Embodiment in GLAIR: A Grounded Layered Architecture with Integrated Reasoning for Autonomous Agents*

Henry Hexmoor
Johan Lammens
Stuart Shapiro

Computer Science Department
226 Bell Hall
State University of New York at Buffalo
Buffalo, NY 14260
hexmoor@cs.buffalo.edu

To appear in proceedings of the Florida AI Research Symposium, Fort Lauderdale, FL

February 1993

Abstract

In order to function robustly in the world, autonomous agents need to assimilate concepts for physical entities and relations, grounded in perception and action. They also need to assimilate concepts for perceptual properties like color, shape, and weight, and perhaps eventually even for nonphysical objects like unicorns. The process of acquiring concepts that carry meaning in terms of the agent's own physiology we call *embodiment*. Unlike current robotic agents, those endowed with embodied concepts will more readily understand high level instructions. As a consequence, these robots won't have to be instructed at a low level. We have developed an autonomous agent architecture that facilitates embodiment of action and perception, and accommodates embodied concepts for both physical and nonphysical objects, properties, and relations.

1 GLAIR

We present an architecture for intelligent autonomous agents which we call GLAIR (Grounded Layered Architecture with Integrated Reasoning). A major motivation for GLAIR and the focus of our attention in this paper is the concept of *embodiment*: what it is, why it is important, and how it can be given a concrete form in an agent architecture. As such our definition is both more concrete and more narrow than the one in [Lak87], for instance. Figure 1 schematically presents our architecture.

Concept learning provides an important motivation for embodiment. Winston's Arch program [Win75] is an early example of a system that learns concepts through examples. This program relies heavily on feature analysis. However, the feature concepts and the concepts of objects learned lack embodiment, as is typical for traditional symbolic AI work. Most of this work never gets implemented in actual agents that can physically interact with their environment (other than via a keyboard and monitor, which is a rather uninteresting case), i.e. the symbols never inhabit a body. It is this kind of approach that Brooks criticizes in papers like [Bro90]. According to Brooks, symbolic representations should be matched

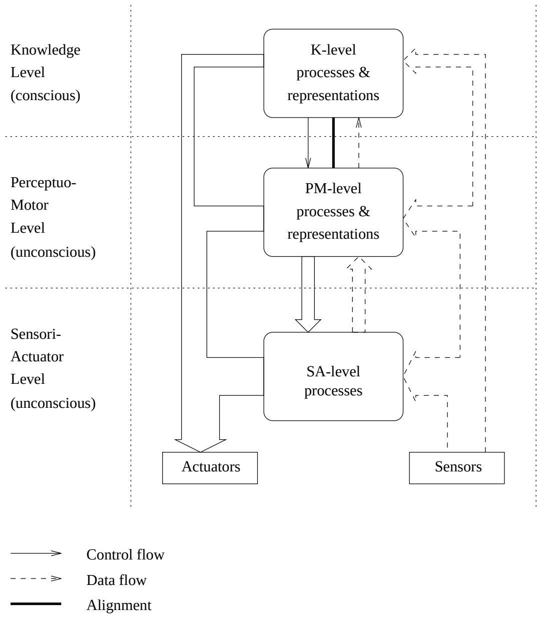

Figure 1: Schematic representation of the GLAIR architecture. Width of control and data paths suggests the amount of information passing through (bandwidth). Sensors include both world-sensors and propriosensors.

to the agent's sensors and actuators, and this grounding provides the constraints on symbols necessary for them to be truly useful. It is this matching of representations to the agent's sensors and actuators that we call embodiment, and how to go about doing this in an architectural framework like GLAIR is the focus of this paper.

The most general motivation behind our work is the desire to be able to "program" a robotic autonomous agent by telling it *what* to do at a high level and have it "understand", rather than telling it *how* to do something in terms of primitive operations, with little or no "understanding".[1] For instance, we want to tell it to go find a red pen, pick it up, and bring it to us, and not have to program it at a low level to do these things.[2] One might say that we want to communicate with the robot at the *speech act* level. To do this, the agent needs a set of general-purpose (relative to its own body) perceptual and motor capabilities along with an "understanding" of these capabilities. The agent also needs a set of concepts which are similar enough to ours to enable easy communication (as [Win75] already mentioned). The best way to accomplish this, in our opinion, is to endow the agent with embodied concepts, grounded in perception and action.

[1] There was a similar motivation behind the move from imperative to logic-based or declarative programming languages, but there are no further similarities.

[2] Retrieving "canned" parameterized routines is still a low-level programming style that we want to avoid.

2 EMBODIMENT DEFINED

We define *embodiment* as the notion that the representation and extension of concepts is in part determined by the *physiology* (the bodily functions) of an agent, and in part by the *interaction* of the agent with its environment. For instance, the extension of color concepts is in part determined by the physiology of our color perception mechanism, and in part by the visual stimuli we look at. The result is the establishment of a mapping between color concepts and certain properties of both the color perception mechanism and objects in the world. Another example is the extension of concepts of action: it is partly determined by the physiology of the agent's motor mechanisms, and partly by the interaction with objects in the world. The result is the establishment of a mapping between concepts of action and certain properties of both the motor mechanisms and objects in the world (what we might call "the shapes of acts").

To be specific, we must introduce a few notions. All concepts that can be represented by an agent fall into two categories: *extensional* and *non-extensional*. We say an extensional concept is *physical* for an agent if the agent is able to identify an extension (example) of the concept in the world. Let's call the agent's act of identification *demonstration*. The demonstration thus involves a referent that is external to the agent's body. An example of a physical object is "a cup". In a room that has at least one cup, the agent should be able demonstrate the concept of a cup by identifying one. Another example of a physical concept is "a container". In a room with a cup in it, the agent should be able to demonstrate the concept of a container by picking out the cup as a container. Other examples of physical concepts can be found in perceptual phenomena like color, shape, size, or weight. The agent should be able to demonstrate objects with different perceptual properties. For example, the agent should demonstrate "a red cup" by picking out a red cup from among cups of different colors. Of course the precise extension of perceptual properties is dependent on the agent's sensors and actuators. If the agent has to demonstrate a concept by using physical (bodily) actions that involve only referents that are integral to the agent's body, we say the concept is *body centered*. Examples of *body centered* concepts are "yawning" and "blinking". We say a concept is *interactive* if demonstrating it involves both body-external and body-integral referents, i.e. the agent has to interact with physical objects. All interactive concepts have physical and body centered components. Examples of interactive concepts are *sitting*, *drinking*, and *driving*.

We call an extensional concept embodied if all of the following conditions hold:

- The concept is *physical*, or *body centered*, or *interactive*.
- The agent can *demonstrate* the concept in the world.

Concepts without extensions, e.g., "a unicorn", can also be embodied. We say a non-extensional concept is embodied if it can be decomposed into embodied concepts and relations. For example, we will consider the concept of "a unicorn" embodied if the concept of "a horse" is embodied, the concept of "a horn" is embodied, and all concepts about "a horse" having "a horn" on its head are embodied. We roughly follow Harnad's definition of elementary grounded versus non-elementary grounded symbols here [Har90].

In our architecture we strive to model agents that can learn to extend their ability to demonstrate extensional concepts. The embodiment of concepts becomes very useful when the agent can demonstrate concepts with variations and in different environments, e.g., driving with different cars, or under various road conditions.

3 AGENTS WITH EMBODIED CONCEPTS

At an abstract level, the way to provide an autonomous agent with embodied concepts is to intersect the set of human physiological capabilities with the set of the agent's potential physiological capabilities, and endow the agent with what is in this intersection. Different agents can use different implementation mechanisms, depending on their particular bodies. To determine an agent's potential physiological capabilities, we consider it to be made up of a set of primitive actuators and sensors, combined with a general purpose computational mechanism. The physical limitations of the sensors, actuators, and computational mechanism bound the set of potential capabilities. For instance with respect to color perception, if the agent uses a CCD color camera (whose spectral sensitivity is usually wider than that of the human eye), combined with a powerful computational mechanism, we consider its potential capabilities wider than the human ones, and thus restrict the implemented capabilities to the human ones. We endow the agent with a color perception mechanism whose functional properties reflect the physiology of human color perception. That results in color concepts that are similar to human color concepts. With respect to the manipulation of objects, most robot manipulators are inferior to human arms and hands, hence we restrict the implemented capabilities to the ones that are allowed by the robot's physiology. The robot's motor mechanism then reflects the properties of its own physiology, rather than those of the human physiology. This results in a set of motor concepts that is a subset of the human one. Embodiment also calls for body-centered and body-measured representations, relative to the agent's own physiology.

We have applied the notion of embodiment to an architecture for an autonomous agent that will be able to communicate at the speech-act level, using human-like concepts. We believe that our approach is general and useful, rather than tailored to any particular task or domain. We call the architecture GLAIR, as mentioned above. Recalling that embodiment is an approach to establishing a mapping between high level concepts on one hand, and properties of the agent's physiology and its interaction with the world on the other hand, we distinguish three levels of abstraction in our architecture. The Knowledge Level is the most abstract level, and incorporates a traditional knowledge representation and reasoning system. The Perceptuo-Motor Level sits in the middle, and is the main locus of embodiment for perception and action mechanisms. The Sensori-Actuator Level is the lowest level, where sensors and actuators are situated, and interactions with the environment take place.

Representation, perception, and generation of behavior are distributed through all three levels. We differentiate conscious reasoning at the Knowledge Level from unconscious Perceptuo-Motor Level and Sensori-Actuator Level processing. Concepts represented at the Knowledge Level are accessible for conscious reasoning and communication with other agents, while representations at the other two levels are not. The levels of our architecture are semi-autonomous and processed in parallel. Conscious reasoning takes place through explicit knowledge representation and reasoning, while unconscious behavior makes use of several different implementation mechanisms. Conscious reasoning guides the unconscious behavior, but the unconscious levels, which are constantly engaged in perceptual and motor processing, can *alarm* the conscious level of important events, taking control if necessary. Control and generation of behavior are layered, and not exclusively top-down or bottom-up. There is a correspondence between terms in the knowledge representation and reasoning system on one hand, and sensory perceived objects, properties, events, and states of affairs in the world and motor capabilities on the other hand. We call this correspondence *alignment*. Behaviors can migrate between levels, e.g. from the Knowledge Level to the Perceptuo-Motor Level. The latter is a case of automating explicitly learned behavior.

4 GLAIR AGENTS

We are developing several agents that conform to the principles of the GLAIR architecture. These agents include a robotic autonomous agent, a video-game playing agent, and a mobile robot agent. Figure 2 schematically presents the structure of one of these GLAIR based agents. The robotic agent incorporates an embodied model of color perception and color naming, and a set of embodied motor capabilities. The video-game agent demonstrates real time behaviors and the inter-level alignment mechanism. The mobile robot agent also incorporates an embodied model of color perception and color naming, and demonstrates emergent behaviors, e.g., pushing a block around. This kind of behavior is often hand-coded in other architectures, e.g. subsumption [Bro90]. The mobile robot agent has first order or "innate" embodied sensations like contact between its body and some other object, and second order or emergent sensations such as moving forward or backward. It also has first order embodied actions like turning its wheels, and second order embodied actions like moving forward or backward (in most cases, actions and sensations come in tightly coupled pairs, and can be considered duals of each other).

All three agents display a variety of integrated behaviors. We distinguish between deliberative, reactive, and reflexive behaviors. Embodied representations at the Perceptuo-Motor Level facilitate this integration. As we move down the levels, computational and representational power is traded off for better response time and simplicity of control. The agent learns from its interactions with the environment. It has a capacity for *engaged* and *disengaged reasoning*. The former occurs when behavior is generated directly while reasoning, in a lock-step fashion. The latter occurs when reasoning is done in a hypothetical mode, not directly generating behavior. Our alignment mechanism allows us to elegantly model both modes of reasoning.

This paper can serve only as an overview for GLAIR and GLAIR-based agents. Details of our implementations and comparisons with competing architectures are given in our technical reports [HN92, HLS92, HCBS93].

5 CONCLUSION

We have defined and motivated *embodied concepts* for autonomous agents. We have also presented an architecture which facilitates the acquiring of embodied concepts by autonomous agents. Our architecture distinguishes itself mainly through its three layers, their different representation mechanisms, and

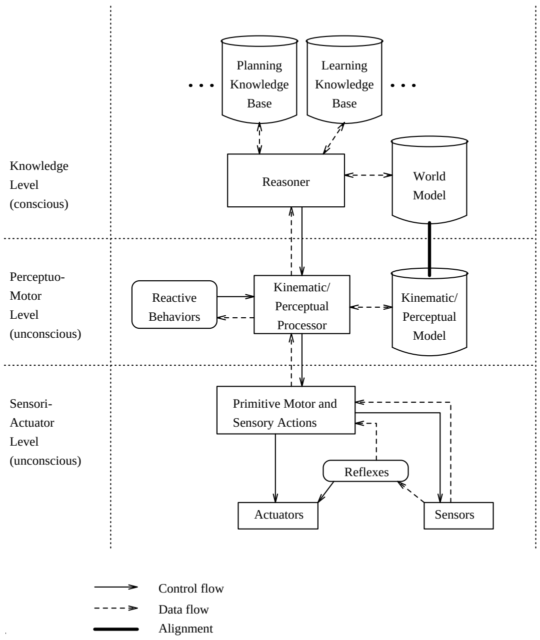

Figure 2: Schematic representation of the structure of a prototypical GLAIR-agent, an agent conforming to the GLAIR architecture.

the mechanisms for aligning the levels. Other significant features are the distinction between conscious and unconscious levels and its implications for the generation of behavior and for communication with the agent. In future work we intend to show how GLAIR as an abstract architecture can be made to model various agents in different domains, all using representations grounded in action and perception.

References

[Bro90] Rodney A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3–15, 1990.

[Har90] Stevan Harnad. The symbol grounding problem. *Physica D*, 42(1-3):335–346, 1990.

[HCBS93] Henry Hexmoor, Guido Caicedo, Frank Bidwell, and Stuart Shapiro. Air battle simulation: An agent with conscious and unconscious layers. In *UBGCCS-93*. Dept. of Computer Science, SUNY at Buffalo, New York, 1993.

[HLS92] Henry Hexmoor, Joe Lammens, and Stuart Shapiro. An autonomous agent architecture for integrating perception and acting with grounded, embodied symbolic reasoning. Technical Report CS-92-21, Dept. of Computer Science, SUNY at Buffalo, New York, 1992.

[HN92] Henry Hexmoor and Donald Nute. Methods for deciding *what to do next* and learning. Technical Report AI-1992-01, AI Programs, The University of Georgia, Athens, Georgia, 1992. Also available from SUNY at Buffalo, CS Department TR-92-23.

[Lak87] George Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago, IL, 1987.

[Win75] Patrick Henry Winston. Learning structural descriptions from examples. In Ronald J. Brachman and Hector J. Levesque, editors, *Readings in Knowledge Representation*, pages 141–168. Morgan Kaufmann, San Mateo CA, 1985 (orig. 1975).