The FMS Cognitive Modeling Group at Carnegie Mellon University

SCALING COGNITIVE MODELS TO ALL LEVELS OF HUMAN ACTIVITY

I. COGNITIVE ARCHITECTURES

Understanding the workings of human cognition remains a fundamental scientific challenge: while the basic building blocks of the brain are well known, as is their general organization, there is no agreement on how cognition emerges from their interaction. Unified theories of cognition consisting of invariant mechanisms and representations, implemented computationally as cognitive architectures, have been proposed as a way to organize the empirical findings and master the complexity of neural systems. If successful, they would also represent the most promising way of engineering artificial systems capable of general intelligence, a.k.a. Strong AI or AGI.

ACT-R is an integrated computational cognitive architecture resulting from decades of cumulative effort by an international community of cognitive researchers. It consists of a modular framework (see figure 1) with the following components: a) procedural and declarative memory modules, including both symbolic and subsymbolic (i.e., statistical) representation and learning mechanisms; b) perceptual and motor modules that incorporate many known human factors parameters and provide principled limitations on the interaction with an external environment; c) a constrained modular framework for incorporating additional factors such as fatigue and emotions that are not currently part of the architecture; and d) asynchronous interaction between modules that assemble small, sub-second cognitive steps into complex streams of cognition to accomplish high-level functionality.

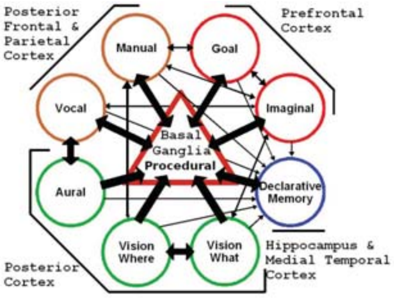
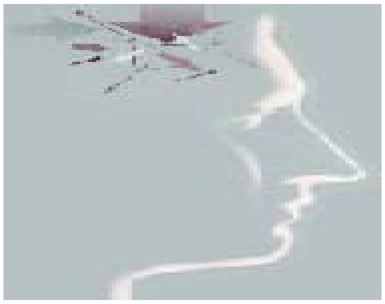


Fig. 1. Overview of the ACT-R Architecture, showing the different modules classified according to brain region.



Models build using the architecture can learn to perform complex dynamic tasks while interacting directly with the same environment as human users. ACT-R can account for all quantitative measures of human performance, from behavioral measures such as response time, percent of correct responses and eye movements, to fine-grained neural measurements such as EEG and fMRI. Hundred of cognitive models have been validated against experimental data, for tasks ranging from performing simple psychology experiments to controlling complex information systems such as air traffic control.

II. RESEARCH AGENDA

ACT-R models have adopted a particular level of abstraction, usually interact with abstract simulation systems, and represent the cognition of a single agent with definite goals. Our current research aims at lifting those limitations by relating to other levels of description of our cognitive system, in particular the neural level, by grounding ACT-R within physical systems such as robots, and by extending the single-model paradigm to multiple models interacting in cognitive networks.

A. Cognitive Neuroscience

The primary challenge in cognitive science is to construct bridges between levels of description similar to those erected in the physical sciences. For instance, when understanding how people gather and process information in decision-making, there are various levels at which these decision processes can be analyzed. Using Marr's tri-level approach to information processing, at the computational level, sensemaking accounts of the decision-making process are formulated in terms of large-scale situational awareness and descriptive knowledge structures.

In contrast, at the implementation level, neurally inspired models provide accounts of brain regions that implement the processes described by sensemaking. Bridging these levels is the algorithmic level, which is captured by functional models such as ACT-R bridging sensemaking goals with neural architecture.

The difficulty lies in translating high-level qualitative sensemaking descriptions to the language of neurological mechanism. We use the ACT-R cognitive architecture to better understand the cognitive processes underlying sensemaking and provide a bridge to the implementational level.

Our most recent project in that direction is a collaborative effort devoted to understanding cognitive biases in the context of geospatial intelligence analysis by building neural models of artificial tasks, with sensemaking as the theoretical framework for information processing.

High-fidelity models of information processing at the neural level are created by our partners at UC Boulder using the Emergent architecture, a framework for developing neural networks that utilize the Leabra learning algorithm.

The role of our group in this project is to create functional cognitive models of the tasks in ACT-R. Doing so serves two purposes. The first is to provide a bridge between sensemaking theory and neural theory. The second is to prototype models of tasks in order to assist the Emergent modelers by providing functional constraints on the neural models. That enterprise is scientifically grounded by the architectural commitments shared by ACT-R and Leabra despite their distinct levels of description.

Mechanisms that were prototyped in ACT-R include perceptual path planning, anchoring and adjustment bias in learning, within-participant probability matching, confirmation bias in information selection and resource allocation in intelligence analysis.

B. *Cognitive Robotics*

A major issue limiting the application of cognitive architectures as general intelligent systems is their lack of physical grounding. Real-world intelligence is grounded in our experience in the physical world. To achieve comparable integration and autonomy in the world such as on robotic platforms, we are extending our collaboration with the Leabra architecture. Specifically, we are developing a hybrid framework called SAL ("Synthesis of ACT-R and Leabra") to leverage the strengths of both architectures. The current instantiation of the SAL framework uses Leabra for perception, leveraging neural adaptivity, while ACT-R handles tasks such as decision-making that rely on symbolic control.

One research goal for the hybrid architecture is to create a system capable of unsupervised learning. To accomplish this, a visual stimulus is presented to Leabra, which then passes the resulting high-level neural representation of the stimulus to ACT-R. ACT-R then determines via declarative memory recalls whether the pattern presented is sufficiently similar to one already known; if not, ACT-R generates an arbitrary label for the stimulus, and issues a command to Leabra to train the visual system on the given stimulus using the ACT-R-generated label. This process is then iterated until a stable labeling of the stimulus on both sides of the hybrid framework is achieved.

Eventually, the hybrid model will learn not just to develop its own categories, but to learn the affordances of objects and to manipulate them. The ongoing development is focused on completing the loop between perception and cognition, and integrating motor functions into the framework. The ultimate goal is a fully autonomous system that learns to perceive its world and act upon it without any assistance.

In parallel, we have also started exploring integrating ACT-R with algorithmic approaches to perception. We use a state-of-the-art object detection algorithm to detect pedestrians walking along a sidewalk. Information from perception is then fed into an ACT-R model that tracks these pedestrians as they make their way along the sidewalk. The model uses these tracks to categorize (and, in some cases, predict) pedestrian behavior as either normal or suspicious. Categorization/prediction works on the basis of the model checking the calculated tracks for the presence of certain features together with other contextual information (such as the presence or absence of certain objects/people). We are currently working on detecting the relevant features automatically from the perceptual data by observing differences between expected and actual behavior.

A long-term goal of this research is to generalize the use of expectations by the model to drive cognitive behavior. In the expectation-driven approach, a model would generate an expected outcome for any proposed action. When this action is performed, the model compares the expected outcome with the real outcome. The result of this comparison is used in selecting future actions. We believe that such an expectation-driven approach has implications all along the cognitive and rational bands in Newell’s Time Scale of Human Cognition, from attention and learning to task selection and planning.

C. *Visual Intelligence*

Beyond basic perceptual grounding, an important focus of our research is on developing models of visual intelligence, namely the capability to learn representations of actions from visual inputs, reason over them, and eventually generate a meaningful description in natural language.

Far from being a bare summation over object features detected in the environment, visual intelligence has to be conceived as a complex phenomenon where perception combines with high-level cognitive processes and conceptual knowledge. A key distinction between this research and the state of the art in computer vision is that the latter exclusively deals with detection and classification of a wide range of objects (nouns in the description of a scene, e.g. “person”, “car”, “ball”, etc.) and features (e.g., position, orientation, shape, direction of motion, etc.) while research in visual intelligence operates at a more composite and unifying level: the goal is to identify the perceptual and cognitive underpinnings for recognizing and reasoning about actions (denoted by verbs in natural language), focusing on the roles played by objects in a scene (e.g., “a person hits a car with a ball”). Human understanding of events results from intertwined perceptual, cognitive and inferential processes: reproducing this capability at the machine level requires a comprehensive infrastructure where optical invariants, low-level perceptual mechanisms and high-level cognitive processes couple with knowledge representations.

In this framework we are working on integrating the ACT-R architecture (cognitive level) with computational ontologies (representational level), aiming at building a hybrid knowledge system where cognitive mechanisms are combined with representational contents. In particular, computational ontologies specify the meaning of (conceptual) representations of the world, encoding the semantics of those knowledge contents in a computational model.

The challenge of this research is to develop a full-fledged system where reasoning capabilities are cognitively grounded and driven by mechanisms of perception, memory and control. In this context, the purely technological problem of augmenting the ACT-R cognitive architecture with suitable ontologies of actions and automatic reasoners depends on the scientific problem of understanding the intertwined dynamics of reasoning, knowledge formation and perceptual mechanisms: in this sense, developing cognitively-inspired visual intelligence systems can also be seen, in a broader context, as a step forward in cognitive science and artificial intelligence.

D. *Economics and game theory*

Traditionally, cognitive models, like participants in psychology experiments, are given specific goals to accomplish. In the real world, however, people are confronted with ill-defined tasks, conflicting incentives, and open-ended learning opportunities.

To embody models with the human ability to set goals and define their own task, we are working on a project aimed at understanding adversarial cognition and motivation. We use strategic games inspired from the field of game theory and enhanced to afford human experimentation. For instance, we are currently developing an experimental paradigm called IPD^3 – Intergroup Prisoner's Dilemma with Intragroup Power Dynamics and Individual Power Drive – that provides a useable interface between humans and computers to play a series of nested repeated games. We have designed the game so that it is solidly grounded in state-of-the-art game-theoretic and socio-cognitive research. This design retains features that are advantageous for experimental purposes (e.g., binary choice, matrix format, computational tractability) while adding features that increase ecological validity (e.g., multiple players, social structure, asymmetries, conflicting motives, and stochastic behavior). We use this paradigm to collect human data in carefully designed experiments. In parallel, we develop cognitive models of human behavior in strategic games. For example, in one empirical study we found that human participants are perfect reciprocators: negative emotions triggered by unreciprocated attempts at cooperation bias subsequent decision-making toward defection. However, a cognitive model of the same game learned that sustained cooperation (even when occasionally unreciprocated) was more effective. The cognitive model did not show the retaliation bias because it lacked affective processing. We are working on developing an affective module for the ACT-R cognitive architecture that will allow us to model the full range of human cognition and emotion in strategic games. Our cognitive modeling approach complements the traditional equilibrium analyses in predicting the effect of game modifications. In repeated games, almost any outcome can be an equilibrium. Game simulations with validated cognitive models as players can be used to narrow down the set of possible equilibria to a limited number of cognitively plausible outcomes, generate specific predictions about human behavior in these games, and provide more believable synthetic characters for games and training.

E. *Network science*

Beyond the interactions of small groups of individuals lies the emergence of civilization from the collective cognitive acts of large-scale societies. Among the most fascinating abilities of human beings is their propensity to verbalize, communicate and adopt ideas within a vast network of social contacts. Human cognitive capabilities are uniquely suited to communication, and they are crucial to the intelligence emerging from human communities. The cognitive and psycholinguistic mechanisms underlying language comprehension and production are still poorly understood. While recent studies paint a picture of how memory and contextualization help humans comprehend a dialogue partner's ideas and individual language, we do not understand whether human memory has evolved to support teamwork and social cognition. In addition to communication, other tasks such as decision-making under uncertainty are crucially enhanced by teamwork, despite the fast and frugal heuristics or performance bounds found in individual human actors.

Cognitive modeling and network simulation techniques have allowed a recent growth in interest for the interaction of cognitive mechanisms with the social environment. Cognitive architectures can characterize the bounded human abilities to recall information, which is key in explaining the interactions of humans in a network.

In recent work, we have shown how individuals adapt their linguistic expressions quickly to their interaction partners, and new communicative conventions soon spread through a network of connected entities, which may be cognitive models of humans or simple software agents.

Cognitive modeling frameworks, validated and refined through careful experimentation, as well as computational tools now allow the larger-scale simulation of human societies and the uptake of existing language resources (corpora) in the quest for the architecture of the human language faculty. We have developed a networked experimentation platform called the Geogame to facilitate large-scale data collection. Datasets collected in real-life situations let us test cognitive and psycholinguistic models. Once validated, they will make better predictions and cover broad ranges of human behaviour. This combination of broad coverage and large-scale simulation requires new computational tools, new methodologies, new datasets and new experimental designs. With ACT-UP, our lab has developed a rapid-prototyping scalable implementation of the ACT-R cognitive architecture, allowing large-scale simulations of underspecified models.

III. ACKNOWLEDGEMENTS

The research described in this profile was supported by grants from AFOSR, DTRA, ARL, DARPA and IARPA to Dr. Christian Lebiere. All views expressed are solely those of the authors and not of the funding agencies.

Contact Information
The FMS Group:
Dir.: Christian Lebiere - cl@cmu.edu
Ion Juvina
Unmesh Kurup
Alessandro Oltramari
David Reitter
Matthew Rutledge-Taylor
Robert Thomson
Jerry Vinokurov
Websites:
http://fms.psy.cmu.edu
http://act-r.psy.cmu.edu