

# Simultaneous Localization and Active Phenomenon Inference (SLAPI)

**Olivier L. Georgeon**

OGEORGEON@UNIV-CATHOLYON.FR

*UR Confluence, Sciences et Humanités (EA 1598) - Lyon Catholic University, France*

**Titouan Knockaert**

TITOUAN.KNOCKAERT@GMAIL.COM

*Université Claude Bernard Lyon 1, LIRIS CNRS UMR5205, F-69622 Villeurbanne, France*

**Editor:** Editor's name

## Abstract

This is the abstract for this article.

**Keywords:** List of keywords

## 1. Introduction

The problem of getting mobile robots to autonomously learn the position of surrounding objects, recognize them, and keep track of their relative displacements is considered by many to be a key prerequisite of truly autonomous robots. Within this framework, the SLAM problem (Simultaneous Localization and Mapping) has been formalized and studied since the 1990s: constructing and updating a map of an unknown environment while simultaneously keeping track of the robot's position within it (e.g., [Taketomi et al., 2017](#)). SLAM algorithms are tailored to the available resources: odometric sensors, sensors of the environment, computational capacities, as well as the landmarks' properties, quantity, and dynamics, and the usage intended for the robot.

When displacements are imprecise and odometric data is not available, when landmarks are not directly identifiable, and below a certain level of scarcity and noise in the sensory data relative to the environment's complexity, it becomes difficult to perform SLAM accurately enough to use the robot for tasks involving complex navigation ([Gay et al., 2021](#)). For such robots, we propose the SLAPI problem: Simultaneous Localization and Active Phenomenon Inference. In contrast with SLAM, SLAPI does not aim at constructing a map to use for navigation. Instead, it aims at organizing behavior spatially in the vicinity of objects to design robots that exhibit intrinsic motivation (e.g., [Oudeyer et al., 2007](#)) such as playfulness and curiosity as they discover and interact with unknown objects. Possible applications may not include delivery tasks but may include entertainment and games, similar to playing with pets.

SLAPI makes no assumption that landmarks can be directly and passively uniquely identified through sensors. The robot must rather actively interact with objects, possibly from different angles and through different modalities of control loops, to categorize and recognize objects, and possibly use them as landmarks. We call this process *active phenomenon inference*, drawing from the work of [Friston et al. \(2021\)](#) on active inference. Here the term *phenomenon* refers to the knowledge of physical objects actively constructed by the robot from its point of view and “as the robot experiences the object through interaction” ([Thórisson, 2021](#)).

We designed a proof-of-concept algorithm to illustrate the SLAPI problem. We demonstrated it in a robot mounted on omnidirectional wheels and endowed with an echo-localization sensor, photosensitive sensors, and an inertial measurement unit, but no camera and lidar. As the robot circles around an object, it constructs the phenomenon corresponding to this object under the form of the set of the spatially-localized control loops that the object affords to the robot. New elements can be subsequently added to this set as the robot improves its knowledge of the object. Results show that the robot drew out a few cries of amusement and endearment from some human observers, which encourages us to keep improving this range of algorithms.

Moreover, we believe that the study of SLAPI problems can shed some light on how animals construct knowledge of objects through sensorimotor interactions, while keeping this knowledge grounded in experience. It can also provide an angle of attack to the more general AI problem of self-motivated open-ended learning in the real world.

## 2. The representational status of sensory data

An autonomous agent faces the necessity to actively infer the existence and the properties of objects in its environment when such existence and properties are not directly registered in sensory data. This raises the question of the *representational status of sensory data*: is sensory data representational or not? This question has been discussed times and again at the philosophical level (e.g., Williford, 2013). Loosely, two hypotheses collide: the hypothesis that sensory data carry information about features of the world, versus the hypothesis that sensory data carry information about the agent’s experience of interaction with the world. We refer to the former as the *representationalist hypothesis*, and to the latter as the *constructivist hypothesis* because it relates to Piaget’s theory of constructivist learning based on sensorimotor schemes (Guillermin and Georgeon, 2022).

SLAPI falls within the constructivist hypothesis because it applies to robots that have coarse sensors that do not provide much descriptive information about the environment. The robot must probe the environment a little bit like a blind person who uses a cane to actively construct a mental representation of its surroundings. Probing experiences consist of control loops during which the robot interacts with the environment. They are triggered by an action intended by the robot and result in an outcome that is informative not of the object itself but of the possibility of interaction afforded by the object to the robot. Figure 1 shows this cycle of interaction. The software selects an action associated with spatial information that specify how the control loop should be enacted in the world. In return, the software receives an outcome associated with spatial information that describes how the control loop has been enacted depending on the actual nature and position of surrounding objects.

Note that the constructivist hypothesis accepts that the outcome may sometimes register features of the environment but only avoids baking this assumption in the algorithm *a priori*. Rudrauf et al. (2017) have also proposed an active inference model related to the constructivist hypothesis. He states that “All we need here is the idea that in one way or another the sensory organs provide an independent source of input and correction for the continually updated world model” (p. 19).



Figure 1: The interaction cycle. Black bullet: the cycle begins with the software selecting an action containing spatial information to enact in the world (right). Black arrowhead: the cycle ends with the software receiving the outcome containing spatial information (left).

### 3. The experimental setup

We use the robot cat of brand Osoyoo<sup>1</sup>, to which we added an inertial measurement unit (Figure 2). We set up an environment with black lines on the floor and various objects that can be detected through the echo-localization sensor. Our goal is to design a robot that explores such environment and constructs knowledge of the objects that it encounters to exhibit behaviors that look interesting to human observers.

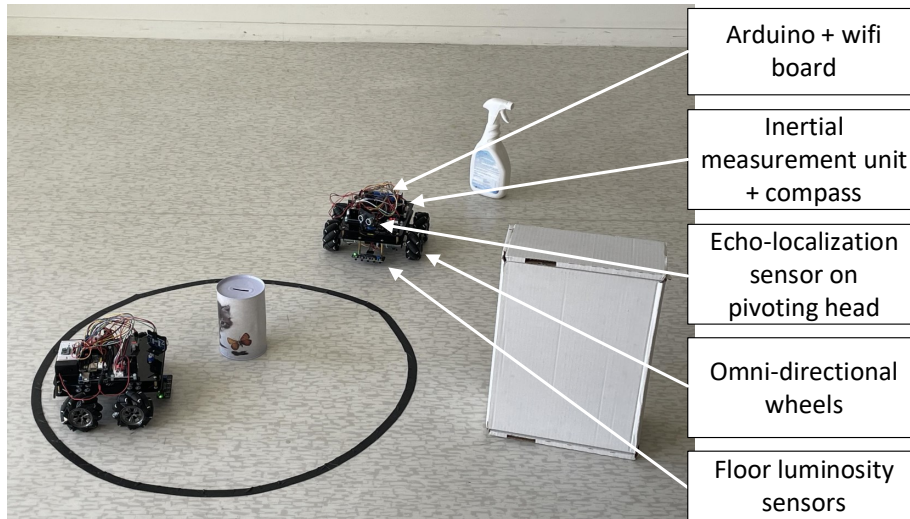


Figure 2: The experimental setup. Osoyoo robots, a black line on the floor, and various objects detectable through echo-localization.

1. <https://osoyoo.com/2019/11/08/omni-direction-mecanum-wheel-robotic-kit-v1/>

#### 4. The software architecture and algorithm

We implemented the cognitive architecture on a remote PC as depicted in Figure 3. This architecture takes the place of the *Software* in Figure 1, whereas the robot plus its environment takes the place of the *World*.

The architecture follows a regular Model-View-Controller design pattern in which the *Memory* plays the role of the database, and the *Workspace* plays the role of the Model. The *Memory* contains the Egocentric memory and the Allocentric memory which are displayed on screen via the *View Controllers* (Figure 3, right). The *Workspace* contains the *Synthesizer* which infers the phenomena, and the *Decider* which selects the next action to send to the robot through wifi.

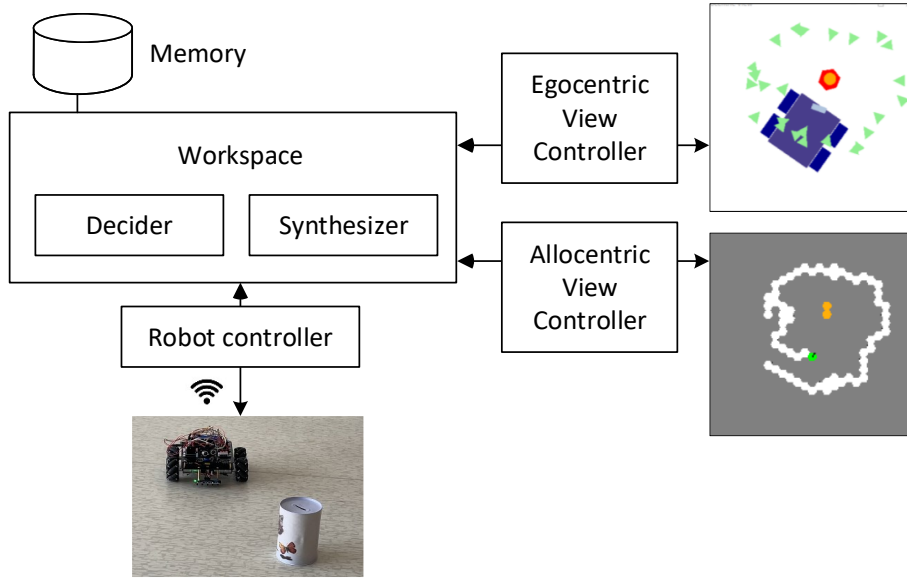


Figure 3: The software architecture implemented on a remote PC. Bottom: the robot receives the action and sends back the outcome through wifi. Top-right: Egocentric memory shows the position of echos (orange circles), the focus of attention (red hexagon), and the previous positions of the robot (green triangles). Bottom-right: Allocentric memory shows the position of the echo (orange cells), the explored cells (white), the position of the robot (green), the unexplored cells (grey).

##### 4.1. The primitive control loops implemented in the robot

A C++ program on the robot’s Arduino board handles the reception of the action, drives the control loop, and then returns the outcome through wifi. Table 1 lists the supported actions and outcomes.

The reception of an action triggers the enaction of the corresponding control loop until its termination condition. For example, the **Forward** action sets the robot in motion. It

has three possible termination conditions: **Default**: Time out of 1 second (approximately 20 cm traveled). **Line detection**: The floor sensor detects a line; the robot retreats back for a few centimeters. **Shock**: The inertial measurement unit detects a strong deceleration indicating a shock with a solid object.

Table 1: Actions available to the robot and their possible outcomes

(a) Actions		(b) Outcomes	
Code	Description	Code	Description
Forward	During 1 sec.	Line left	Floor sensors
Backward	or until shock	Line front	cross luminosity
Shift left	or line	Line right	threshold
Shift right	detection	Shock	Violent deceleration
Turn left	$\pi/4$	Echo lost focus	No echo where expected
Turn right	$-\pi/4$	Echo left	
Head scan	$[-\pi/2, \pi/2]$	Echo right	Direction and range of the nearest echo
		Echo far left	
		Echo far right	
		Echo far front	
		Echo close front	
		Default	No line, no echo

Besides the action code, the cognitive architecture sends two more values to the robot: the coordinates of a focus point, and an estimated speed. The robot’s program uses these to keep the robot’s head aligned towards the focus point during the robot’s displacement. This helps the robot keep tracks of objects over various moves. It also gives the human observer the impression that the robot keeps its attention to a particular object, making it look more alive.

The robot also returns additional information to the cognitive architecture: the yaw, azimuth (angle relative to the north) and duration of the various phases of the control loop. The cognitive architecture uses this information to update the spatial memory based on the robot’s displacement. Moreover, the robot’s program controls the head so that it aligns itself towards the nearest echo. The robot then returns the nearest echo measure along with the head direction. The cognitive architecture uses this information to mark the position of echos in spatial memory. This will be used to infer the presence of an object in the area. Table 2 summarizes the data exchanged between the PC and the robot.

Table 2: Dialogue between the PC and the robot through wifi

PC to Robot	Action code, focus coordinates $(x, y)$ , estimated speed $(x, y)$
Robot to PC	Outcome code, echo distance, head direction, yaw, azimuth, duration

## 4.2. The egocentric memory

The egocentric memory stores the robot’s *experiences* in a coordinate system centered on the robot. It is inspired by the brain’s egocentric cells located in the superior colliculus (Grievès and Jeffery, 2017).

An experience is a data structure that contains the action, the outcome, the position in space relative to the robot, and the timestamp. The position in space corresponds to where the interaction took place, for example under the floor sensors for a line detection, or at the estimated origin of an echo. The positions of experiences are updated with the robot’s displacement on each interaction cycle.

Of course, errors accumulate causing a drift in the position of experiences, making them unreliable after a few cycles. In the screen display (Figure 3, top right), we make them fade away and disappear after 5 cycles.

## 4.3. The allocentric memory

The allocentric memory stores a spatial representation in a coordinate system relative the environment. It is inspired by *grid cells* in the hippocampus of the mammalian brain (Grievès and Jeffery, 2017). Like natural grid cells, we implemented an hexagonal cell grid (Figure 3, bottom right). Each cell stores the list of experiences in a particular place, and can have one of the following status:

<b>Unknown</b>	No experience has been attempted here.
<b>Occupied</b>	The robot is placed here.
<b>Empty</b>	The robot has been here in the past.
<b>Experience</b>	One or several experiences have been placed here.

Allocentric memory is used to construct a local spatial representation of the objects with which the robot is interacting. This representation is copied to a data structure called a *phenomenon*.

We use the term phenomenon in its common sense: the perception by a cognitive being of “something” in the environment. In computer science, we comply with Thórisson (2021)’s definition of “Any useful grouping of a subset of spatio-temporal patterns experienced by an agent in an environment”.

In our implementation, the phenomenon data structure stores a local map of experiences afforded by the object. This map is a subset of egocentric memory converted to a coordinate system centered on the phenomenon. This process of phenomenon inference is performed by the Synthesizer.

## 4.4. The synthesizer

The synthesizer constructs a representation of an object by reading experiences from egocentric memory and filling and updating allocentric memory. It works in coordination with the Decider (Section 4.5) to explore the object through different directions.

The synthesizer estimates the robot’s position in allocentric memory through path integration. It infers the presence of an object from repeated experiences in the same area. Once an object has been identified with enough experiences, it can be isolated in allocentric memory and copied to a phenomenon data structure.

#### 4.5. The decider

### 5. Results

### 6. Conclusion

This example is meant to illustrate a more profound epistemological assumption that the world *in itself* cannot be accessed through direct representational data. This idea goes back to Kant, relates to Piaget’s developmental psychology and constructivist epistemology, and has echos in modern physics. Karl Friston said that addressing this problem may open the door to the third wave of AI (? , time code 22:66).

### References

- Karl Friston, Rosalyn J. Moran, Yukie Nagai, Tadahiro Taniguchi, Hiroaki Gomi, and Josh Tenenbaum. World model learning and inference. *Neural Networks*, 144:573–590, December 2021. ISSN 0893-6080. doi: 10.1016/j.neunet.2021.09.011. URL <https://www.sciencedirect.com/science/article/pii/S0893608021003610>.
- Simon Gay, K vin Le Run, Edwige Pissaloux, Katerine Romeo, and Christ le Lecomte. Towards a Predictive Bio-Inspired Navigation Model. *Information*, 12(3):100, March 2021. ISSN 2078-2489. doi: 10.3390/info12030100. URL <https://www.mdpi.com/2078-2489/12/3/100>. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- Roddy M. Grieves and Kate J. Jeffery. The representation of space in the brain. *Behavioural Processes*, 135:113–131, 2017. ISSN 0376-6357. doi: <https://doi.org/10.1016/j.beproc.2016.12.012>. URL <https://www.sciencedirect.com/science/article/pii/S0376635716302480>.
- Mathieu Guillermin and Olivier Georgeon. Artificial Interactionism: Avoiding Isolating Perception From Cognition in AI. *Frontiers in Artificial Intelligence*, 5, 2022. ISSN 2624-8212. doi: 10.3389/frai.2022.806041. URL <https://www.frontiersin.org/article/10.3389/frai.2022.806041>.
- Pierre-Yves Oudeyer, Fr dric Kaplan, and Verena V. Hafner. Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286, April 2007. ISSN 1089-778X. doi: 10.1109/TEVC.2006.890271. URL <http://ieeexplore.ieee.org/document/4141061/>.
- David Rudrauf, Daniel Bennequin, Isabela Granic, Gregory Landini, Karl Friston, and Kenneth Williford. A mathematical model of embodied consciousness. *Journal of Theoretical Biology*, 428:106–131, September 2017. ISSN 00225193. doi: 10.1016/j.jtbi.2017.05.032. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022519317302540>.

Takafumi Taketomi, Hideaki Uchiyama, and Sei Ikeda. *Visual SLAM algorithms: a survey from 2010 to 2016*. 2017. Publication Title: IPSJ Transactions on Computer Vision and Applications.

Kristinn R. Thórisson. *The 'Explanation Hypothesis' in General Self-Supervised Learning*. International Workshop in Self-Supervised Learning, 2021.

Kenneth Williford. Husserl's Hyletic Data and Phenomenal Consciousness. *Phenomenology and the Cognitive Sciences*, 12(3):501–519, 2013. doi: 10.1007/s11097-013-9297-z. Publisher: Springer Verlag.