Self-Awareness in Autonomous Systems summary

A. Matakos, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.0.1

Artificial Intelligence & Information Analysis Lab

Self-Awareness in Autonomous Systems

- **Self-awareness definition**
- Self-aware systems
- Cognitive architecture

Introduction

- What is Self-Awareness?
- Some definitions first
- Self-Awareness is:
  - “A conscious knowledge of one's own character and feelings”
  - “An awareness of one's own personality or individuality”
- A physical process from a Neurology standpoint
  - We’re not sure how exactly self-awareness arises in Ego-things
- A mental process from a Psychology standpoint

What is Self-Awareness?

- Nassier argues about 5 levels of self-knowledge:
  - ➢ The ecological self
  - ➢ The interpersonal self
  - ➢ The extended self
  - ➢ The private self
  - ➢ The conceptual self
- These layers are developed in parallel from infancy
- Self-awareness arises by the interaction of the various levels of self-knowledge

What is Self-Awareness?

- Many things around the subject are still unclear
- Its foundations are not proven to exist
  - Some Neurologists argue it's an illusion
- There is no genuine theory of the underlying principles and methods that give rise to true Self-Awareness
- This makes implementation ever so harder
- Some propose to ground it in the solid theoretical framework of Integrated Information Theory

Technical Difficulties

- Extremely demanding processing capabalities in order to run live
- High complexity of inter-connected algorithms and systems
- Training requires huge amounts of properly refined data
- Real-world implications which require:
  - ➢ Extreme reliability
  - ➢ Accuracy
  - ➢ Conflict-resolving ability

Motivations for Self-Awareness in robots

- The benefits are endless
- The path to AGI / Superintelligence goes through machines that are Self-Aware
- Conscious, meta-selfaware Ego-things
  - ➢ Can assist humans in every aspect of their lives
- Huge advancements for
  - ➢ Medicine
  - ➢ Industry / IT sector
  - ➢ Biology, Physics, Computer Science, Engineering, etc.

Self-Awareness in Autonomous Systems

- Self-awareness definition
- **Self-aware systems**
- Cognitive architecture

Cognitive Architectures

Some existing cognitive architectures:

- ACT – R (2004) (Adaptive Control of Thought-Rational)
  - ➢ Classical rule-based system
  - ➢ (facts + events) organized with (production rules + procedures)
  - ➢ Memory contains “chunks” of data
  - ➢ Rules applied to chunks lead to:
    - o Robot action in the world
    - o Change in declarative memory
  - Chunks that are frequently selected have a higher chance of being selected again

Cognitive Architectures

- ACT – R (2004) (Adaptive Control of Thought-Rational)
  - New rules are created from trying to achieve goals
  - Costs and Success rates of the rules converge through an optimization procedure
- SOAR (2006) (State, Operator And Result)
  - Aims to model human cognition
  - Knowledge represented by production rules
  - Rules conditions are matched to working memory, where attributes and values are encoded
  - “Chunking” + reinforcement learning produce new rules

A theoretical approach

- “Levels of Networked Self-Awareness” (2018) L.Esterle, J. Brown
- Tackles the issue of live runtime of SISSY systems
- Self-Awareness alone is not enough
- Self-Aware systems must be aware of the self-awareness of other systems.
    - ➢ Just like humans, e.g. a baby falls into inanimate objects until it learns these objects are not self-aware
- These systems create networks

Another approach

- "Towards Self-Aware Robots" (2018), by Chatila R, Renaudo E, Andries M, Chavez-Garcia R-O, Luce-Vayrac P, Gottstein R, Alami R, Clodic A, Devin S, Girard B and Khamassi M
  - A more sophisticated and hands-on approach
- Self-awareness must first rely on perception of self as different from the environment and from other agents
- This is achieved via "Perception and Affordance learning"
- The robot develops sensorimotor representations and not just exteroceptive representations
  - Perception is thus not an isolated process
  - The robot learns that the same object can have different meaning to different individuals
  - A link between self-awareness and situation-awareness is formed

Perception and Affordance learning

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

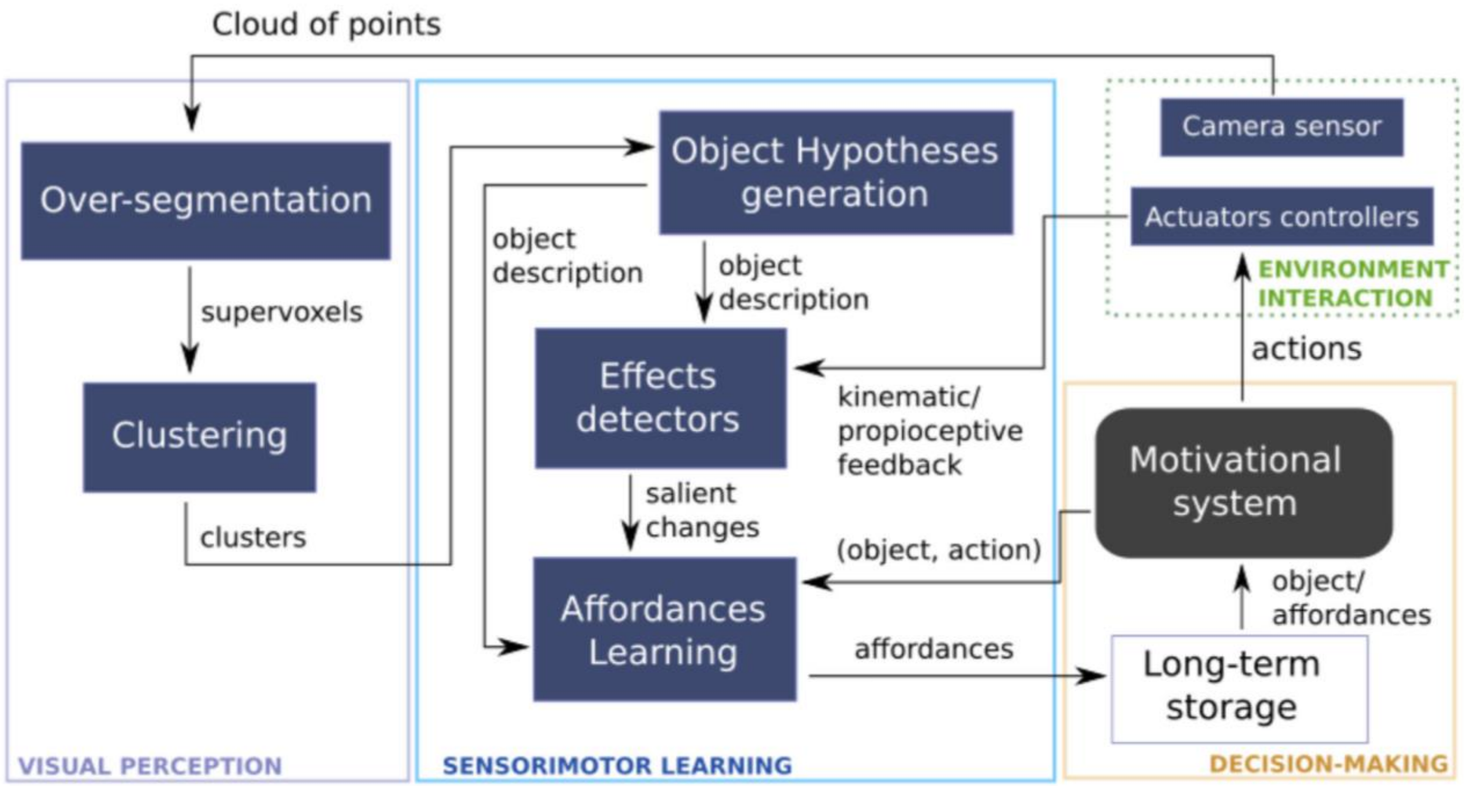

FIGURE 1 | Architecture of the proposed sensorimotor approach for scene affordance learning.

Perception and Affordance learning

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

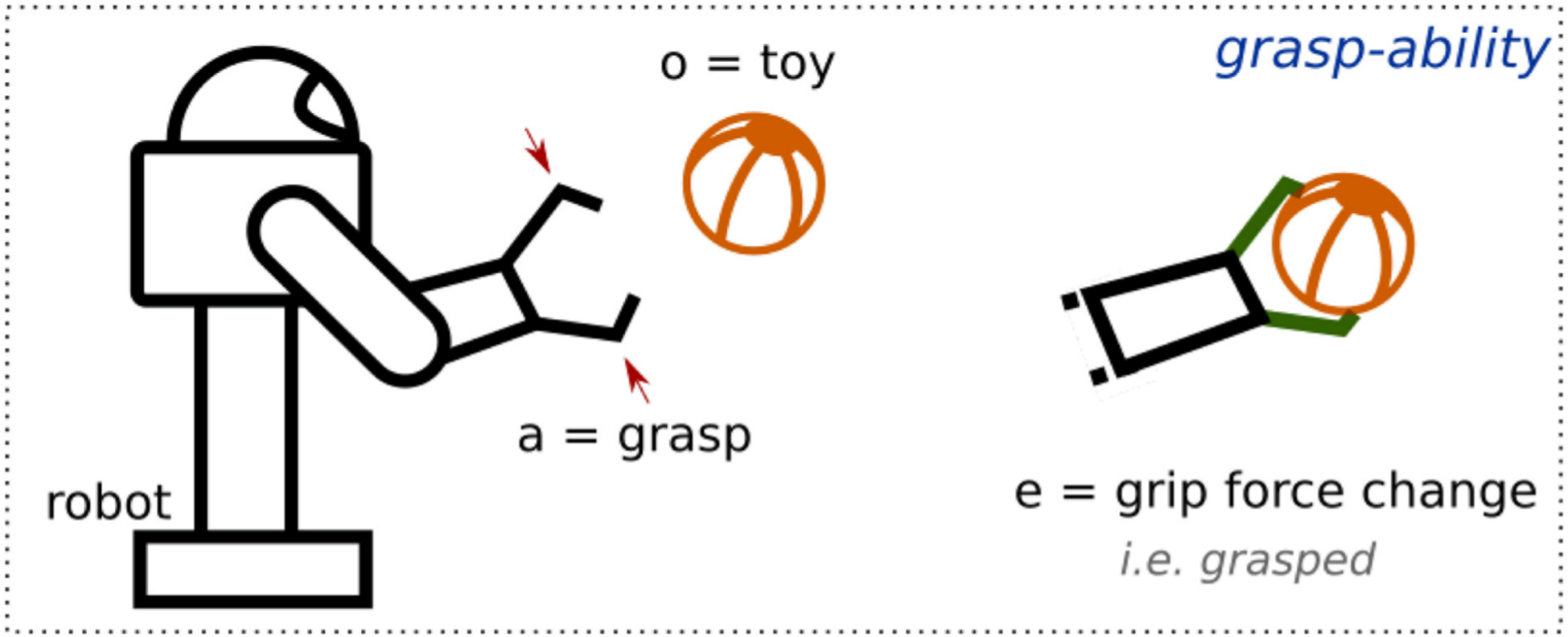

FIGURE 3 | Representation of the *grasp-ability* affordance relation.

Learning Actions and Plans

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

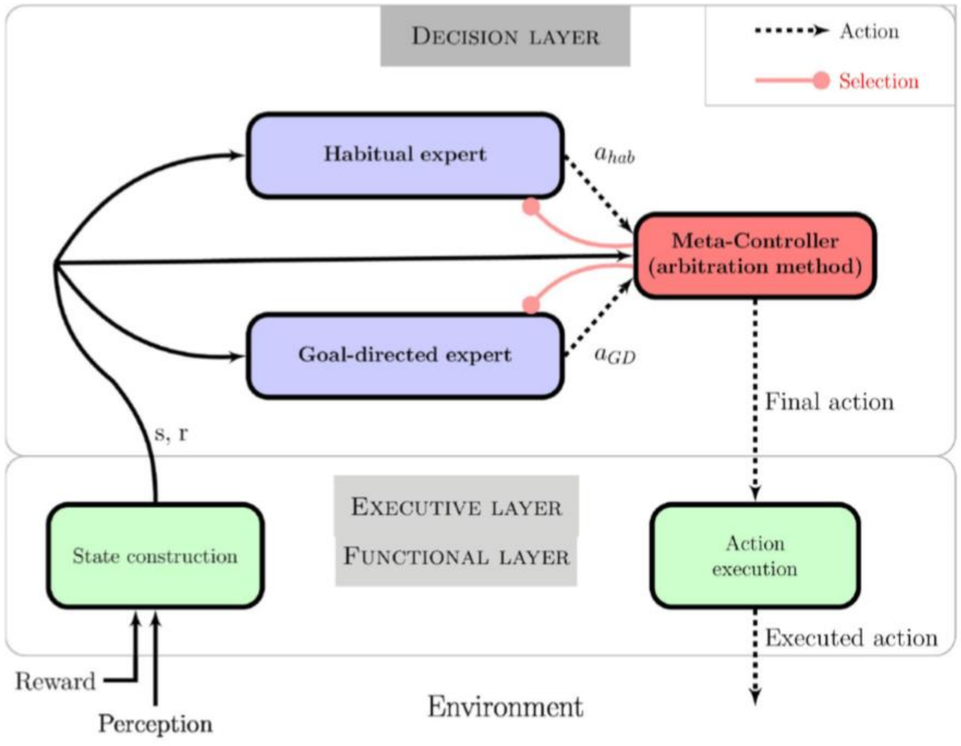

FIGURE 4 | Global action selection architecture composed of two decision systems implementing corresponding behaviors: the goal-directed expert is a model-based RL algorithm whereas the habitual expert is a model-free RL algorithm. The meta-controller is in charge of monitoring different expert information, giving control to one of the two. The reward information comes from the motivational system and represents the goal of the task.

Artificial Intelligence & Information Analysis Lab

Joint Human-robot action

- A PR2 robot sharing with a human the goal of cleaning a table:

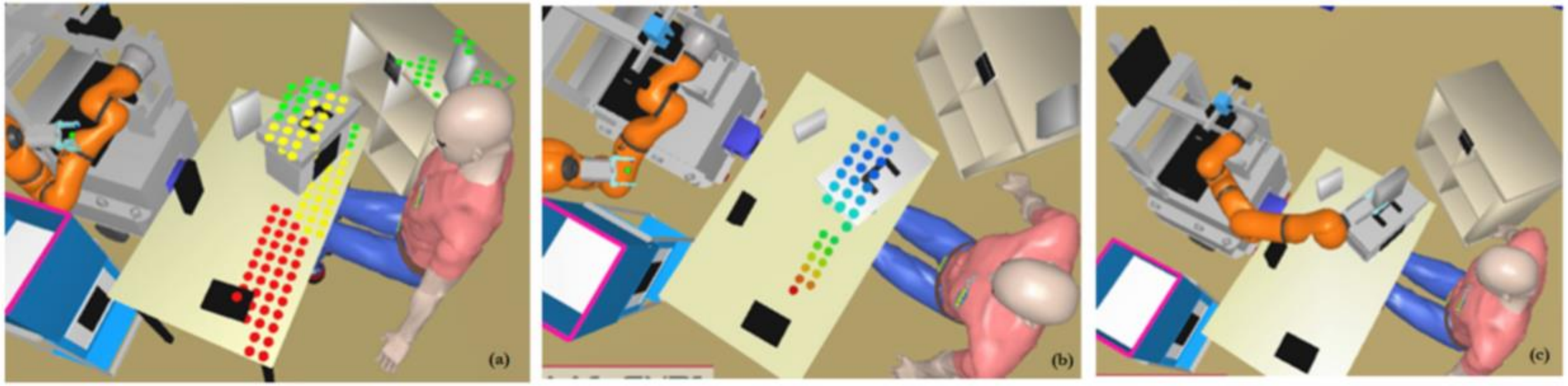

FIGURE 8 | Task of making an object accessible by the human to the robot (Pandey et al., 2013): **(a)** Places on the support planes where the human can put something with least effort. **(b)** Weighted points where the robot can support the human by taking the object. **(c)** The planner found a possible placement of the object on the box from where it is feasible for the robot to take. Note that, because of the object-closeness based weight assignment, this placement also reduces the human's effort to carry the object.

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

Joint Human-robot action

FIGURE 9 | Initial state of the world in the Clean the table scenario. In this task, the robot and the human share the goal of cleaning the table together.

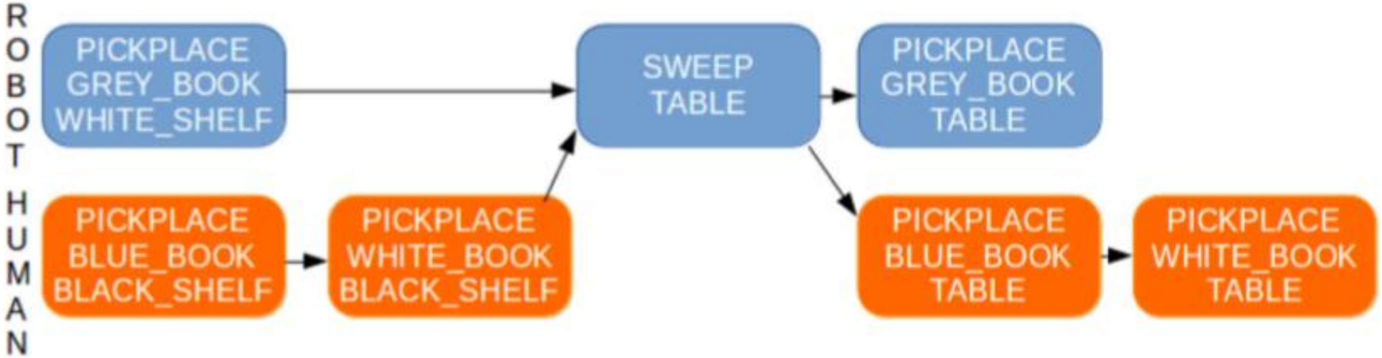

FIGURE 10 | Shared plan computed by the robot to solve the joint goal: first removing the three objects (books) that are located on the table, then sweeping the table in order to clean it and finally placing the objects back on the table. While cooperatively achieving the task, the robot will be able to detect and assess correctly why the human partner stays idle, for instance in cases where, due to a momentary absence, the human may have missed the fact that robot has swept the table.

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

Motivational System

- Implementing motivation for the robot
  - ➢ Consists of goals which are associated with rewards
    - Goals can be chained or individual
    - Goals can be permanently active or not
  - ➢ Goals are achieved through policies
- Motivational architecture:

1. Handles motivations
2. Computes possible policies for each motivation
3. Predicts the behavior of each policy and its effect on motivations
4. Predicts the effects of a chain of policies
5. Finds an optimal arrangement of the policies, using a reward maximizing function

Decision-making system

- Based on the Motivational system, a Decision-making system is organized in modules:
  - ➢ Intentional module, where objectives are in the form of motivations
  - ➢ Operational module, which computes policies
  - ➢ Deliberation module
- The Deliberation module is the "brains" of this model:
  - ➢ It provides the goal that will reward the operational model
    - o This in turn triggers the corresponding motivation, and creates policies
  - ➢ Then it computes the effects of these policies on all motivations

Decision-making system

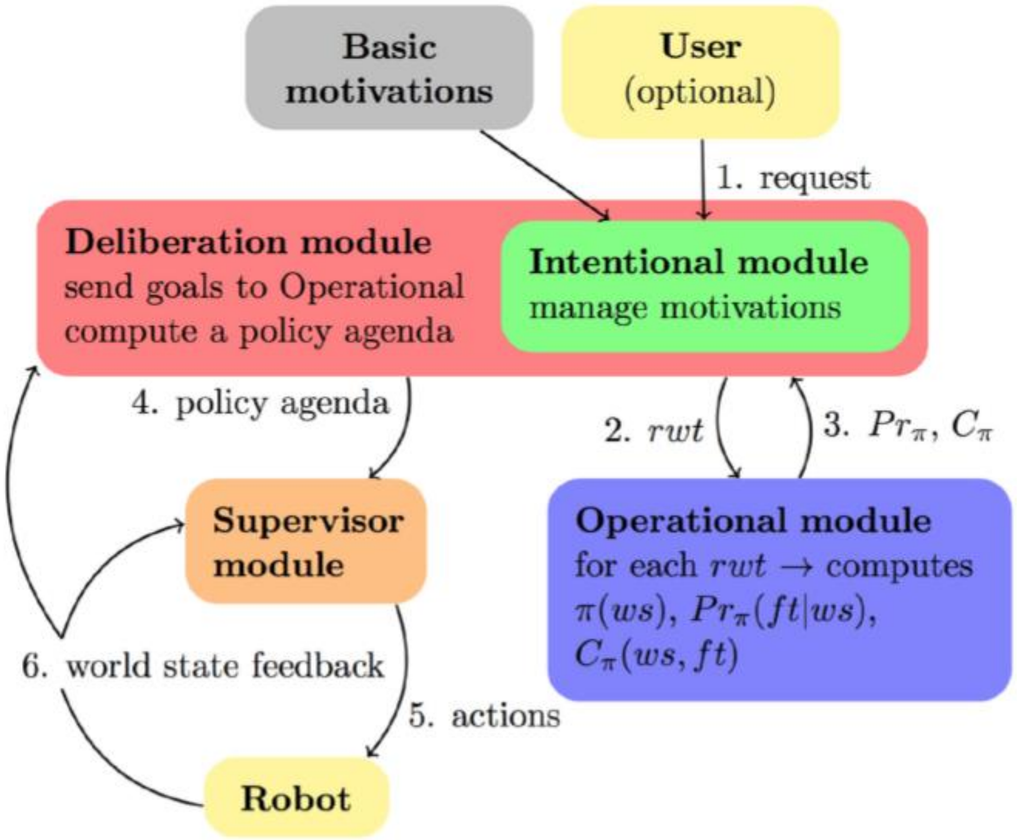

FIGURE 11 | Decision-making architecture including operational, intentional and deliberation modules. The deliberation module implements a meta-reasoning capability.

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

Self-Awareness in Autonomous Systems

- Self-awareness definition
- Self-aware systems
- **Cognitive architecture**

The Cognitive Architecture

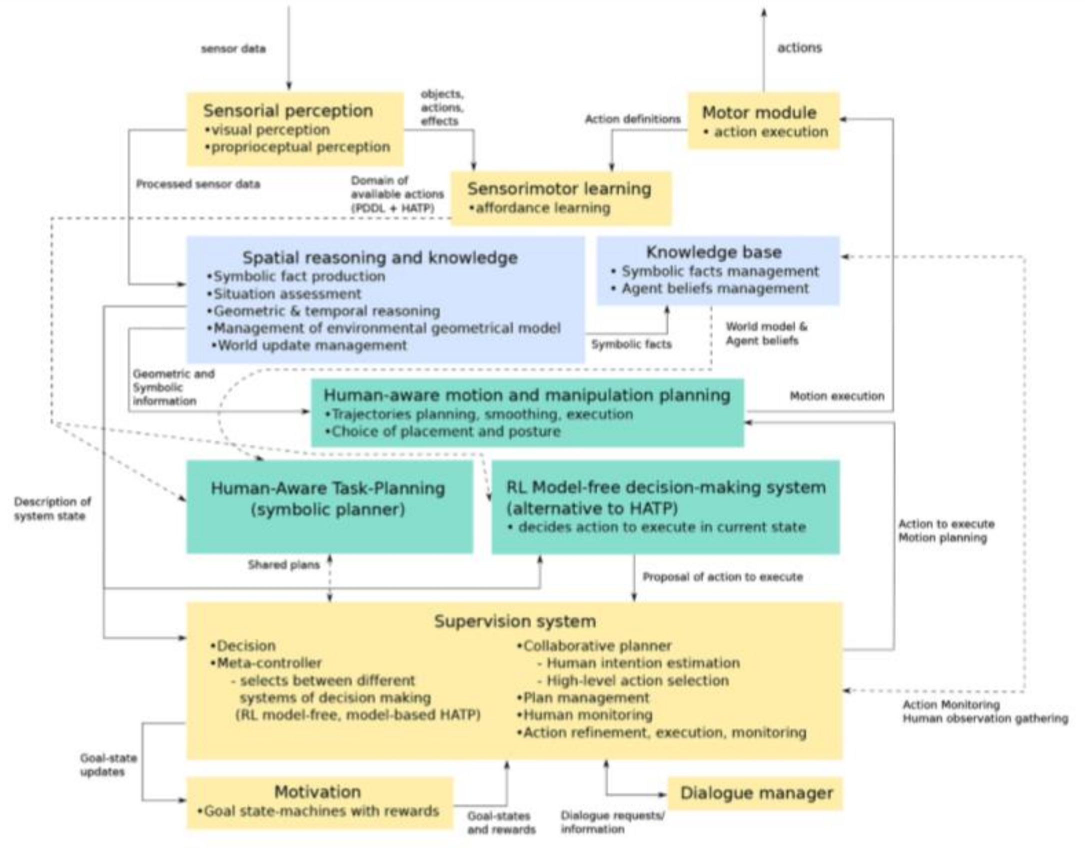

FIGURE 12 | The global cognitive architecture employed in the RoboErgoSum project. Blue modules are responsible for generating and managing the symbolic knowledge. Decision-making modules are shown in green. Solid and dashed lines are used only to improve diagram readability where lines cross, and are otherwise identical in meaning.

Source:https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full

Bibliography

1. Ulric Neisser (1988) Five kinds of self-knowledge, Philosophical Psychology, 1:1, 35-59, \\DOI: 10.1080/09515088808572924 https://doi.org/10.1080/09515088808572924
2. Chatila R, Renaudo E, Andries M, Chavez-Garcia R-O, Luce-Vayrac P, Gottstein R, Alami R, Clodic A, Devin S, Girard B and Khamassi M (2018) Toward Self-Aware Robots. Front. Robot. AI 5:88. doi: 10.3389/frobt.2018.00088 https://www.frontiersin.org/articles/10.3389/frobt.2018.00088/full
3. Schilling, K., De Lafontaine, J. & Roth, H. Autonomy capabilities of European deep space probes. Auton Robot 3, 19–30 (1996). https://doi.org/10.1007/BF00162465
4. The AI Revolution: The Road to Superintelligence, by Tim Urban https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html
5. Lukas Esterle, John NA Brown, Levels of Networked Self-Awareness DOI 10.1109/FAS-W.2018.00054 https://ieeexplore.ieee.org/document/8599561

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr