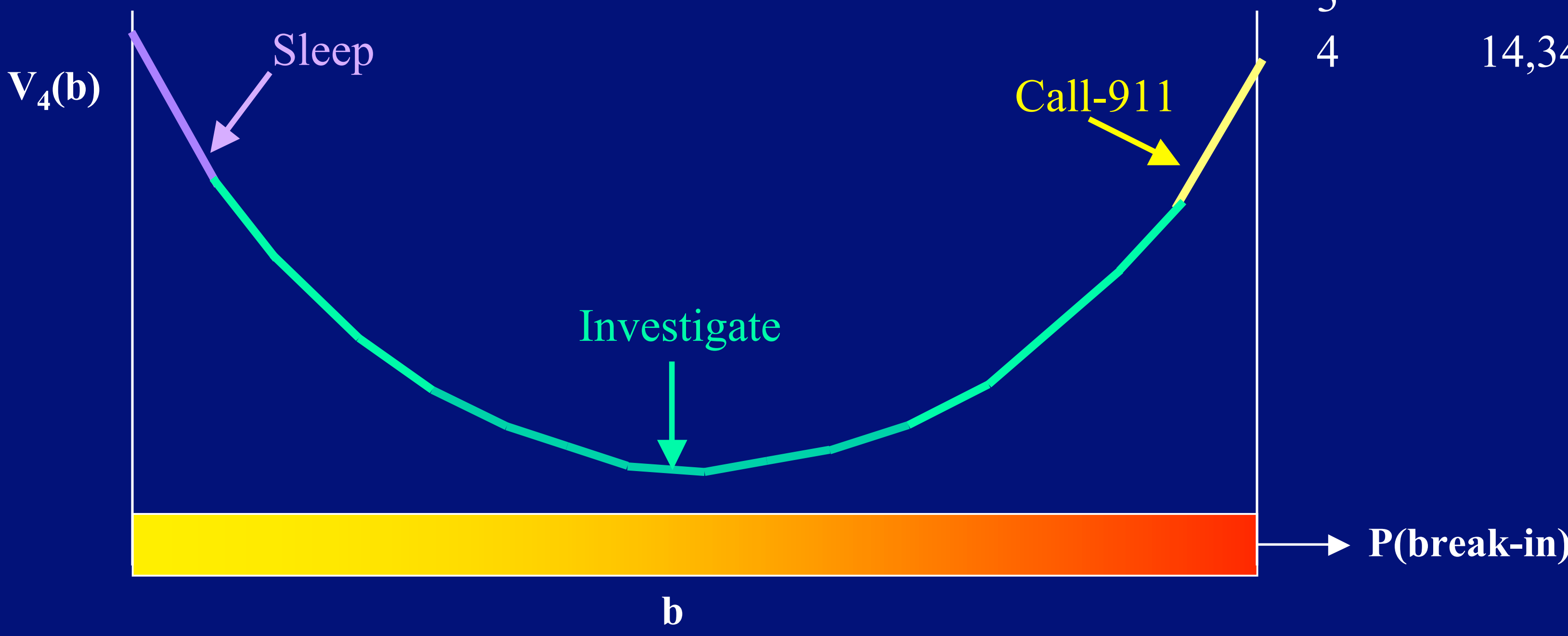
Plan length	# vectors
1	3
2	27
3	2187



Optimal POMDP solving

- Simple problem: *2 states, 3 actions, 3 observations*

Plan length	# vectors
1	3
2	27
3	2187
4	14,348,907



How many vectors for this problem?

10^4 (navigation) x 10^3 (dialogue) states
1000+ observations
100+ actions

Pictures courtesy of Sebastian Thrun.

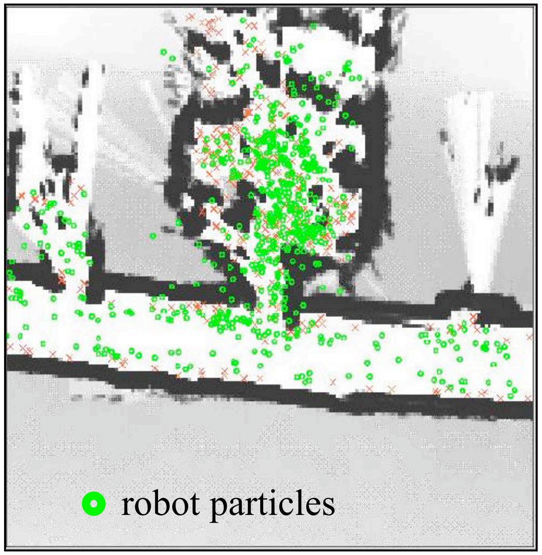
The curse of history

Policy size grows exponentially with the **planning horizon**:

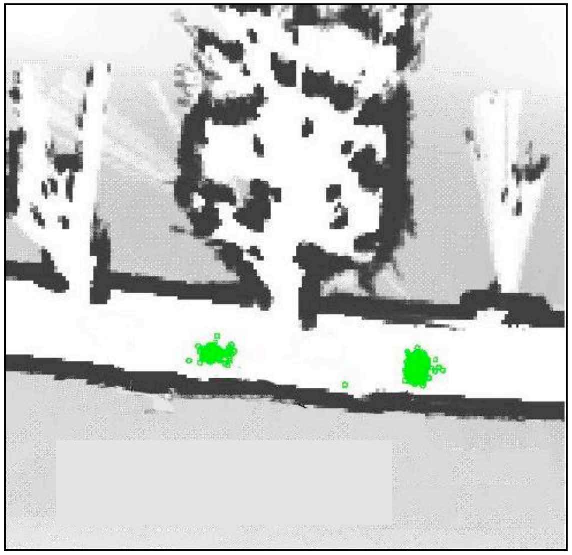
$$\boxed{\Gamma_n} = O(A \boxed{\Gamma_{n-1}}^{Z})$$

Where n = planning horizon
A = # actions
Z = # observations

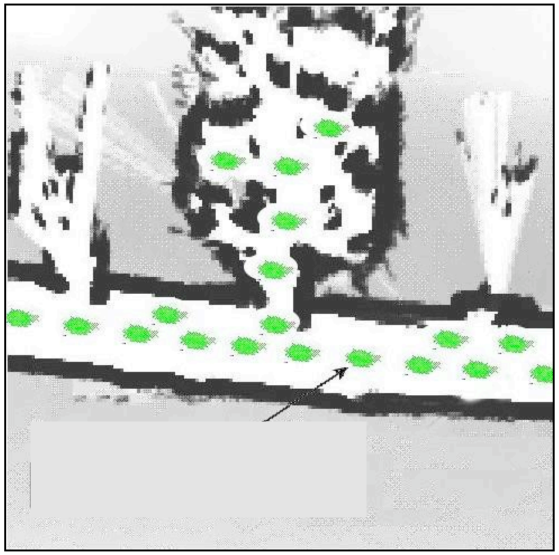
Exact solving assumes all beliefs are equally likely



Uniform belief



Bi-modal belief



N-modal belief

INSIGHT: No sequence of actions and observations can produce this N-modal belief.

Pictures courtesy of Nicholas Roy.

Talk outline

- Uncertainty in plan-based robotics
- Partially Observable Markov Decision Processes (POMDPs)
- Exploiting geometric structure
  - » Point-based value iteration (PBVI)
- Exploiting hierarchical control structure
  - » Policy-contingent abstraction (PolCA+)

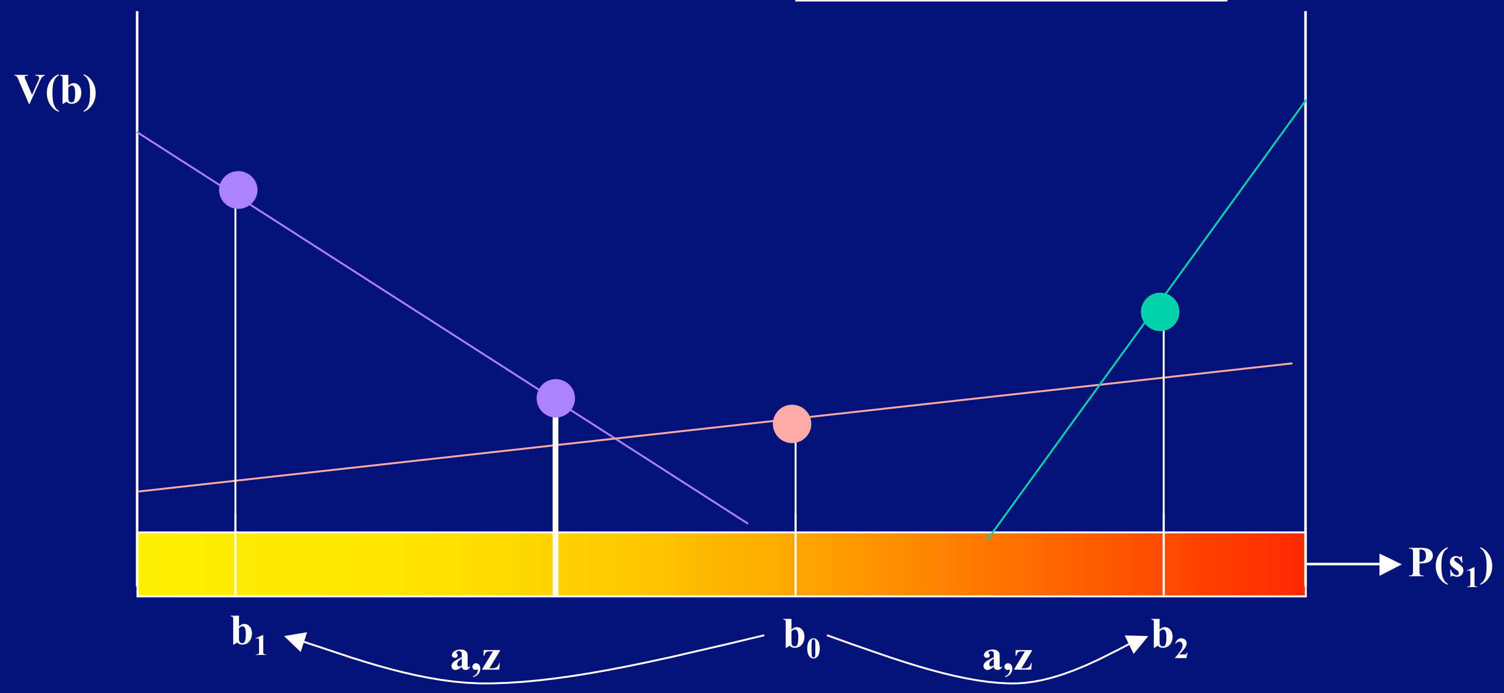
A new algorithm: *Point-based value iteration*

Approach:

Select a small set of belief points ⇒ ***Use well-separated, reachable beliefs***

Plan for those belief points only ⇒ ***Learn value and its gradient***

Pick action that maximizes value ⇒ $V(b) = \max_{\alpha \in \Gamma} (\alpha \cdot b)$



The curse of history - revisited

Policy size:

$O(|A| |\Gamma_{n-1}|^{|Z|})$

Update time:

$O(|S| |A| |\Gamma_{n-1}|^{|Z|})$

↓

Policy size:

$O(|B|)$

Update time:

$O(|S| |A| |Z| |\Gamma_{n-1}| |B|)$

The anytime PBVI algorithm

- Alternate between:

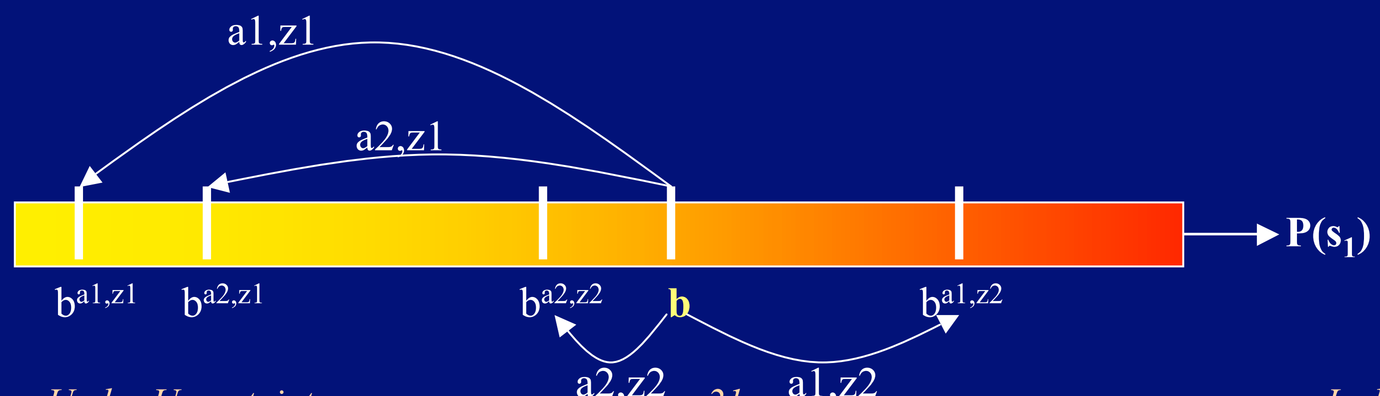
  1. Growing the set of belief point

  2. Planning for those belief points

- Terminate when you run out of time or have a good policy.

Belief selection in PBVI

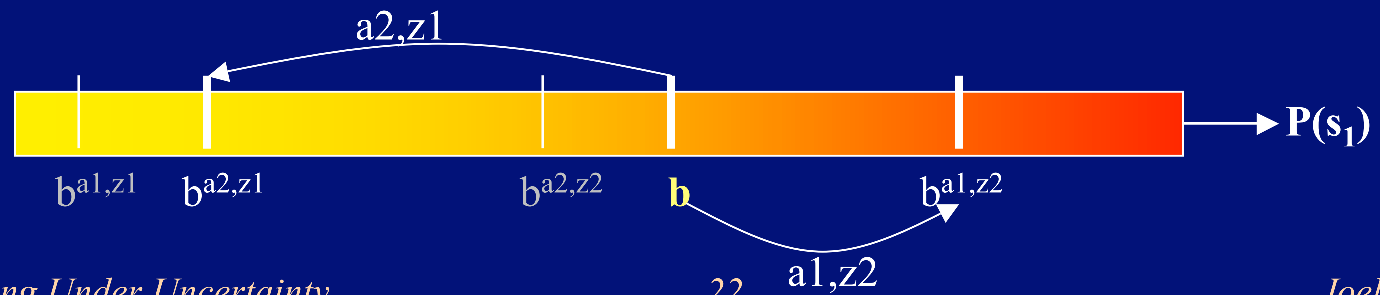
1. Focus on reachable beliefs.



Belief selection in PBVI

1. Focus on reachable beliefs.

2. Focus on high probability reachable beliefs.

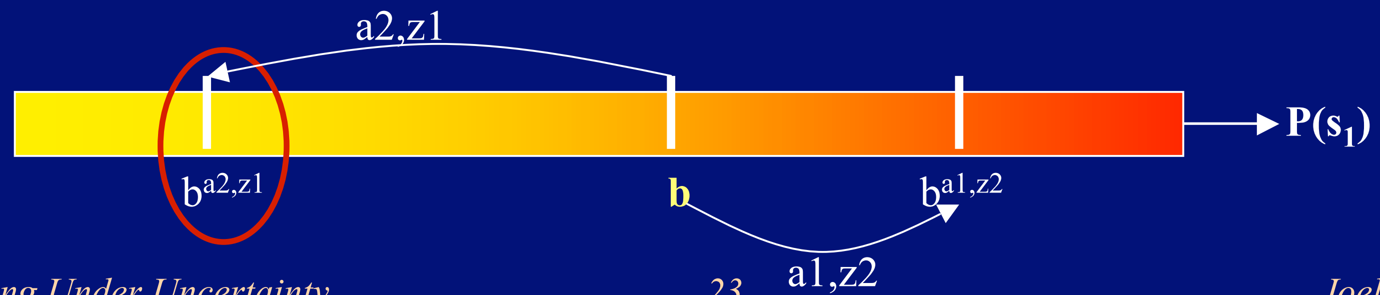


Belief selection in PBVI

1. Focus on reachable beliefs.

2. Focus on high probability reachable beliefs.

3. Select well-separated high probability reachable beliefs.



Theoretical properties of PBVI

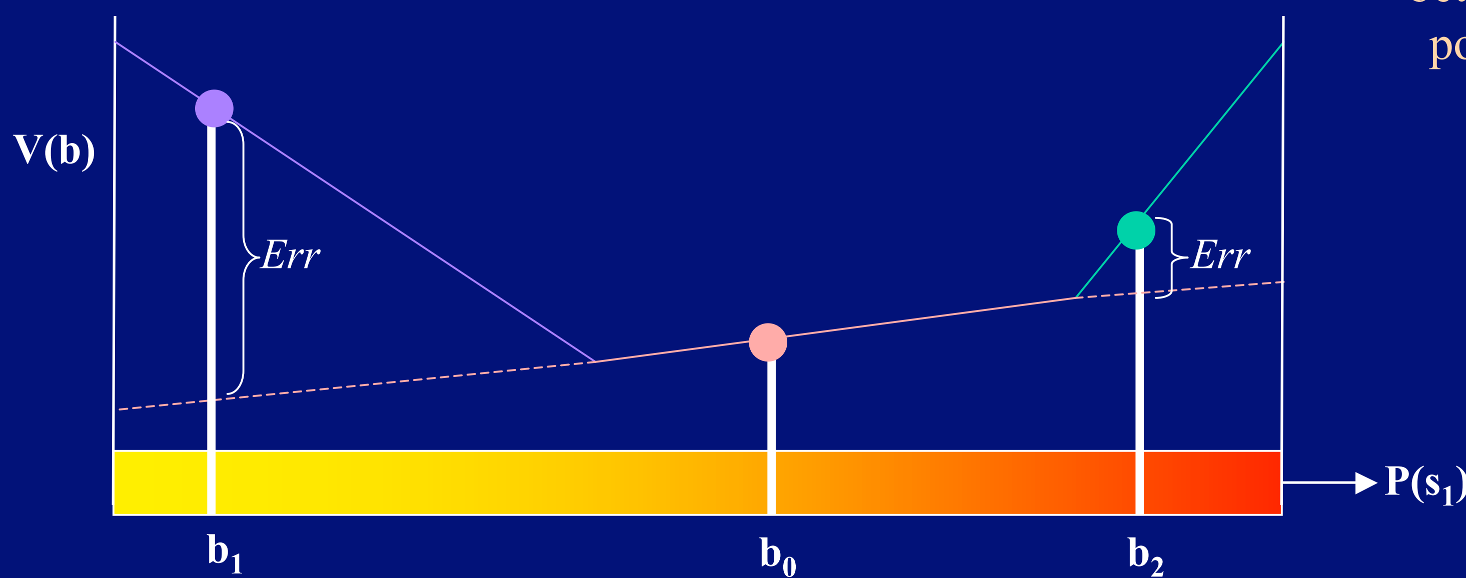
<u>Theorem:</u> *For any set of belief points B and planning horizon n, the error of the PBVI algorithm is bounded by:*

$$\| V_n^B - V_n^* \|_\infty \le \left[\frac{(R_{\max} - R_{\min})}{(1-\gamma)^2} \right] \max_{b' \in \Delta} \min_{b \in B} \| b - b' \|_1$$

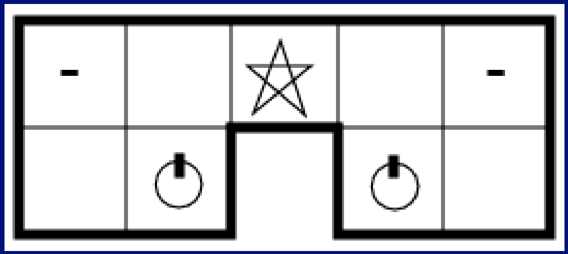
Where Δ is the set of reachable beliefs
B is the set of all beliefs

error bound

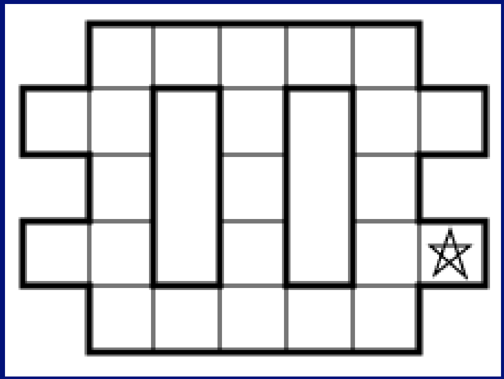
distance between points



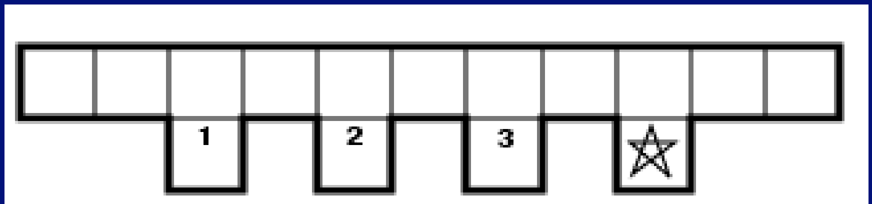
Empirical results on well-known POMDPs



Maze1: *36 states*



Maze2: *92 states*



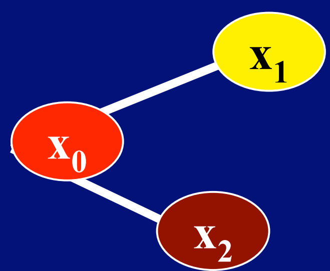
Pictures from Littman, Cassandra and Kaelbling, 1995.

Maze3: *60 states*

Classes of value function approximations

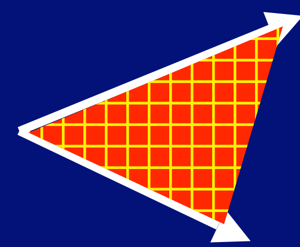
1. No belief

[Littman&al., 1995]



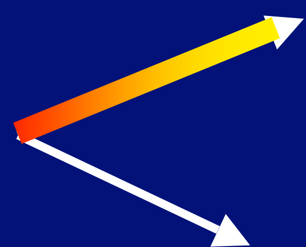
2. Grid over belief

[Lovejoy, 1991; Brafman 1997; Hauskrecht, 2000; Zhou&Hansen, 2001]



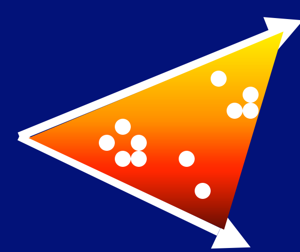
3. Compressed belief

[Poupart&Boutilier, 2002; Roy&Gordon, 2002]



4. Sample belief points

[Poon, 2001; Pineau&al, 2003]



Performance on well-known POMDPs

Method	REWARD			TIME (sec)			# Belief points		
	Maze1	Maze2	Maze3	Maze1	Maze2	Maze3	Maze1	Maze2	Maze3
No belief *[Littman&al., 1995]*	0.20	0.11	0.26	0.19	1.44	0.51	-	-	-
Grid *[Brafman., 1997]*	0.94	-	-	-	-	-	174	337	-
Compressed *[Poupart&al., 2003]*	0.00	0.07	0.11	>24hrs	>24hrs	>24hrs	-	-	-
Sample *[Poon, 2001]*	2.30	0.35	0.53	12166	27898	450	660	1840	300
PBVI *[Pineau&al., 2003]*	2.25	0.34	0.53	3448	360	288	470	95	86

Additional results not shown *[Smith&Simmons, 2004; Spaan&Vlassis, 2004]*.

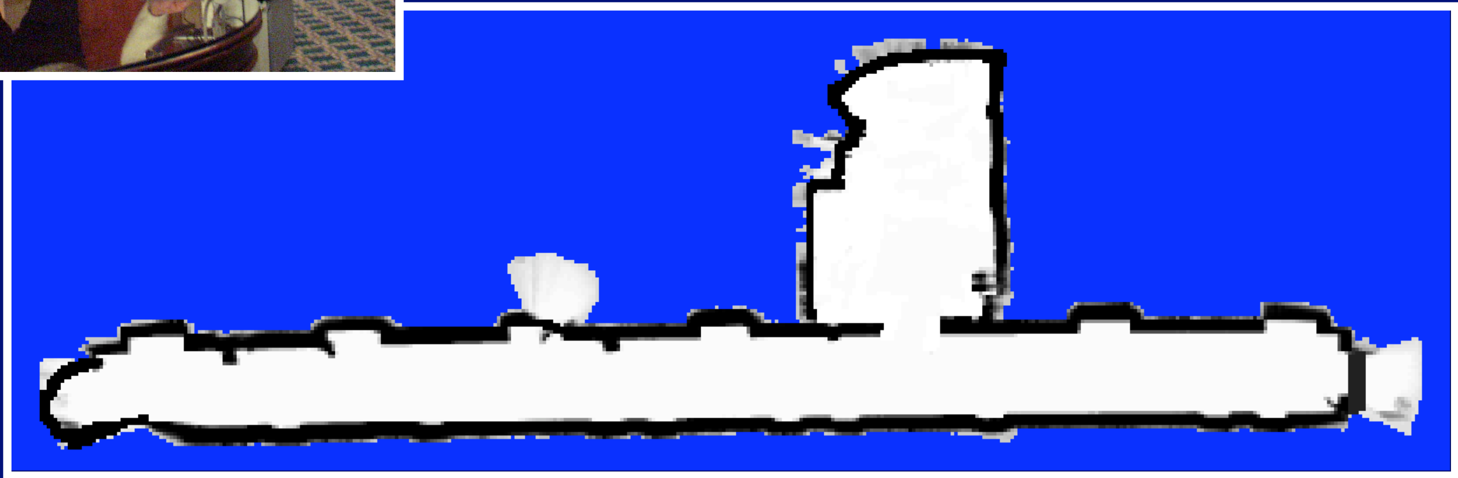
PBVI in the Nursebot domain

Objective: Find the patient.

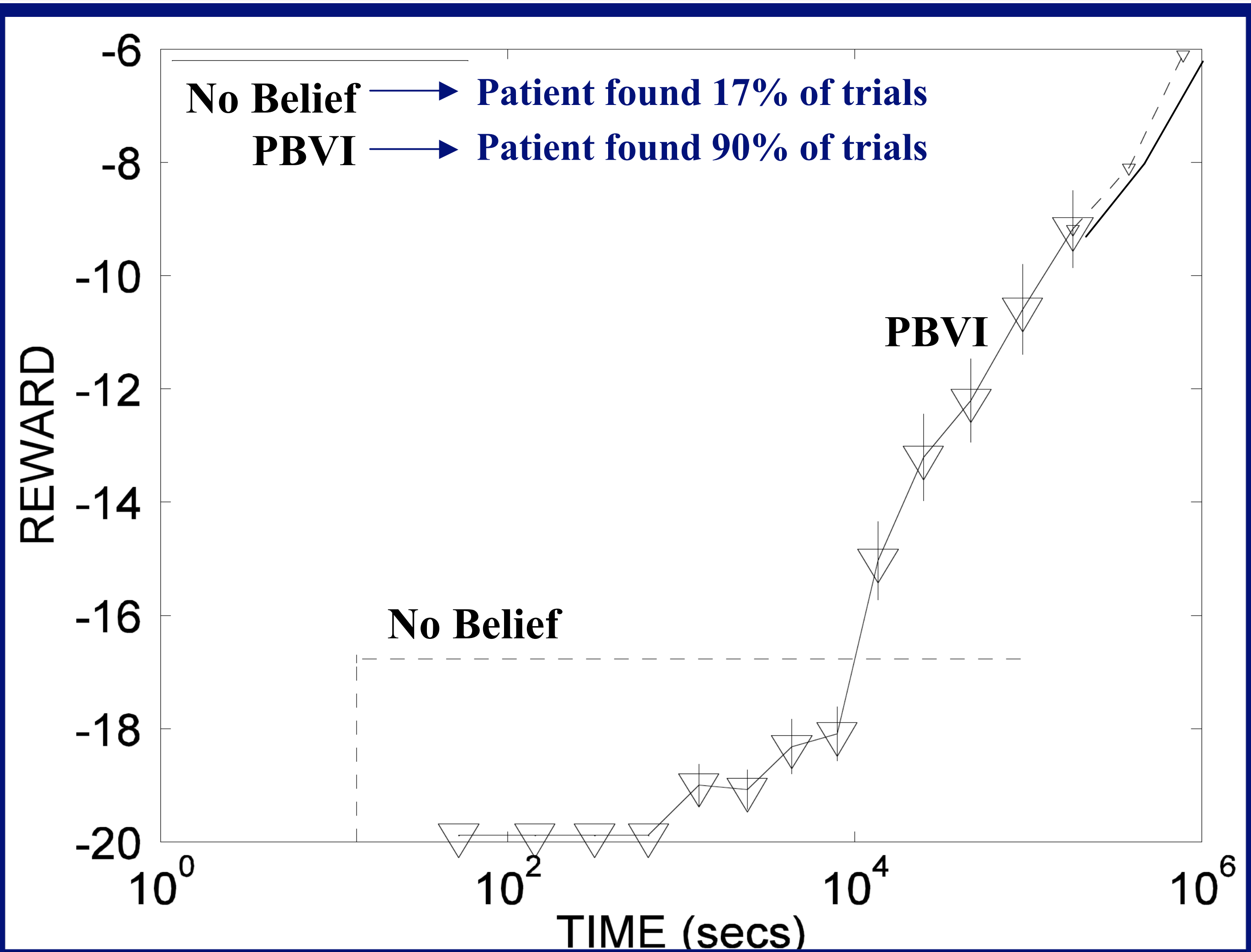
State space = RobotPosition × PatientPosition

Observation space = RobotPosition + PatientFound

Action space = {North, South, East, West, Declare}

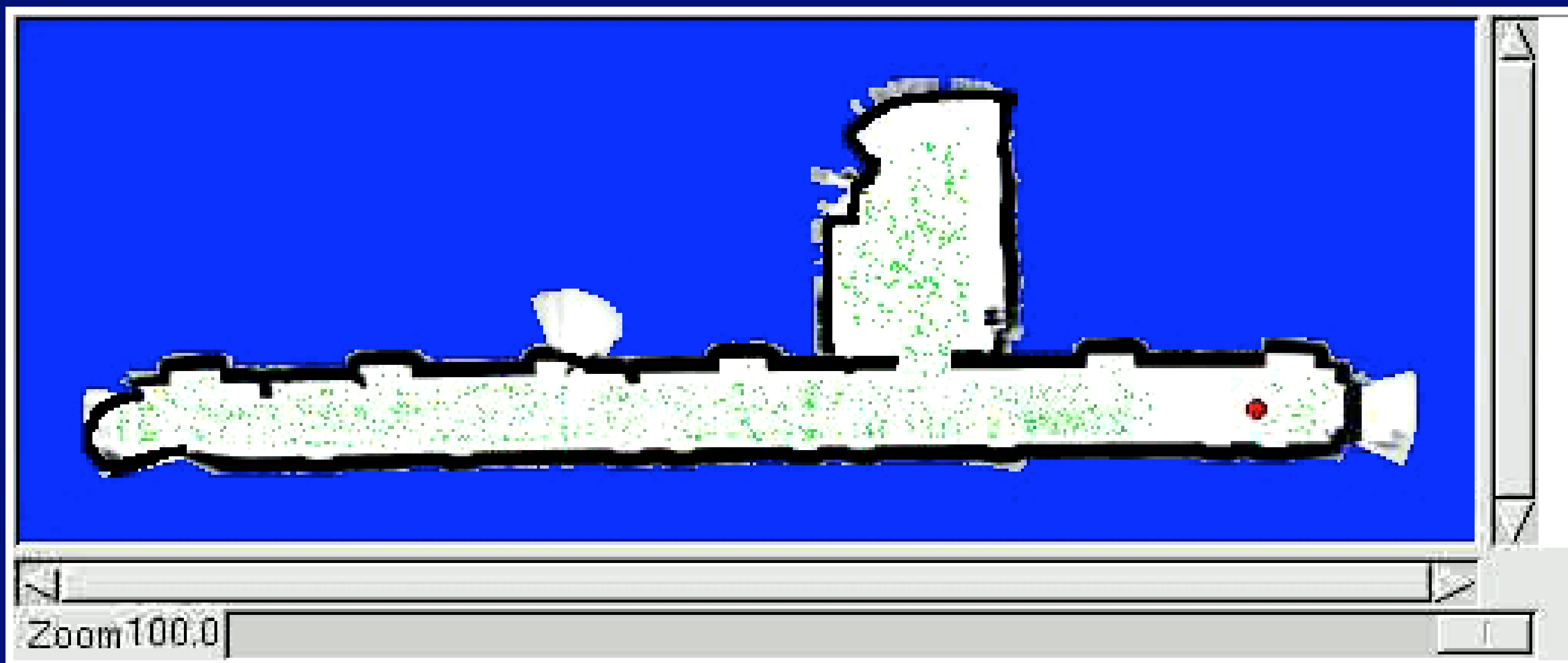


PBVI performance on *find-the-patient* domain

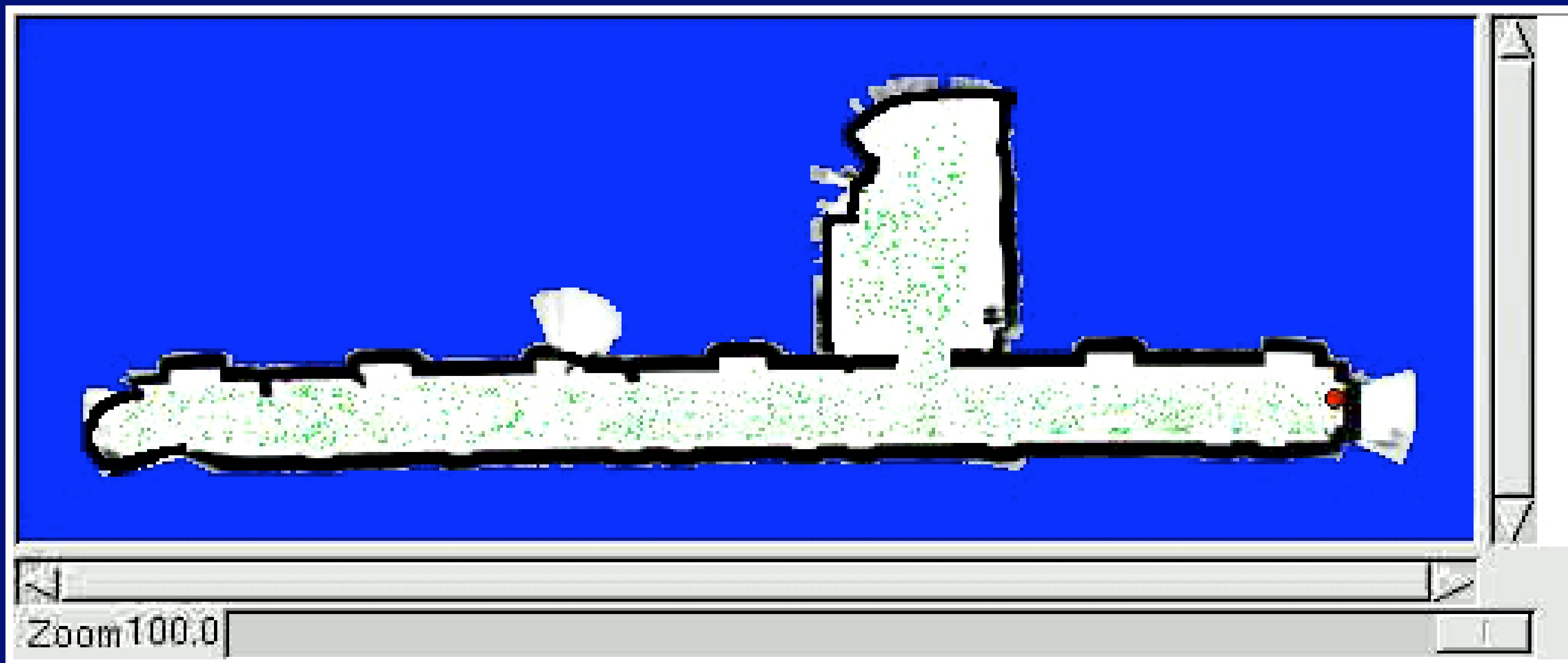


Additional results not shown *[Poupart&Boutilier, 2004; Smith&Simmons, 2004; Spaan&Vlassis, 2004]*.

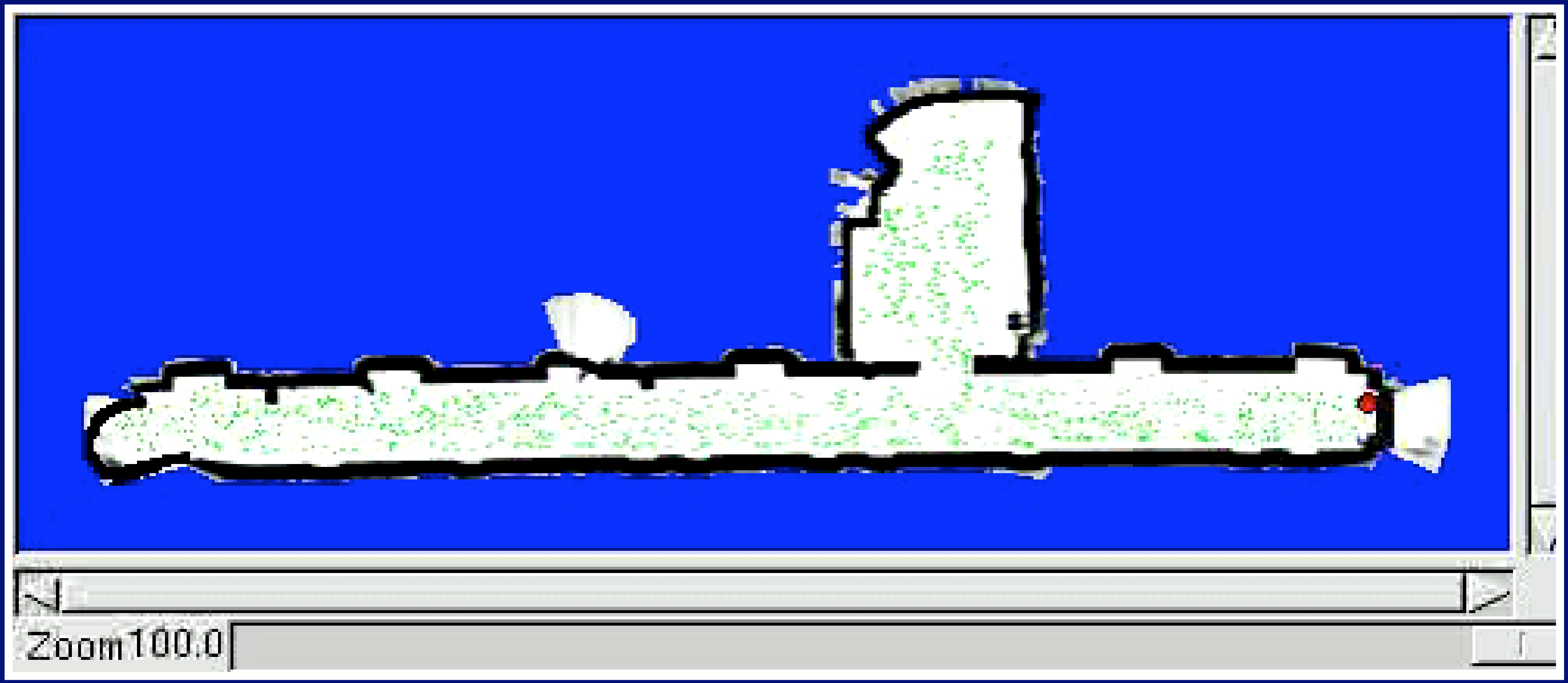
Policy assuming full observability



PBVI policy with 3141 belief points



PBVI policy with 643 belief points



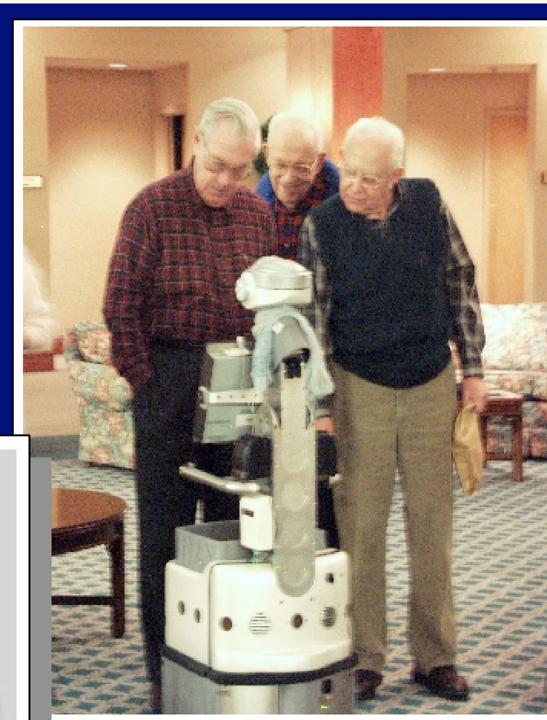
Contributions of the PBVI algorithm

- Algorithmic:
  - New belief sampling algorithm.
  - Efficient heuristic for belief point selection.
  - Anytime performance.
- Experimental:
  - Outperforms previous value approximation algorithms on known problems.
  - Solves new larger problem (1 order of magnitude increase in problem size).
- Theoretical:
  - Bounded approximation error.

Back to the big picture

How can we go from 10^3 states

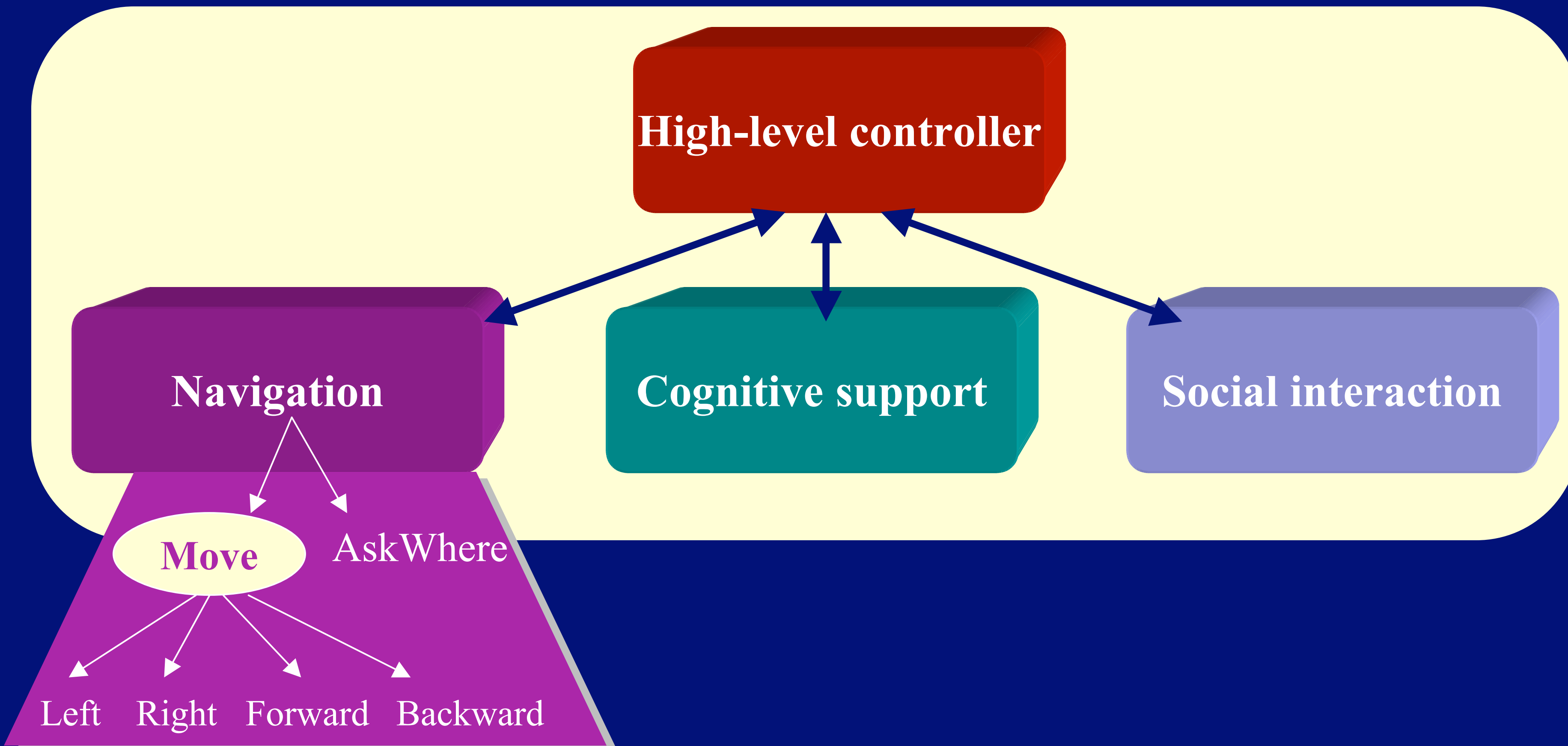
to real-world problems?



Pictures courtesy of Sebastian Thrun.

Structured POMDPs

⇒ **Many real-world decision-making problems exhibit structure inherent to the problem domain.**



Structured POMDP approaches

Factored models

[Boutilier & Poole, 1996; Hansen & Feng, 2000; Guestrin et al., 2001]

Idea: Represent state space with multi-valued state features.

Hierarchical POMDPs

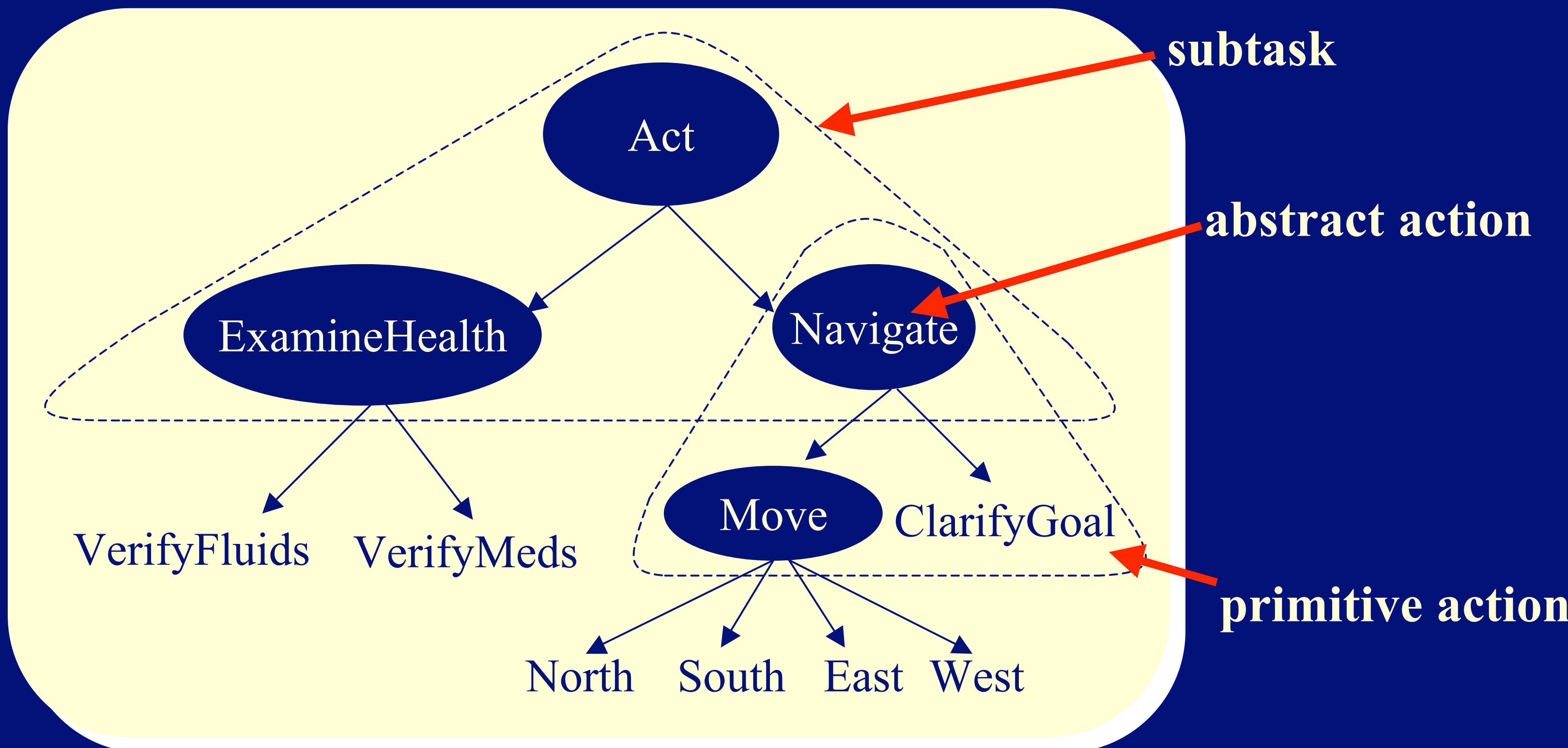
[Wiering & Schmidhuber, 1997; Theocharous et al., 2000; Hernandez-Gardiol & Mahadevan, 2000]

Idea: Exploit domain knowledge to divide one POMDP into many smaller ones.

Talk outline

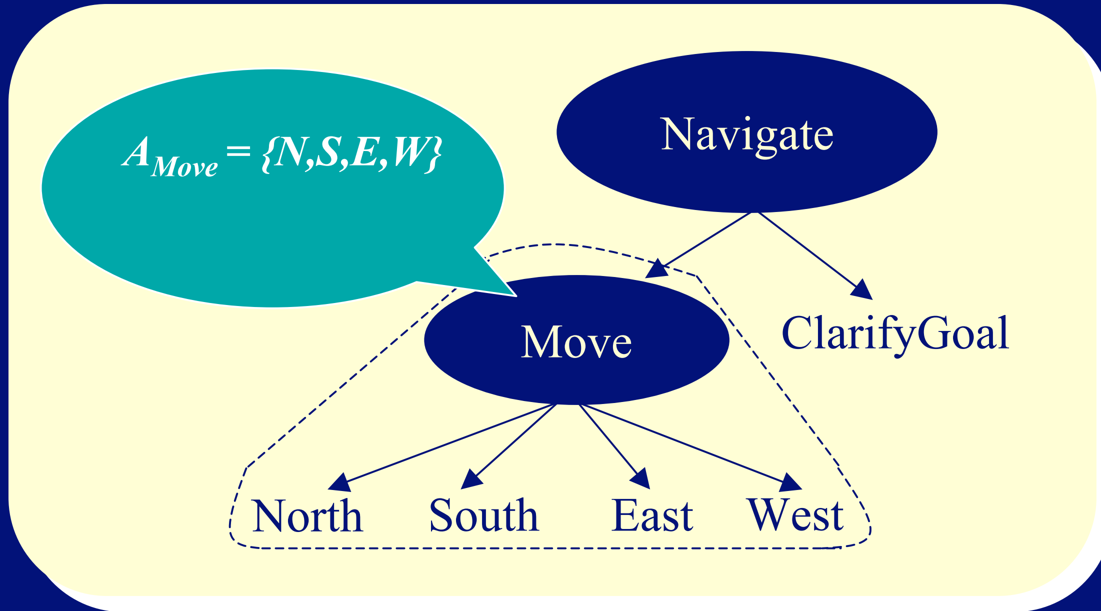
- Uncertainty in plan-based robotics
- Partially Observable Markov Decision Processes (POMDPs)
- Exploiting geometric structure
  - » Point-based value iteration (PBVI)
- Exploiting hierarchical control structure
  - » Policy-contingent abstraction (PolCA+)

A hierarchy of POMDPs



PolCA+: Planning with a hierarchy of POMDPs

Step 1: Select the action set

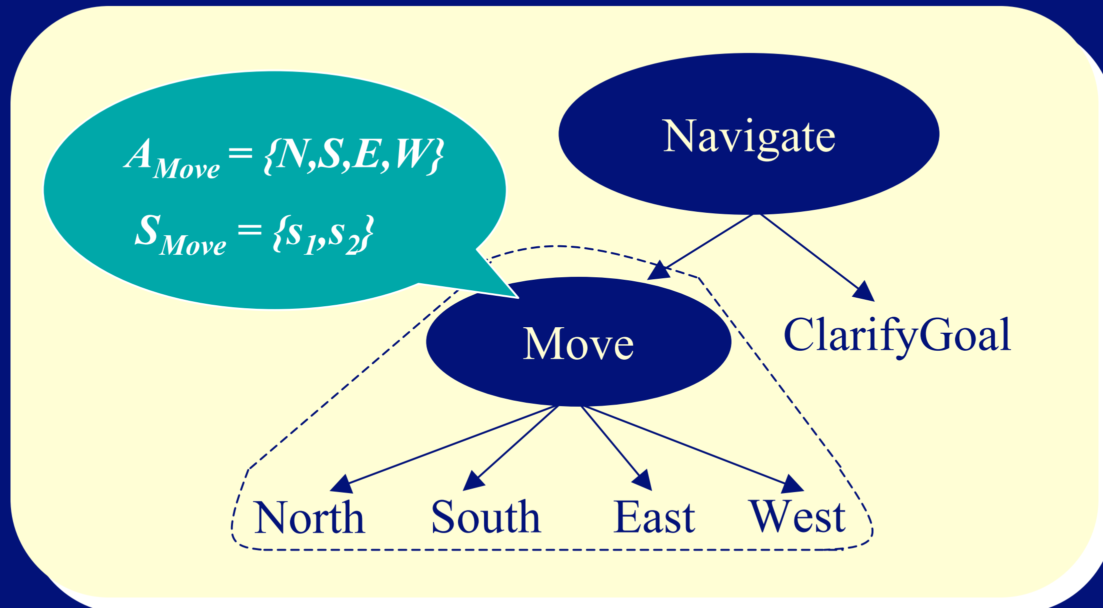


ACTIONS
North
South
East
West
ClarifyGoal
VerifyFluids
VerifyMeds

PolCA+: Planning with a hierarchy of POMDPs

Step 1: Select the action set

Step 2: Minimize the state set



ACTIONS

North
South
East
West
ClarifyGoal
VerifyFluids
VerifyMeds

STATE FEATURES

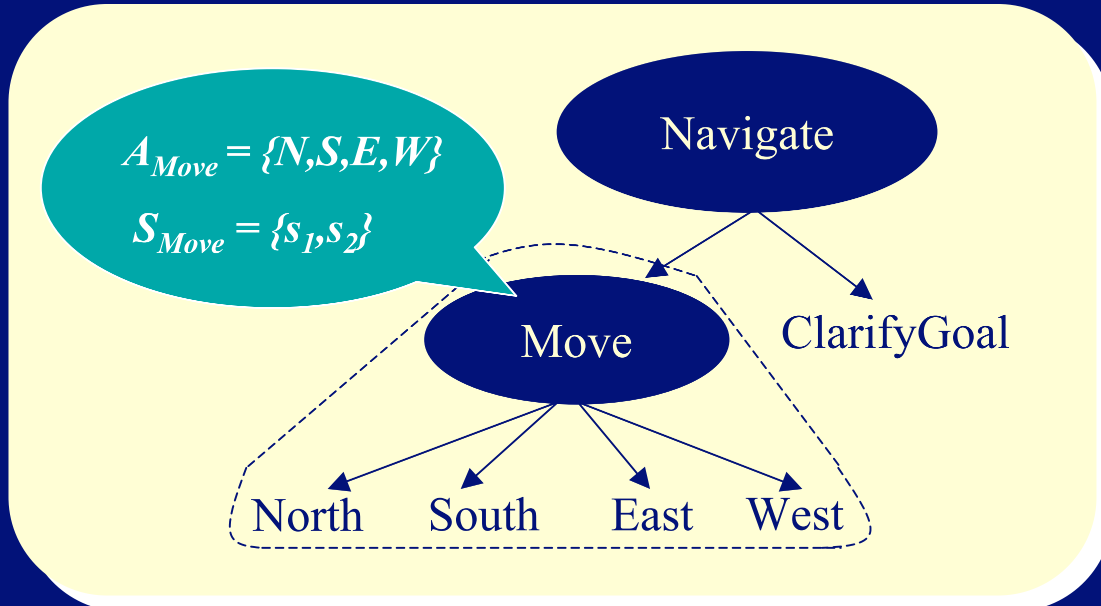
X-position
Y-position
X-goal
Y-goal
HealthStatus

PolCA+: Planning with a hierarchy of POMDPs

Step 1: Select the action set

Step 2: Minimize the state set

Step 3: Choose parameters



ACTIONS

North
South
East
West
ClarifyGoal
VerifyFluids
VerifyMeds

STATE FEATURES

X-position
Y-position
X-goal
Y-goal
HealthStatus

PARAMETERS

$\{b_h, T_h, O_h, R_h\}$

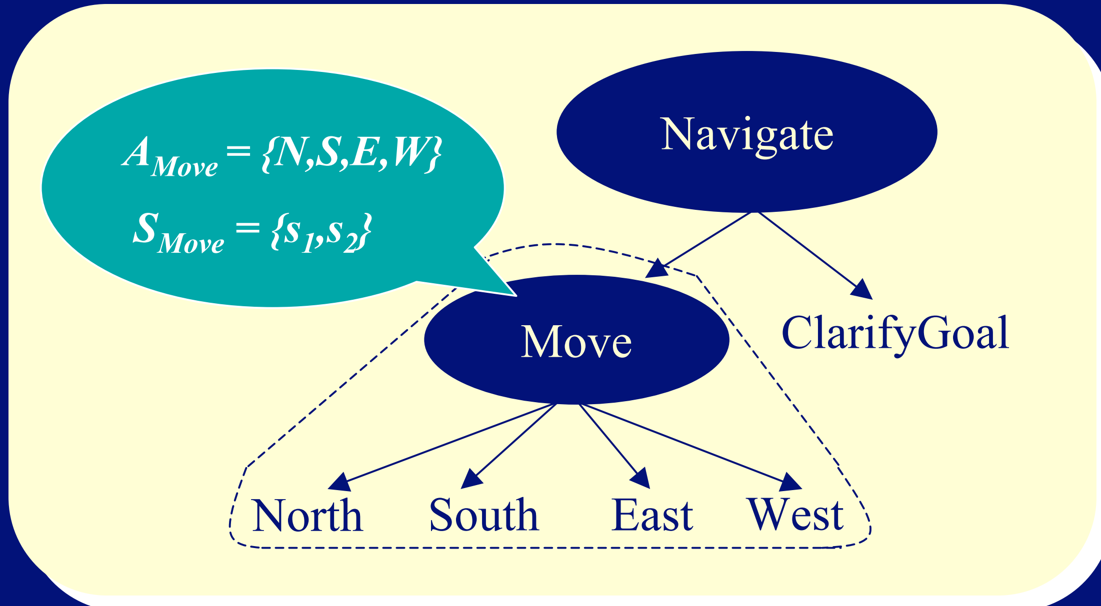
PolCA+: Planning with a hierarchy of POMDPs

Step 1: Select the action set

Step 2: Minimize the state set

Step 3: Choose parameters

Step 4: Plan task ***h***



ACTIONS
North
South
East
West
ClarifyGoal
VerifyFluids
VerifyMeds

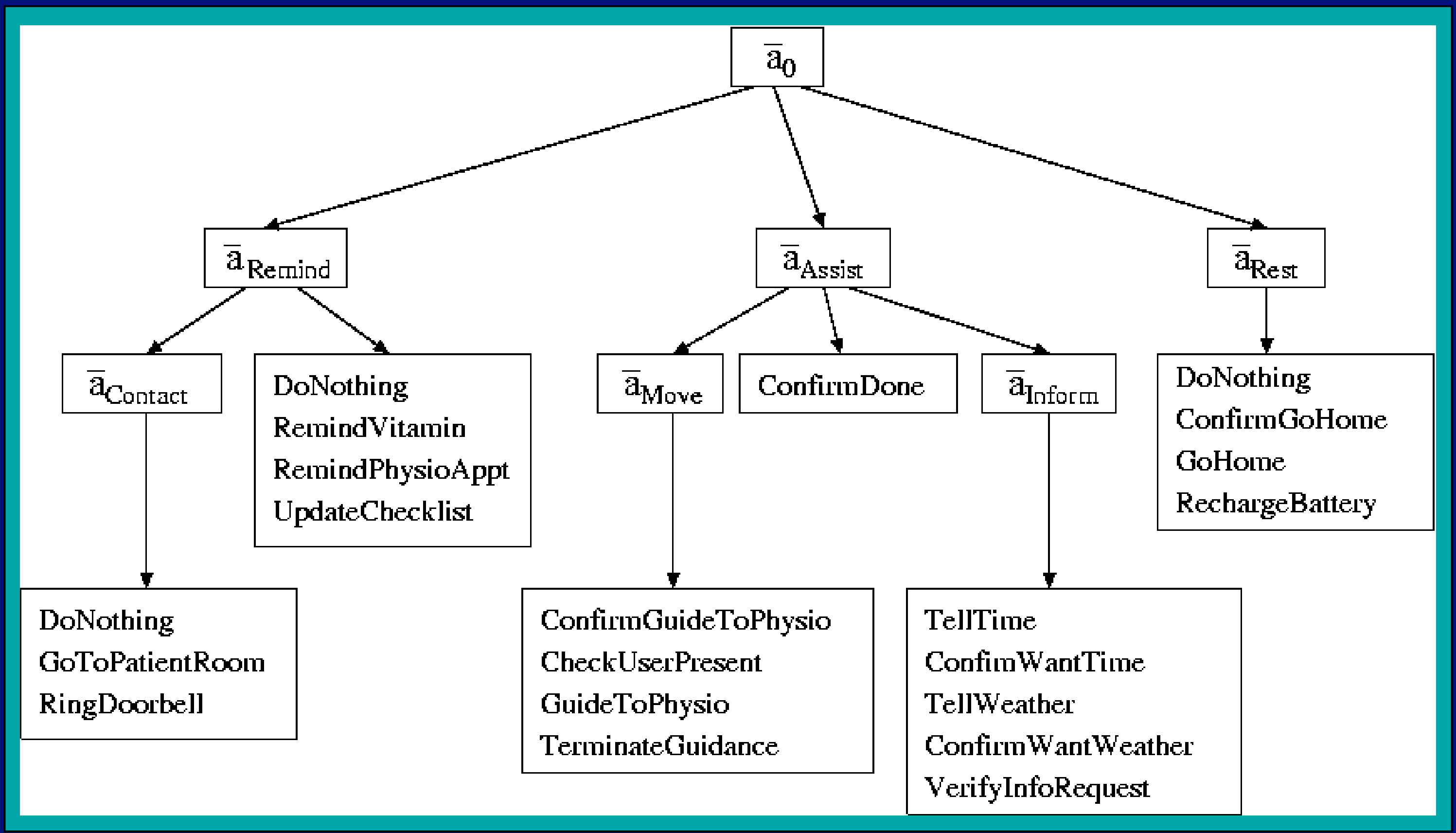
STATE FEATURES
X-position
Y-position
X-goal
Y-goal
HealthStatus

PARAMETERS
$\{b_h, T_h, O_h, R_h\}$

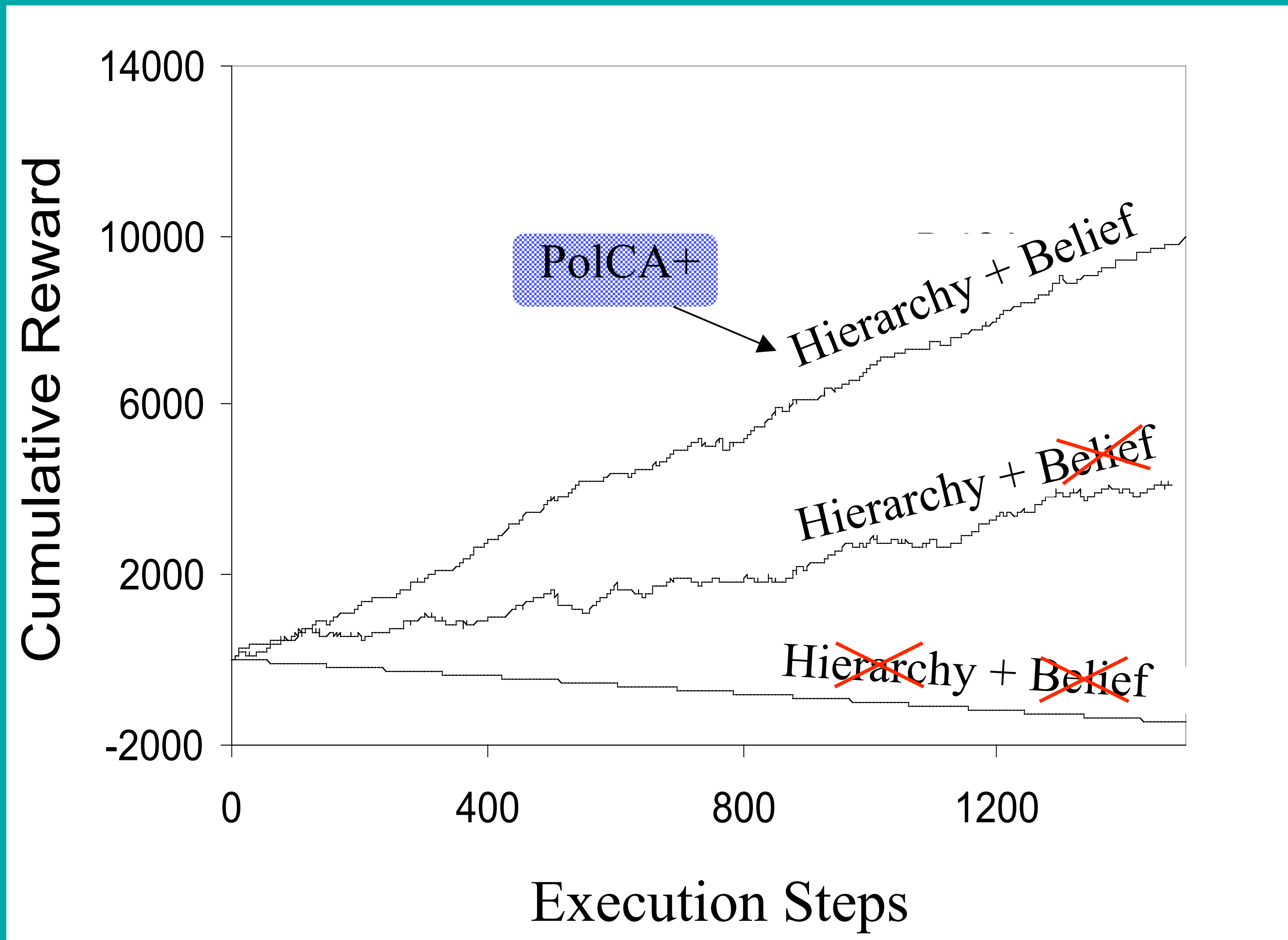
PLAN
π_h

PolCA+ in the Nursebot domain

- ***Goal***: A robot is deployed in a nursing home, where it provides reminders to elderly users and accompanies them to appointments.



Performance measure



Contributions of the PolCA+ algorithm

- Algorithmic:
  - New hierarchical approach for POMDP framework.
  - Automatic state and observation abstraction for POMDPs.
- Novel POMDP applications:
  - High-level robot control architecture.
  - Robust dialogue management.
- Theoretical:
  - For special case (fully observable), guarantees recursive optimality.

Summary

- Exact planning under uncertainty is hard.

→ Structure can help.

- Geometric structure (PBVI):
  – Solve “large” POMDPs by exploiting spatial distribution of beliefs.

- Hierarchical control structure (PolCA+):
  – Solve large POMDPs by divide-and-conquer.

Acknowledgments

Advisors:

Geoff Gordon, Sebastian Thrun.

Thesis committee:

Craig Boutilier, Michael Littman, Matthew Mason, Andrew Moore.

Nursebot team:

Greg Armstrong, Jacqueline Dunbar-Jacob, Sara Kiesler, Judith Matthews, Michael Montemerlo, Nicholas Roy, Martha Pollack.

Administrative support:

Jean Harpley, Suzanne Lyons Muth, Sharon Woodside.

Friends and family.

A visit to the nursing home

Nursebot Pearl

Assisting Nursing
Home Residents

Longwood, Oakdale, May 2001
CMU/Pitt/Mich Nursebot Project

Questions?

Future work

Belief-point planning:

- How can we handle domains with multi-valued state features?
- Can we leverage dimensionality reduction?
- Can we find better ways to pick belief points?

Hierarchical planning:

- Can we automatically learn hierarchies?
- How can we learn (or do without) pseudo-reward functions?

More generally:

- Incorporating parameter learning / user customization.
- More extensive field experiments.