Object Anchoring (and Symbol Grounding)

Joachim Hertzberg

Osnabrück University
and DFKI Robotics Innovation Center

Overview

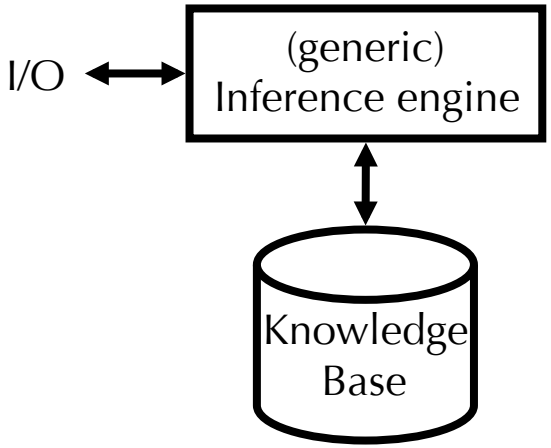
1. Object Anchoring – The Problem
2. Object Anchoring Issues in Plan-Based Robot Control

Overview

1. **Object Anchoring – The Problem**
2. **Object Anchoring Issues in Plan-Based Robot Control**

AI Had a Dream: Knowledge-Based Robotics

Knowledge-Based SW System

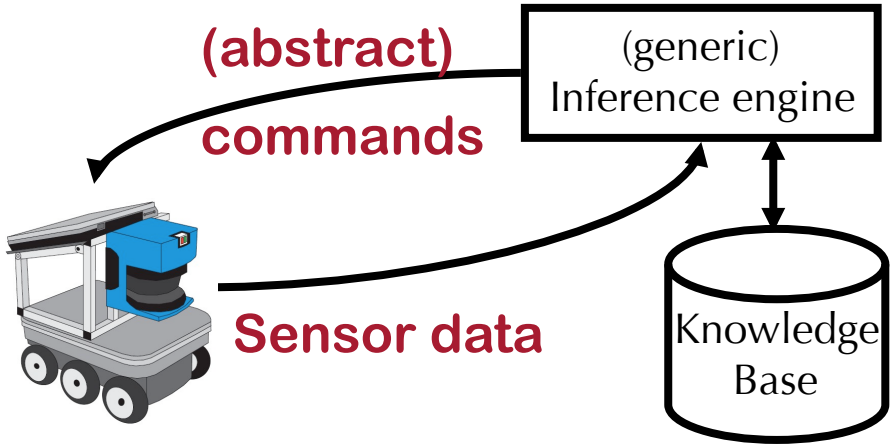


For instance, expert systems

- Knowl. Engineer fills KB off-line
- User gives input, interprets output

For instance, robots

- Knowl. Engineer fills KB off-line
- KBR input comes from sensors
- KBR output controls robot action



Knowledge-based robots need to translate sensor data into symbols and inferences into control!

Symbol Grounding

S. Harnad: The Symbol Grounding Problem Physica D 42:335–346, 1990
cogprints.org/3106/01/sgproblem1.html

How is symbol meaning to be grounded in something other than just more meaningless symbols?

Is that an Important Issue?

- Some (AI) say: Nay – a technical problem at best!
- Some (Philosophy, Cog.Sci.) say: That is the very issue which makes an artificial intelligence impossible in principle!
- Some (AI, Cog.Sci., Robotics, JH) say: That is currently among the most exciting and relevant points for basic research in AI!

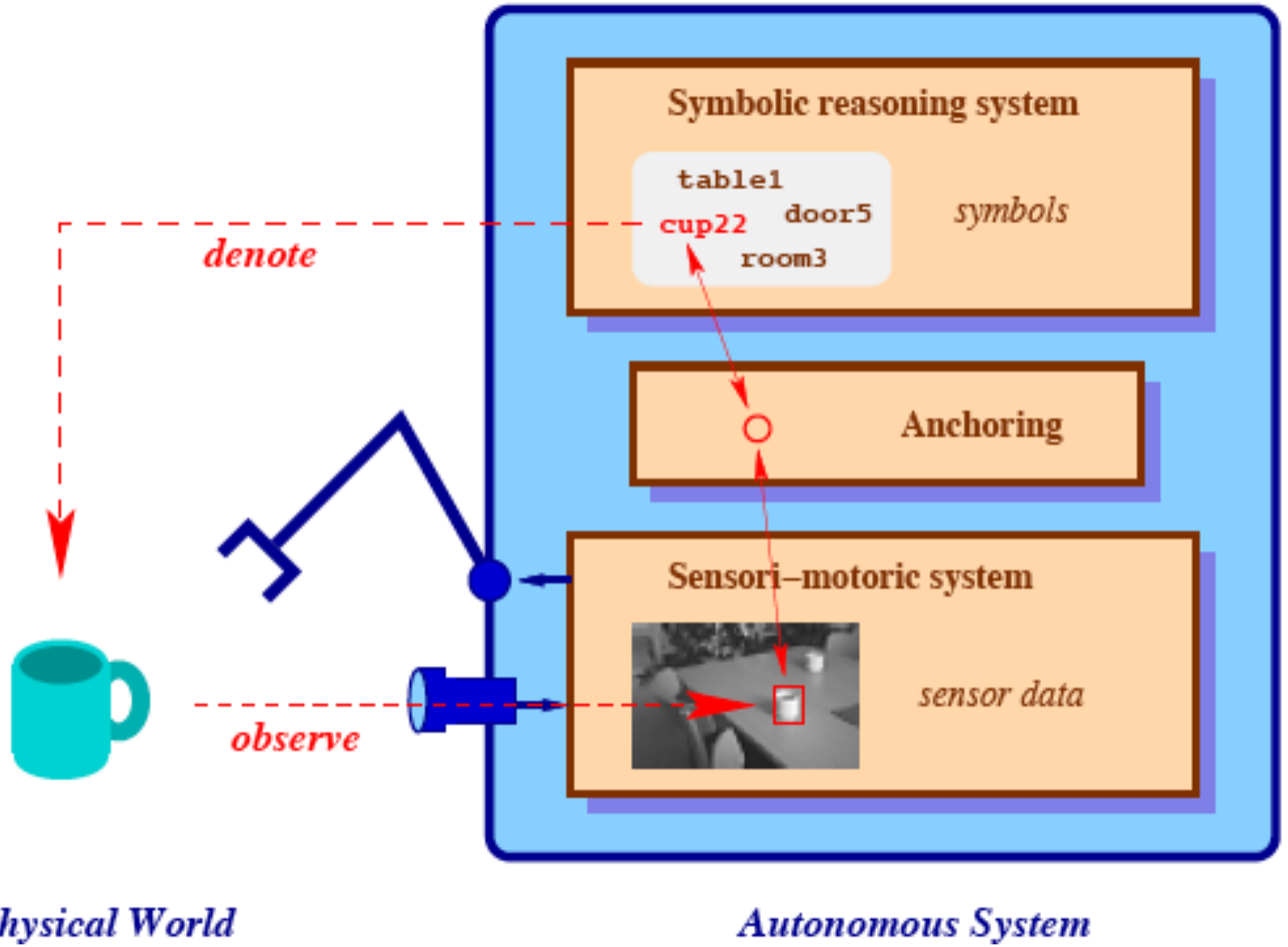
S. Coradeschi, A. Loutfi, B. Wrede:
A Short Review of Symbol Grounding in Robotic and Intelligent Systems.
Künstl. Intell. 27:129–136, 2013, http://www.aass.oru.se/~sci/SI-review-final.pdf

Roboticists' Specialty: Object Anchoring

S. Coradeschi, A. Saffiotti: An Introduction to the Anchoring Problem
Robotics&Aut.Syst, 43(2–3):85–96, 2003, www.aass.oru.se/~asaffio/Papers/ras03.html

- **Anchoring**: "the process of creating and maintaining the correspondence between symbols and sensor data that refer to the same physical objects"
- **Anchoring problem**: "the problem of how to perform anchoring in an artificial system"
- Specializes general symbol grounding: Only physical objects
  e.g., no abstract entities ("weather, happiness"), no properties/relations ("red, smarter than"), no events/actions ("cooking, foundation of Rome")
- Assumption about system architecture: Symbol processing ("reasoning") and sensor data processing are disjoint processes
  e.g., no geometric reasoning on analog representation

The Architecture of Object Anchoring



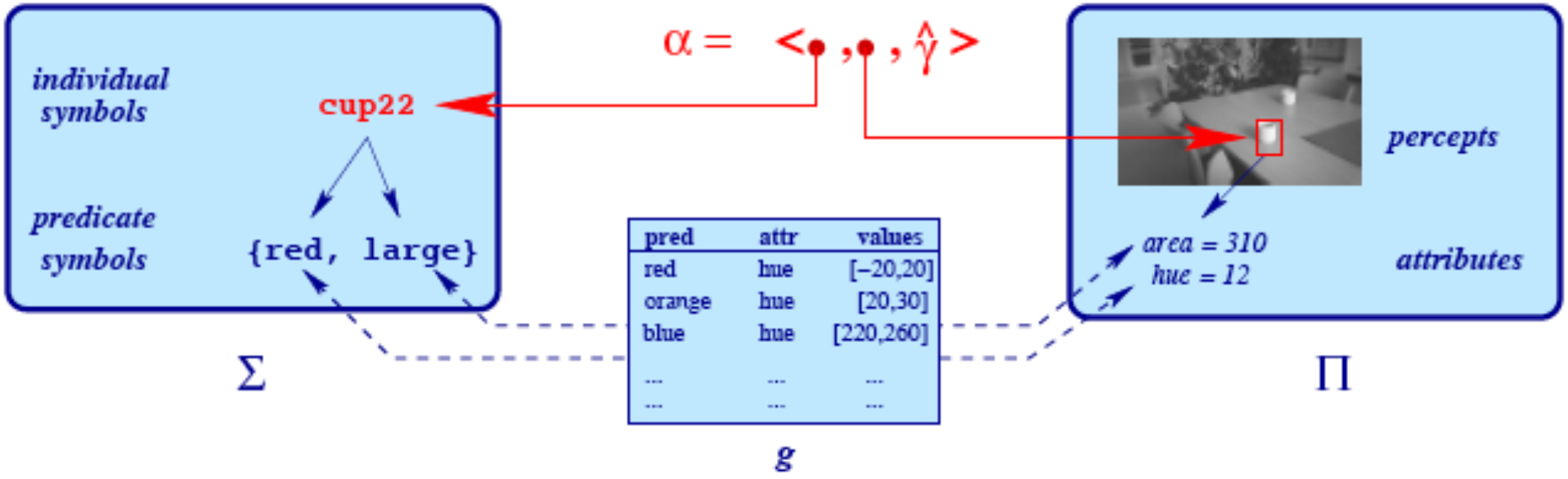
Object Anchoring: Components

Σ: Logical language

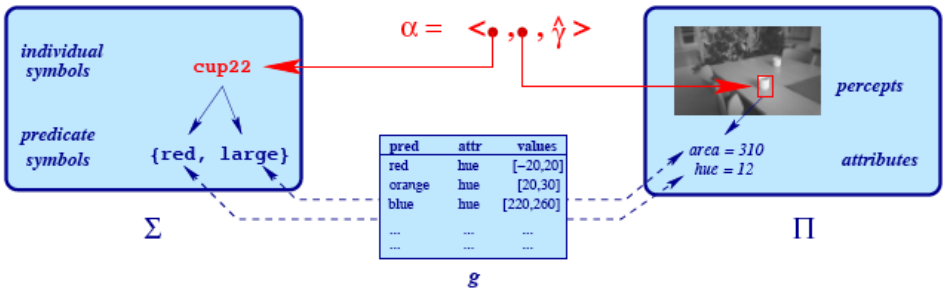
Π: ("**Perceptual System**"): Sensors

g: ("**predicate grounding relation**"): Relation for defining correspondences between predicates from Σ and fitting values of attributes observed in Π

α: the (object) **anchor** …



The Anchor in Object Anchoring



- Anchor α: Data structure: Pointer to symbol (in Σ) and sensor data (in Π);

- **Signature** γ: Estimation(!) of recent attribute values for identification in the sensor data
  - gets extrapolated when the object is out of sight
  - serves for tracking and/or reacquiring it
  - e.g. color, position, size, speed, …

Functions in Object Anchoring

- **Find**: Install anchor for given(!) symbol in Σ, which is compatible with g and anchors a recent percept from Π (cf. “scene labeling”: Find object of given type)
- **Track**: By alternating prediction of signature and observation, watch over time some object in the percepts
- **Reacquire**: Recover object, which was out of sight for some time. (Use recent anchor signature to that end!)
- “Additional functionalities will probably be needed for different types of anchoring processes, for instance, bottom-up anchoring”
- Who tells the individuals for which to make a new anchor from others? (Human enters my office – anchor? Bug flies into my office – anchor?
- Anchors need to exist for arbitrary instances of a class! (“An (arbitrary) mug on the shelf”)
- “Throw-away anchors” are needed (“The mug I am using right now”)

Have we Seen Something Similar Before?

Remember Bayes Filters?

$$\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) = \alpha \cdot \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}) \cdot \int_{\mathbf{x}_t} \left[\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t) \cdot P(\mathbf{x}_t|\mathbf{e}_{1:t}) \right]$$

Sensor model = estimation of object appearance in sensor data

Change model = pred. of next state

Localization can be understood as grounding the symbol Pose(x,z,β) or Pose(x,y,z,α,β,γ), resp.

- Expl. for more general symbol grounding – no physical object
- The complete localization literature deals with that grounding!
- Unfortunately, algorithms for *pose grounding* (aka. localization) cannot be generalized to grounding other symbols!

Overview

1. Object Anchoring – The Problem
2. **Object Anchoring Issues in Plan-Based Robot Control**

Issues

- Anchoring seems to go well together with knowledge representation in ontologies (Description Logic) and with plan-based robot control
- But there are (at least) two issues:
  - **object identity**
  - **aggregated objects**

What is Wrong with …?

Guest: “A mug of coffee, please!”

Waiter Robot: “Let me check in my recent KB…
Do you wish to get it in `mug-1`, `mug-2`, or `mug-5`?
I am sorry, `mug-3` and `mug-6` are in use;
`mug-4` was broken three months ago.”

G: “??? … What’s the difference? I don’t care! … Mug five!”

R: “They are identical, Sir, except for their identity. As you wish!”

… two minutes later …

R: “I am sorry, `mug-5` got used unintentionally by human staff.
I have secured `mug-2` here.
Would it be all-right if I served it filled with coffee?”

G: *… gets up to leave restaurant …*

R: “Would `mug-1` suit you better? …
Hello!?
WHAT IS WRONG WITH `mug-2`?!”

Some (not all) Objects are Equal

- Some everyday objects are designed for looking/being special
- Some (most?) are designed as equal mass products
- Those have all the **identical predicate grounding relation**
- They **cannot be discriminated** from sensor readings
- They **need not be discriminated** for all practical purposes
- Don’t try anchoring in this hopeless and useless case; **but** …

Functional Identity

- … don’t drink from my mug!
- Identity is often defined functionally (“my mug”)
- The object then has to be anchored based on features that are not object-intrinsic (e.g., spatial relations: “my mug” is the mug by “my place”)
- Functional identities can be transient wrt. the individual physical object (e.g., “my mug” goes away when the table is cleared)

How can transient functional identity be used both efficiently and formally sound in (representation, reasoning, and) object anchoring?

Functional Identity in KR&R for Robots

- … is an **open issue** (to my knowledge):
  - Waiter gets any mug with coffee from the counter (there may be several)
  - Mug becomes “guest-17’s mug” when served to her
  - identity released when clearing the table
- How would it be handled in a DL ontology (A-Box)?
  - Tying temporary roles to individual object could work – but does not work for a robot due to sensorial equality of objects (see above)
- How would it be handled in a planner?
  - Propositional planning inadequate
  - Using variables (schema or logical) inadequate
- Modeling as a set of resources appears to be most adequate
  - Single resource gets claimed (➔ temporal identity) and released

Aggregate Objects …

- … can be represented in DL
- Can be transient functional ("my cover")
- Then consist of parts that can, but need not be trans. funct.
  (include transient functional plate etc., but individual "mug-BVB")
- Must be anchored by perceiving the required parts …
- … but some parts may be missing or overlooked
  (Cover, but without spoon and glass)
- Perception becomes abductive process: from detected parts and aggregate hypothesis abduce existence of missing parts
- Highly heuristic!
  (From a single spoon, you could, but should not abduce dinner table!)

Scene Interpretation According to Neumann/Möller

B. Neumann, R. Möller: On Scene Interpretation with Description Logics
in: Christensen & Nagel (eds.): Cognitive Vision Systems, 2006

A **Scene Interpretation** is a consistent theory (ABox+TBox) in a Description Logic over the TBox of the defined concepts, based on labelings of sensor data

Incompleteness of Scene Interpretations

- Objects need not be part of aggregates
  (e.g.: There is a plate that is not part of a cover)
- Objects need not be instantiated in the most specific way
  (e.g.: There is a piece of silverware that is not identified as either knife or fork or spoon)
- Not all parts of compound objects need be completely instantiated (e.g.: the cup of a particular cover is not instantiated)

Functions in Scene Interpretation

- **Aggregate Instantiation**: Given objects, pool them into one aggregate (for logicians/GOFAI-ists: a form of **abduction**!)
  Example: Summarize recognized instances of types plate, knife, fork, saucer into an instance of cover (lacking a cup)
- **Instance Specialization**: Refine object into instance of one of its subclasses
  Example: Breakfast plate rather than plate
- **Instance Expansion**: Instantiate more parts of an aggregate instance
  Example: Assign cup instance to previously cup-less cover instance
- **Instance Merging**: Identify allegedly different instances of the same class as one single physical object
  Example: Contour and texture module have identified and instantiated 1 knife each in neighboring positions – call it 1 object!

Perception = Hallucination + Control

Perception is controlled hallucination. **Max Clowes, 1971**

Heuristic decisions in scene interpretation

- Choose data to work on
  (image region, elementary object/s, aggregate/s)
- Choose type of interpretation step
- Choose preferred way of executing this step
  (**Example**: Specialize into what? Expand by what?)

Sum up: Some Open Points in Neumann/Möller

- There is **no complete implementation** (for all that I know) – not for the image processing application,
  not to mention for a robot
- How to **choose possible scene interpretation steps**
  (aggregate instantiation, …)
  and their possible parametrization/application at any time?
  (N/M: probabilistic approach; blackboard architecture looks plausible)
- Nothing is said about **object anchoring**: How do we care for a continuous identity of objects over time, even if they have vanished from the sensor data for some time?

Thank you for your time!