Symposium on Learning and Motivation in Cognitive Architectures

Final Report for ONR Grant N00014-03-1-0536

Daniel Shapiro and Pat Langley
Institute for the Study of Learning and Expertise
2164 Staunton Court, Palo Alto, CA 94306

The Symposium on Learning and Motivation in Cognitive Architectures was held at Stanford University on March 22 and 23rd, 2003. Over 50 researchers attended the meeting, of which 14 participants, all known for their work in this area, presented talks on their recent results. The symposium fostered a "working atmosphere" in which scientists from two very distinct perspectives – cognitive science and artificial intelligence – could meet to share views and approaches. The role of emotion and motivation in guiding artificial agent behavior was discussed, as were a variety of approaches for representing and acquiring motivated behavior. Speakers were encouraged to devote half their time to an overview of their architecture and the other half to ways in which it addresses issues in learning and/or motivation.

We organized the symposium into talks and discussions on two consecutive days, where the first emphasized mechanisms for representing and learning motivated behavior within existing architectures and the second considered more strongly the role of emotions. We also dedicated a session of the second day to the issues of modularity and scalability in cognitive architectures. Below we summarize the contents for each speaker's presentation in the order they were given. This information, together with slides and references to what each author judged to be his most relevant paper, can also be found on the symposium Web site at `http://www.isle.org/symposia/cogarch/`.

1. Learning in Soar: Beyond Chunking. *John Laird* (University of Michigan) reviewed the Soar architecture and the constraints inherent to incorporating learning into such a framework, using the mechanism of chunking to illustrate his ideas. He discussed his reasons for investigating the addition of new learning mechanisms to Soar and presented a preliminary design and implementation of episodic and reinforcement learning. Laird concluded with a discussion of the promises and challenges of integrating these learning mechanisms.

2. Learning in ACT-R: Revisiting Chunking Without Subgoals. *Richard Lewis* (University of Michigan) reviewed the latest version of ACT-R, which includes a mechanism for production composition, abandons architectural support for subgoaling, and incorporates perceptual/motor components. He presented preliminary results from a simple HCI task model that involves learning of hierarchically controlled behavior. This model acquires task-specific production rules via a general interpretive process working over a declarative goal hierarchy. Learning gradually collapses the hierarchy and creates new control symbols in a way quite similar to Rosenbloom and Newell's work on chunking, but without architecturally distinguished subgoals. The model is supported qualitatively by new psychological data that provide evidence for the hierarchical control of behavior and its gradual collapse. Lewis also drew tentative lessons from his work for symbolic cognitive architectures more generally.

3. Exploring the Interaction of Implicit and Explicit Processes in Skill Learning. *Ron Sun* (University of Missouri-Columbia) reported on Clarion, an architecture that captures a wide range of quantitative phenomena about the interaction between the acquisition of implicit embodied skills and more explicit, abstract knowledge. By combining a bottom-up approach that learns implicit knowledge and then builds explicit knowledge upon that base, this framework accounts for ways in which interaction between these two processes improve or hamper learning. Sun also described experiments with human subjects designed to explicate further the interaction between implicit and explicit processes and to evaluate his theory of the human cognitive architecture.

4. EPIC: An Embodied Cognitive Architecture for Modeling Human Cognition and Performance. *David Kieras* (University of Michigan) described EPIC, an architecture that consists of a multithread-capable production system surrounded by perceptual-motor modules that represent fundamental properties of human performance. Models built with EPIC reveal the critical role of perceptual-motor constraints and executive strategies in multitask performance, and they offer a new account for the role of working memory and learning in both simple and multiple-task skills. Kieras also discussed potential learning mechanisms that would augment the existing framework.

5. Goals and Threats: Motivations in CIRCA. *David Musliner* (Honeywell Laboratories) described the Cooperative Intelligent Real-Time Control Architecture, including its core planning system and several recent variations. This framework was designed to control autonomous systems operating in hazardous or mission-critical environments, including adversarial domains. To meet real-time deadlines, the architecture provides guaranteed timeliness and formally verifiable correctness. Musliner also discussed CIRCA's motivation mechanisms, which rely on the specification of goals and threats to system safety.

6. Representing and Reasoning About Resources in APEX. *Michael Freed* (NASA Ames Research Center) discussed APEX, an agent architecture designed specifically to cope with time pressure and uncertainty. For example, air traffic control consists almost entirely of routine activity, and its complexity arises largely from the need to manage multiple tasks like guiding a plane to a destination airport, which involves issuing a series of standard turn and descent authorizations. Since such routines must be carried out over tens of minutes, the task of handling any individual plane must be interrupted periodically to handle new arrivals or to resume an ongoing task. The APEX framework incorporates and builds on multitask management capabilities found in previous systems, but it also introduces novel features.

7. Hierarchical Architectures for Controlling Real-World Agents. *David Kortenkamp* (TRAC Labs) reported on 3T, a hierarchical, hybrid control architecture that attempts to capture the strengths of both deliberative and reactive control approaches. The framework incorporates a continuous/reactive control layer, a symbolic/deliberative control layer, and a middle layer that mediates between continuous and symbolic control. The 3T architecture has been used to control over a dozen robotic and non-robotic applications. Kortenkamp described each control layer in detail, including its representations and reasoning

mechanisms. 3T supports adjustable autonomy, which means that user input can occur at any control layer, producing either fine-grained or coarse-grained external goals, which serve as the source of the architecture's motivation.

8. Emotions in the CogAff Architecture. *Aaron Sloman* (University of Birmingham) introduced CogAff, a framework for integrated architectures that combines different forms of representation and mechanisms. He presented an instance of this framework, H-Cogaff, a conjectured architecture for human-like systems that incorporates a variety of concurrently active components and accommodates many features of human mental function, including emotions and learning. H-Cogaff supports more varieties of emotion and learning than are normally considered, along with many affective states, including desires, pleasures, pains, attitudes, and moods. These ideas have implications both for applications of AI, such as digital entertainment or learning environments, and for scientific theories about human minds and brains.

9. The Architectural Role of Emotions in Cognitive Systems. *Jonathan Gratch* (University of Southern California) claimed that emotion is nature's solution to challenges faced in the design of intelligent systems. Findings in psychology and neuroscience have overturned views that emotion conflicts with rational thought and have suggested mechanisms through which emotion helps an organism adapt to its environment. Gratch discussed how an architectural examination of these findings lets one abstract the function that emotion plays in human information processing and map it to issues in agent architectural design. He reviewed early architectural models from Simon and Johnson-Laird and Oatley, but reexamined their issues in light of recent psychological findings. Cognitive appraisal theory, which dominates recent psychological thought on emotion, emphasizes the adaptive function of emotion and the close relationship between cognition, emotion, and motivation. He described this theory in some detail to give a flavor of how contemporary research on human emotion can influence autonomous agent design and address central concerns in the development of cognitive systems.

10. A Value-Driven Architecture for Intelligent Behavior. *Daniel Shapiro* (Institute for the Study of Learning and Expertise) described Icarus, an integrated cognitive architecture for intelligent agents in which affective values play a central role. The framework incorporates long-term and short-term memories for concepts and skills, and it includes mechanisms for recognizing concepts, calculating reward, nominating and selecting skills based on expected values, executing those skills in a reactive manner, repairing these skills when they fail, and abandoning skills when their expected values are low. Shapiro illustrated these processes with examples from the domain of automobile driving, and he related Icarus' assumptions to five principles for the design of cognitive architectures.

11. Engineering Cognitive Architectures for Comprehensiveness, Scalability, Stability, and Ease of Use. *Randolph Jones* (Colby College) presented a perspective on cognitive architectures that arose from his experience in developing "heavy" agents that are long-lived and exhibit high degrees of competence in complex environments. Industrial-strength heavy agents add a number of requirements to the list that already exists for cognitive architectures. In particular, such agents may be applied across a wide variety of tasks and situations that require many architectural options. The underlying architecture must scale to support

agents that run efficiently, for long periods, with very large knowledge bases, and it must be engineered to provide a robust and stable platform for applications with low fault tolerance. Finally, to be successful, it must be easy to use by architectural researchers, agent developers, subject-matter experts, and end users. He argued that these concerns should be designed into the architecture from the start and discussed some initial efforts toward such a design.

12. Challenge Problems for Cognitive Architectures. *Robert Neches* (University of Southern California) provided participants with a tour of an extremely large-scale application, with concommitent challenges to research on cognitive architectures in the areas of planning, attention management, representation and understanding, and interactions among components. He also touched on his earlier work with PRISM, a formalism for specifying production-system architectures in terms of decomposable modules.

13. Web Agents: Learning and Modularity. *Steve Minton* (Fetch Technologies) described his recent research on agent-based systems that can access and interact with World Wide Web sites, including several systems that act as "intelligent assistants" for information gathering and planning tasks. This work has focused on developing methods for automatically learning the structure of Web sites, so that an agent can extract data and/or execute transactions against a site. One of the critical aspects of our research concerns how to decompose the learning problem into modular, learnable subcomponents. Minton's talk focused on the issues of modularity and abstraction, including how these challenges influence the design cognitive architectures for use in constructing Web agents.

The second day was closed off by *Robert Balzer* (Teknowledge), who presented general observations about the previous symposium talks. His analysis emphasized the need for making cognitive architectures both scalable and modular, the first so they can address real-world problems and the second so they can be incorporated into integrated systems that take advantage of their various strengths. Balzer's commentary led into an open discussion on these topics, including how to address these factors in future research.

Collectively, the presentations outlined an interesting synthesis of motivational constructs and implementation technologies. Researchers from the cognitive science tradition emphasized the functional role that emotion plays within an architecture to orchestrate adaptive response, whereas researchers from artificial intelligence illustrated mechanisms for capturing motivations within agent models and for acquiring them through learning. The underlying representations included declarative goals, decision-theoretic value functions, and relative priorities over tasks.

One especially encouraging aspect of the symposium was an emergent interest in a systems perspective on cognitive architectures. This was evidenced by the fact that discussions consistently returned to the the challenges involved in building general, scalable, and modular architectures as a means of supplying exportable functionality. Several talks focused directly on this theme.

In summary, the Symposium on Learning and Motivation in Cognitive Architectures encouraged the interchange of knowledge between two rather disparate communities that are nevertheless unified by their interest in integrated architectures for intelligent agents. Informal discussions during the meeting also suggested promising directions for future research in this increasingly important area of artificial intelligence and cognitive science.