Project Joshua Blue: Common Sense via Common Experience

Sam S. Adams, Steve Burbeck, Nancy Alvarado, Craig Latta

IBM Research
PO Box 12195
3039 Cornwallis Road
Research Triangle Park, NC 27709
ssadams@us.ibm.com

Abstract

Our position related to 'Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems' is threefold. First, that conscious awareness of experience in humans is bootstrapped from sensory/motor interactions within the first 12 months following conception. Second, that human developmental physiology indicates the appropriate behavioral metalevel to focus work on connecting an artificial mind to its sensory/motor embodiment. Third, that this perspective leads to specific architectural and representational choices in the design of systems capable of acquiring human-like common sense.

Research Statement

Common sense is the critical missing component in achieving ultimate success in both machine learning and natural language understanding. We strongly believe that common sense is rooted in common experience. Human experience is grounded in years of sensory/motor interactions in a rich environment that includes other humans in addition to the rest of the physical world. Conscious awareness of that experience is bootstrapped from sensory/motor interactions within the first 12 months after conception. We believe that this developmental pathway is essential to achieving human-like understanding in any substrate, whether biological or silicon.

Our approach to achieving common sense in computers depends heavily on understanding and emulating human cognitive developmental processes within a unified architecture for perception, emotion, learning, cognition, and interaction with the world. To accomplish this we are developing such an architecture while integrating relevant subsystems (vision, coordination, audition, etc) developed in other groups in addition to our own development efforts. We intend to develop and test systems in both virtual and physical environments.

Our near term metrics focus on proving experimentally that our system can achieve domain independent, self motivated acquisition of common sense knowledge and effective exploitation of that knowledge in a rich, interactive environment. Our long-term goal is to create the first system that passes an unrestricted "3 year old Turing Test". We believe that passing such a test will empirically establish that we have indeed achieved human-like understanding or common sense within a computer system.

Bootstrapping a Mind from Sensory/Motor Experience

At conception, our lifetime of experience begins at a primitive biochemical level. As cell growth and differentiation ensues, our experience becomes more varied as newly specialized cell types emerge with different biochemical behavior and sensitivities. After only a few days of development, ensembles of similar cells generate emergent behaviors that start the blastosphere on the road toward tissue and organ development. These tissues interact at a supercellular level, generating yet another metalevel of behavior and sensitivities. Feedback loops abound at each level, from intracellular chemistry to systems of organs, with numerous interactions and interdependencies between the elements of each level as well as interactions across these emergent levels of behavior. The dynamical behavior at each level bootstraps the emergent capabilities for the next level. The goal of individual cellular homeostasis becomes dominated by tissue homeostasis, which gives way to organ homeostasis that eventually yields to the overall goal of organism homeostasis. Individual elements at any one level may be suboptimized or even sacrificed for the survival or optimization of a higher level. At some unknown point during this process, a level of behavior and sensitivity emerges that becomes reflectively aware, or conscious, of its own state. This consciousness now begins the slow process of learning how to manage its own state via interaction with its environment through the sensors and effectors that have developed previously. Previous levels were simply reactive to the changes in their environment. The conscious level is the first to be proactive, to intentionally manipulate its environment to improve its performance in meeting its needs.

Emergent Behavior and the *Tabula Rasa*

Given the bootstrapping process above, what can we say about the character of the behaviors and the knowledge

structures they generate at each emergent metalevel? At all levels, each new layer of emergent behavior innately provides a metaontology for the knowledge that can be generated and utilized at that level. This metaontology is directly encoded in the behaviors themselves and knowledge generated at this level is restricted by the execution of those behaviors. The knowledge structures (both genotypes and phenotypes) generated at each level are themselves compliant phenotypes of the metaontology encoded in the emergent behavior of that level. This implies that while there is lower level encoded knowledge in existence at the birth of a new emergent level, there is no *a priori* knowledge at the new level until it is generated via experience. So while it is true that when looking across metalevels that there is no *tabula rasa*, each new level of emergent behavior begins its operational life with a clean slate, at least at its own level.

Conscious Behavior: Escaping the Shackles of Genetics

So what makes consciousness so different from the lower levels of behavior that support it? A computer analogy provides useful insights. The behavior of any computer is constrained by a combination of its hardware, software, and the information available to be processed. One difference between hardware behavior and software behavior is that hardware behavior is grounded in the physics of the electronics, while software behavior is grounded in the "physics" or behavior of the hardware. This analogy can be stretched to the information input to the software as well, since the resulting information output from the system is constrained or "grounded" in the "physics" of the software. This situation is analogous to the emergent behavioral levels discussed above in that each prior level constrains the behavior of the next because it provides the implementation of the primitive functions of that behavior. The main difference from this limited perspective between the computer system and the biological mind is that the behaviors and knowledge at each level of the computer example directly encoded by their designer. They are not emergent. Of course, there are software techniques for generating data and software, and even techniques for generating new hardware designs, but with very few exceptions, there are no new layers of behavior and knowledge generated, only modifications to the existing layers. But what does this have to do with consciousness? The answer lies in the understanding that while software does indeed implement behavior and orchestrates the hardware functions to support a specific information processing application, software itself is information. This allows applications to be developed whose ongoing behavior is the result of self-modifying software, where the information produced by a software program becomes part of the program itself and changes its future behavior. The computer science concepts of "interpreters", "virtual machines" and “genetic algorithms” are examples of this behavior. Again, what does this have to do with consciousness? We believe that the major difference between the first conscious level and the previous levels on which it rests is that this is the point where the system's behavior escapes its "genetics”. For the first time, awareness of and reflection upon its internal states produces intentional changes in its own "programming" to hopefully improve its ability to meet its needs. This new capability provides a major qualitative difference in behavior at the initially conscious level, and provides the fundamental architecture for the layers of behavior and knowledge that emerge throughout its experience. We believe these new emergent layers constitute what we know as human memory, learning, and cognition. We believe that the layer directly below the initial conscious level represents the sensor and effector interface between the "mind" and the body and its environment. The task of anchoring cognition to environmental interactions falls directly on this initial conscious level.

Architecting Consciousness

Given the forgoing analysis, what are the fundamental behaviors and metaontology that emerge to support human consciousness? We believe we have identified a significant number of these behaviors. We have also discovered a representational approach that enables direct implementation of these behaviors within a unified architecture for perception, emotion, learning, cognition, and interaction with the world. We intend to present this architecture at the symposium, with specific attention given to this representational approach. We also intend to present initial results from our experimental implementation of this architecture.