Artificial General Intelligence

2. Concept-Centered Representation

Dr. Pei Wang

Dr. Patrick Hammer

Dr. Robert Johansson

Digital Futures (KTH, SU, RISE)

Stockholm, Fall 2022

Describing a system

An information-processing system consists of:

- goals (motivations, drives, tasks, problems, …)
- actions (operations, responses, solutions, …)
- knowledge (beliefs, skills, programs, …)

A system can be described as

- a theory in a human language
- a model in a formal language
- an implementation in a computer language

Formal description

- The purpose of formalization
- Common approaches in formalization:
  - Numerical vectors and functions
  - Grammar and inferential rules
  - Data structures and algorithms
- These approaches are equivalent in principle, but with different extents of naturalness and efficiency for a specific job

Reasoning system

The logic part:

- A language for knowledge representation
- A set of inference rules on the language
- A semantics for interpretation and justification

The control part:

- A memory structure for knowledge and tasks
- A mechanism for task and knowledge selection

Properties of reasoning systems

- Separation of the meta-level and object-level
- Separation of the logic and the control
- Generality of the grammar and inference rules
- Explainability and justifiability

Issues:

- The association with "symbolic AI"
- The heritage of mathematical logic
- The criterion of validity under AIKR

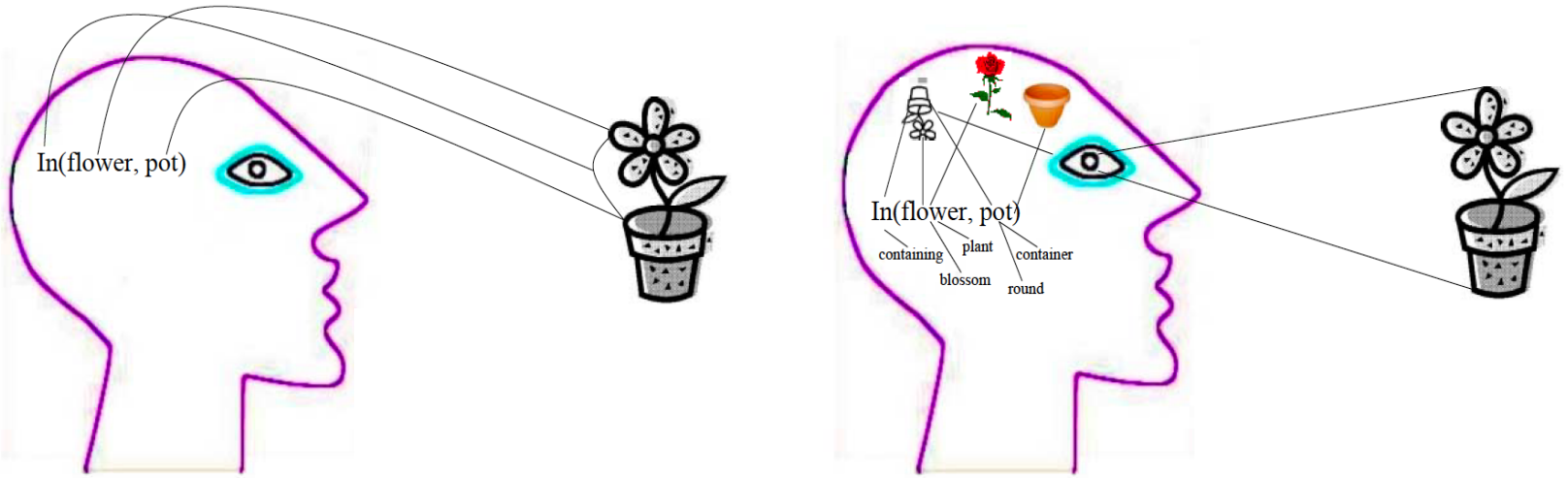
Conceptualizing experience

- Experience as a history of interaction
- Concepts as abstract segments of experience that are identifiable and manipulatable within system
- Concepts link the past, present, and future
- Concepts summarize experience
- Knowledge as conceptual relations
- The same experience may be conceptualized differently in different situations

The meaning of a concept

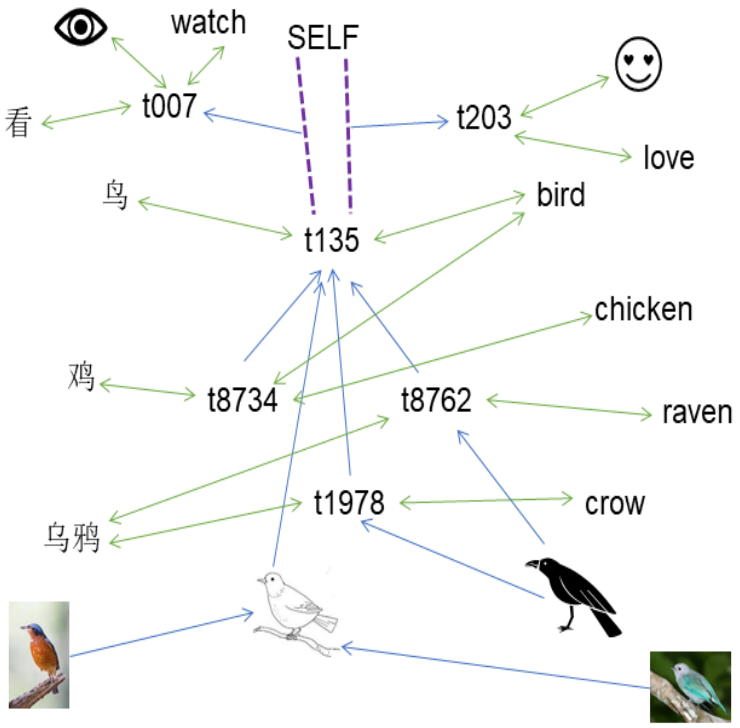
- What a concept means to a system at a moment depending on the experience it represents
- The meaning of a concept is in its experienced relations with other concepts, including what it can be associated with temporally, spatially, and abstractly
- A concept has an internal identifier, but is not a symbol representing an external object or event

The tale of two semantics



- Different ontological assumptions
- "Chinese room" and "Symbol grounding"
- Perception as assimilation-accommodation

Memory as a concept network



- Every concept is named by a term
- Terms can be sensory, operational, linguistic, internal, compound
- Various extents of abstraction
- No symbol/subsymbol

Truth-value as evidential support

- Knowledge summarizes experience abstractly, so uncertainty must be represented and processed
- Empirical truth-value as degree of belief according to evidential support
- Each statement states a conceptual relation which can be experienced repeatedly
- In idealized cases, each time the relation is either *true* or *false*, with respect to the experience

Measurements of uncertainty

- With the idealized cases as units, the amounts of positive and negative evidence are w^+ and w^-
- *Frequency*: $f = w^+ \,/\, w^-$
- *Confidence*: $c = w \,/\, (w + k)$
- Future evidence $k \geq 1$, as a personality parameter
- *Frequency interval* in the near future:

$$[l, u] = [w^+ \,/\, (w + k),\ (w^+ + k) \,/\, (w + k)]$$

- *Ignorance*: $i = u - l = k \,/\, (w + k)$

Unity of the measurements

Table 3.1. The mappings among measurements of uncertainty.

to\from	$\{w^+, w\}$	$\langle f, c\rangle$	$[l, u]$ (and i)
$\{w^+, w\}$		$w^+ = k \times f \times c/(1-c)$ $w = k \times c/(1-c)$	$w^+ = k \times l/i$ $w = k \times (1-i)/i$
$\langle f, c\rangle$	$f = w^+/w$ $c = w/(w+k)$		$f = l/(1-i)$ $c = 1-i$
$[l, u]$	$l = w^+/(w+k)$ $u = (w^+ + k)/(w+k)$	$l = f \times c$ $u = 1 - c \times (1-f)$	

Comparison with probability

- Axiomatic definition of probability
- Interpretations: *objective*, *subjective*, *logical*
- Normativity: consistency, "Dutch Book"
- "Human cognitive biases"
- Second-order and imprecise probability
- Limitation of Bayesian learning
- Accuracy and non-numerical representation
- Inconsistency: inevitable and manageable

Suggested Readings

- Ernest Davis, Knowledge Representation
- Wilfrid Hodges, Model theory
- John R. Searle, Minds, Brains, and Programs
- Stevan Harnad, The symbol grounding problem
- Pei Wang, Non-Axiomatic Logic, Chapter 2, 3
- Pei Wang, Rigid Flexibility, Chapter 7, 8