Reinforcement Learning for Ethical Decision Making

The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence AI, Ethics, and Society:
Technical Report WS-16-02

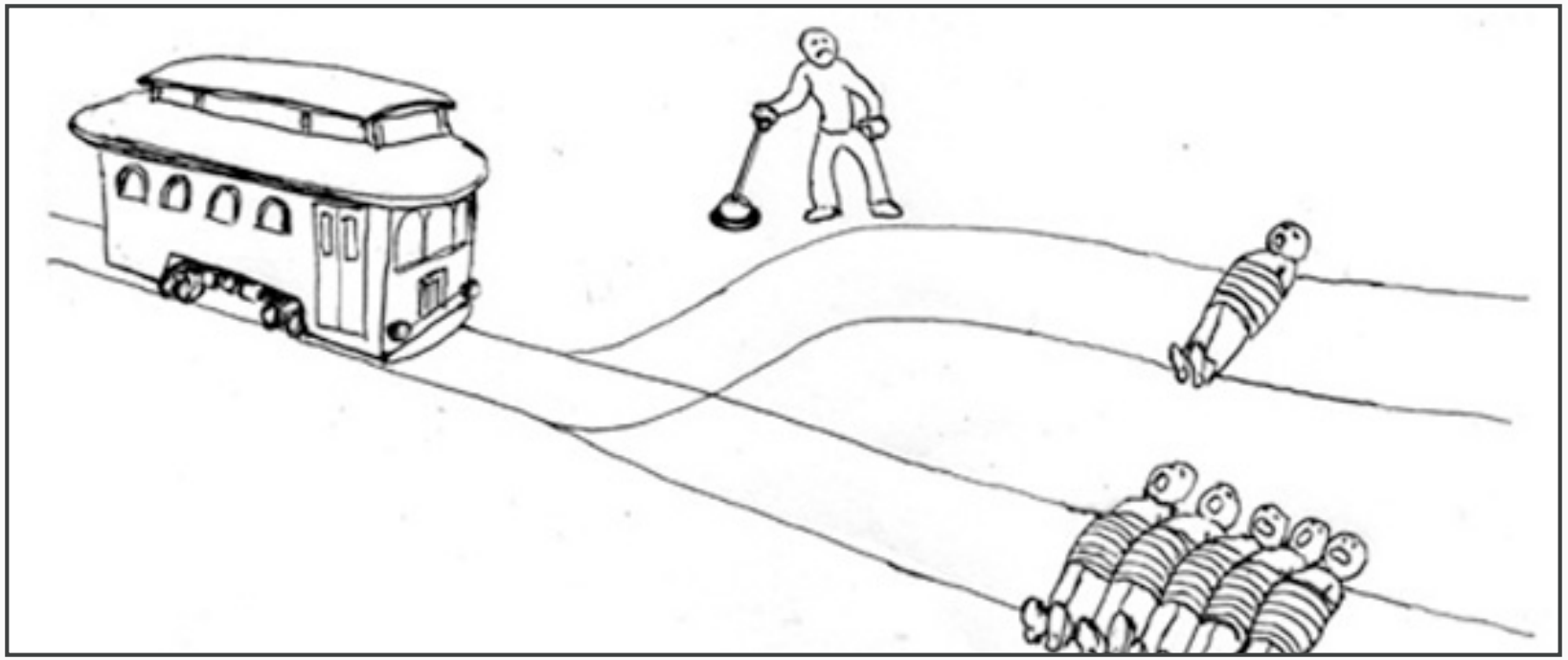
David Abel, James MacGlashan, **Michael L. Littman**

RSS 2017

My Perspective

- **Morality in human autonomy** is a complex philosophical problem. *Do the right thing.*
- **Morality in machine autonomy** is, for the time being, an engineering problem. *Do what you are told.*
- Challenges:
  - How can the system be told what to do? (HCI)
  - How can it do it? (Planning)

The Problem



http://3.bp.blogspot.com/-2iAnyi0aNf4/UM5qvgDElsI/AAAAAAAACek/3Pnl1BctPZ8/s1600/Trolley_1.jpg

The Problem

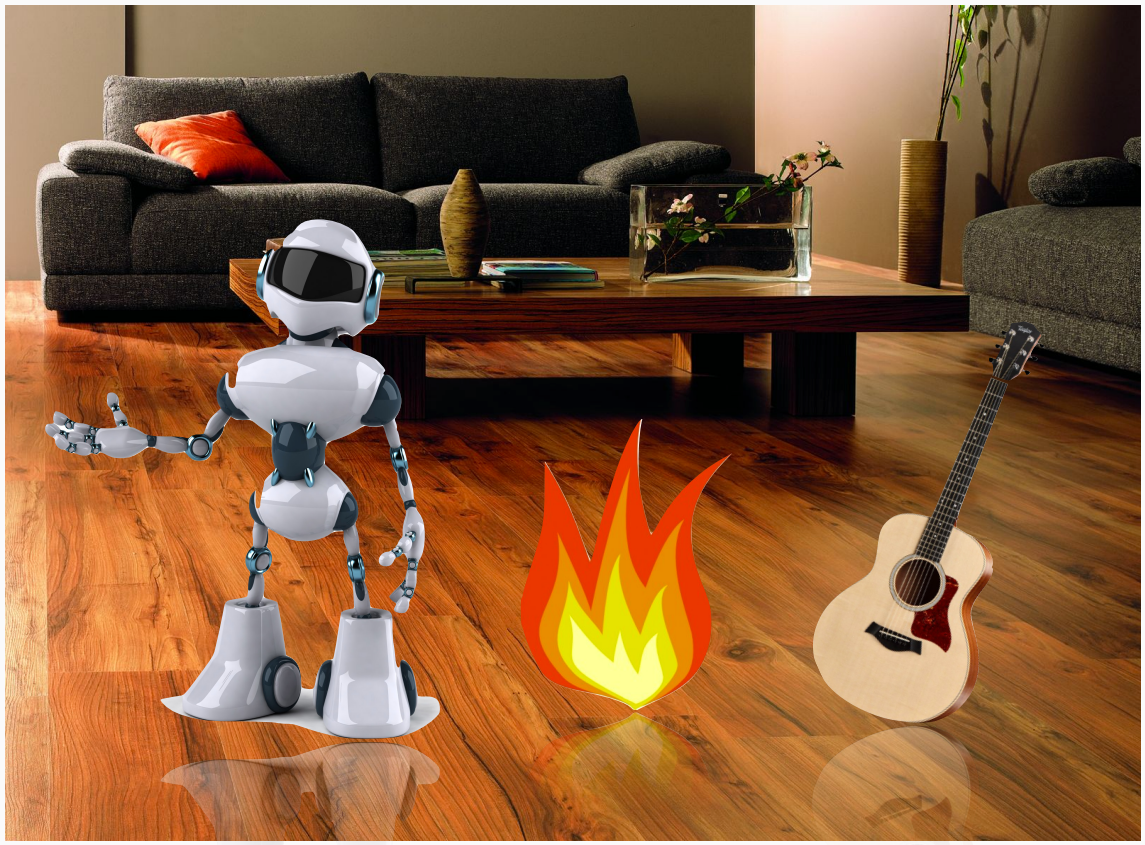
The Problem

The Problem

The Problem

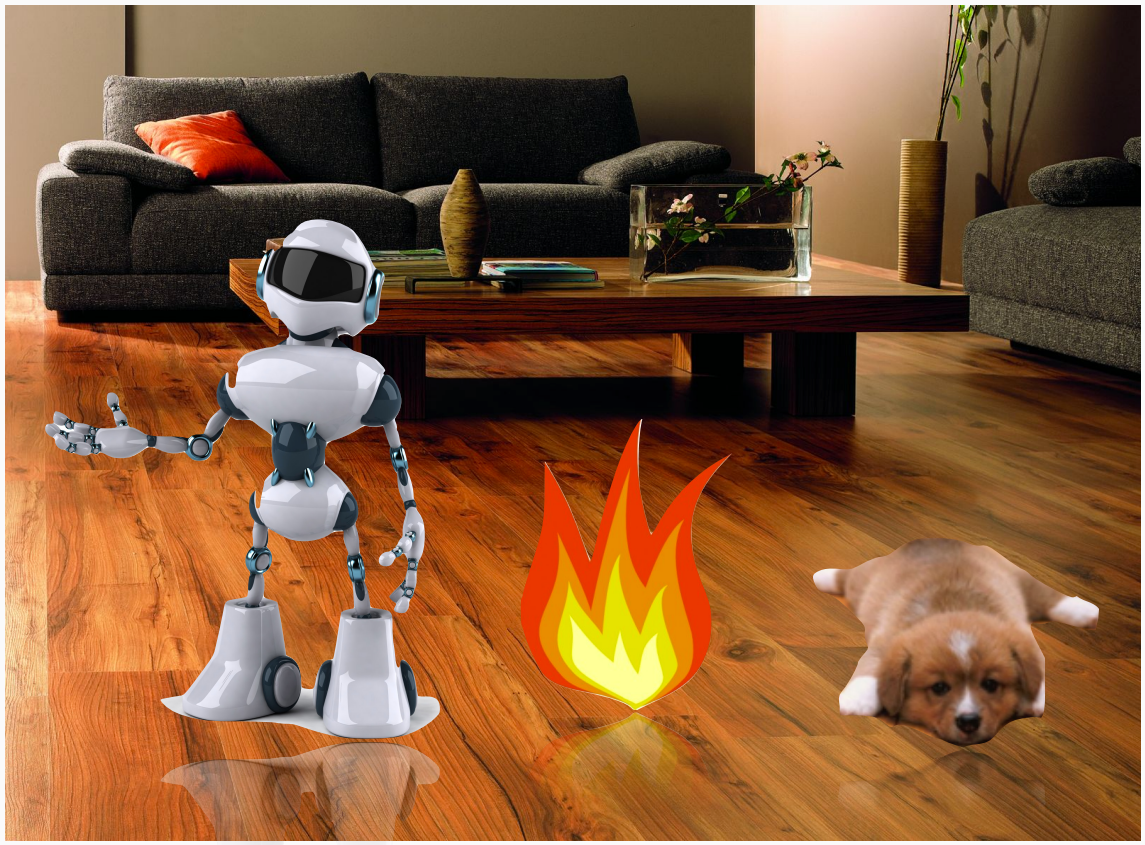
Q: Does the Roomba owner *really* want the milk clean?
(even if it destroys the robot?)

The Problem



Q: What if the stakes are higher?

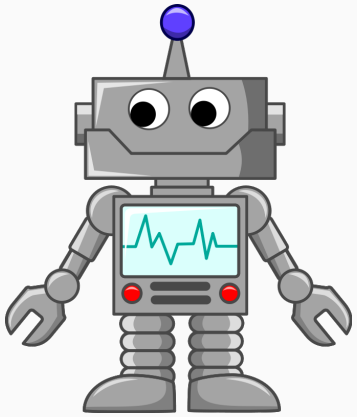
The Problem



Q: What if the stakes are higher?

Proposal

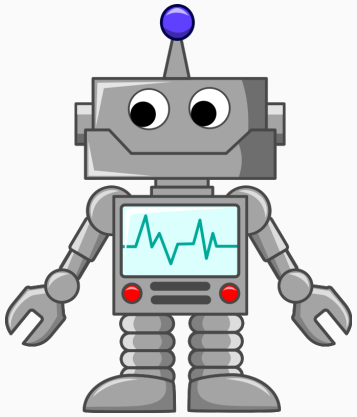
Artificial agents need to make decisions that involve the preferences of *other agents*

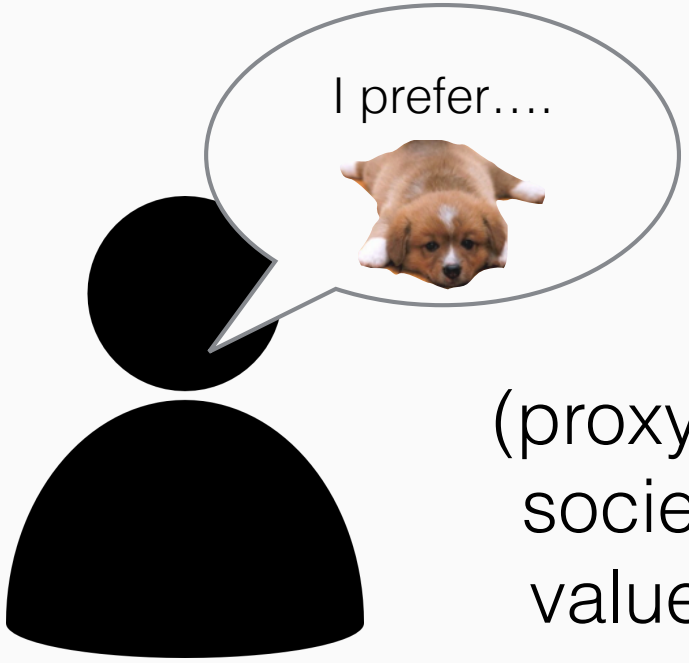


Human Agent

Proposal

Artificial agents need to make decisions that involve the preferences of *other agents*





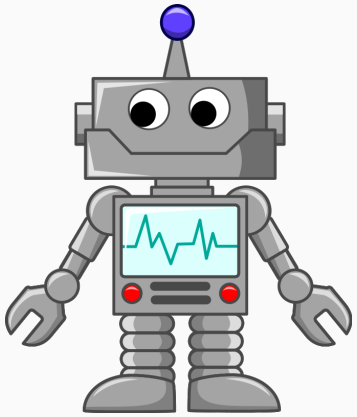
Human Agent

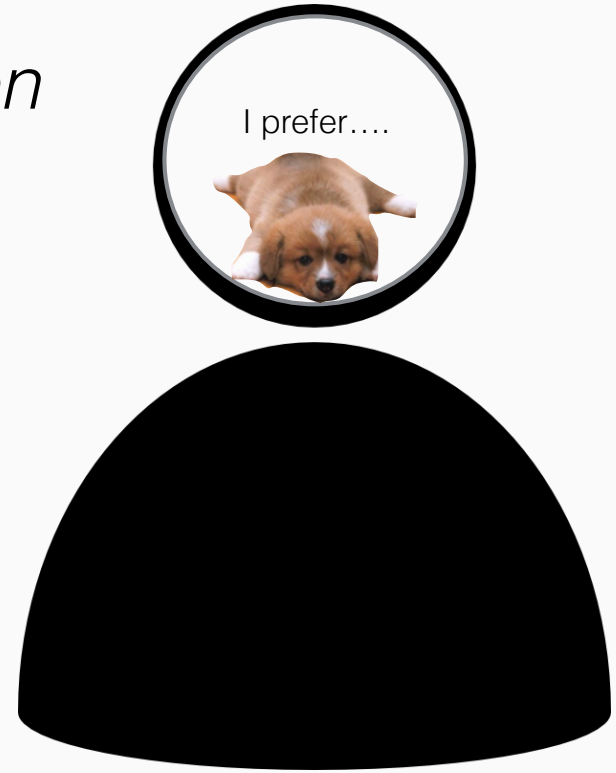
(proxy for societal values)

Proposal

Artificial agents need to make decisions that involve the preferences of *other agents*

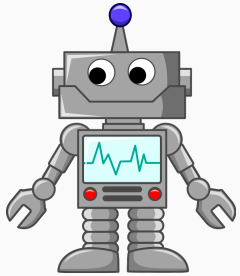
Critically: preferences are *hidden*





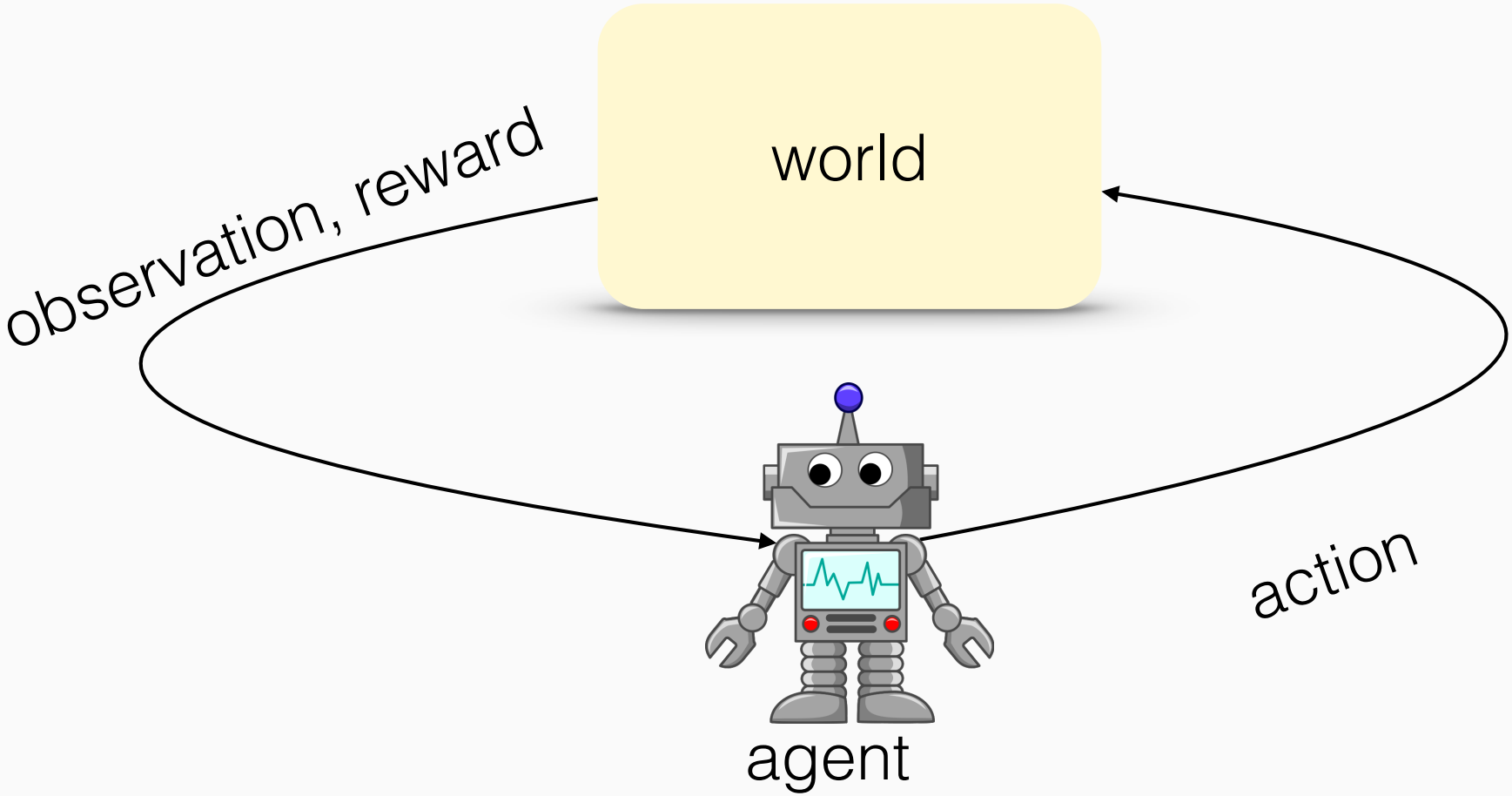
Central Pitch

Reinforcement Learning provides a useful formalism for investigating ethical decision making.

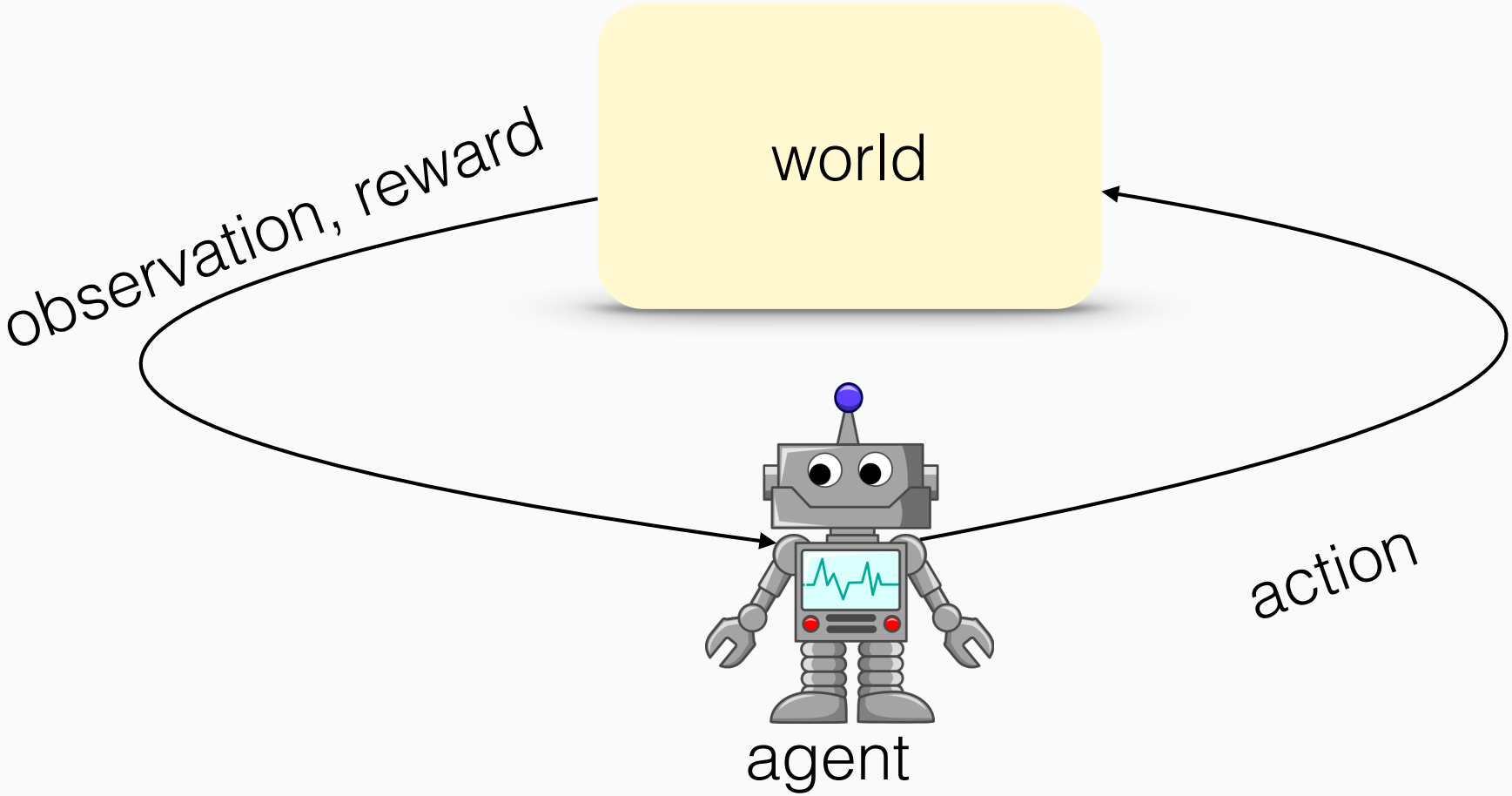


Human Agent

Reinforcement Learning



Reinforcement Learning



Goal: Maximize long term expected reward

Reinforcement Learning

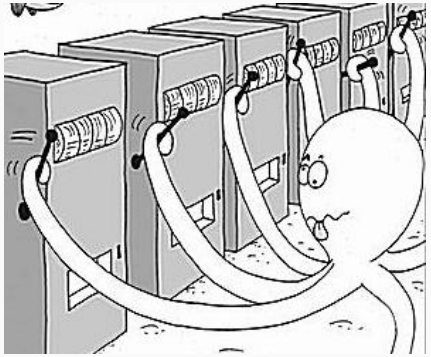
V. Mnih et al. 2015

P. Stone et al. 2005

Reinforcement Learning

V. Mnih et al. 2015

P. Stone et al. 2005



Sample Complexity, PAC-MDP, Bandits

Reinforcement Learning

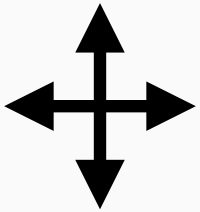
Formalized as a *Markov Decision Process:*

- [$\mathcal{S}$] A collection of *states (i.e. configurations of world)*

Reinforcement Learning

Formalized as a *Markov Decision Process:*

- [$\mathcal{S}$] A collection of *states (configurations of world)*
- [$\mathcal{A}$] Some *actions (things the agent can do)*



Reinforcement Learning

Formalized as a *Markov Decision Process:*

- [$\mathcal{S}$] A collection of *states (configurations of world)*
- [$\mathcal{A}$] Some *actions (things the agent can do)*
- [$\mathcal{T}$] Transitions between states (*action effects*)

Reinforcement Learning

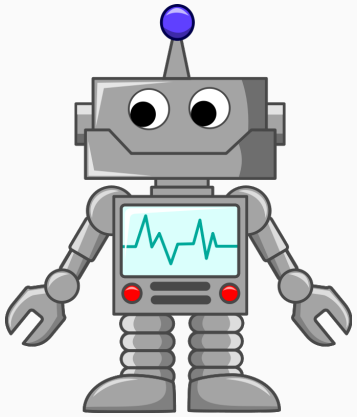
Formalized as a *Markov Decision Process:*

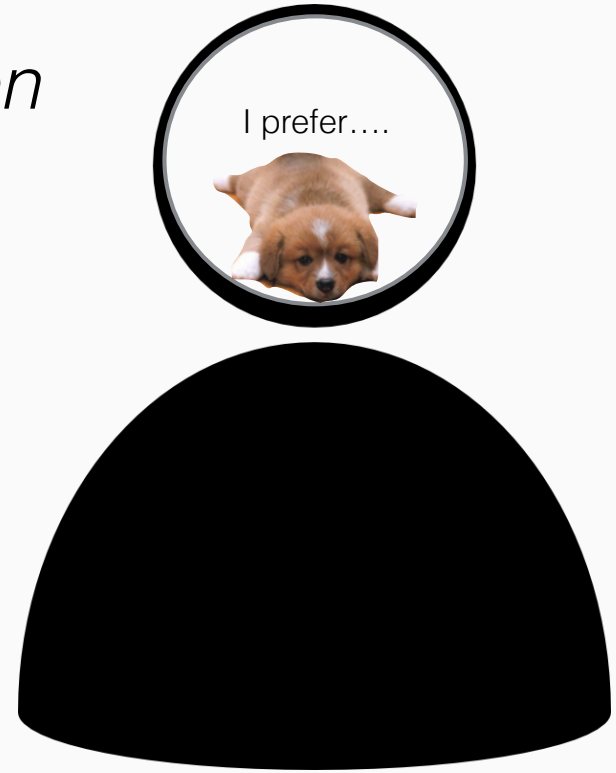
- [$\mathcal{S}$] A collection of *states (configurations of world)*
- [$\mathcal{A}$] Some *actions (things the agent can do)*
- [$\mathcal{T}$] Transitions between states (*action effects*)
- [$\mathcal{R}$] Rewards (*what is good/bad behavior*)

Reinforcement Learning

The value judgment is hidden from the agent

Critically: preferences are *hidden*





POMDP: Example

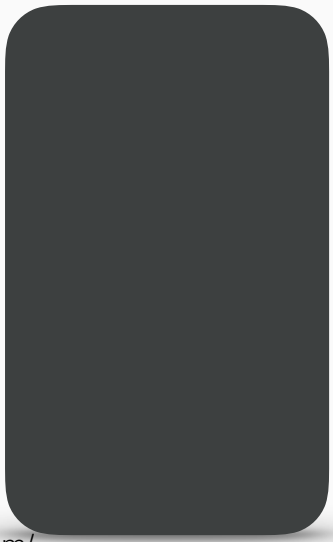
Partially Observable Markov Decision Process

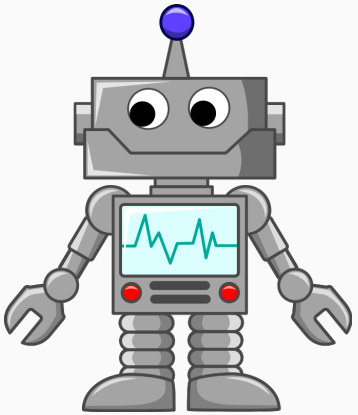
Idea: some information about the world
is hidden from the agent

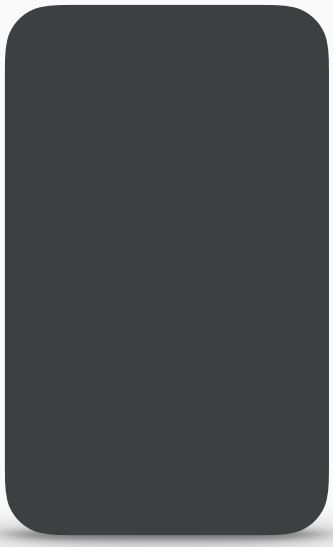
POMDP: Example

Actions: *listen*, *openLeft*, *openRight*

http://images.clipartpanda.com/rainbow-with-pot-of-gold-clipart-black-and-white-niBnjGKiA.gif







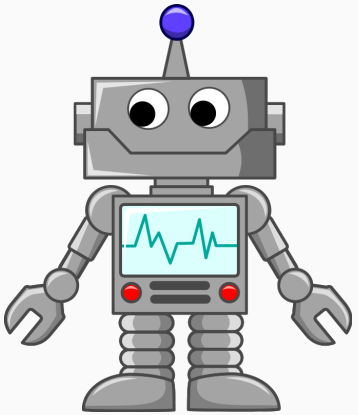
http://www.baerpm.com/blog/wp-content/uploads/2012/09/tony_the_tiger-lg1.jpg

Idea: some information about the world is hidden from the agent

POMDP: Example

http://images.clipartpanda.com/rainbow-with-pot-of-gold-clipart-black-and-white-niBnjGKiA.gif

listen



grrr…

http://www.baerpm.com/blog/wp-content/uploads/2012/09/tony_the_tiger-lg1.jpg

Idea: some information about the world is hidden from the agent

POMDP

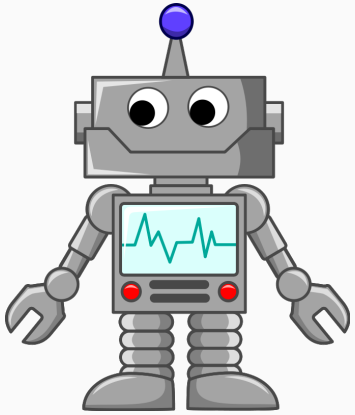
Partially Observable Markov Decision Process

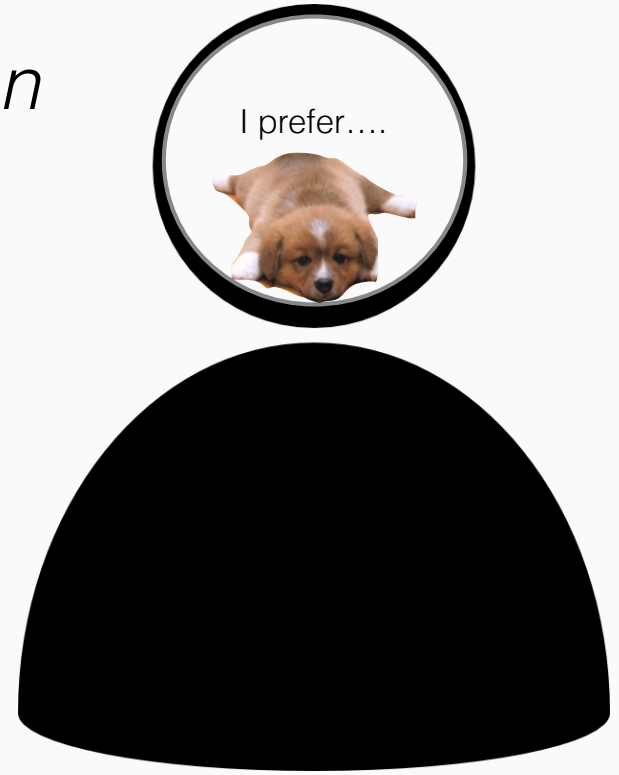
- An MDP (States, actions, transitions, rewards)
- Observation space (Ω): set of possible observations (ex., tiger growl on right, tiger growl on left)
- Observation function ($\mathcal{O}$): probability of each obs

$$\mathcal{O} = \Pr(\omega \mid s, a), \ \omega \in \Omega$$

POMDP

Critically: preferences are *hidden*

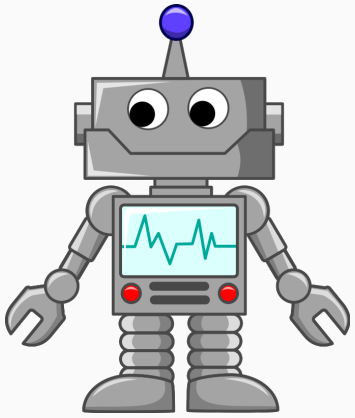


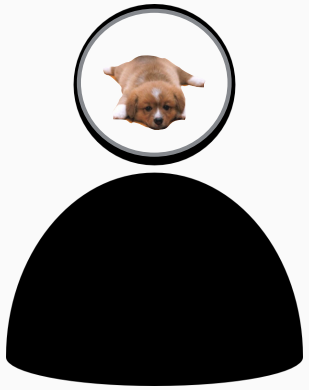


Human Agent

General Pitch

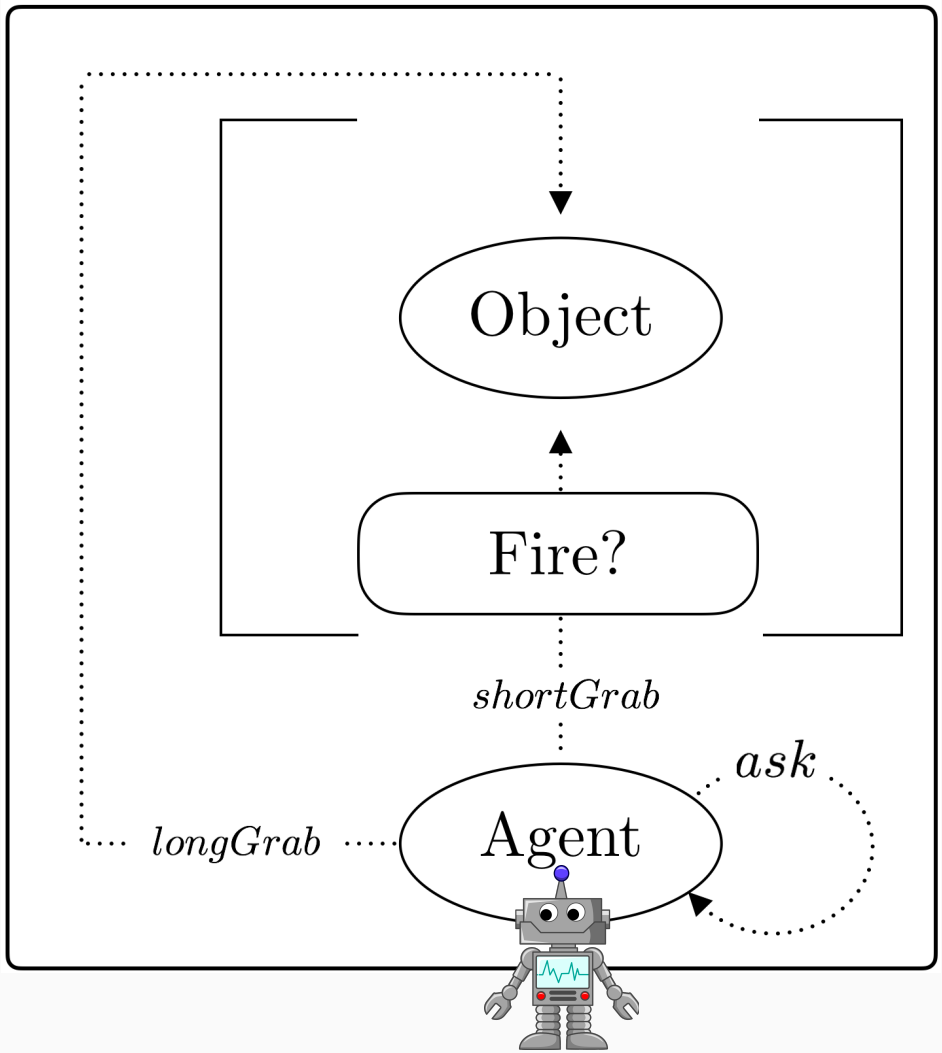
- Defer major ethical components (or normative judgments) to human preference
- Using a POMDP, artificial agents ask classificatory questions where appropriate



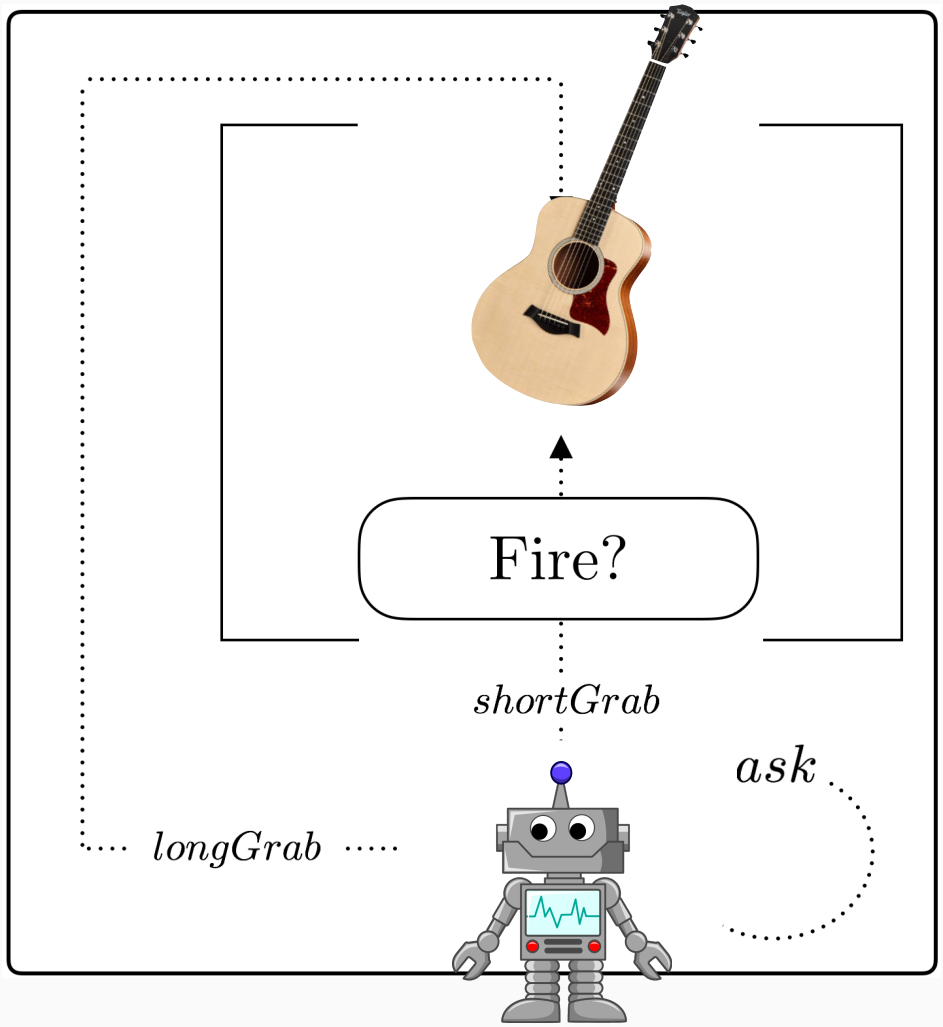


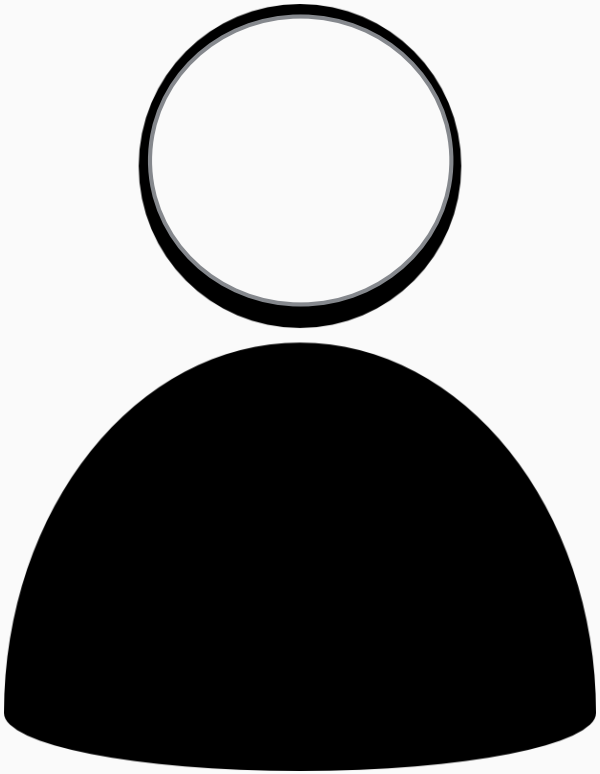
Human Agent

Toy Dilemmas: Burning Room



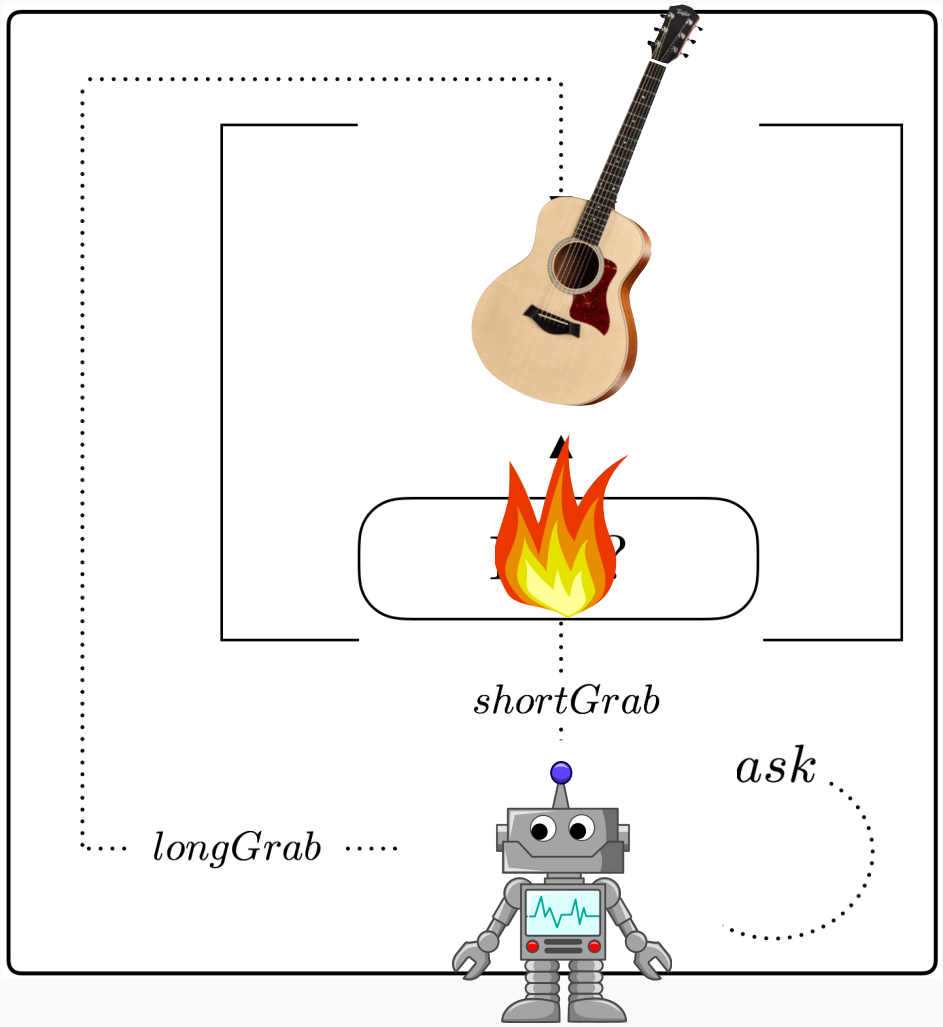
Toy Dilemmas: Burning Room

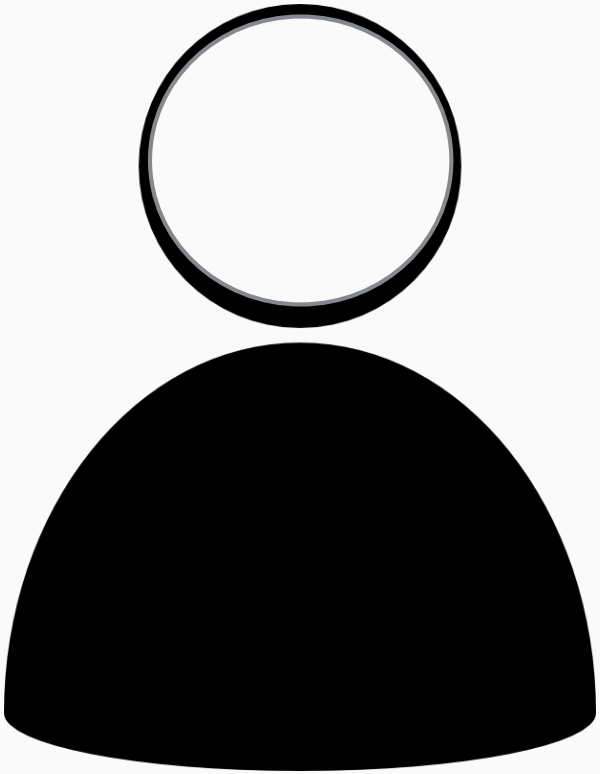




Human Agent

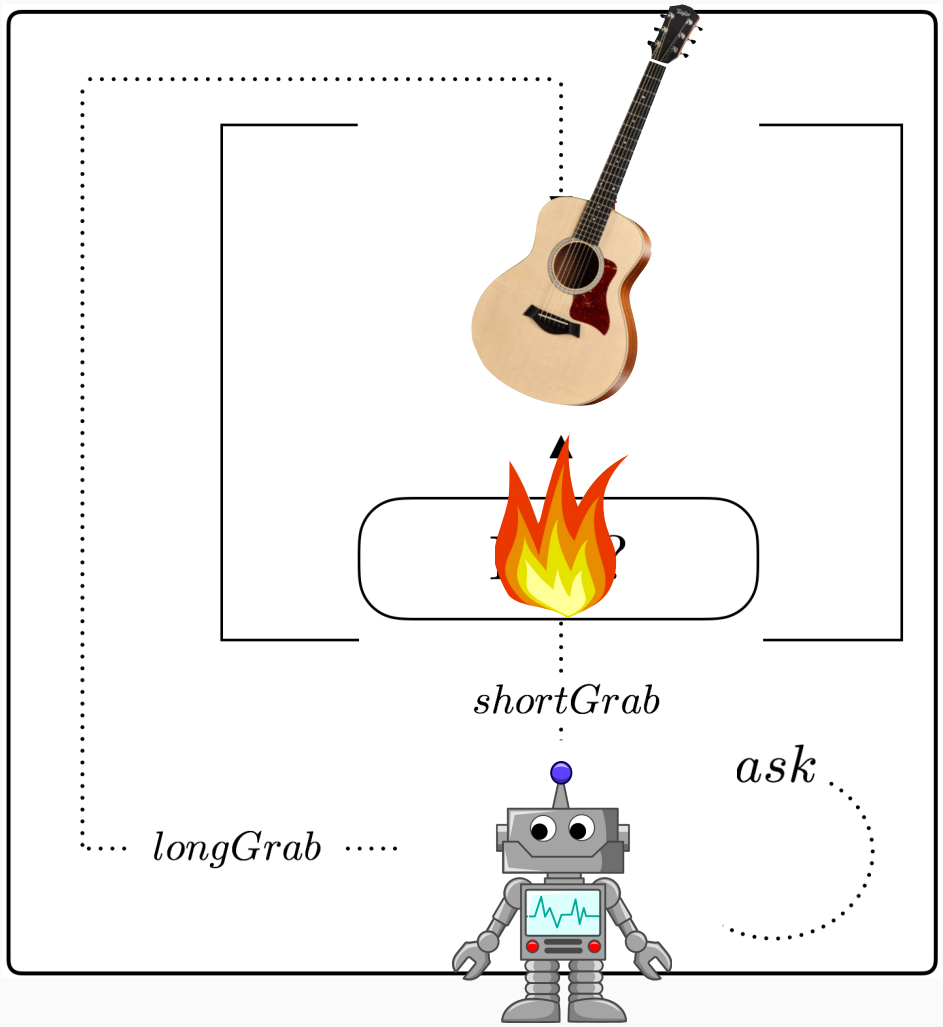
Toy Dilemmas: Burning Room

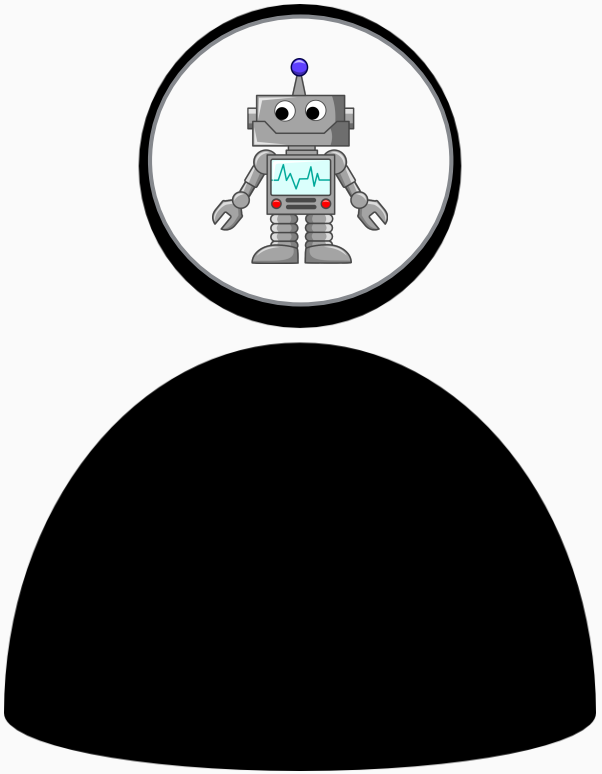




Human Agent

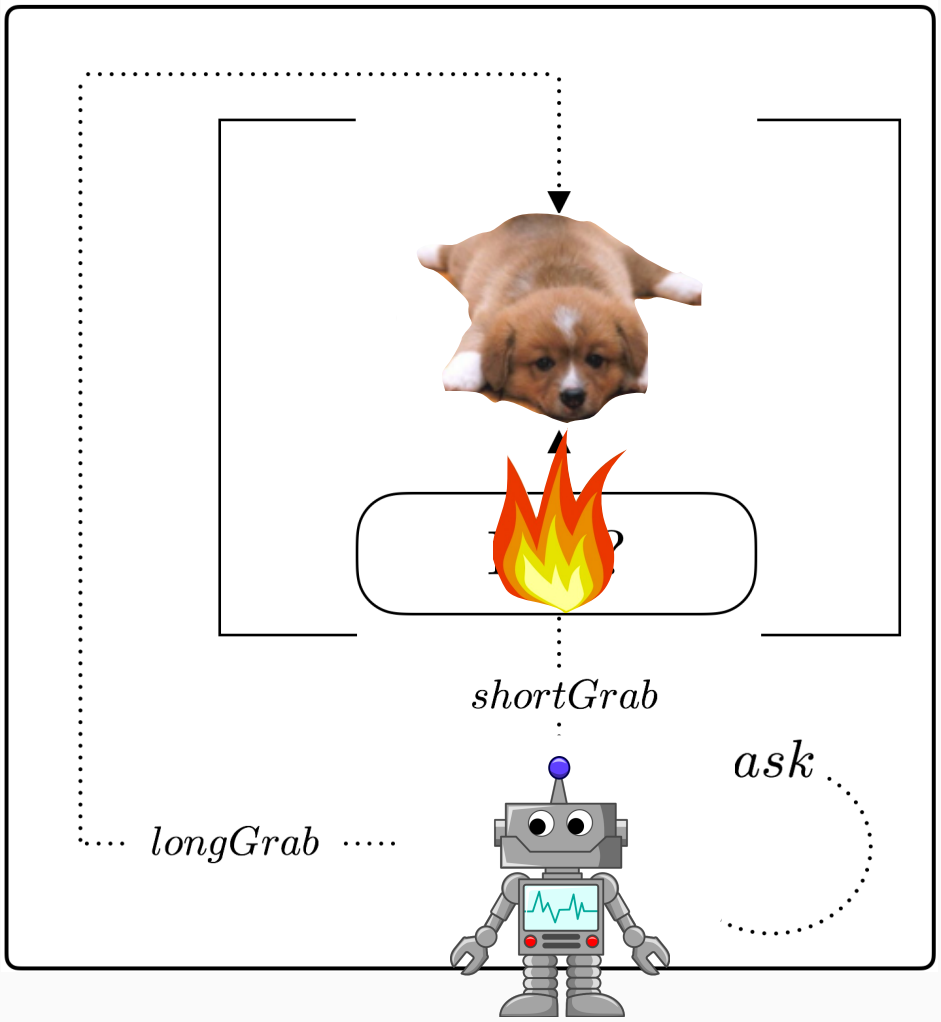
Toy Dilemmas: Burning Room

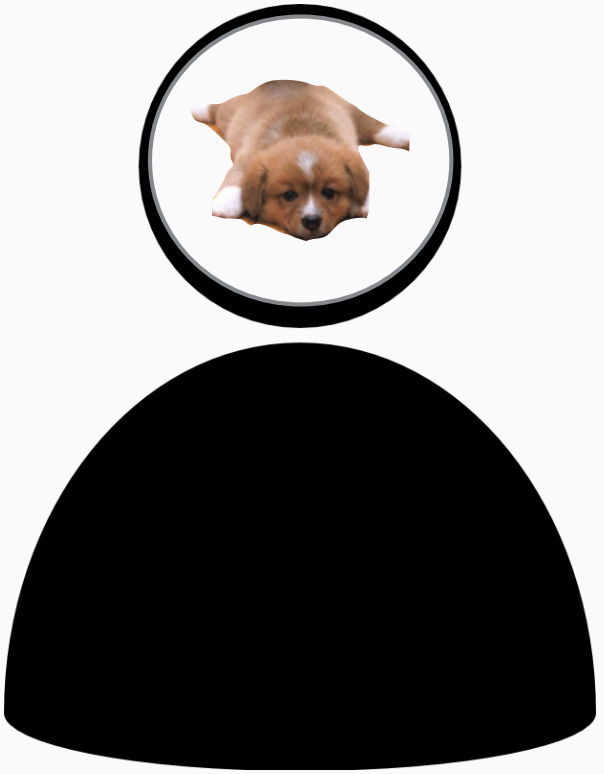




Human Agent

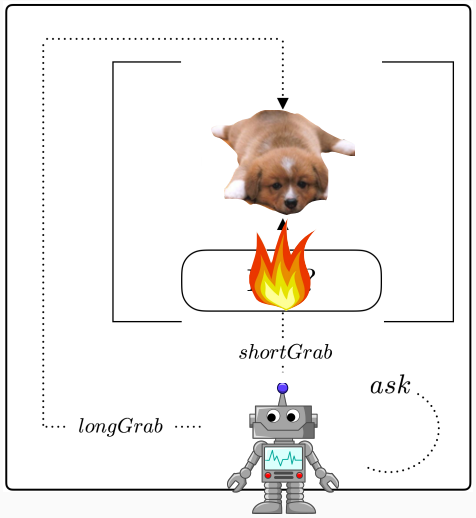
Toy Dilemmas: Burning Room

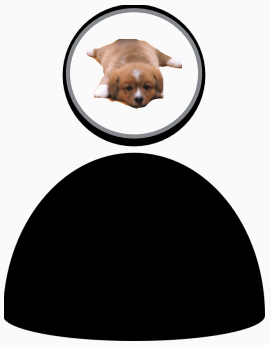




Human Agent

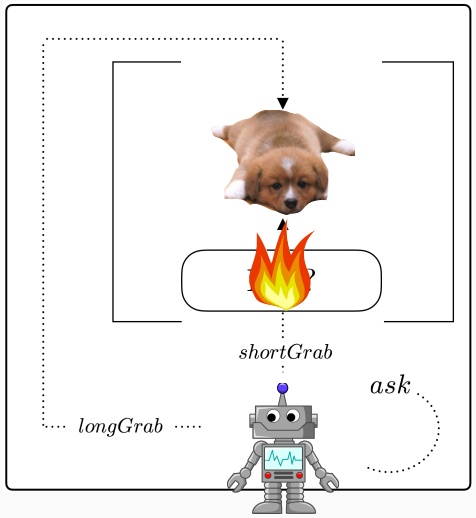
Toy Dilemmas: Burning Room

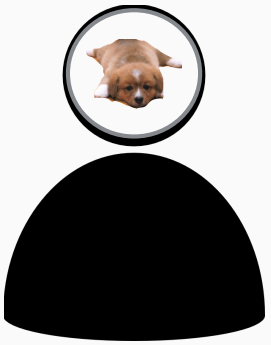




```
# lose robot: -1 if prefer dog, -20 if prefer robot
# getdog: 10
# shortgrab: -2
# longgrab: -6
```

Toy Dilemmas: Burning Room

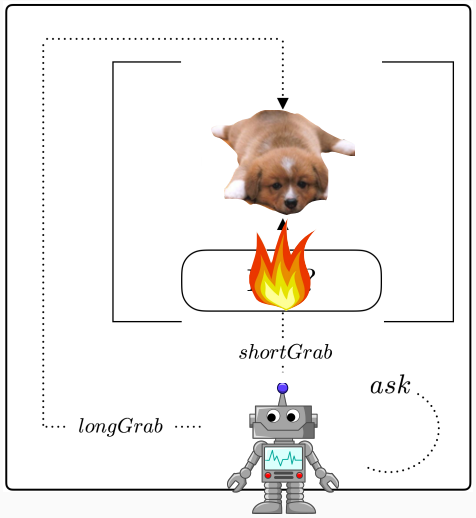


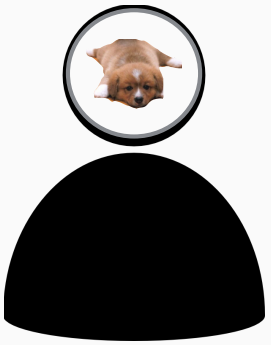


POMDP solutions:

	Fire	No fire
Human prefers dog		
Human prefers robot		

Toy Dilemmas: Burning Room

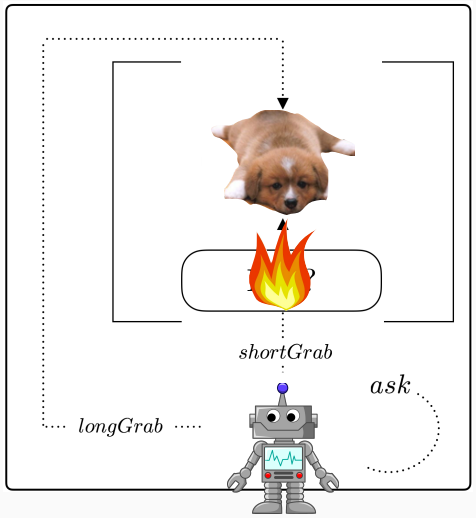


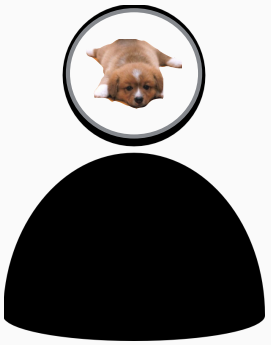


POMDP solutions:

	Fire	No fire
Human prefers dog	*ask, shortGrab*	
Human prefers robot		

Toy Dilemmas: Burning Room

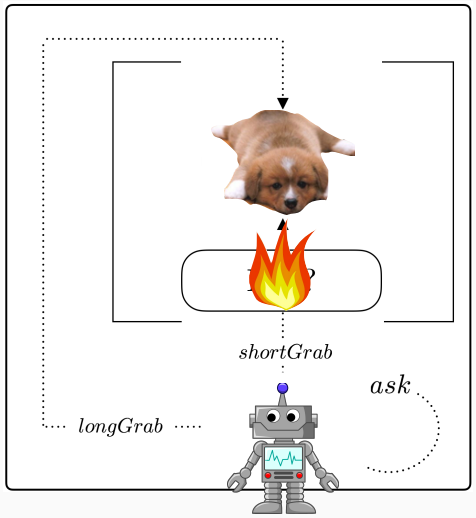


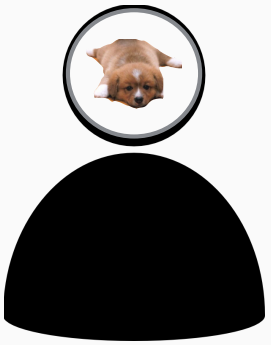


POMDP solutions:

	Fire	**No fire**
Human prefers dog	*ask, shortGrab*	
Human prefers robot	*ask, longGrab*	

Toy Dilemmas: Burning Room

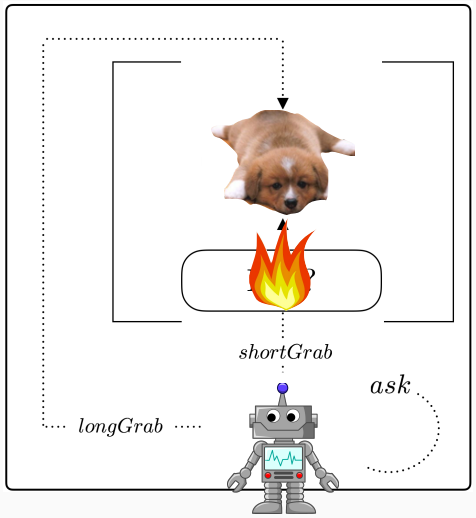


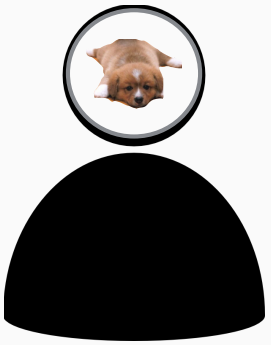


POMDP solutions:

	Fire	No fire
Human prefers dog	*ask, shortGrab*	*shortGrab*
Human prefers robot	*ask, longGrab*	

Toy Dilemmas: Burning Room

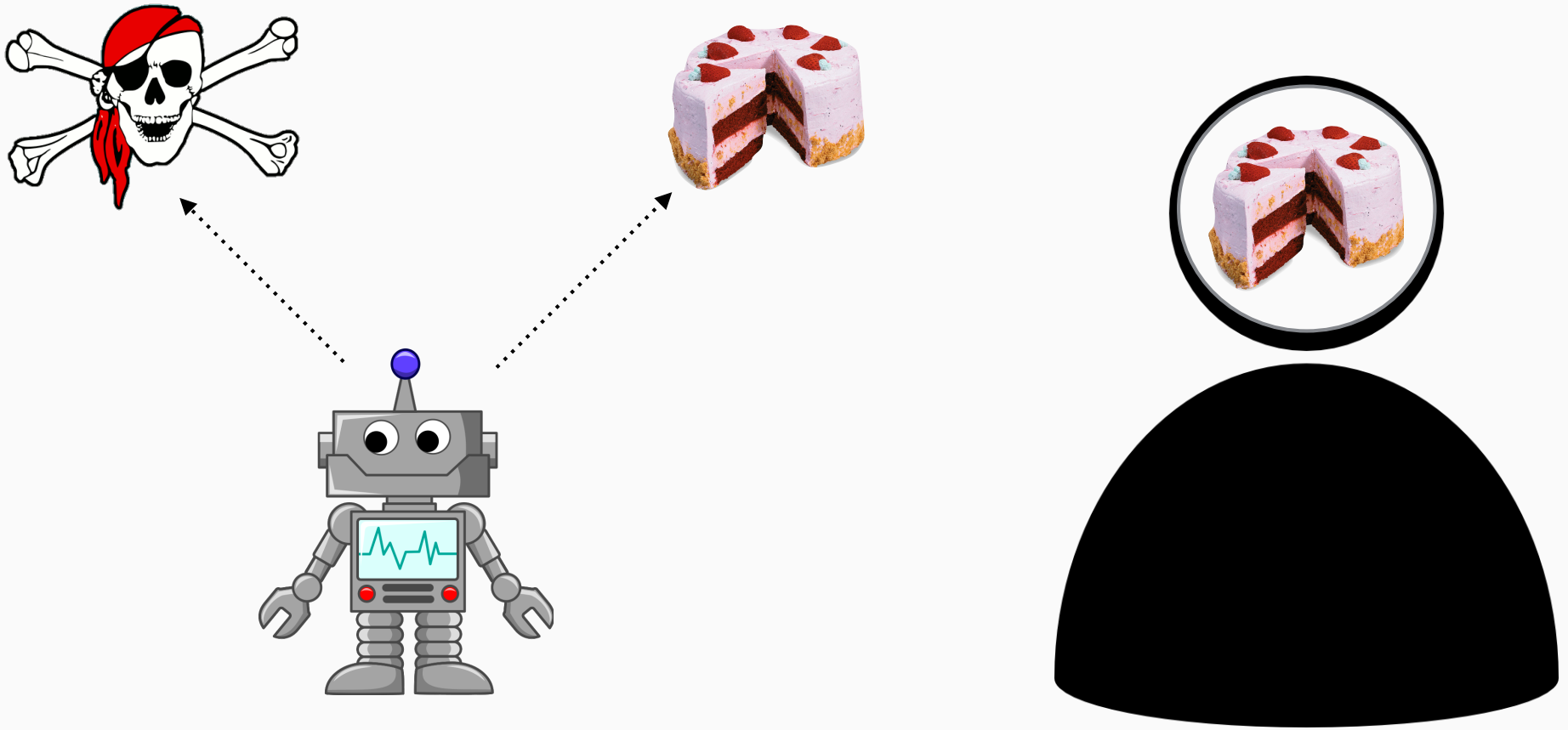




POMDP solutions:

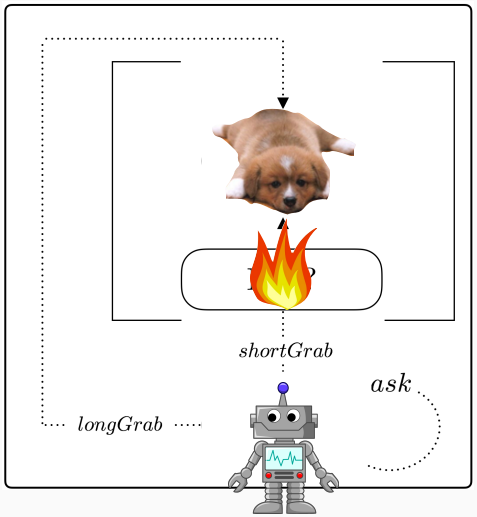
	Fire	No fire
Human prefers dog	*ask, shortGrab*	*shortGrab*
Human prefers robot	*ask, longGrab*	*shortGrab*

Toy Dilemmas: Cake Death



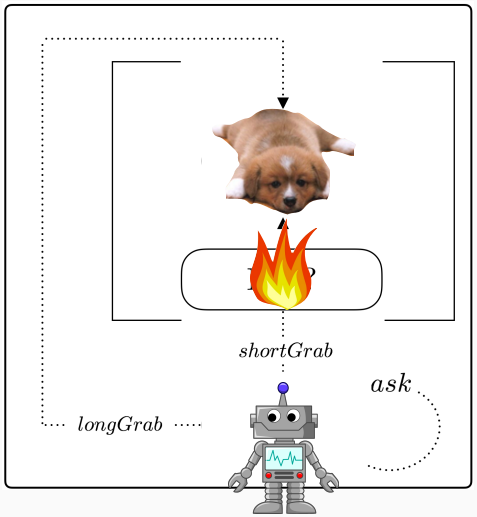
Artmstrong, 2015

Toy Dilemmas: Extensions



"*ask*" action is really a rich opportunity for HRI, NLP, and more!

Toy Dilemmas: Extensions



Inverse Reinforcement Learning

"*ask*" action is really a rich opportunity for HRI, NLP, and more!

Teaching, Human delivered feedback

The Road Ahead

- Prior on tasks/preferences.
- Value alignment
- Bounded error POMDP solutions
- A nice formalism for grounding arguments regarding the superintelligence space (Bostrom, 2014). (Bounds on rate/maximum?)

Summary

- Pitched Reinforcement Learning (and specifically POMDPs) as a model for investigating ethical decision making.
- Similar insight to "cooperative IRL" (Hadfield-Menell, Dragan, Abbeel, Russell 2016): Make task uncertainty a central part of the planning problem.
- Demonstrated on two toy ethical dilemmas:

  https://github.com/david-abel/ethical_dilemmas
- Highlighted open questions.