From Vague Instructions to Executable Robot Plans

Representation of, and Reasoning about Action Verbs

Michael Beetz

Institute for Artificial Intelligence
Universität Bremen

27. Juni 2013

Autonomous Mobile Manipulation
Where Are We? Where Are We Going?

Where we are

doing
everyday manipulation

- we can do selected tasks
- with selected objects/tools
- under selected conditions

because of perception, reachability, manipulation, reasoning, integration

Where we are going

mastering
everyday manipulation

naturally taskable robotic agents that can perform *human-scale* everyday manipulation activities

- prepare breakfast, set the table, clean up, . . .

in any (reasonable) context

The Challenge

Clean the living room!

[Eric Berger, Stanford Univ]

The “Semantics” of Actions

Flip it!

The “Semantics” of Actions

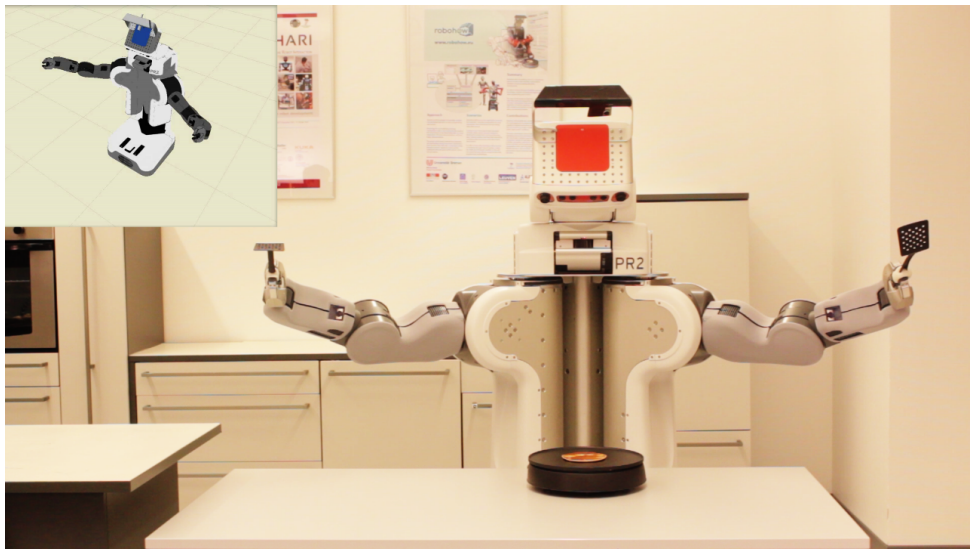
Flip it!

The behavior that achieves this instruction

The “Semantics” of Actions

Flip it!

In the context of pancake making:



+ action awareness

Mastering Everyday Manipulation is Knowledge-intensive!

Common Knowledge

plan parameterization

\- vague instruction

= knowledge required

Mastering Everyday Manipulation is Knowledge-intensive!

Common Knowledge

plan parameterization

\- vague instruction

= knowledge required

Everyday Activity

Anderson [1995]: An everyday activity is a complex task that is

- common and mundane $\rightsquigarrow$ routine
- requires and produces great deal of knowledge and experience
- adequate or satisficing performance

Vagueness is a ressource for flexible, robust, and efficient action!

Action Descriptions

Concept

- given action descriptions (vague)
- plausible action description refinement
- effective action description refinements

Example

“flip the pancake”

(an action
(type flip)
(object-acted-on *pancake*))

Action Descriptions

Concept

- given action descriptions (vague)
- plausible action description refinement
- effective action description refinements

Example

(an action
(type action)
(object-acted-on *pancake-mix-container*)
(effects (on (some stuff
(type pancake-mix)
(form circular)
(diameter 15cm)))
oven))

Action Descriptions

Concept

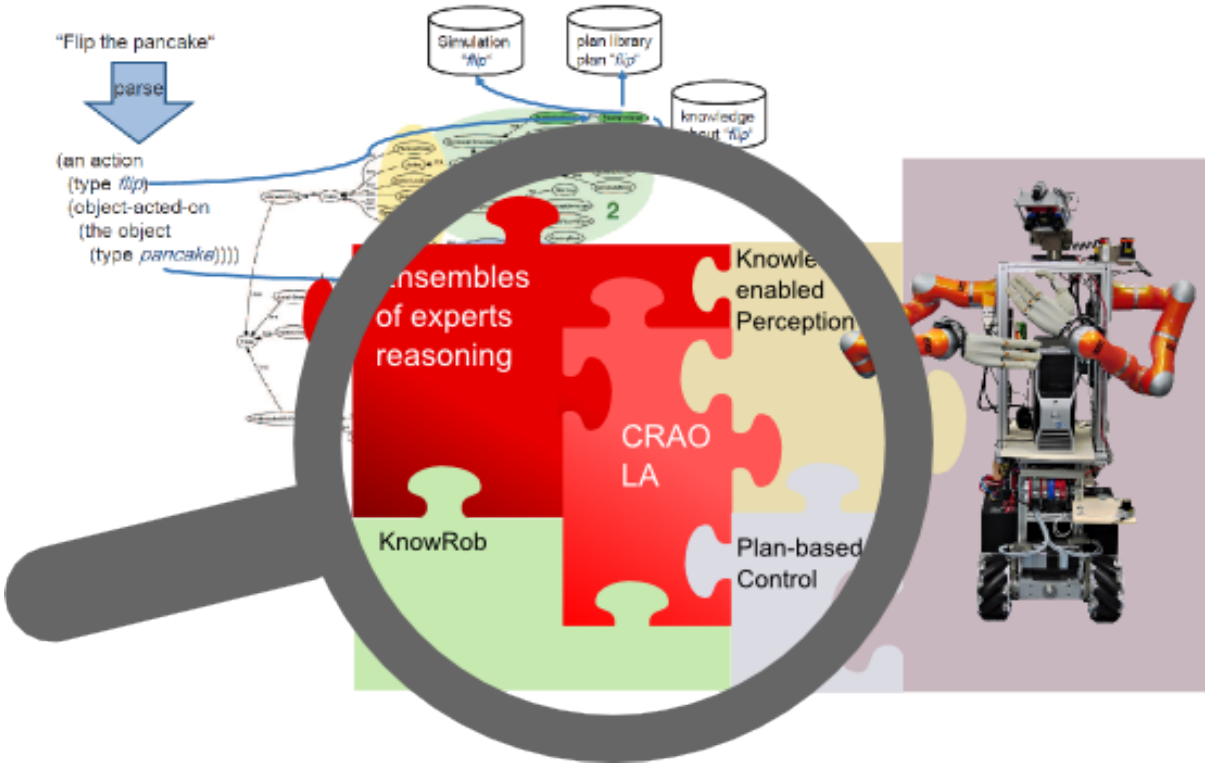
- given action descriptions (vague)
- plausible action description refinement
- effective action description refinements

Example

```
(an action
      (type flip)
      (object-acted-on pancake)
      (object-acted-with spatula)
      (movement-plan
            (:tag approach-oven
                  (movement
                        (constraints ...)))
            (:tag push-under
                  (movement ...))
            (ordering-constraints
               (:approach-oven :push-under))))
```

CRAOLA

A Cognitive Robot Assistant that Observes, Learns, and Advises



CRAOLA will be a Siri/CALO/Watson-like system (ensembles of experts, hypothesis generation and analysis, hypothesis ranking)

CRAOLA handles

- multiple descriptions of the same entity
- incomplete (vague), ambiguous, inconsistent, wrong, inaccurate descriptions
- make descriptions ***effective***

The ease of formulating plans for challenging manipulation tasks critically depends on the expressive power of entity descriptions and the "action intelligence" for their interpretation.

The "Semantics" of Actions

for each **action verb *av*** we have

$$\langle \mathbf{KB_{av}}, \mathbf{sim_{av}}, \mathbf{Descr_{av}}, \mathbf{PL_{av}} \rangle$$

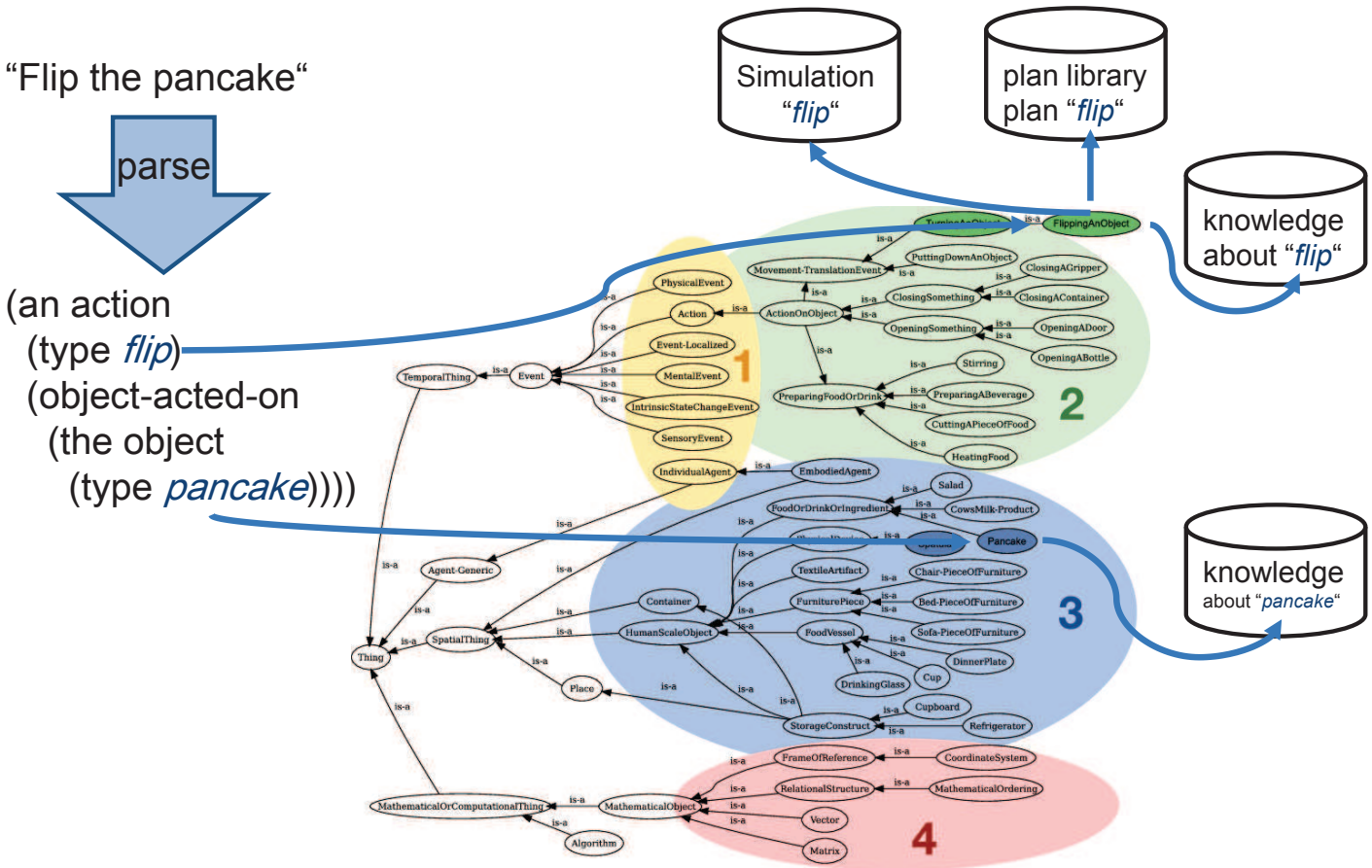
with

$\mathbf{KB_{av}}$	knowledge base for *av*s
$\mathbf{sim_{av}}$	tools for imagining *av* descriptions and plan behavior
$\mathbf{Descr_{av}}$	language for describing spacializations of *av*
$\mathbf{PL_{av}}$	plan language for *av* descriptions

+ ensemble of expert reasoning methods that hypothesize, test, and rank

"Knowledge" Associated with Action Verbs

$\langle KB_{av}, sim_{av}\ Descr_{av}, PL_{av} \rangle$



CRAOLA Inferring Missing Information

$\langle \text{Descr}_{av}, \text{PL}_{av}, \text{KB}_{av}, \text{sim}_{av} \rangle$

"Flip!"

CRAOLA Inferring Missing Information

$\langle \text{Descr}_{av}, \text{PL}_{av}, \text{KB}_{av}, \text{sim}_{av} \rangle$

"Flip!"

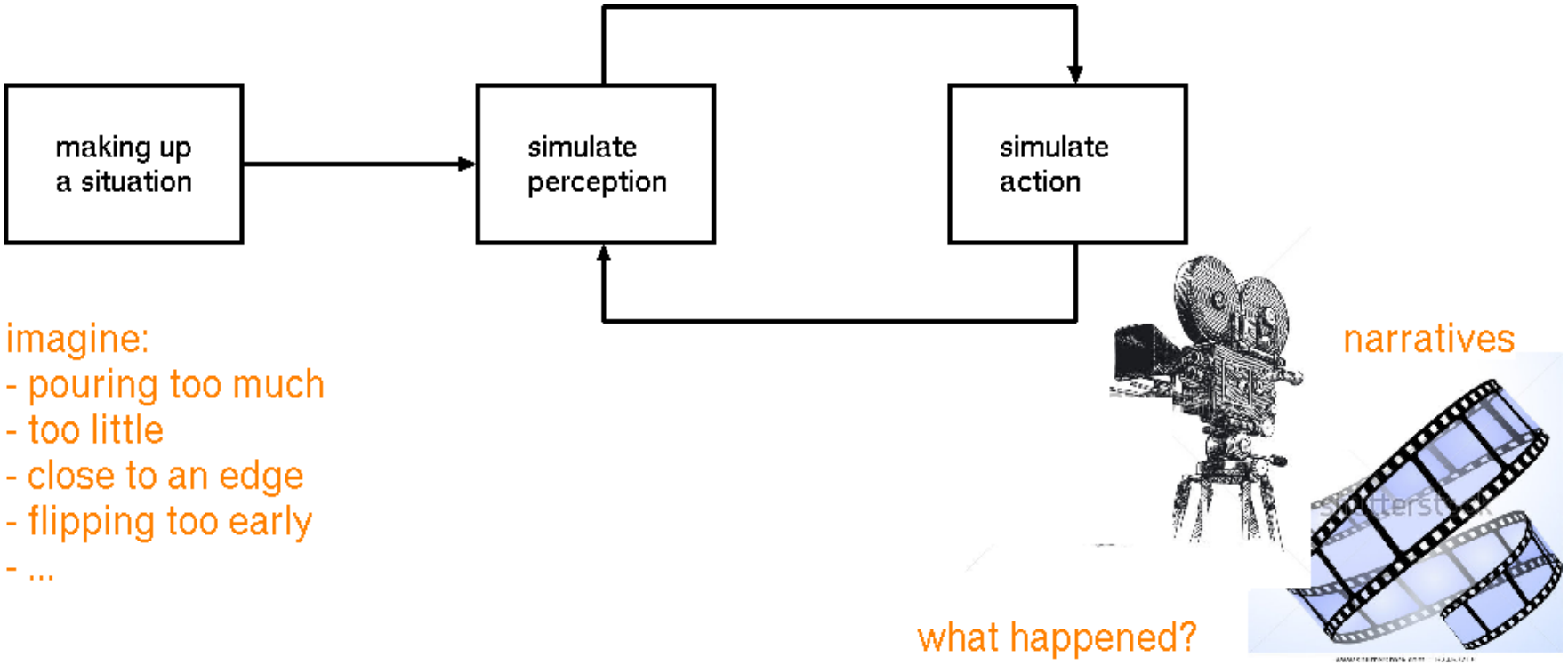
given the objects:

infer:

- what is the role of  in a flipping action
- what is the role of  in a flipping action

Simulation Theory of Cognition

$\langle \text{Descr}_{av}, \text{PL}_{av}, \text{KB}_{av}, \text{sim}_{av} \rangle$

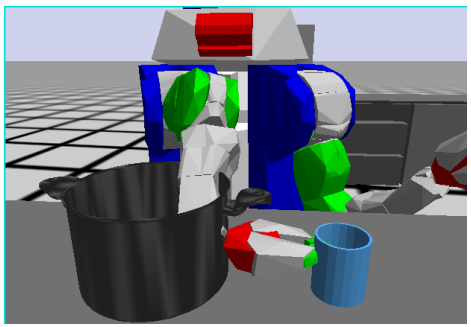


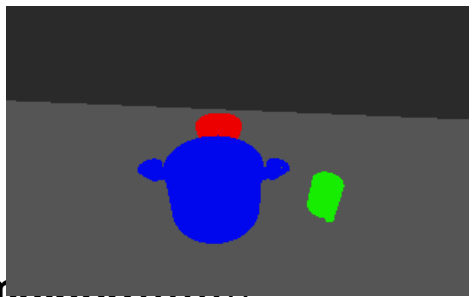
[Hesslow, Feldman, Narayanan, Winston]

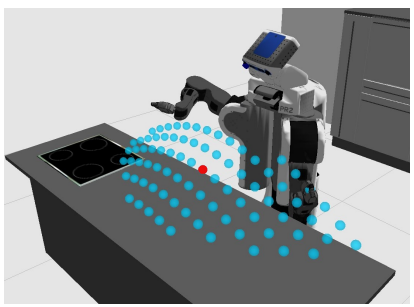
Parameterizing Actions with their Effects

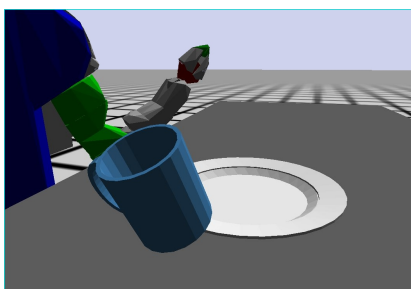
Put the pancake mix away

```
(perform (an action
           (type put-away)
           (object ?obj = (the object
                               (type pancake-mix)))
           (destination ?loc = (a location
                                  (on counter)
                                  (stable ?obj)
                                  (reachable t)
                                  (visible-for James)
                                  (not (hindering (the activity
                                                       (type pancake-making)))))))
```





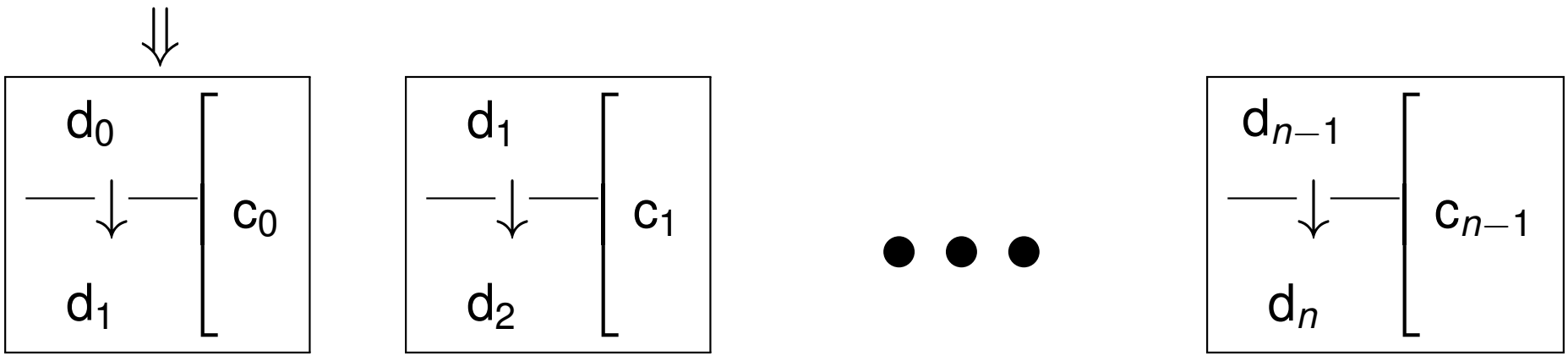




Knowledge-based Description Transformation

$\langle \text{Descr}_{av}, \text{PL}_{av}, \text{KB}_{av}, \text{sim}_{av} \rangle$

d_0 = vague description *av*



d_n = effective description

Transformation of Action Descriptions

$\langle \text{Descr}_{av}, \text{PL}_{av}, \text{KB}_{av}, \text{sim}_{av} \rangle$

?AD ==
an action
of type ?AV
with ?DESC

↓

an action
of type ?AV
with ?DESC
extended with ?REF

1. *?PS* is a plan schema for action verb *?AV*
2. *?NRs* are the needed parameters in the PRAC of *?AV*
3. *?GRs* are the roles of the PRAC of *?AV* that are given in the action description *?AD*
4. *?REF* is the pair $\langle ?NRs, ?Vals \rangle$ such that $?Vals = argmax_{?Vals} P_{flip}(?NRs \mid ?GRS)$

Transformation of Action Descriptions
based on Observation

?AD ==
an action
of type ?AV
with ?DESC

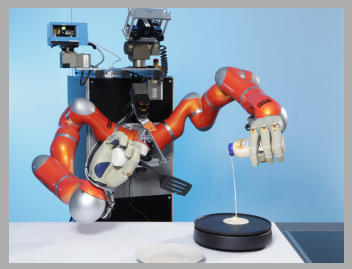
↓

an action
of type ?AV
with ?DESC
extended with ?TPCs

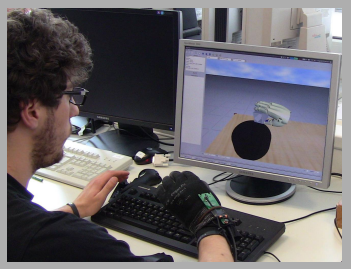
1. if event *?EV* occurs over interval *?TI*
2. *?EV* is of event type *?AV*
   and generates trajectory *?TRAJ*
3. *?AV* has the movement phases *?MPs*
4. segmenting *?TRAJ* into *?MPs* results
   in *?SEGs*
5. *?TPCs* is the set of pairs $\langle ?P, ?TP \rangle$ such tha
   ?P is a movement phase and
   ?TP is the toolpose at the end of *?SEG*
   ?TPCs are the respective
   movement constraints

Action Intelligence

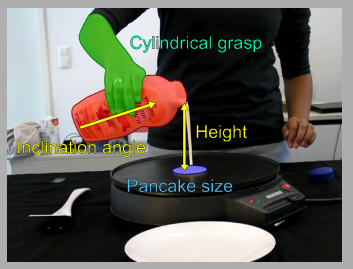
Learning from Experience



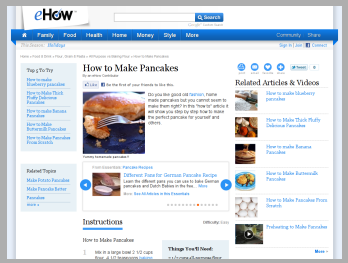
Learning from Computer Games



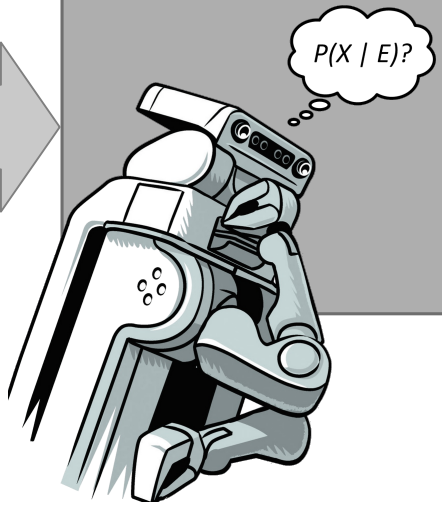
Interpretation of Human Behavior



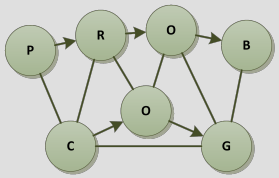
Web-enabled Knowledge Acquisition



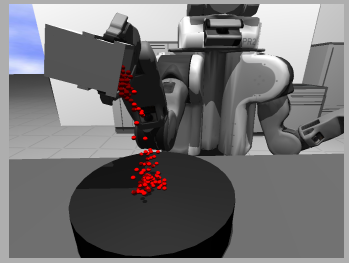
Probabilistic Robot Action Cores (PRAC)



Probabilistic First-Order Models

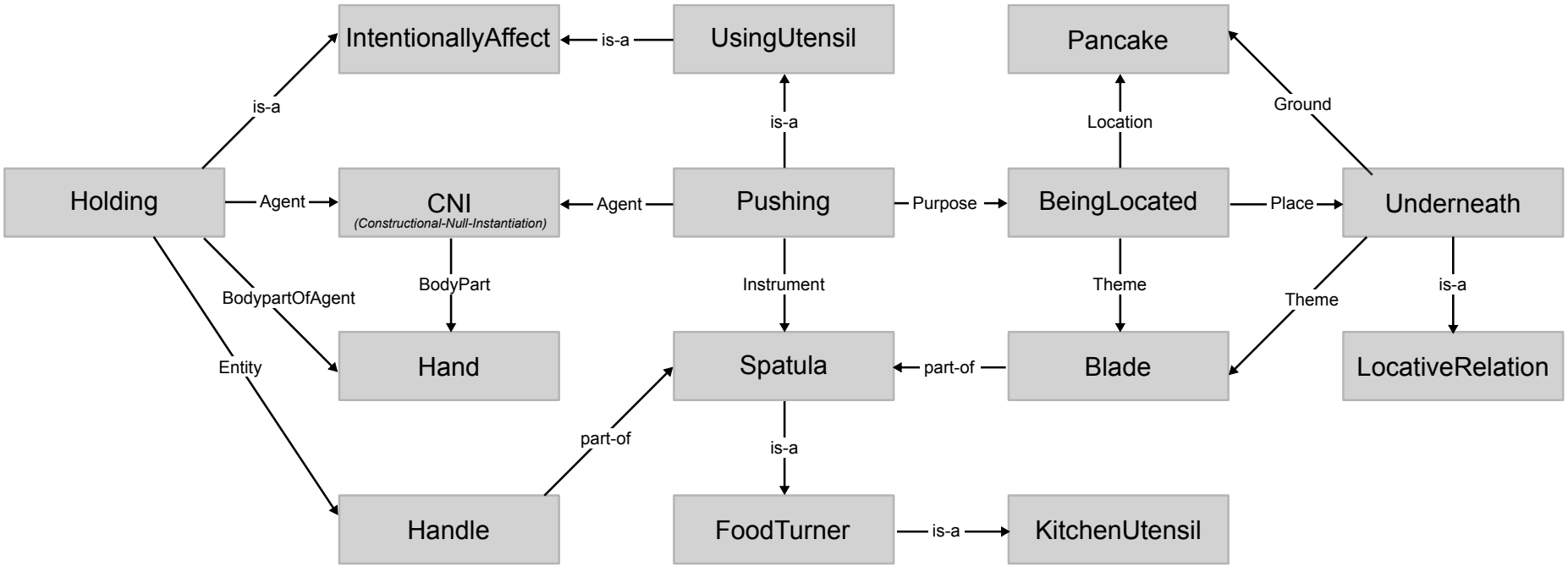


Prediction/Envisioning of Effects

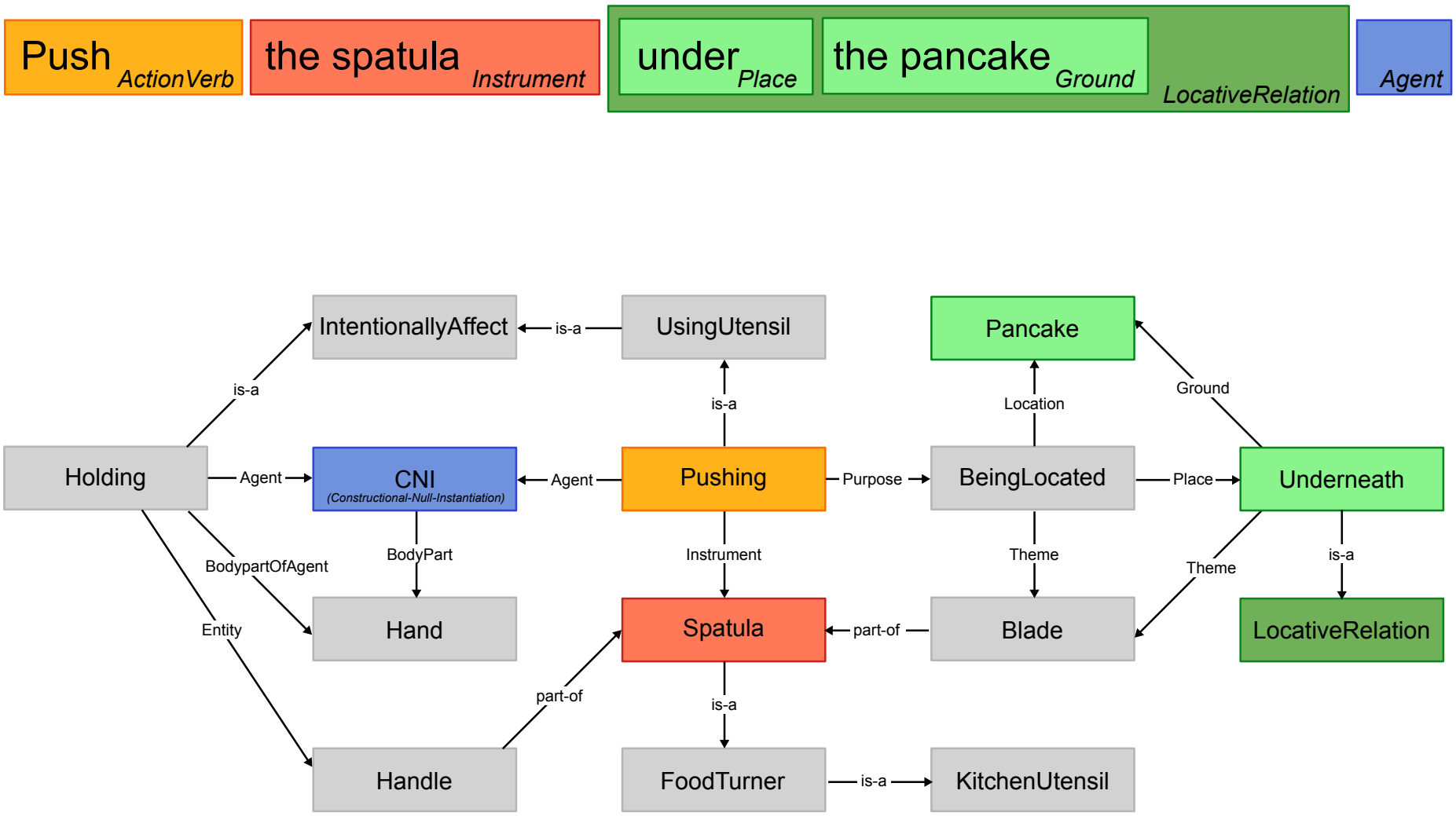


Action Intelligence
Action Specification

Action Core: Set of inter- and intraconceptual relations that constitute an abstract event type, assigning an action role to each entity that is affected by an action verb.



Action Intelligence
Action Core



PRAC
Probabilistic Robot Action Cores

Relational Nature of Knowledge in NL

(WordNet Taxonomy, FrameNet Semantic Relations)

+

Uncertainty in NL

↓

Statistical Relational Learning

- Use probabilistic first-order models for combining two powerful knowledge representation formalisms:

  First-order logic and probability theory

- Model the joint over all semantic relations and the class taxonomy:

$$P(\mathit{ObActedOn}, \mathit{Instr}, \mathit{Dest}, \mathit{Source}, \ldots \mid \mathit{is\text{-}a}, \mathit{part\text{-}of})$$

Probabilistic Robot Action Cores

Understanding what is given and infer what is missing

Example Queries:

- ”Fill Milk into a bowl.”→ Do what to which objects?
  Action Core: *FluidFlowTranslation*
  Action Roles: *Agent*, *Theme*, *Source*, *Goal*, ...

$$\underset{o_1', o_2' \in \{o_1, o_2\}}{\arg\max} \; P\left(Theme(o_1'), Goal(o_2') \mid isa(o_1, Milk), isa(o_2, Bowl) \right)$$

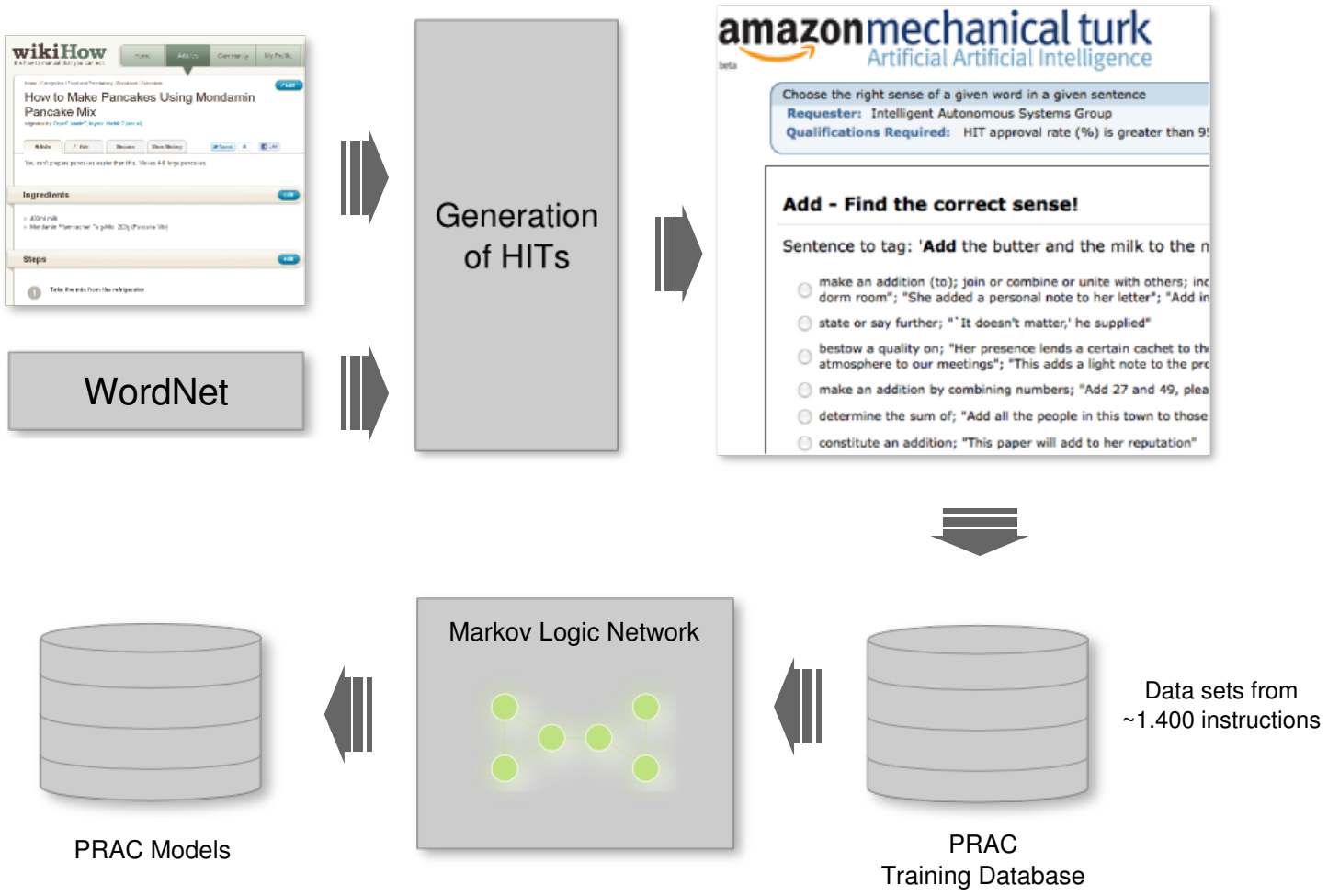
$$= \{o_1' = o_1, o_2' = o_2\}$$

- ”Fill the pot with water.”→ Where do I get the water from?
  Action Core: *FluidFlowTranslation*
  Action Roles: *Agent*, *Theme*, *Source*, *Goal*, ...

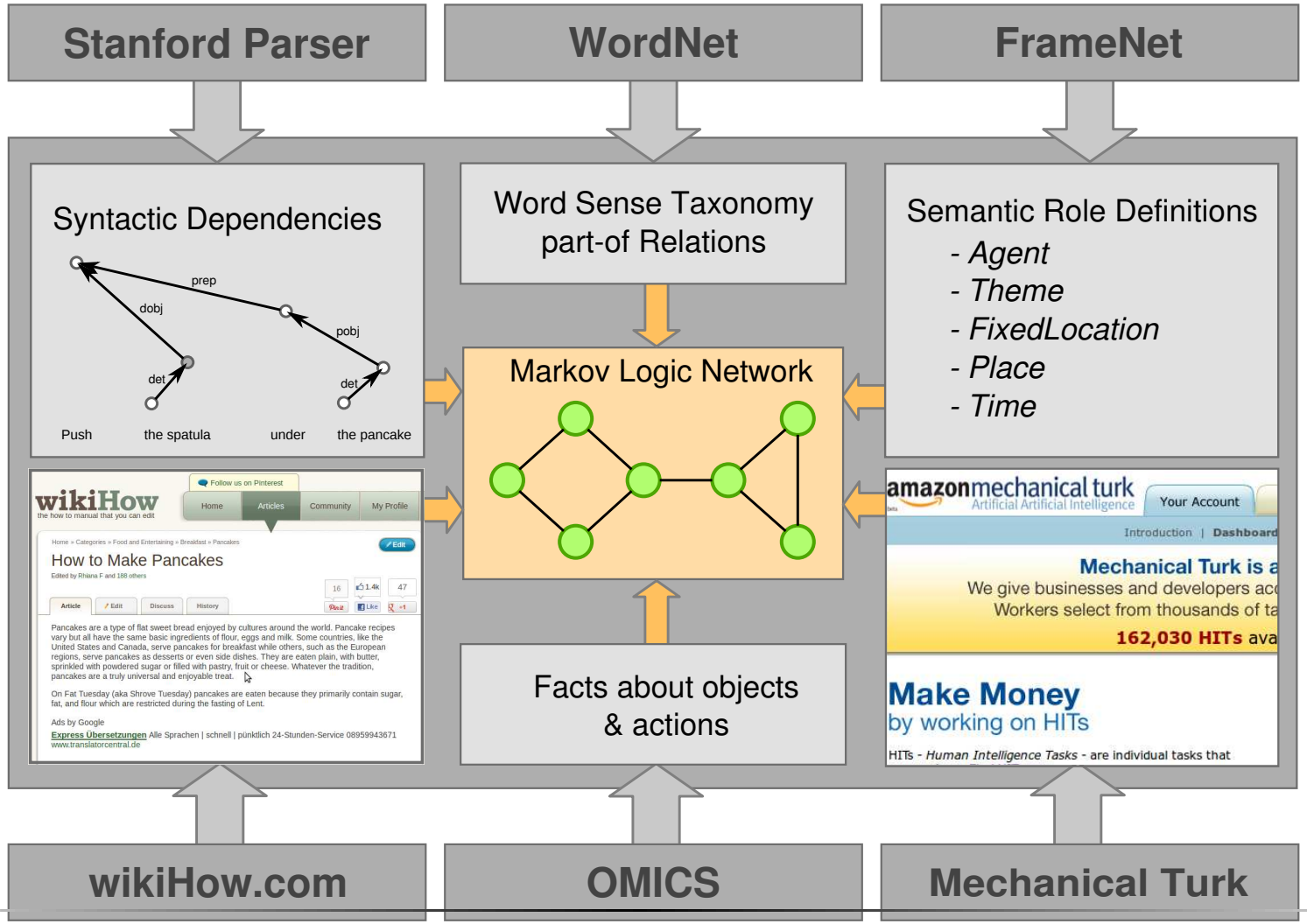
$$\underset{c \in Concepts}{\arg\max} \; P\left(isa(s, c) \mid isa(w, Water), isa(p, Pot), Theme(w), Goal(p), Source(s) \right)$$

$$= TapFlowControlDevice$$

Data Acquisition Pipeline

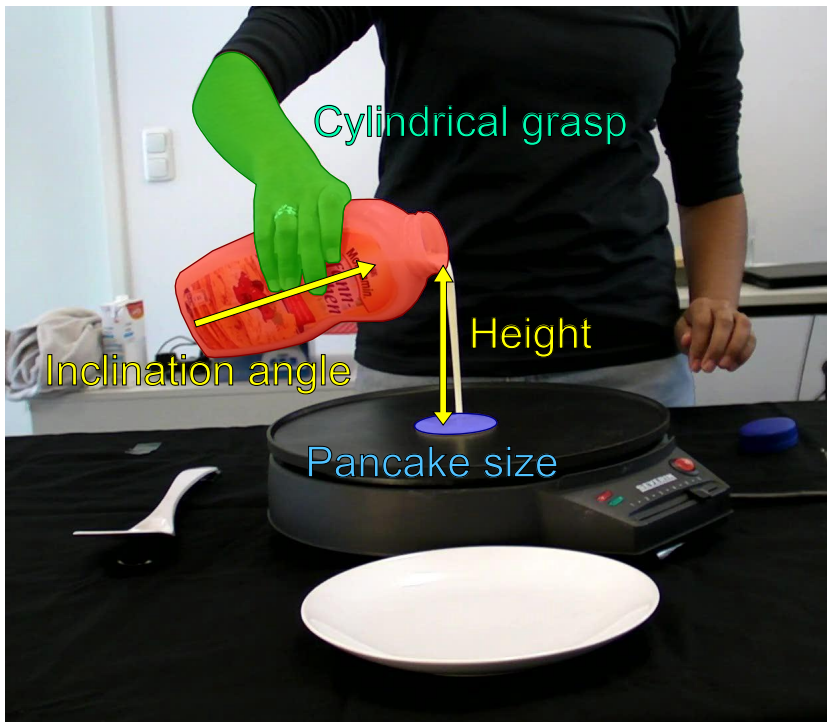


Implementation
Learning from Web Resources



PRAC Inference
Example

Combining Textual with Observational Knowledge




```
(occurs (an action
          (type flip)
          (trajectory (a trajectory ...)))
        (the time-interval
          (start-time ...)
          (end-time ...)))
```

Combining Textual with Observational Knowledge

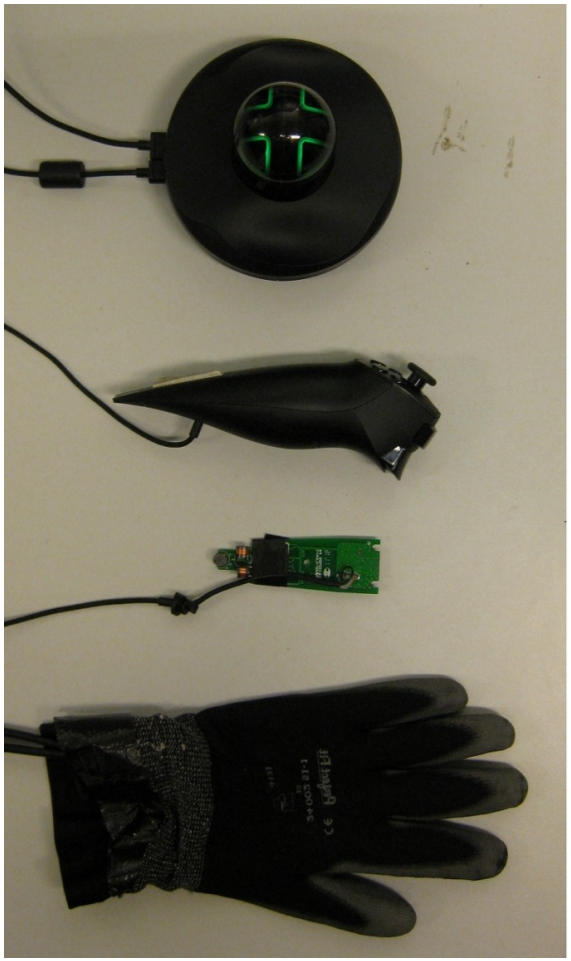
courtesy of [Argyros, et al], RoboHow

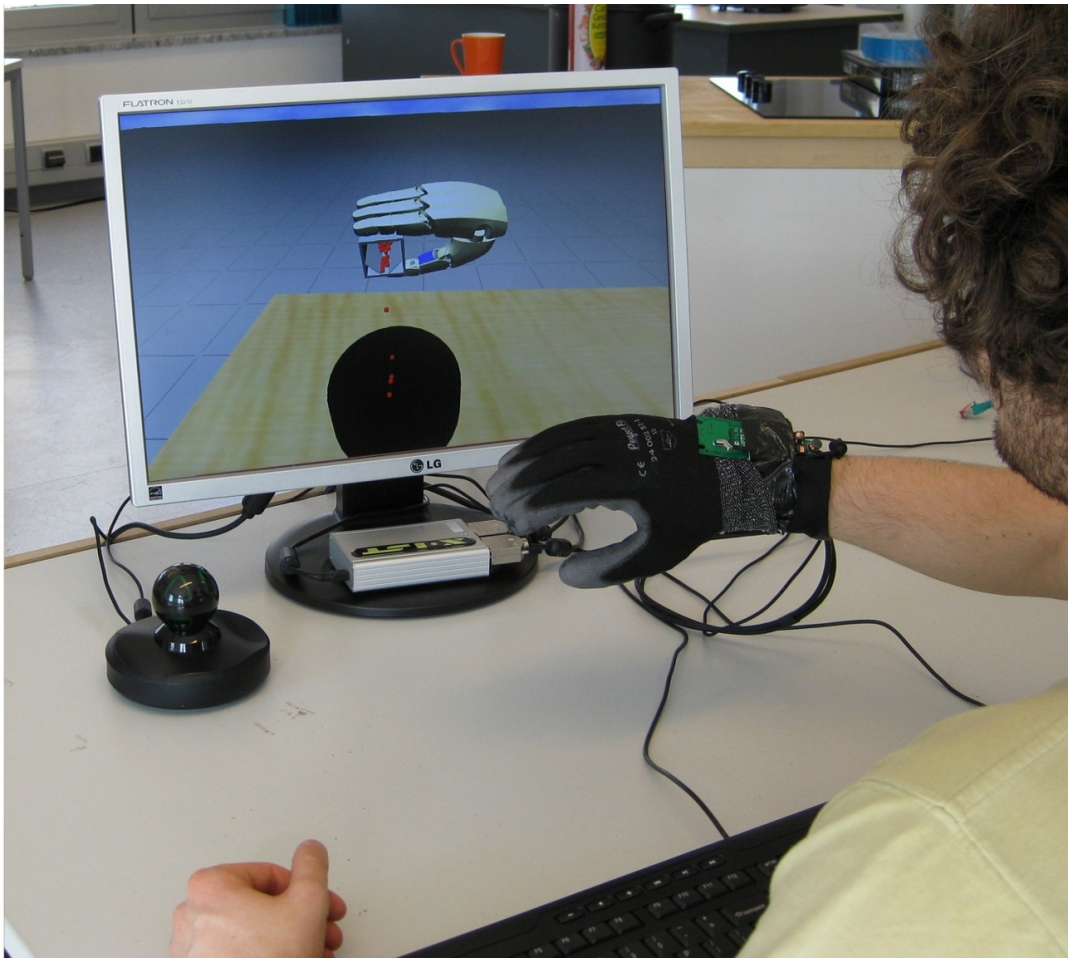
```
(occurs (an action
              (type flip)
              (trajectory (a trajectory . . . )))
         (the time-interval
              (start-time . . . )
              (end-time . . . )))
```

Combining Textual with Observational Knowledge

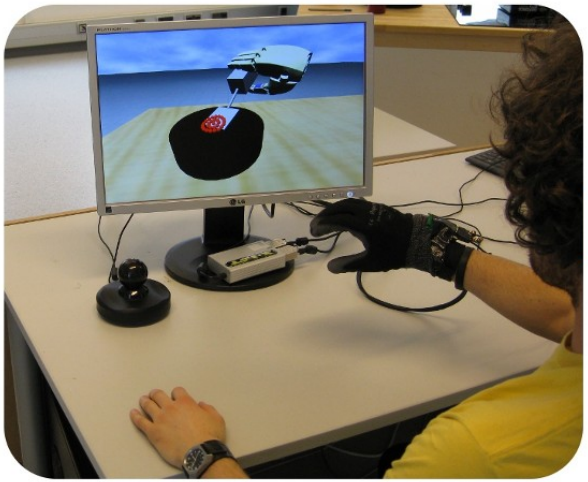
```
(occurs (an action
            (type flip)
            (trajectory (a trajectory . . . )))
        (the time-interval
            (start-time . . . )
            (end-time . . . )))
```

Learning about Actions with GwaPs



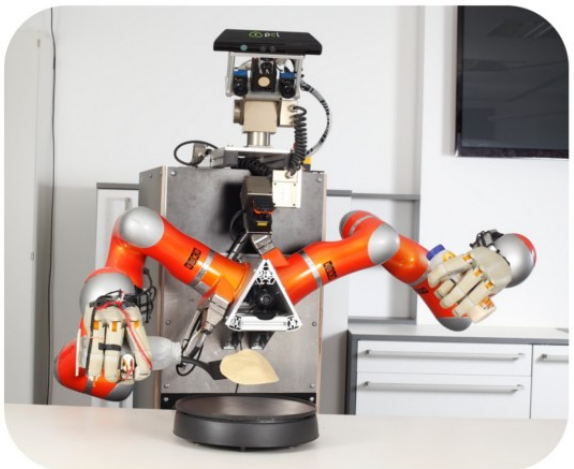


Learning Symbolic Narratives from Games

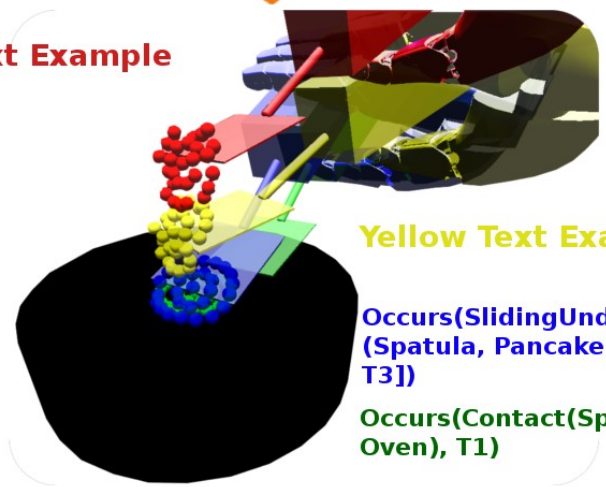


$$SimulatorValue\ \overbrace{(position(o, pos),}^{Function}\ \overbrace{t}^{Time\ point}\ ,\ \overbrace{tl}^{Timeline}\ )$$

. . .



Red Text Example

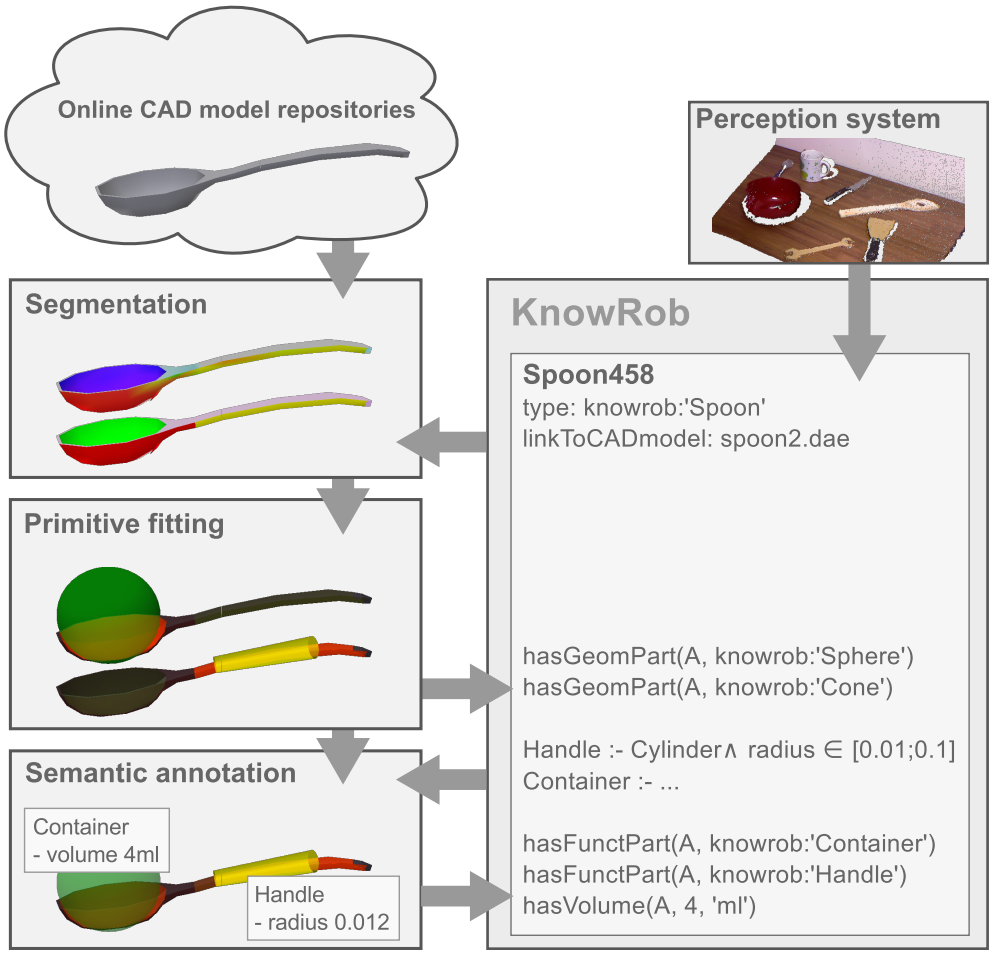


Yellow Text Example

Occurs(SlidingUnder (Spatula, Pancake), [T2, T3])

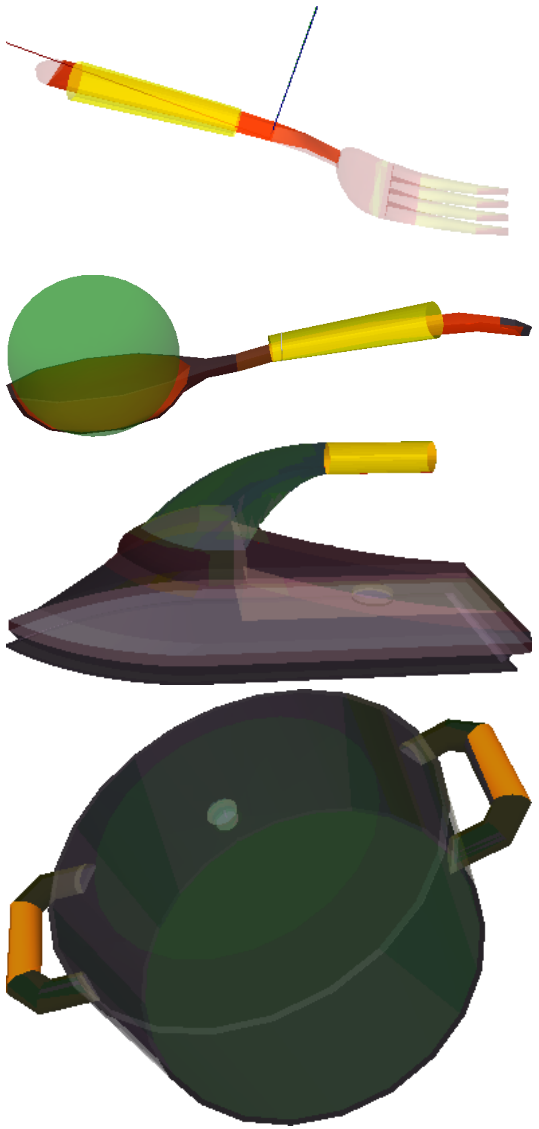
Occurs(Contact(Spatula, Oven), T1)

(an object-part (function handle)
(part-of (an object (type spoon) (source perception))))

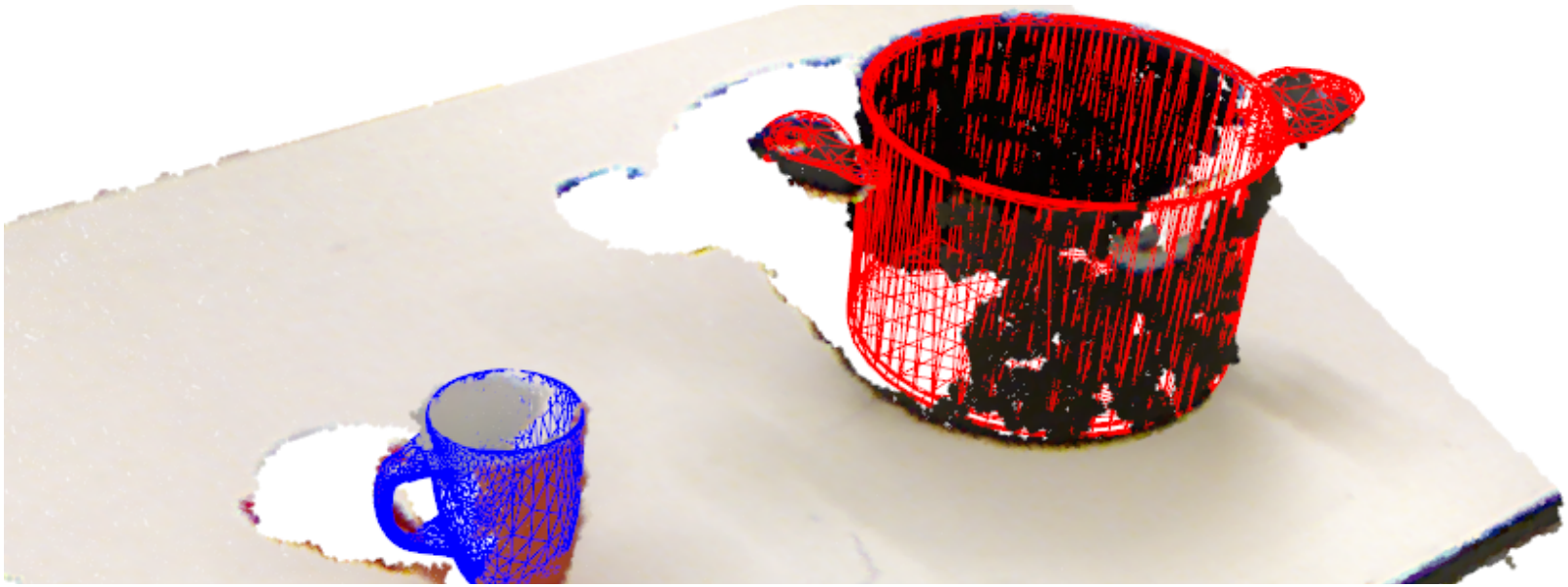


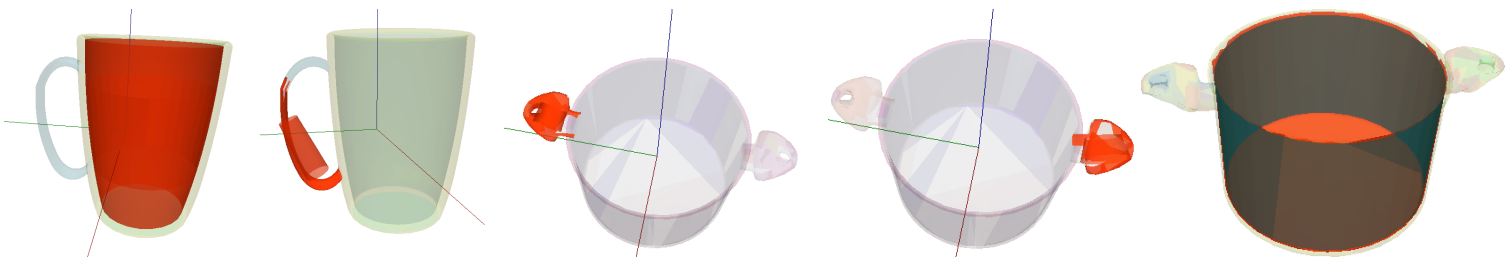
Semantic & geometric object models

- Translate abstract descriptions like “grasp the spatula at the handle” to geometric primitives
- Combined semantic and geometric object models:
  - *Geometric knowledge:* mesh segments, coordinates, motions
  - *Semantic knowledge:* types, properties, parts, functions
- Automatically extracted from CAD models by mesh segmentation

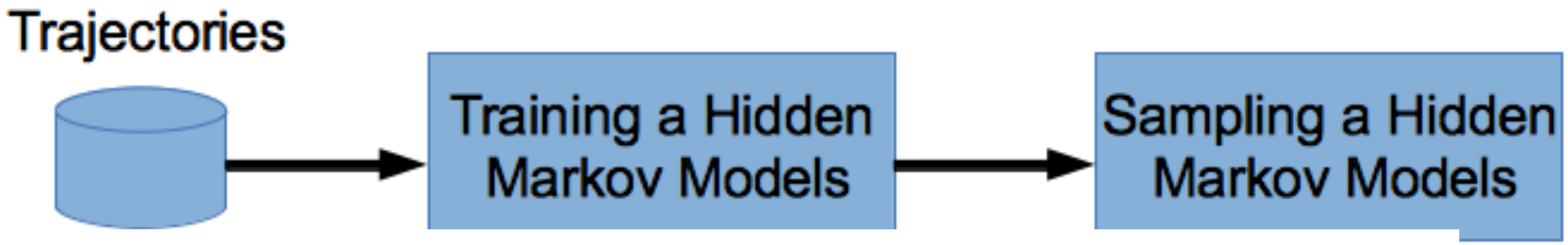


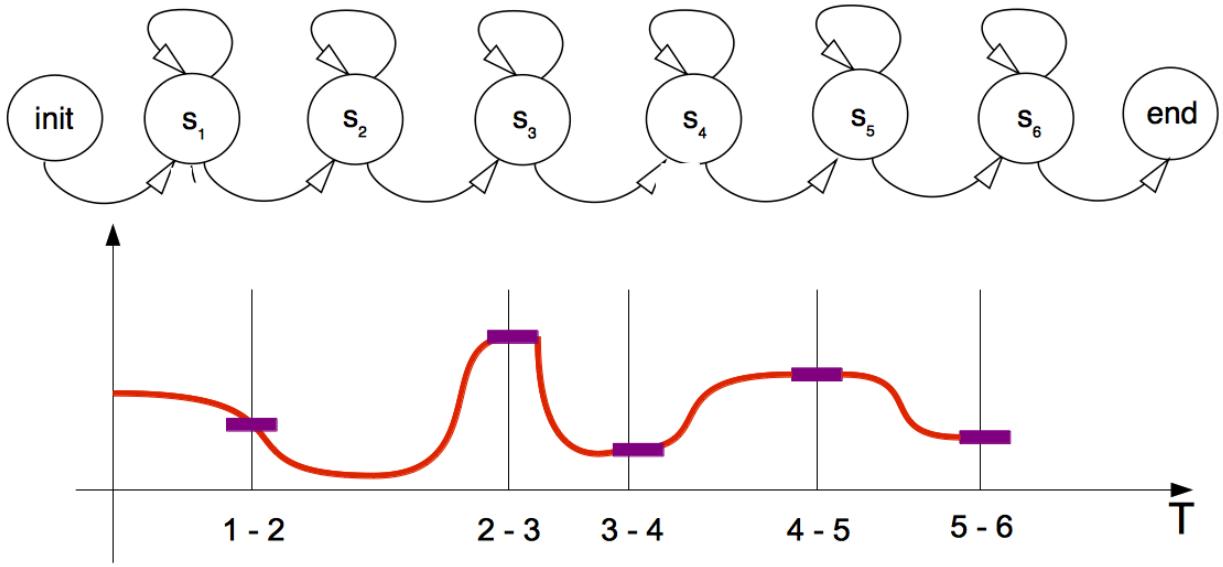
Reasoning about real-world objects



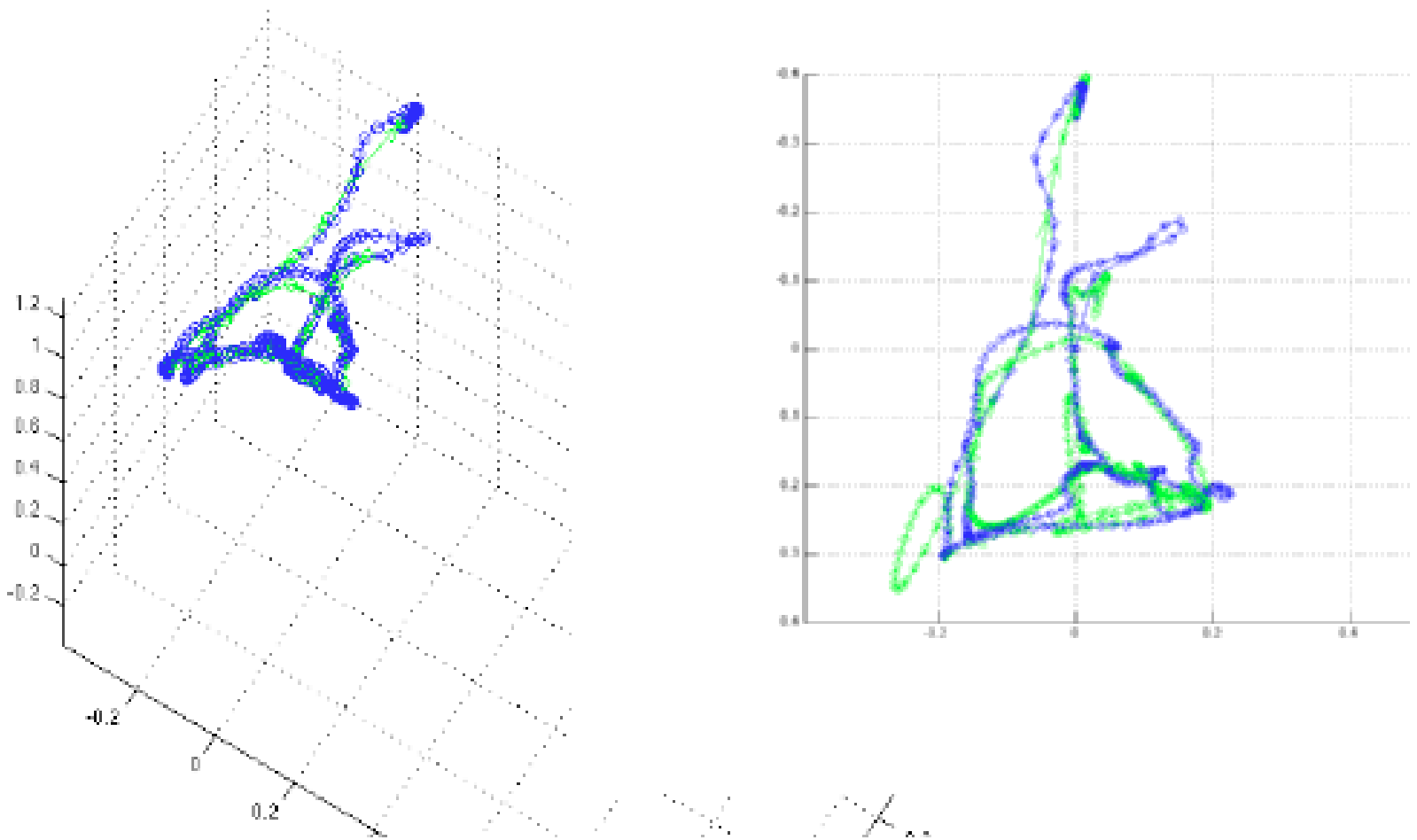


Learning Movement Results

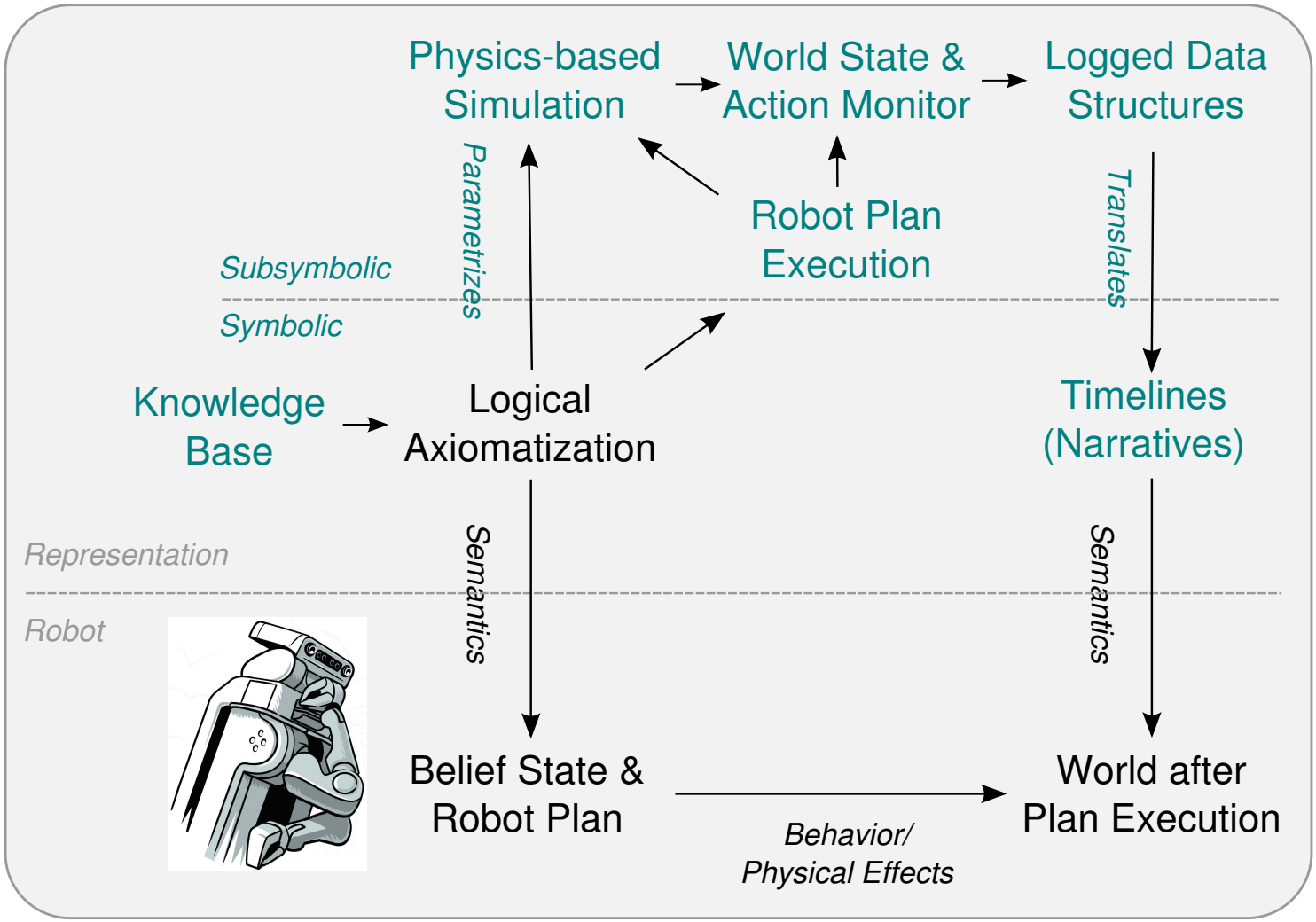




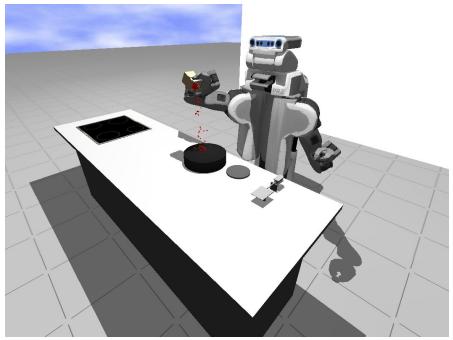
Results of HMM Learning

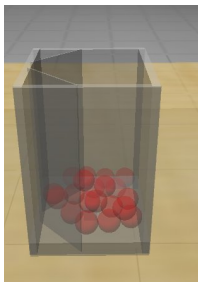


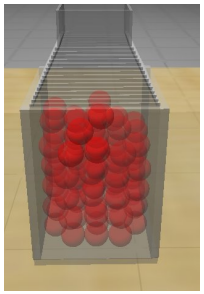
Envisioning Action Effects

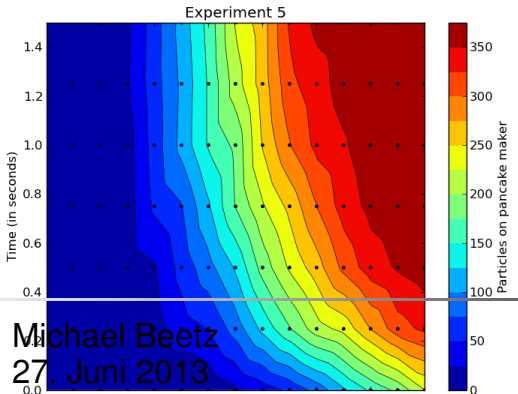


Learning from Physics Simulation
Predicting the size of a pancake

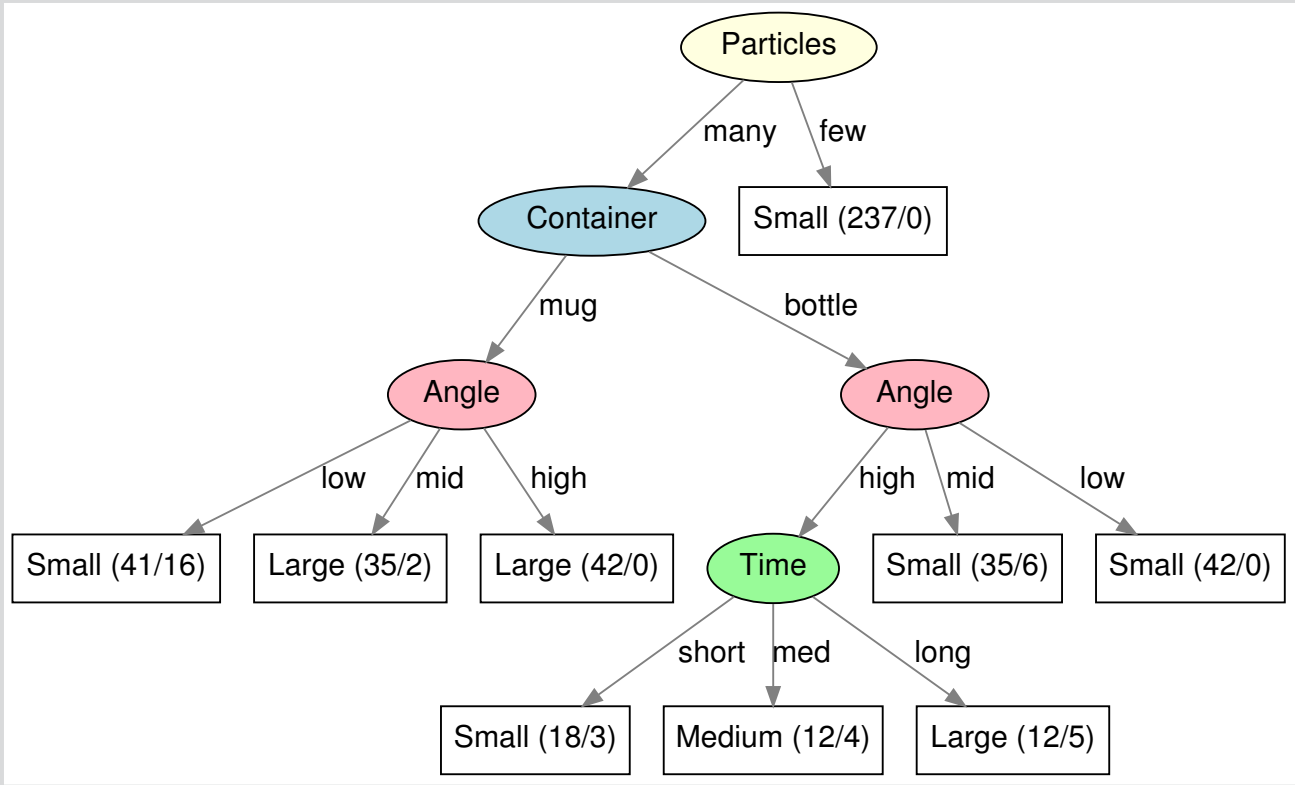








Learned decision tree



Conclusions

- “action intelligence” requires the capability of performing vaguely described actions and tasks
- “action intelligence” is knowledge intensive

$$\langle \mathrm{KB}_{av}, \mathrm{sim}_{av}, \mathrm{Descr}_{av}, \mathrm{PL}_{av} \rangle$$

- CRAOLA (Cognitive Robot Assistant that Observes, Learns, and Advises)
  - PRAC learning and reasoning
  - complementing PRACS with
    - Games with a Purpose
    - Simulation-based reasoning

Thank you for your attention!

http://www.robohow.eu & http://www.acat-project.eu

ROBOHOW and ACAT are supported by the European Commission as part of the Cognitive Systems and Robotics Initiative in the Seventh Framework Programme FP7/2007-2013.