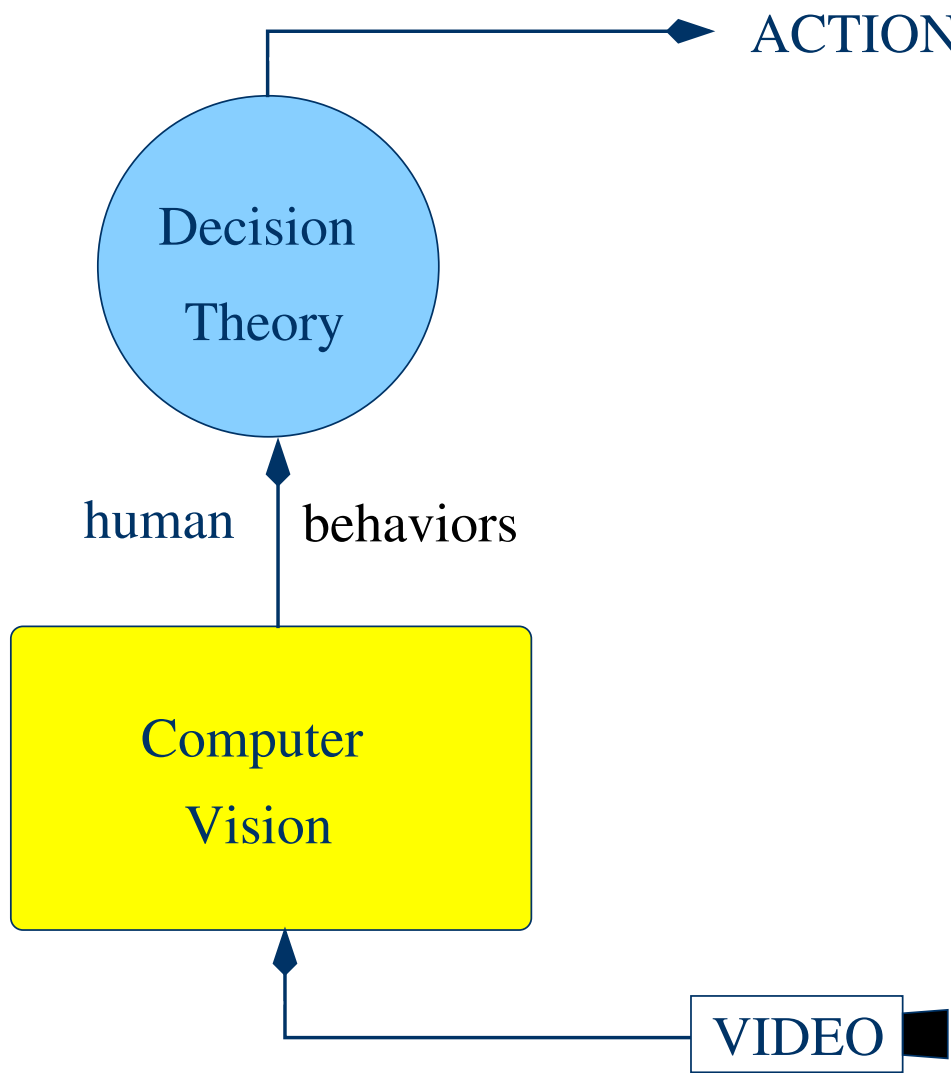


Learning Models of Human Behavior using a Value Directed Approach

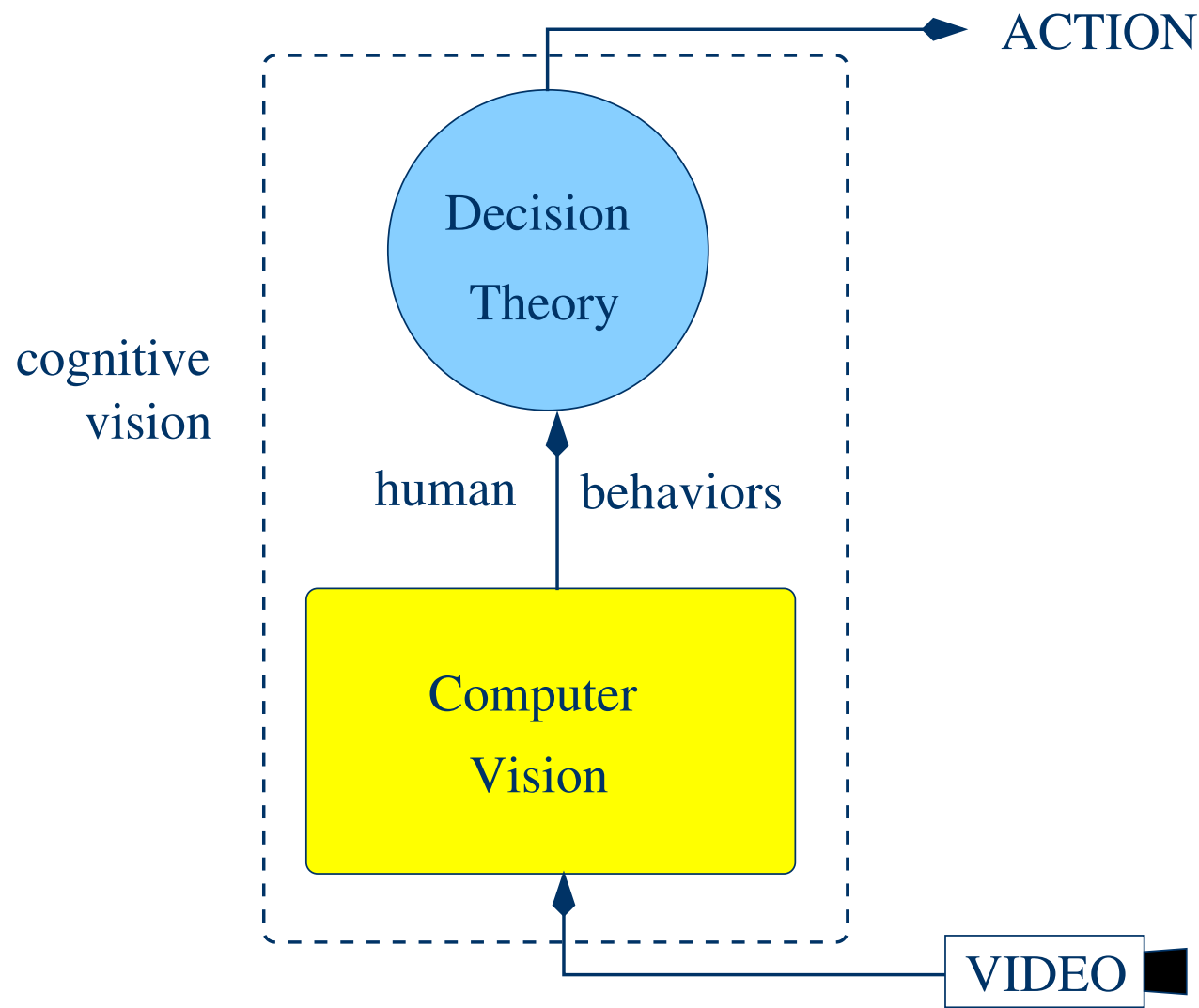
Jesse Hoey
Computer Science Department
University of Toronto

<http://www.cs.toronto.edu/~jhoey/>

Motivation: Modeling Human Behaviors



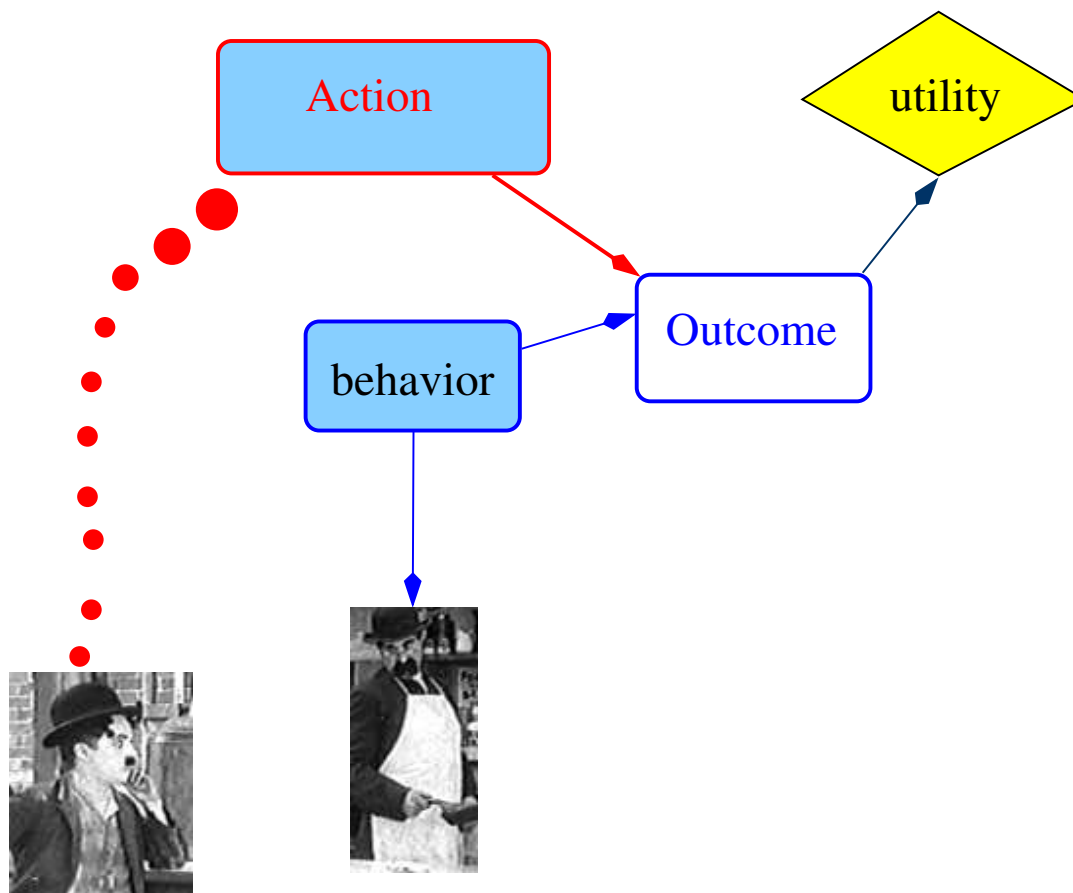
Motivation: Modeling Human Behaviors



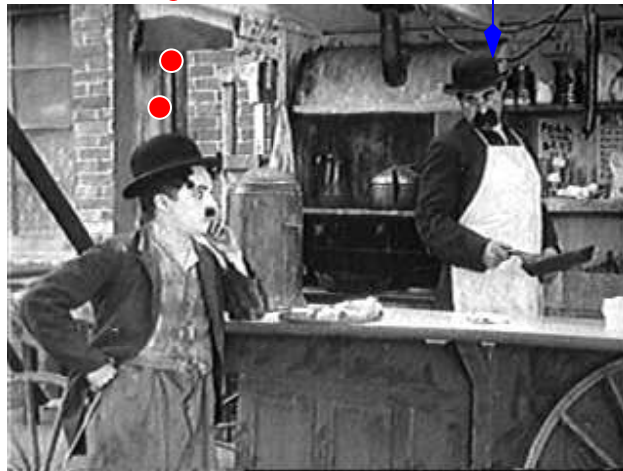
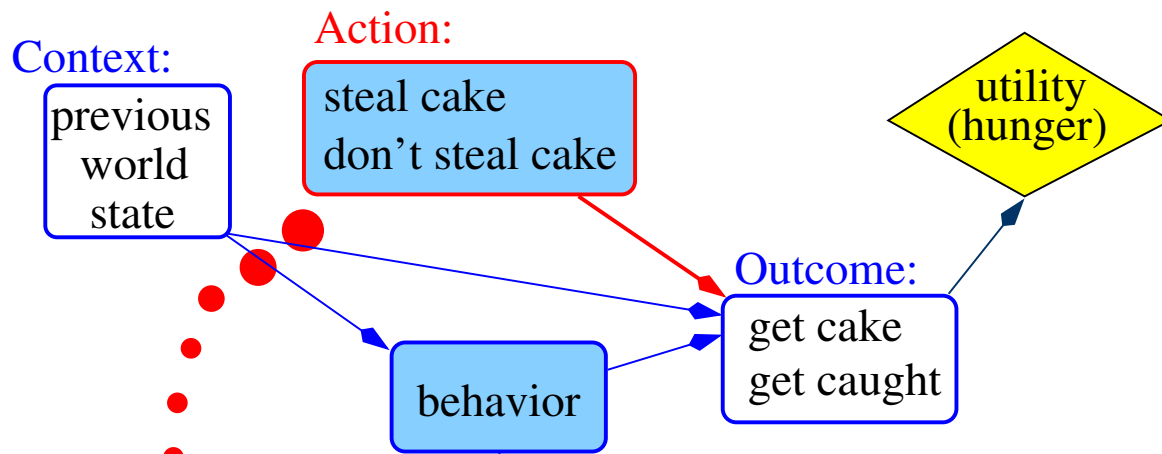
POMDPs for Human Behavior Understanding



POMDPs for Human Behavior Understanding



POMDPs for Human Behavior Understanding



Partially
Observable
Markov
Decision
Process

Overview

- POMDPs for Display Understanding in Context
- Computer Vision: Modeling video sequences
 - spatial abstraction
 - temporal abstraction
- Learning POMDPs
- Solving POMDPs
- Value-Directed Learning
- Experiments
 - Robot Control
 - Card Matching Game
- Conclusions, Current & Future Work

Partially Observable Markov Decision Processes (POMDPs)

A POMDP is a probabilistic temporal model of agent interacting with its environment :

a tuple $\langle S, A, T, R, O, B \rangle$

S : finite set of unobservable states

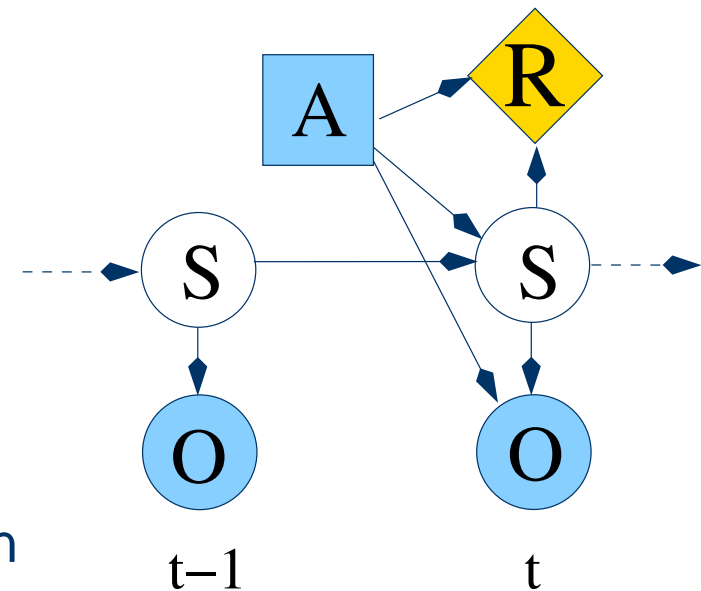
A : finite set of agent actions

$T : S \times A \rightarrow S$ transition function

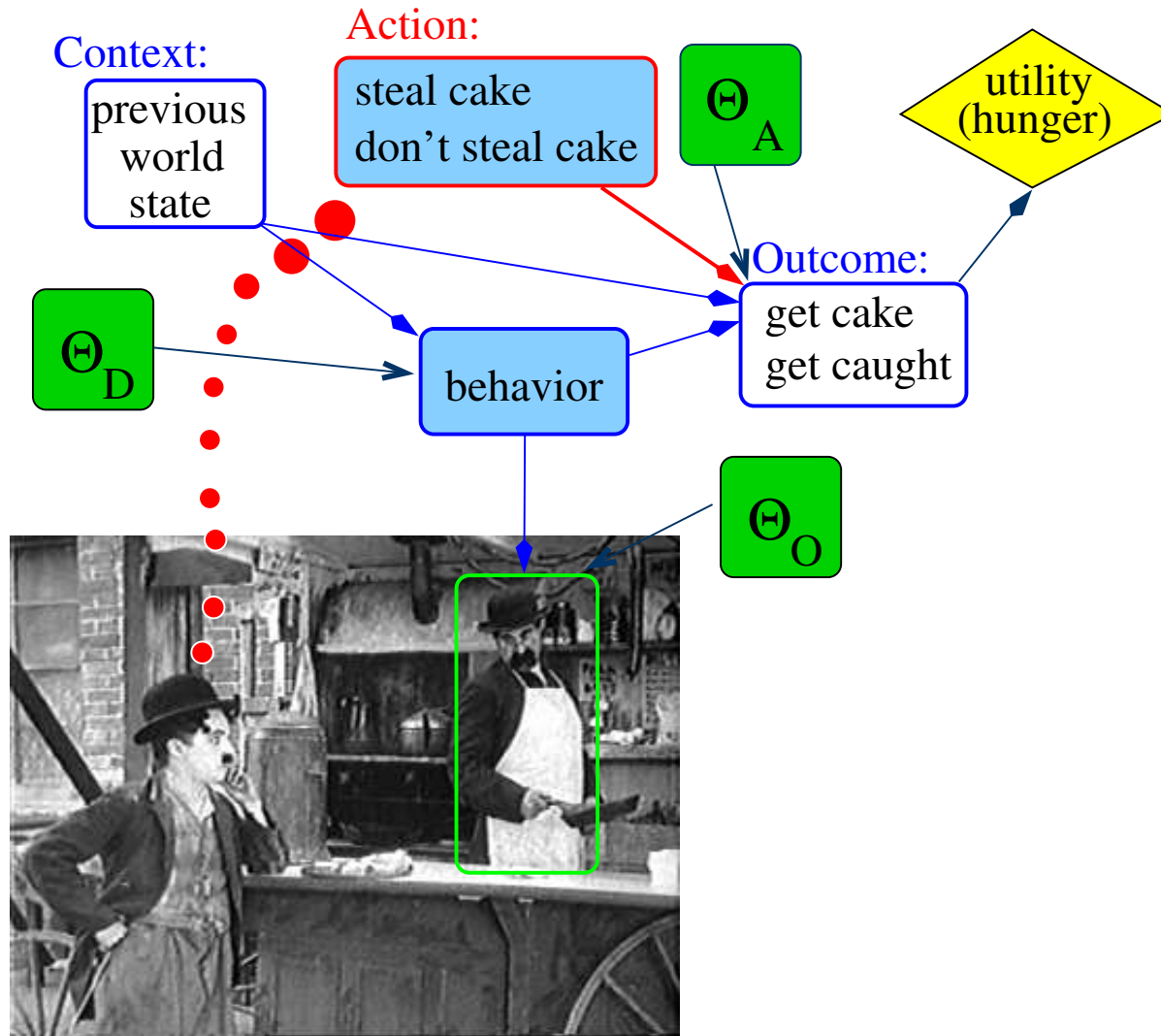
$R : S \times A \rightarrow \mathcal{R}$ reward function

O : set of observations

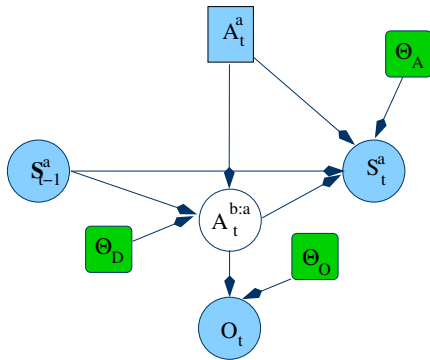
$B : S \times A \rightarrow O$ observation function



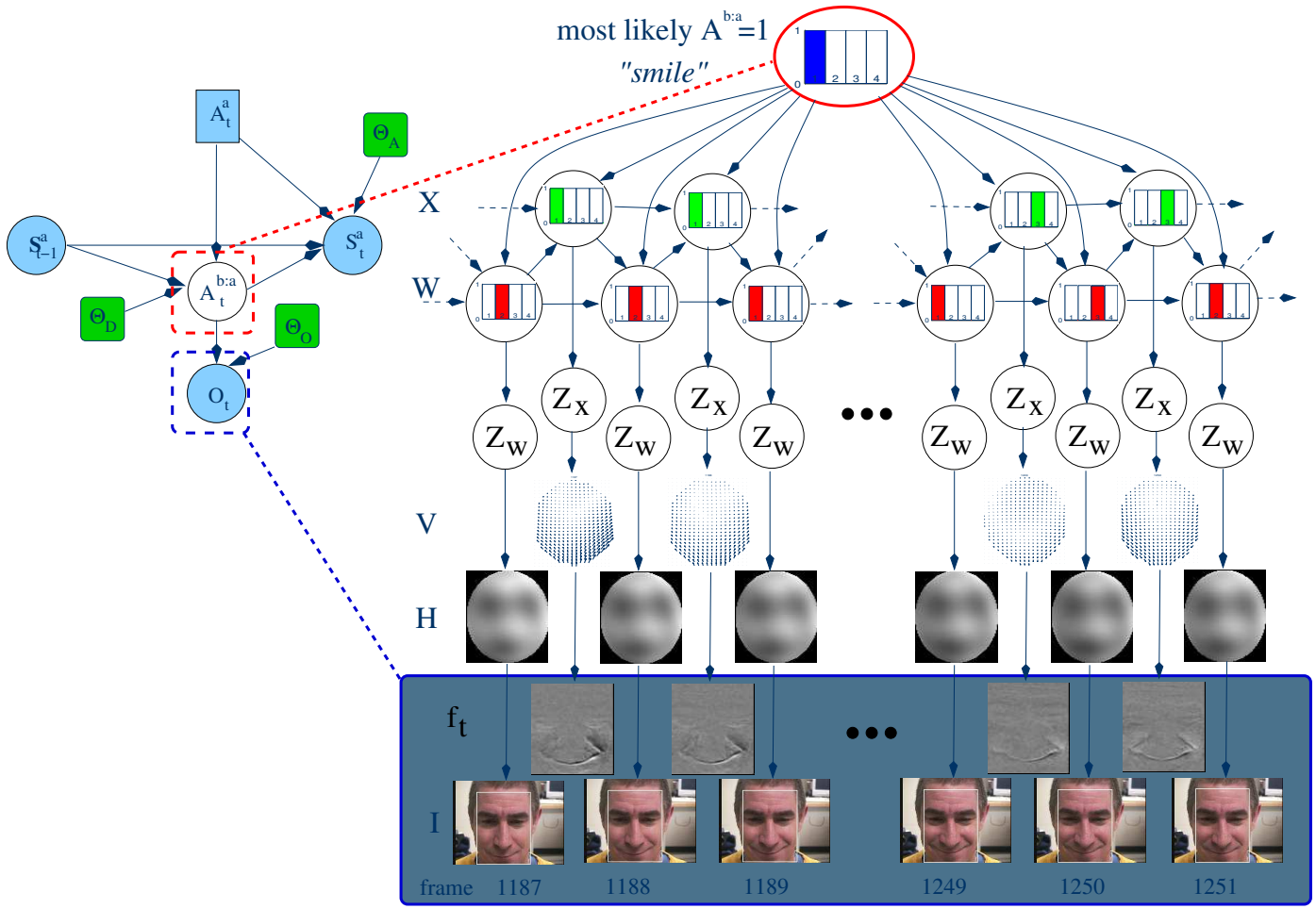
POMDPs for Human Behavior Understanding



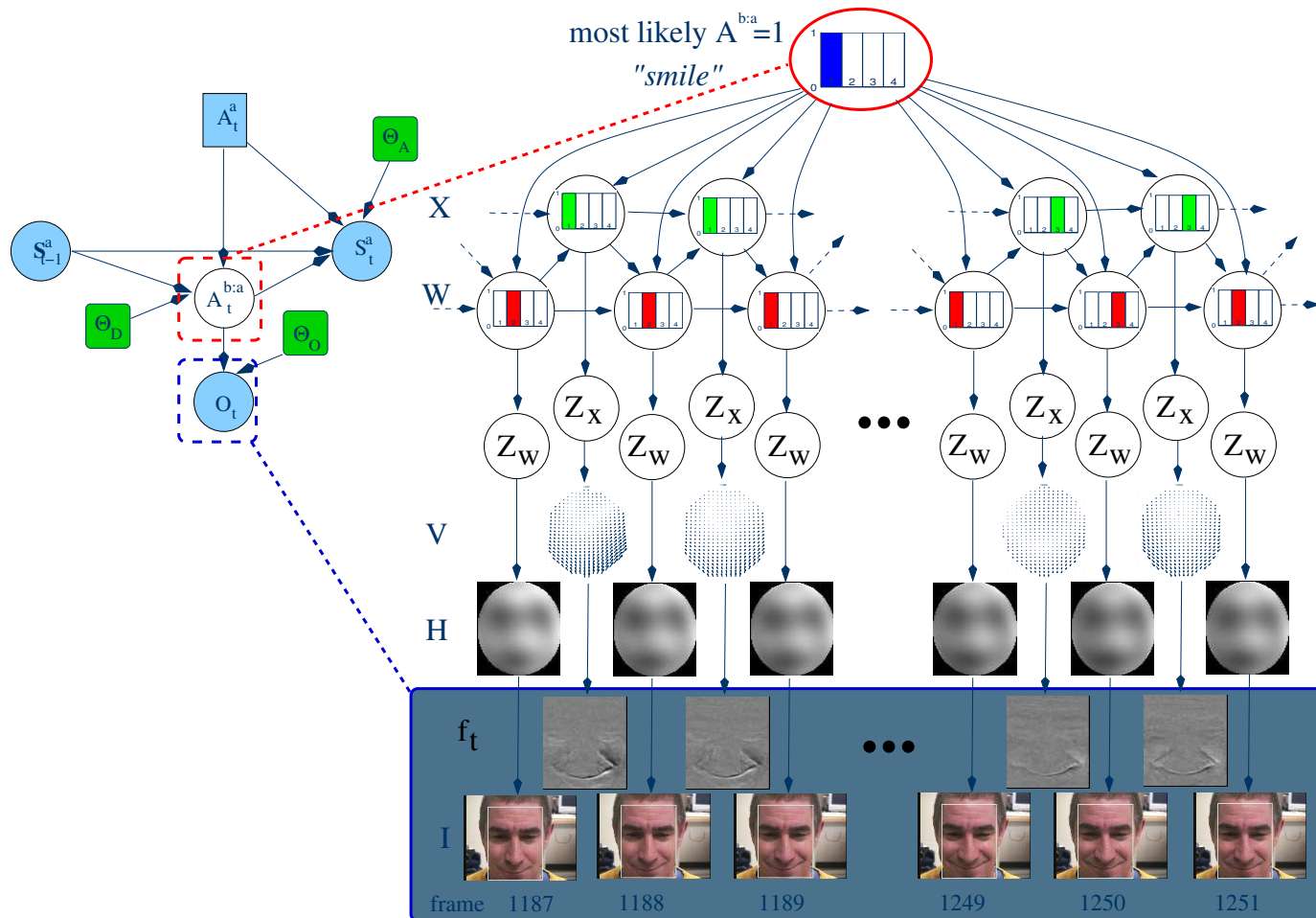
Output Model



Output Model

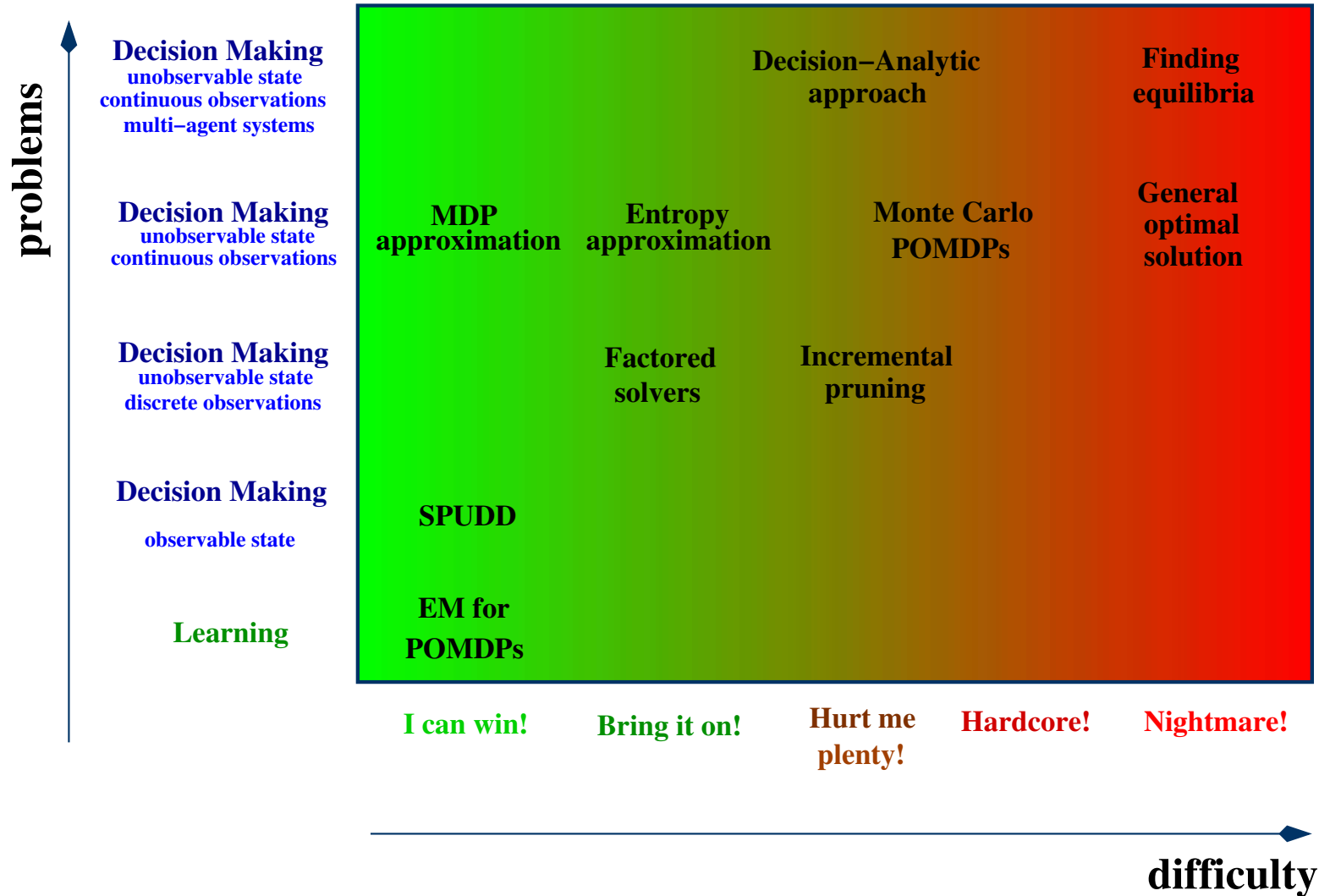


Output Model

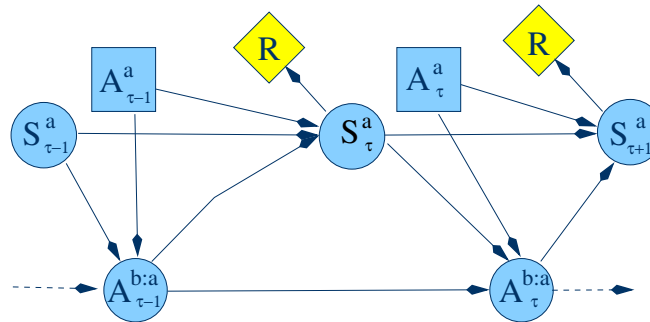


$$P(O|A^{b:a}) = \sum_{ij} P(I_T|W_{T,i}A^{b:a})P(\nabla f_T|X_{T,j}A^{b:a}) \sum_{kl} \Theta_{Xijkn} \Theta_{Wjklm} P(X_{T-1,k}W_{T-1,l} \{O\}_{1,T-1} | A^{b:a})$$

Solution Techniques



Solving the Model



MDP Approximation: Assume $A^{b:a}$ is observable

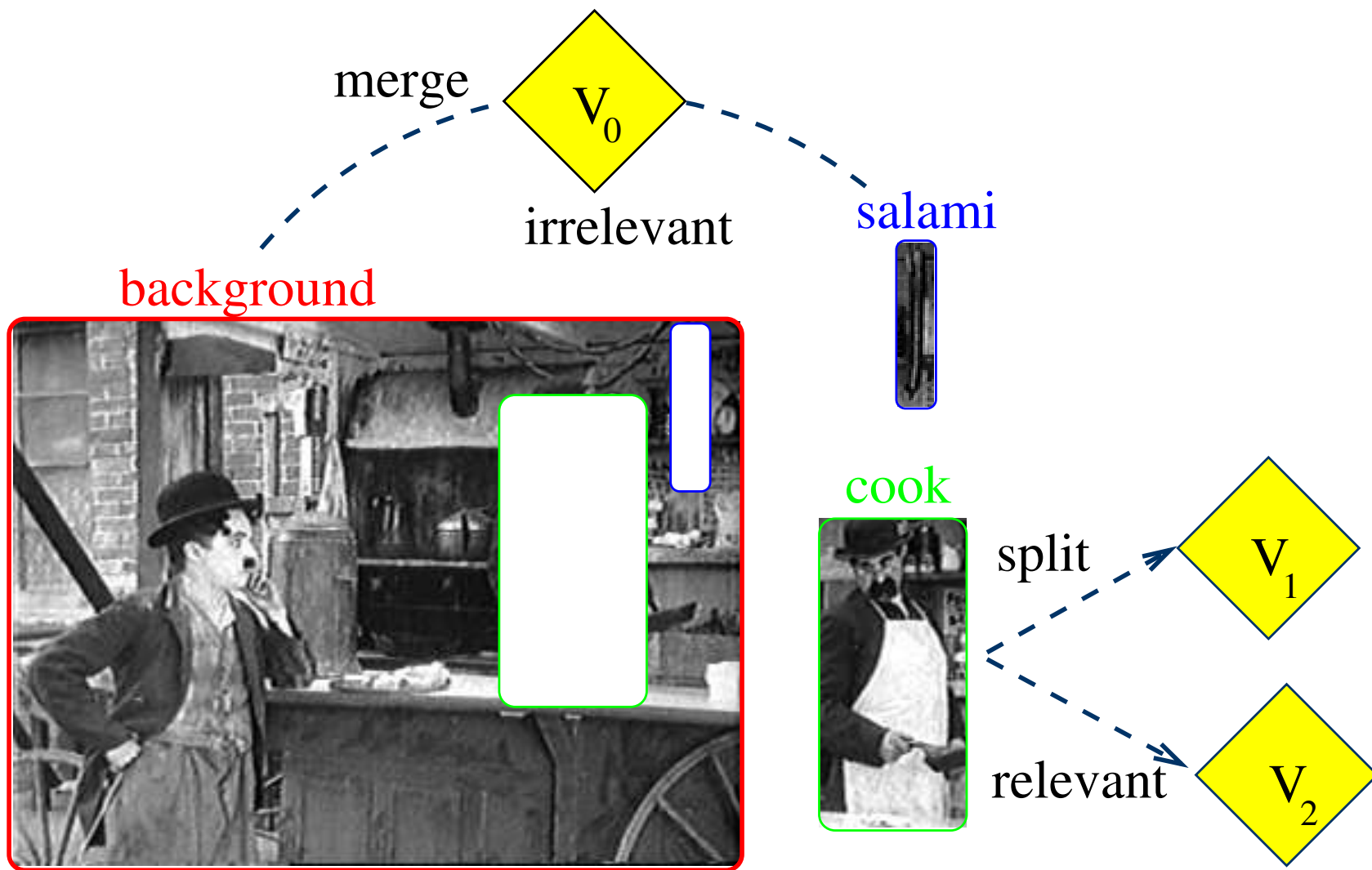
Dynamic Programming: Value Iteration

$$V^{n+1}(s) = R(s) + \max_{a \in \mathcal{A}} \left\{ \sum_{t \in \mathcal{S}} Pr(t|a, s) \cdot V^n(t) \right\}, \quad V^0 = R$$

n-stage to go Policy: actions that maximize expected value

$$\pi^n(s) = \arg \max_{a \in \mathcal{A}} \left\{ \sum_{t \in \mathcal{S}} Pr(t|a, s) \cdot V^n(t) \right\}$$

Value Directed Structure Learning



Value Directed Structure Learning

State merging:

repeat

- 1.learn the POMDP model
- 2.compute value functions for behaviors
- 3.compute distance between value functions
- 4.if policies agree,
merge behaviors closest in value

until number of behaviors stops changing

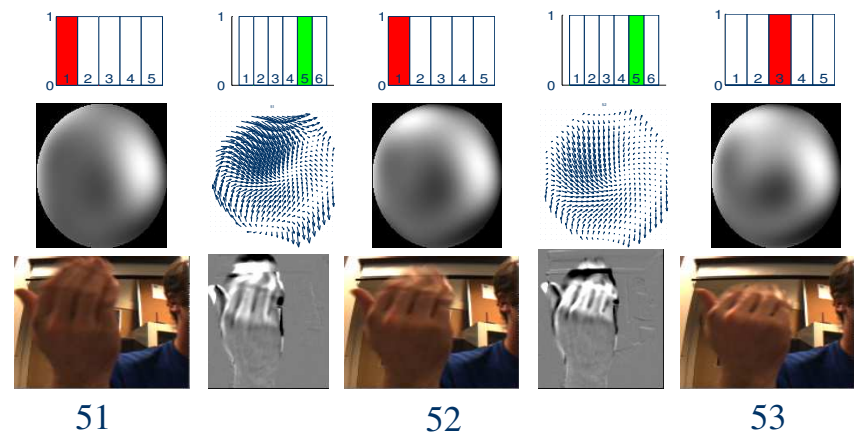
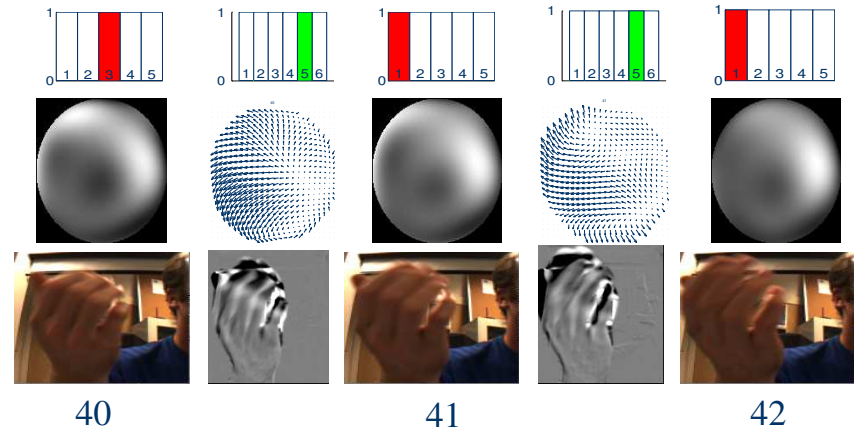
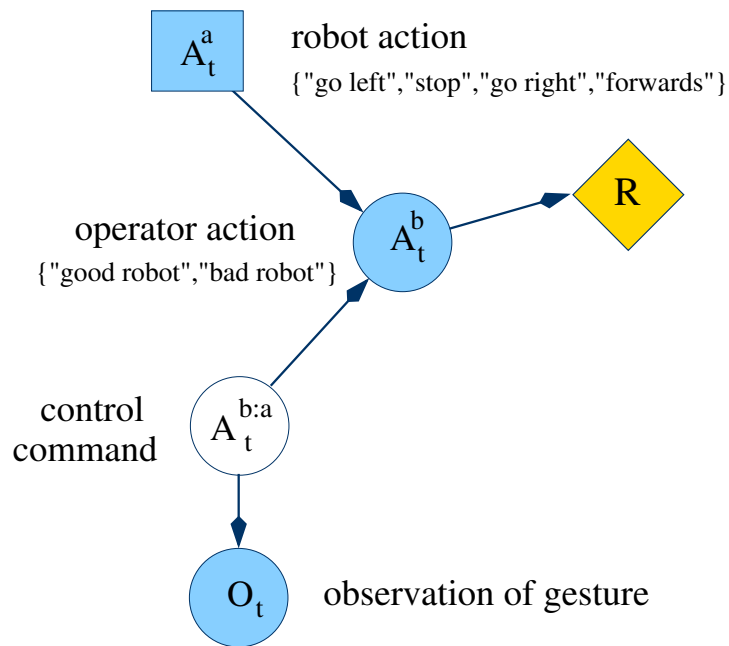
State splitting

repeat

- 1.learn the POMDP model
- 2.examine states for predicitive power - entropy?
- 3.Split behaviors which predict
different outcomes

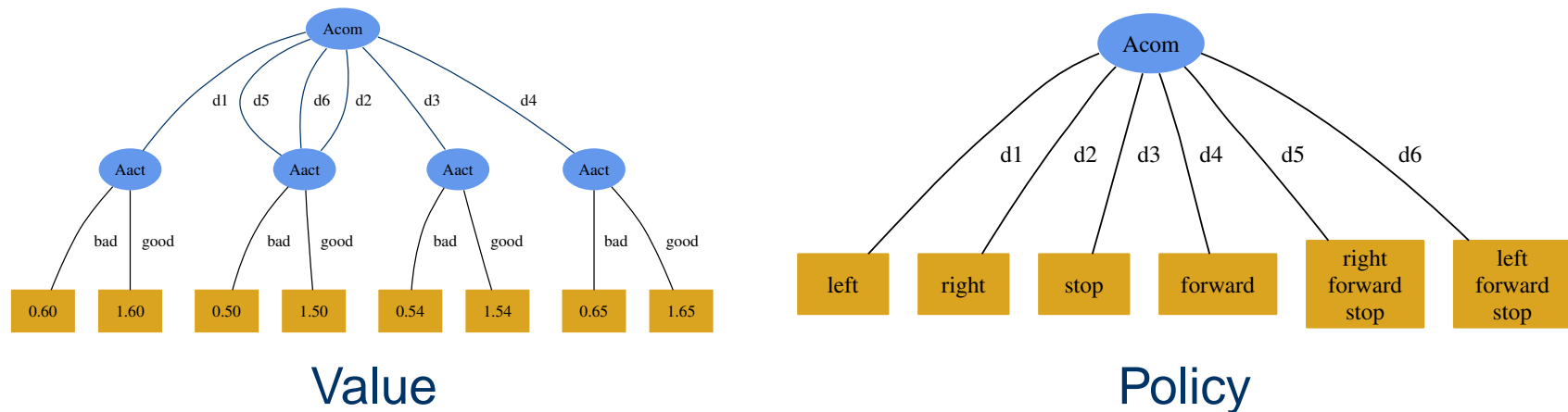
until number of behaviors stops changing

Experiments: Robot Control Gestures



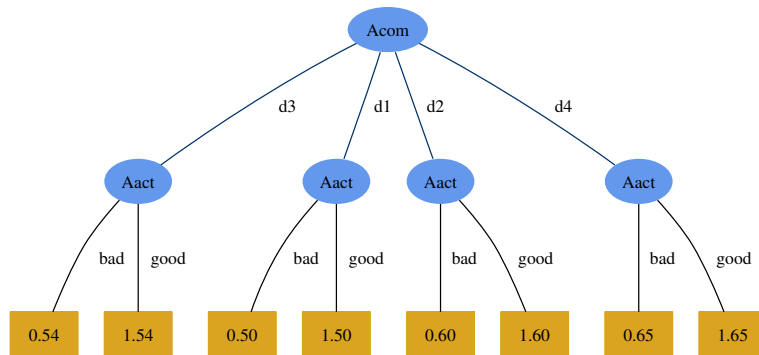
Value-Directed Structure Learning

Part of Value function & policy for robot control:

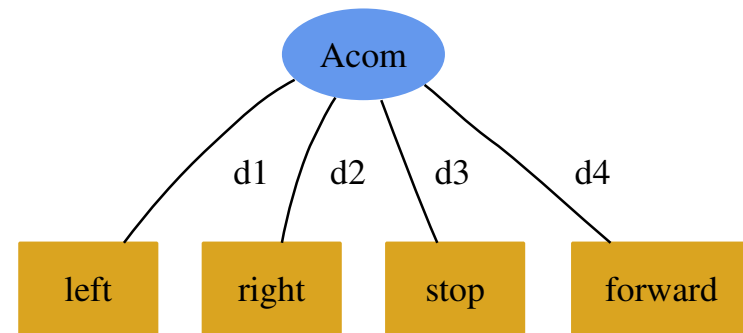


- Some states of $Acom$ are redundant
- detect & merge using state aggregation in the policy and value function
- re-compute policy

Value-Directed Structure Learning



Value



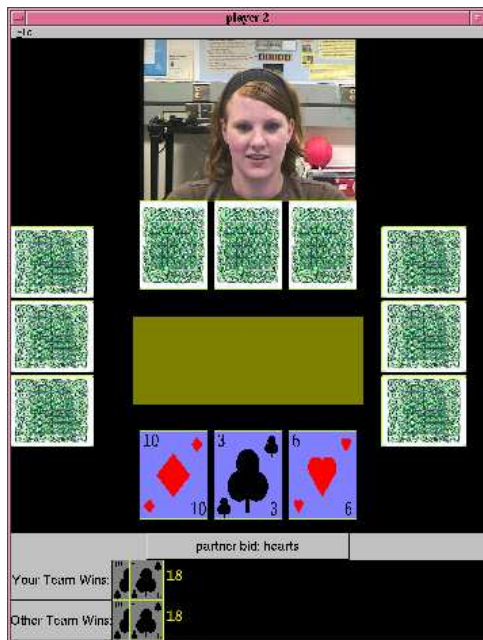
Policy

- Leave-four-out cross validation (12 times)
- Take actions and accumulate rewards
- Success rate: $47/48 = 98\%$ or 11/12 correct policies
- Merges to 4 states all 12 times.

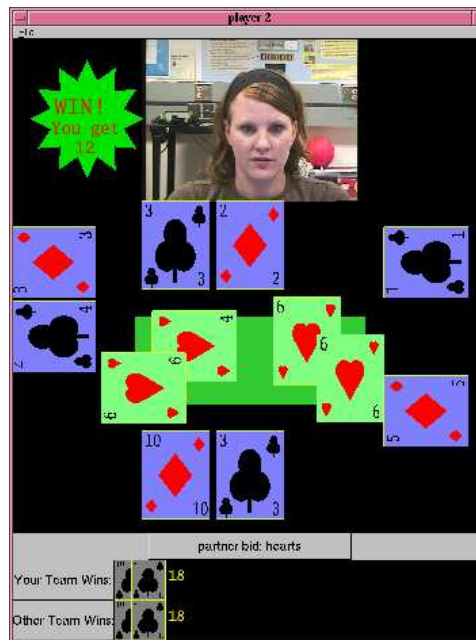
Experiments: Card Matching Game

Cooperative two player game

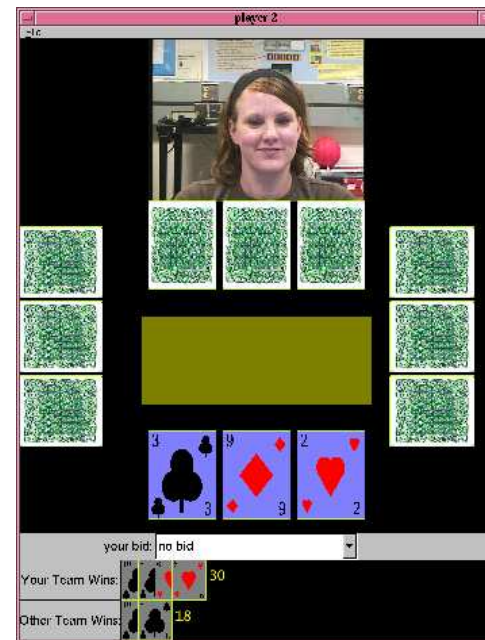
Goal: match cards



stage 1



stage 2



stage 3

Card Matching Results

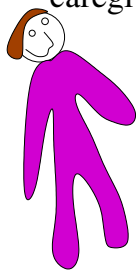
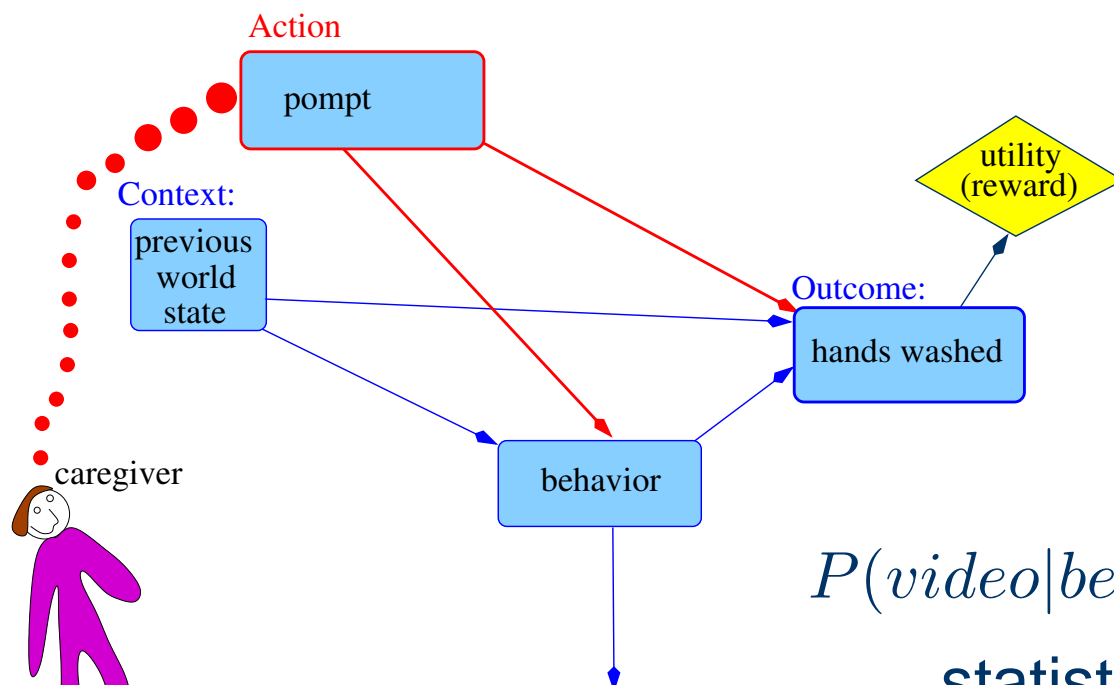
3 behaviors identified: nodding, shaking, null

Predicts:

- 6/7 human actions in test data
- 19/20 human actions in training data.

Errors: lack of POMDP data, temporal segmentation problems

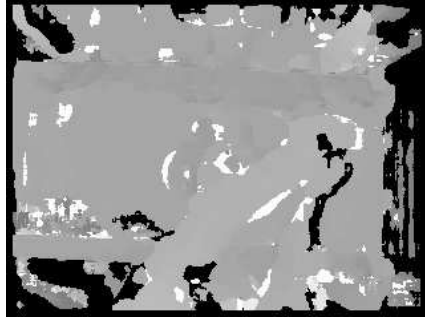
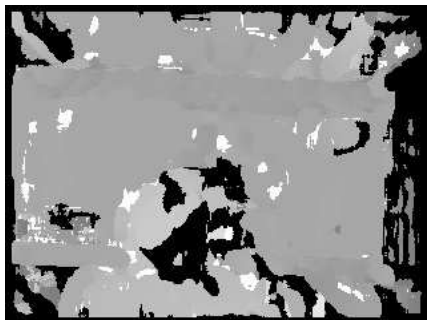
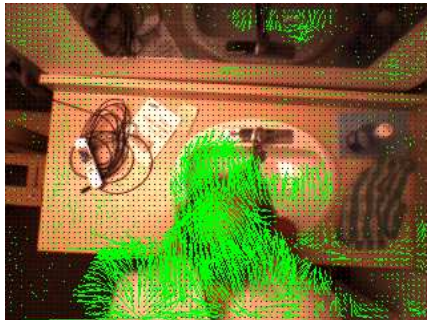
Handwashing Behavior Understanding



$$P(\text{video}|\text{behavior})$$

statistically significant
value directed

Difficult Cases



self-occlusion

objects appear to merge

object occlusion

Conclusions

- Computer Vision + Probabilistic Models + Decision Theory
- Learning *purposeful* human behavior models from *unlabeled* data.
- System is general and portable - no reliance on expert knowledge
- Applications: HCI, surveillance, assisted living, driver support
- Future work
 - Spatial segmentation and representation + tracking
 - Multimodal observations
 - Temporal segmentation
 - POMDP solutions
 - Value-directed learning (Hoey & Little, CVPR 2004)