MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1277

February 1991

Imagination and Situated Cognition

Lynn Andrea Stein

Abstract

A subsumption-based mobile robot is extended to perform cognitive tasks. Following directions, the robot navigates directly to previously unexplored goals. This robot exploits a novel architecture based on the idea that cognition uses the underlying machinery of interaction, imagining sensations and actions.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contracts N00014-85-K-0124 and N00014-89-J-3202 and in part by the System Development Foundation.

1 Introduction

This paper is concerned with a concrete example of the integration of higher-level cognitive AI and lower-level robotics. Robotic systems are embodied: their central tasks concern interaction with the immediately present world. In contrast, cognition is concerned with objects that are remote—in distance, in time, or in some other dimension. We exploit the architecture of a particular robotic system to perform a cognitive task, by *imagining* the subjects of our cognition.

We suggest that much of the abstract information that forms the meat of cognition is used not as a central model of the world, but as virtual reality. The self-same processes that robots use to explore and interact with the world form the interface to this information. The only difference between interaction with the actual world and with the imagined one is the set of sensors and actuators providing the lowest-level interface.

Consider, for example, the following tasks. In the first, a pitcher and bowl sit on a table before you. You lift the pitcher and pour its contents into the bowl. Now consider your actions in reading the preceding example. In all likelihood, you formed a picture *in your mind's eye* of the tabletop, pitcher, and bowl. You simulated the pouring. In the virtual world that you created for yourself, you sensed and acted. Indeed, there is evidence in the psychology literature that such "imagings" are accompanied by activity patterns in the visual cortex, resembling those observed during actual vision. This virtual reality, your imagination, is precisely the goal of our programme.

2 A Robot that Explores

Toto [Mataric, 1990] is a mobile robot capable of goal-directed navigation. It is implemented on a Real World Interface base augmented with a ring of twelve Polaroid ultrasonic ranging sensors and a flux-gate compass. Its primary computational resource is a CMOS 68000. Its software simulates a subsumption architecture [Brooks, 1986].

Toto's most basic level consists of routines to explore its world. Independent collections of finite state machines implement such basic competencies as obstacle-avoidance and random walking. Wall-following—"maze exploration"—emerges as the result of this collection of lowest-level behaviors.

A second layer, above the wall-following routines, implements a fully distributed "world modeler." This behavior is implemented as a dynamic graph of landmark recognizers. Landmarks correspond to gross sonar configurations (e.g., *wall left*) augmented with compass readings. Rough odometry is used to aid in recognition of previously visited landmarks. Each time a novel landmark is recognized, a new graph node allocates itself, making graph connections as appropriate. The

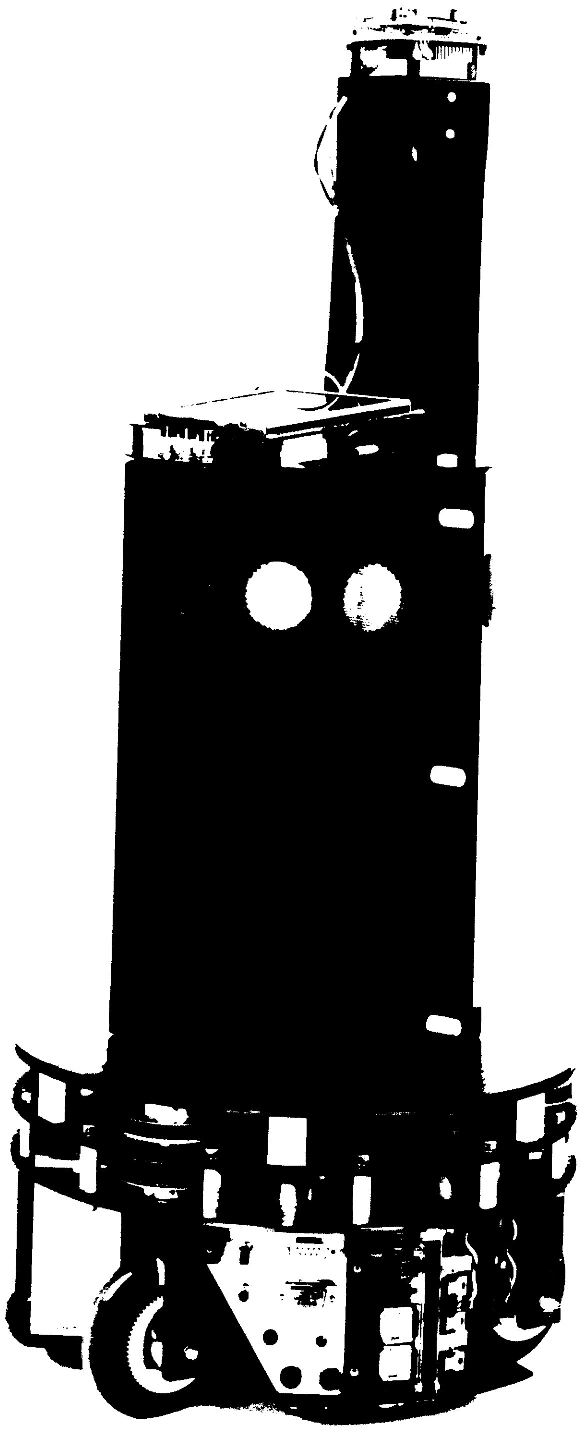

Figure 1: Toto.

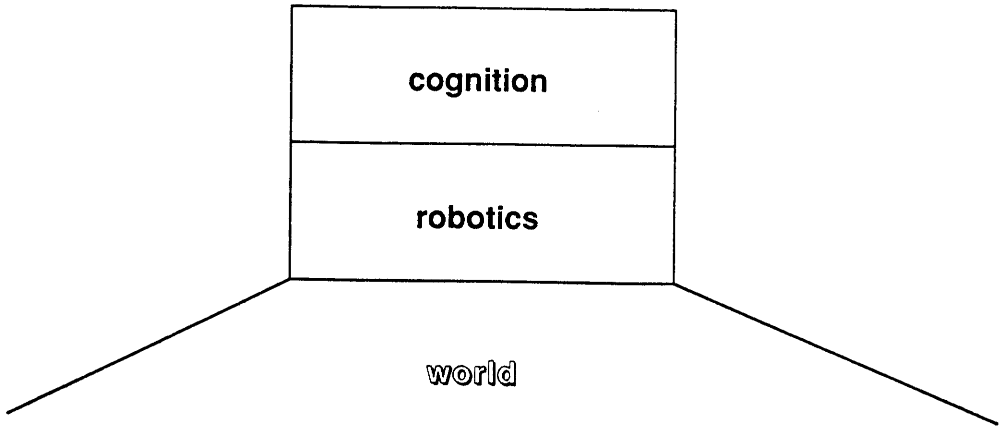

Figure 2: Traditional architecture.

resulting behaviors form an internal representation of the environment.

Finally, Toto accepts commands (by means of three buttons) to return to previously recognized landmarks. When a goal location is specified, Toto's landmark graph uses spreading activation to determine the appropriate direction in which to head. Activation persists until Toto has returned to the requested location. Throughout, Toto's lowest level behaviors enforce obstacle avoidance and corridor traversal, and Toto's intermediate layer processes landmarks as they are encountered.

Toto's landmark representation and goal-driven navigation are cognitive tasks, involving internal representation of the external environment. This represents a qualitative advance in the capabilities of subsumption-based robots. Nonetheless, this internal representation is accessible only through interaction with the world. Toto cannot reason about things unless it has previously encountered them. In the next section, we describe a simple modification to Toto's architecture that allows Toto to represent previously unvisited landmarks.

3 Exploring the Unknown

Previous approaches to cognition in robotic systems have implemented more intelligent behaviors as higher levels of control. In the MetaToto project, we have taken a different approach. The existing machinery that implements Toto's core provides a strong base for cognitive tasks. It is limited, however, in being able to conceptualize only what has been physically encountered.

MetaToto is an extension of Toto's core behavior that accepts directions to navigate to a goal not previously encountered. Toto's goal-directed navigation routines are implemented in terms of its existing internal representation, and it is

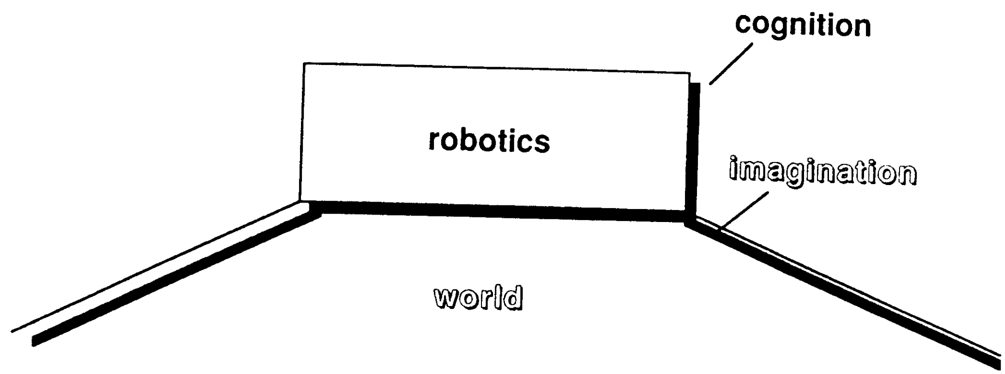

Figure 3: Proposed architecture.

impossible even to ask that Toto visit an unexplored location: Toto has no concept corresponding to locations it has not encountered. The primary task for MetaToto, then, is the representation of landmarks that have simply been described.

Our approach to architecture is to reuse Toto's existing mechanisms in adding this new skill to MetaToto. Where Toto must encounter a landmark, MetaToto merely envisions that landmark. That is, MetaToto takes the landmark description and *imagines* what that landmark would "feel" like: what sonar readings it might evoke, what MetaToto's compass might indicate, etc. We claim that cognition is often simply imagined sensation and action.

In the traditional architecture, cognition rests on top of robotics: robotics provides an intermediary between the external world and a central "cognition box." This approach has led to widespread belief that the two problems can be studied independently, and that technology and research will ultimately meet at the interface between cognition and robotics. Unfortunately, there is little agreement even as to what constitutes this interface.

In contrast, our view suggests that cognition is simply the robotic architecture applied to imagined stimuli. That is, the interface between robotics and the immediate world is multiplexed to provide a second, low-level interface between robotics and imagination. The robot senses and acts in this imagined world precisely as it does in the actual world.

4 Implementing Imagination

If cognition is largely imagined sensation and action, then the difficult tasks for implementing cognition are simulating sensors and actuators, and modeling the appropriate feedback through the imagined world. Both tasks have been attempted in other contexts. The relative success of the approach here relies on some critical assumptions about the nature of the robot's interface with the world and hence with imagination.

4.1 Sensing and Acting

Toto relies on qualitative, rather than quantitative, information about the world. In part, this means that it does not matter if Toto has an occasional anomolous sonar reading. More significantly, it means that moderate inaccuracies in the physical sensors and actuators are not merely tolerated, but expected. Toto's decisions are based on gross judgements (e.g., *dangerously close*) and measurements averaged over time.

Second, Toto relies on constant feedback from the world, and constant interaction with the world. In contrast to traditional planners, which decide on a course of action and then pass control to an executer, Toto "continually redecides what to do" [Agre and Chapman, 1987]. This serves as a form of protection from major errors: any incorrect actions will be recognized and corrected before they can become disasterous. As a result, Toto need not worry about plans gone awry.

Both of these properties mean that MetaToto's simulation of the sensors and actuators need not be accurate. Sonars are simulated using simple ray projection. Angles are approximated. Still, the inaccuracy of MetaToto's imagination are little worse than the variance between two runs of the actual robot, and close enough to allow construction of the appropriate landmark graph.

4.2 Imagination vs. World Models

A second aspect of the architecture bears on the simulation of feedback through imagination, rather than through the world. Feedback through the world has been a strength of reactive systems, and imagination removes that aspect of the architecture. In this sense, it represents a step towards the more traditional world models of classical planning systems.

Imagination differs from classical world models, however. Imagination is ephemeral. MetaToto need only know the sensations that occur *now*. Where Toto "continually redecides what to do," MetaToto continually re-imagines the world. Thus, while world models persist and require maintenence, imagination can be reconstructed on the fly.

In addition, cognition requires imagining only the *relevant* details. That is, only those aspects that bear on things immediately sense-able must be imagined. Because the interface between robotics and imagination is at the level of sensation, rather than in terms of higher-level predicates, we do not need a model of the global properties of the world. Only that which is imagined to be immediately accessible must be simulated.

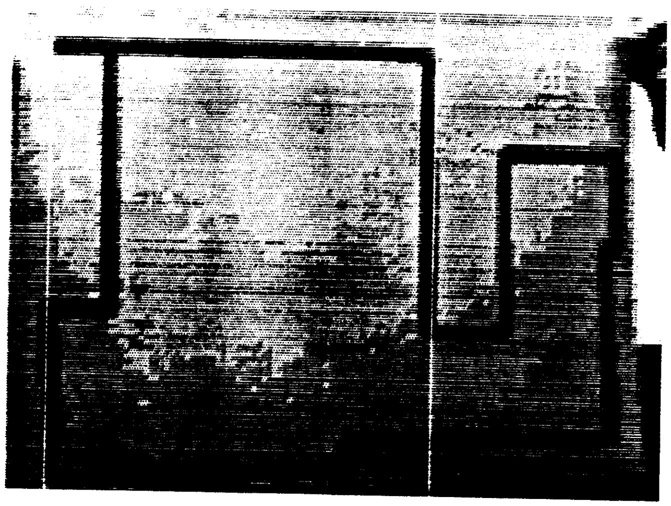

Figure 4: Floor plan, as seen by MetaToto.

5 MetaToto

The initial implementation of MetaToto takes directions in the form of a floor plan. A floor plan—as seen by MetaToto's camera—is shown in figure 4. The use of a geometric communication language facilitates certain of the simulation aspects of MetaToto's imagination. In section 6, we discuss a more verbal communication language.

MetaToto is implemented on the same hardware as Toto, using largely the same software. The modifications to Toto's software involve only the creation and integration of an imagination system. The entire system allows the robot to perform all tasks of which Toto was previously capable, plus the additional cognitive exploration of physically unseen environments.

MetaToto's imagination uses a photographed floor plan of the environment it is to explore. Rather than looking at the plan from above, however, MetaToto imagines that it is located in a particular place in the plan. Virtual sensors describe what it "feels" like to be at that location: what sonar and compass readings MetaToto might receive if physically present. MetaToto imagines sensing and acting in the floor plan much as Toto would sense and act in the actual world, with much the same effect. The routines that sense and act in the imagined world are precisely the same as those that would sense and act in the actual world; they differ only by calling the imagined sonar rather than the real. In this manner, MetaToto explores the floor plan, building the same internal representation of landmarks as Toto would create in its explorations of the environment.

Once MetaToto has completed its exploration of the floor plan, it is capable of goal-directed navigation in the world. However, unlike Toto, MetaToto can go to places that it has only imagined, and not actually encountered. Because the landmark graph has been created by the same mechanisms that are used in exploring the world, MetaToto cannot distinguish those generated by its imagination and those actually encountered. Should the floor plan prove to have been incomplete or inaccurate, MetaToto will simply augment its internal representation as it explores the uncharted area of the actual world.

6 Following Directions

MetaToto's use of a geometric representation for communication facilitates the simulation aspects of imagination. Humans, however, are capable of understanding verbally imparted directions. While this is in some senses an unfair task for MetaToto, it is nonetheless achievable.

Giving MetaToto directions is "unfair" in the sense that humans give humans directions in anthropocentric terms. We speak of "the second left" or "the corner" because these are the landmarks in terms of which we represent the world. MetaToto has no notion of left turns or corners; instead, it represents the world in terms of sonar and compass readings. Thus, to make this task fair in MetaToto's terms, we ought to speak of such landmarks as "the second extended short sonar reading on left and right simultaneously."

Nonetheless, MetaToto could understand the anthropocentric landmarks in much the same way as it uses the floor plan. What, after all, does it "feel" like to explore these landmarks? The simulation aspect may be more complicated, but the task is essentially the same. For example, the landmark "the second left" corresponds to the following (imagined) sensations:

short sonar left
long sonar left
short sonar left
long sonar left

By imagining this sequence, MetaToto could construct an internal representation corresponding to that which would be encountered while seeking the second left. Directions, although more remote than geometric representation, still have a natural analog in terms of imagined sensation.

7 Conclusion

Unlike previous "cognition boxes," MetaToto is distinguished only by the set of sensors and actuators in which the behaviors ground out: when imagining, MetaToto seizes control of the sensor and actuator control signals, and substitutes interaction with the floor plan. Rather than a "higher level reasoning module," MetaToto is a lowest level interface to an alternate (imagined) reality.

MetaToto achieves by embodied imagination the cognition-intensive task of reading, understanding, and acting on the knowledge contained in a floor plan; and MetaToto does this using entirely Toto's existing architecture, with the sole addition of the virtual sensors and actuators required for navigation of the floor plan. Although MetaToto is only a simple example of imagination, we are hopeful that experiences with MetaToto will lead to more sophisticated use of imagination and virtual sensing, and to the development of truly embodied forms of cognition.

Acknowledgements

This paper could not have been written without the help of Ian Horswill, Maja Mataric, and Rod Brooks.

References

[Agre and Chapman, 1987] Philip E. Agre and David Chapman. Pengi: An implementation of a theory of activity. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 196–201, Seattle, Washington, July 1987. Morgan Kaufmann Publishers, Inc.

[Brooks, 1986] Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, April 1986.

[Mataric, 1990] Maja Mataric. A distributed model for mobile robot environment learning. Technical Report 1228, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, Massachusetts, May 1990.