Grounded Concept Development Using Introspective Atoms

Eric Berkowitz and **Brian Mastenbrook**
1400 N. Roosevelt Blvd.
Schaumburg, IL 60173
eric@cs.roosevelt.edu
chandler@acm.roosevelt.edu

Abstract

In this paper we present a system that uses its underlying physiology, a hierarchical memory and a collection of memory management algorithms to learn concepts as cases and to build higher level concepts from experiences represented as sequences of atoms. Using a memory structure that requires all base memories to be grounded in introspective atoms, the system builds a set of grounded concepts that must all be formed from and applied to this same set of atoms. All interaction the system has with its environment must be represented by the system itself and therefore, given a complete ability to perceive its own physiological and mental processes, can be modeled and recreated.

Introduction

ASPARC is a symbolic system (Brooks 1991; Etzioni 1993) but the focus is on simulating genuine spatial existence and system components appropriate for physical implementation (Bailey 1997). Beyond the pragmatic issue in research there is another reason for using a simulated organism. Actions are always taken with a view toward intended or at least expected outcome. As such, humans and an artificial entity require the facilities to imagine and therefore to maintain a completely cerebral (Epstein 1992) representation of the world that allows *ad hoc* portrayal of actions in a given situation (Berkowitz 2000; Johnson 1987; Lakoff & Johnson 1999). Reasoning or communicating at various levels of abstraction requires the ability to maintain many such representations. People do this all the time; hence the phrase *look before you leap*. To allow this, an organism will always require some representation of its world that allows such cerebral modeling, completely removed from the mechanisms of the real world. Imagining jumping into an empty pool does not make it happen in the real world.

INTROSPECTIVE ATOMS

ASPARC is built around the existence of introspective atoms. ASPARC begins by developing symbols that directly represent each of the processes that ASPARC can invoke on its world (and by definition on itself) and the sensory feedback it is capable of perceiving. Theses symbols are the introspective atoms and combine self-knowledge (Leake 1995) with world knowledge in a single perception. This creates a system where there is a distinct boundary between the physiological layer and the reasoning layer of the system but the reasoning layer has a representation for every phenomenon in its world. Thus:

- Every atom in the reasoning system has a direct representation in the physical world as an observable item or event.
- Every observable item or event that the system can perceive must be represented as an atom or a collection of atoms.
- The system must include itself in its observable world.

The introspective atoms constitute a form of protocol by which the reasoning layer interacts with the physical layer and, by the definition above, this protocol must be completely inclusive. That is, in a system in which reasoning processes are defined to be transparent meaning that all data representation is representation of reasoning processes, then any processes that are not transparent must be atomic and must be grounded. If domain specific knowledge is knowledge about actual interaction with the world in a given domain and the set of introspective atoms represents all processes by which the system can interact with its environment the the result is a system in which a domain is not added as external enhancements but represented directly in the memory using the same atoms as basic perceptions (Figure 1). ASPARC's memory is designed around a hierarchy that permits relationships based on specification and inclusion creating layers of semantic abstraction (Saitta & Zucker 1998) A memory can consist of a sequence of other memories or be defined as a specialized case of another memory. This is not a traditional inheritance relationship since no inheritance takes place. Rather it is an organizational structure implying similarities between memories. These similarities are used by ASPARC when deciding where to store a new memory and in a similar fashion, when seeking out a memory that matches a given new situation.

PHYSIOLOGY

In order to facilitate the development of symbols for atomic actions representing ASPARC's perception of itself and its environment including performing its most basic embodied actions, such as extending a foot or hand, ASPARC's simulated body was designed with joints and a fairly full range of

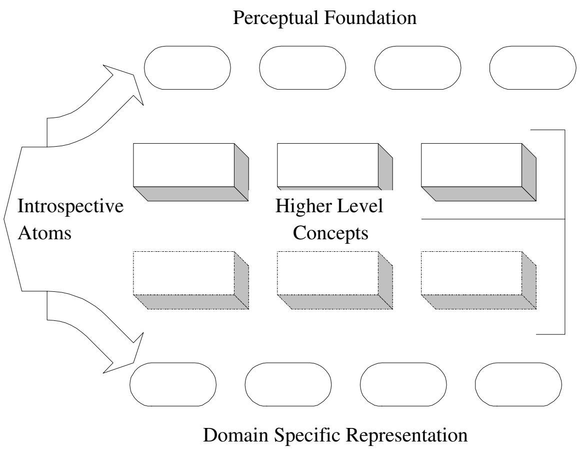

Figure 1: Basic Perception and Domain Knowledge as Atoms

motion. In order to distinguish one type of movement from another, such as stepping forward or stepping sideways, the system was given controllable knee and hip joints. ASPARC steps forward by first stretching out a leg and then pulling the rest of its body forward in a manner very similar to a human step. As it does this it receives feedback from all of its systems and parts. Attempting to stick out its leg causes pressure to be felt in its joints and then the leg moves, or perhaps is obstructed in which case it does not move and pressure is felt in its foot. All of this feedback is sent into a serialized stream. This stream is used to build memories, to activate existing memories (reminding) and to activate higher level patterns (words and concepts).

META-SPATIAL PERCEPTION

Making ASPARC able to begin from no memories at all, without any artificial layer that would break down the link from memories to introspective atoms, continually brought up the question of how to frame the memory without imposing a predetermined structure that would result in a system that is totally artificial in nature. The goal was to create a system that could ground higher level concepts in introspective atoms but the problem was how to begin to perceive the initial events; how to put the initial concepts into an empty memory. The solution to this problem was to ground ASPARC's initial reasoning and perception, in a meta-spatial perception. As in earlier systems (Berkowitz 2000; 2001), the theory that human perception is spatially grounded was modeled, but it was determined that for this perception to frame the most basic memories it must be by nature and not nurture. Spatial perception is part of how our minds are designed or have evolved, not learned behavior. The mapping of events onto paths in space is not because movement along a path is a basic memory. It is because this form of perception has become part of what we are biologically, most likely as a result or millions of years of existence in a world that follows a constant set of physical laws. ASPARC was therefore given basic instinctive mechanisms based on perceiving everything in a completely cognitive, empty three dimensional plane. For ASPARC, basic physical motion in reality is not the underlying foundation for perception; rather, it too is perceived in terms of a meta-spatial mapping into an instinctive meta-spatial perception that is an evolved systemic manifestation of a system with physical existence in its world. ASPARC does not first learn of movement and then decide to use it as the underpinning for its reasoning and symbolic representation of its world. From its very first moments it perceives everything it experiences in terms of itself. Without this foundation there is no perception.

INITIAL LEARNING

ASPARC's first experiences are formed by sending sets of processes from those supported by its physiology, to its limbs and parts and recording the events that follow. This brings up the question of whether the concept of cause and effect are learned or part of the meta-spatial perception. In the world of experience, cause and effect could possibly be taught through rote repetition of events and outcomes. Unfortunately that would still leave the question of why an entity, even seeing reality *B* following reality *A* one million times would assume that it would necessarily follow again the next time it encounters reality *A*. This research takes the position that cause and effect are part of the instinctive meta-spatial perception. Given any particular state of a meta-space, any movement necessarily creates a new state in that meta-space. This is not to say that the meta-space imparts some knowledge of Newtonian physics. When a child holds an apple over the edge of its tray and drops it to the floor, it learning that letting go of the food results in it falling (which results in mommy getting upset). This is knowledge, not instinct and is not part of the meta-space. What is part of the meta-space is the instinctive knowledge that an act will have some result. Training and/or learning teach us what the result will be and that the result of a given action, *ceteris paribus*, is going to be the same however all this is framed by the instinctive packaging in the memory and mind of cause and consequence. ASPARC's initial training sessions were run by having it continuously perform single random acts and recording the memories. In order to teach it the result of each act without side effects, ASPARC was continually returned to an identical steady state and then made to perform a single random act. It quickly became apparent, that with the system's increasing abilities, unfeasible amounts of time were required for the random teaching method to ensure that ASPARC had undergone every possible experience. The random training method was replaced with a directed, combinatoric training scheme that ensured that the system would have the maximum number of experience in the minimal amount of time. ASPARC is directed through a process of experiences. Each act is evaluated using the most basic criterion in the meta-space: given the existence of a self and goal, does this act change their relationship; that is, does it cause change as its basic perceptual level by modifying the meta-space.

Each possible action ASPARC is capable of performing is tried. The actions are sorted into two groups, those that change the relationship to the goal in the meta-space and those that do not. The system then tries every pairwise combination the act that singularly did not affect the goal to see if, when performed as a temporally set, they do meet this criterion. The action of extending a foot does not change the relationship to the goal but the sequence of extending a foot and the pulling the body upright over the appropriate (stepping) hip does. Through this learning process, ASPARC acquires the knowledge that within its set of possible actions only extending an arm changes it relationship with a goal. Within the set of pairwise combinations it learns the following actions for affecting its relation to the goal.

FROM STEPPING TO WALKING

First attempts at getting ASPARC to move from stepping to walking were based on a path metaphor from ASPARC's predecessor system. This path was represented by and defined as a set of ordered locations in space and the acts that cause motion between them. ASPARC could be placed in space two locations from the goal and then made to take two steps. It would then learn a method of resolving a situation where it is two steps from the goal and could recognize this situation and reuse this method. But how about a new situation where it is three locations from the goal? Four? Five hundred? Where the distance can not be determined? The natural approach to this was to get ASPARC to recursively decompose longer distances into collections of shorter ones. It was even believed that by doing so ASPARC could acquire methods of basic arithmetic. All such attempts failed. Further investigation revealed that these failures were due to the fact that implicit in what appeared to be the lowest level facilities ASPARC might be given, there were many subtle assumptions and hidden levels of abstraction. The proposed conceptual decomposition required understanding of numbers as magnitudes, ordinals, and intervals. It became clear that such skills could not be granted to ASPARC without losing the grounded nature of the system.

ASPARC's training was then redesigned around a foundation which we now believe to be devoid of subsumed levels of abstraction; the concept of infinity. This is not an understanding of the meaning of infinity; rather, it is a mode of thinking.

CREATING THE BASE CONCEPTS

Infinity based perception gives a definition of walk that does not suffer from any of the problems described above. It has no hidden levels of abstraction, and does not rely on any higher level mathematical concepts. Further, this definition allows ASPARC to recognize a "walk" of any length and to create plans for walking without needing any *a priori* knowledge. Being five locations from the goal does not require a five step solution in any explicit way. Being away from the goal by any, even indeterminate number of locations, requires a single generalized solution of walk. In fact while the discussion up till this point used the concept distance defined by a number of "locations" in space this is actually the inverse of how ASPARC perceives the situation. Its basic understanding, being infinity based, divides the possible relationships with the goal into only two: the unitary *at* and the infinite *away*. The single infinite set of *away* problem-states are all addressed by the single infinite solution "alk". Discrete measures of distance are actually a level of abstraction built above this perception. Distance can only be defined after walk, as the finite number of steps in a given walk. Thus the concept of a finite set is added to the concept of unitary and infinite in a natural grounded way. In a human child's development also, its first mathematical concepts are nominal. Numbers are unique tokens or names. The child can distinguish 2 from not-2 just like he/she can distinguish mommy from not-mommy and me from not-me. There is an implicit but not conscious awareness of finite versus infinite that is only formalized much later in the child's study of mathematics (Elkind 2001; Mix, Huttenlocher, & Levin 2002; Spelke & Tsivlin 2001). Earlier systems we developed used artificial bootstrapping defined walk in terms of traversing steps along a path in space. This was possible since the definition of path was entered as bootstrapped data and did not need to be acquired naturally. ASPARC demonstrates that during natural concept development "path" is actually an abstraction of walking. A path is defined as the perceived collection of experiences collected during a walk. It has both a physical manifestation and a experiential one that are intimately linked. The basic prototype for all action, the path, is created in the meta-space by assigning meaning to a specific collection of movements.

NEW CONCEPTS, NEW MEANING THROUGH NEW MEMORIES

During its training, ASPARC develops a new complex concept with the following sequence of sub-concepts:

1. A memory of ASPARC "away" from the goal both in feeling and in the imagery.
2. A memory of $step^+$
3. A memory of $step^+ \Rightarrow failure$
4. A memory of ASPARC "at" the goal

ASPARC has, through its own existence, built a concept of "do-something" grounded as described in Greene and Chien (1993), a concept which as shown Berkowitz and Greene (2001) can be used to to form the foundation for spatial reasoning. Further, the do-something concept is completely defined in terms of introspective primitives and as such can be re-enacted by the system as well as used for reasoning. Whereas before ASPARC's lexicon was limited to concepts defining its basic actions and the meta-concept "go," it now has, through the use of this new memory, a term for performing a specific act.

FAILURE AND INQUIRY BASED ON INTROSPECTIVE ATOMS

ASPARC can detect total failure, that is, its inability to locate any memory to apply to its current situation. When such an event occurs, ASPARC will interactively inquire "What should I do next?" The response to this must be one or more symbols representing re-enactable memories, that is, concepts whose meanings are grounded in ASPARC's introspective primitives. ASPARC will act out the instructions it received and the complete memory of the event will be added; thus, ASPARC can reenact the same memory in the same situation without inquiry. While ASPARC is told what to do, its memory of the event is of its own performance of those instructions in terms of its own introspective primitives. ASPARC can communicate with a person at whatever level of conceptual development its memory is currently at. New information can be given to ASPARC built on combinations of symbols representing concepts that currently exist in the system's memory. In another experiment, if ASPARC is told to "go" but this time the goal, rather than being directly in front it, is now both in front of the system and off to the left, ASPARC will begin to go expecting this to be another case of $step^+ \Rightarrow at-goal$ as is now stored in its memory. ASPARC will proceed forward but when it should be at the goal it indeed is not. ASPARC has learned of a second dimension to its world, that is, while it previously paid attention only to those facts that defined its world as a linear construct, it now learned one example of its world being a plane. It still has no concept of a third dimension (up and down). Detecting the failure of all existing memories to form plans appropriate for this situation ASPARC requests input. While to the human observer the goal is off to ASPARC's left, ASPARC has no memory of dealing with such a situation. What ASPARC does have is a collection of atoms representing the basic actions it can do with its body and concepts built upon them. Two of the atoms represent turning its left leg outward and aligning it body. Together they form a concept labeled with a symbol whose English representation is 'turn-left' and that is completely grounded in ASPARC's existence. Told to perform the act 'turn-left' ASPARC does so utilizing its meaning. Matching the resultant situation to its memory, ASPARC turns and arrives at its goal. The result is that ASPARC now has a memory of arriving at a goal offset from it on two axes. It also has the concept of a multi-part action. While before a single task was represented as $step^+ \Rightarrow at - goal$ a task can now be represented as $step^+ \Rightarrow failure \Rightarrow action \Rightarrow step^+ \Rightarrow at - goal$. Part of an action can itself be an action that does not lead, in itself, to the goal.

OBSERVED REASONING AS DATA

ASPARC learns by watching the results of its own reasoning procedure or that of others. This creates the ability to add knowledge for any domain that can be represented as introspective atoms. If the set of atoms is complete, any domain the system can reason about can be represented. ASPARC observes itself operating in its world using the same mechanisms with which it observes its world. That is, its introspective mechanisms are identical to its outward mechanisms, and as such it learns about itself as part of its world. New levels of concepts include new perceptions of itself and its abilities. By learning about its own processes it can recursively enhance its own capabilities and perceptions.

DISCUSSION

ASPARC's predecessor worked with built-in atoms designed by its developers. It was believed at the time that these atoms were the most basic representations possible for the system's base concepts. However, when given the ability to develop its own atoms in a completely empty memory, ASPARC's definitions are very different and demonstrates many of the hidden concepts which, rather than constituting atoms, are actually built upon them. Examples are the concept of path and objects. In the earlier systems, *path* was placed in the memory as an ordered collection of *objects* upon which all concepts of *doing* could be built. It was believed that this represented base concepts. ASPARC's autonomous ability to develop its own atoms demonstrates that this was incorrect and reveals the levels of hidden concepts in this original definition. ASPARC builds the concept of a path out of two locations without higher level concepts such as ordinality. Also, the concept of a step is represented in ASPARC's memory as a movement between locations not objects. ASPARC's grounded definition for step revealed the hidden higher level concepts subsumed in earlier attempts at priming the memory, such as the implied concepts of objects, order, and contiguity. ASPARC's concept of a path involves two locations, not object. The path is defined in terms of moving between these locations and the steps, defined earlier as objects, are now to ASPARC, the act of stepping, that is, the smallest episodic concept that changes its location. ASPARC's definition is truly grounded in introspective atoms requiring no prior definition of objects, order, contiguousness or other concepts. It forms a far more natural foundation for the mapping of actions.

CONCLUSION

ASPARC develops concepts appropriate for direct translation into human thought and reasoning based on a system that organizes its own memories based on atomic concepts emanating from its perception of its own existence. ASPARC develops these concepts as a foundation for communication and reasoning as an emergent phenomenon (Steels 1996c; 1996b) but based on the assumption that the framework for spatial perception is biological in people and should be endemic to the perception and communication systems of artificial entities. By framing their perceptions in the cognitive model (Davidson 1992) to match our own, we frame their concept development to facilitate translation to our own while keeping the concepts grounded for the simulated entity. That is, unlike completely self organizing concept development (Steels 1996a), while the concepts are emergent, the mechanisms for organizing them are predetermined. We believe that even in systems that attempt to completely autonomously develop concepts and allow them to organize completely independently the mechanisms for facilitating the concept acquisition and organization, without which nothing would happen, always imply some element of preordained structure, whether or not the designers focus on them. ASPARC demonstrates the acquisition of basic concepts for reasoning and communication using introspective atoms and how these atoms can form the foundation of complex concepts. By developing a memory in this way we believe we will be able to develop ASPARC's reasoning and communication abilities around generalized algorithms and the agent's own physiology that gave the memories their complete initial meaning. Additional directed training of ASPARC will, we hope, continue to yield such concepts as well as the minimal prerequisite set of reasoning abilities required to understand them and their grounded meaning.

References

Bailey, D. 1997. *A Computer Model of Embodiment in the Acqusition of Action Verbs*. Ph.D. Dissertation, EECS Dept., University of California at Berkeley, Berkeley, CA.

Berkowitz, E. G. 2000. *Metaphor-Based Reasoning for Software Agents and Heterogeneous Robots Using Body-Based Feeling Sequences*. Ph.D. Dissertation, Department of Computer Science Illinois Institute of Technology, Chicago, IL.

Berkowitz, E. G. 2001. Body based reasoning using a feeling-based lexicon, mental imagery and an object-oriented metaphor hierarchy. In *Proceedings of the 14th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI 2001*, 47 – 56.

Brooks, R. 1991. Intelligence without representation. *Artificial Intelligence* 47:139 – 159.

Davidson, P. 1992. Concept acquisition by autonomous agents: Cognitive modeling versus the engineering approach. *Cognitive Studies* 12.

Elkind, D. 2001. Early childhood education developmental or academic. *Education Next*.

Epstein, S. L. 1992. The role of memory and concepts in learning. *Minds and Machines* 2:239 – 265.

Etzioni, O. 1993. Intelligence without robots: A reply to brooks. *AI Magazine* 14(4):7 – 13.

Greene, P. H., and Chien, G. H. 1993. Feeling-based schemas, neuralarchitectures, and distributed ai: From schema assemblage to neural networks. In *Proceedings of the 1993 Workshop On Neural Architecture and AI*, 112 – 115. Los Angeles, CA: University of Southern California.

Johnson, M. 1987. *The Body in the Mind: The Bodily Basis of Meaning Imagination, and Reason*. Chicago, IL: University of Chicago Press.

Lakoff, G., and Johnson, M. 1999. *Philosophy in the Flesh*. New York, NY: Basic Books.

Leake, D. 1995. Representing self-knowledge for introspection about memory search. In *Proceedings of the AAAI Spring Symposium on Representing Mental States and Mechanisms*.

Mix, K. S.; Huttenlocher, J.; and Levin, S. C. 2002. *Quantitative Development in Infancy and Early Childhood*. New York, NY: Oxford University Press.

Saitta, L., and Zucker, J.-D. 1998. Semantic abstraction for concept representation and learning. In *Symposium on Abstraction, Reformulation and Approximation SARA98*.

Spelke, E. S., and Tsivlin, S. 2001. Initial knowledge and conceptual change: space and number.

Steels, L. 1996a. Emergent adaptive lexicons. In Maes, P., ed., *Proceedigns of the Simulation of Adaptive Behavior Conference*. Cambridge MA: The MIT Press.

Steels, L. 1996b. A self-organizing spatial vocabulary. *In Artificial Life Journal* 2.

Steels, L. 1996c. Self-organizing vocabularies. In Langton, C., ed., *Proceedings of Alife V*.