Challenges and Opportunities in Interactive Cognitive Systems

Pat Langley

Silicon Valley Campus
Carnegie Mellon University

Department of Computer Science
University of Auckland

Thanks to Paul Bello, Ron Brachman, Ken Forbus, John Laird, Allen Newell, Paul Rosenbloom, and Herbert Simon for discussions that helped refine ideas in this talk.

Artificial Intelligence
Then and Now

The Vision of Artificial Intelligence

The field of artificial intelligence was launched in 1956 at the Dartmouth meeting; its audacious aims were to:

- Understand the mind in computational terms;
- Reproduce all mental abilities in computational artifacts.

This view continued through the mid-1980s, but recent years have seen emergence of very different goals.

Why have most AI researchers retreated from the field's initial aspirations? What happened? How should we respond?

Early Emphases in AI Research

- Knowledge representation
  - encoding the meaning of complex natural language
  - flexibility and power found in human reasoning
- Problem solving and planning
  - general methods guided by (nonadmissable) heuristics
  - targeted the flexibility seen in human problem solving
- Natural language processing
  - structural processing with strong links to psycho/linguistics
  - emphasis on deep language understanding and generation
- Machine learning
  - incremental methods that learn as rapidly as humans
  - interest in reasoning, language, and problem solving

Current Emphases in AI Research

- Knowledge representation
  - focus on restricted logics that guarantee efficient processing
  - less flexibility and power than found in human reasoning
- Problem solving and planning
  - relies on extensive search and emphasize processing speed
  - bears little resemblance to flexible problem solving in humans
- Natural language processing
  - statistical methods with few links to psycho/linguistics
  - emphasis on tasks like information retrieval and extraction
- Machine learning
  - statistical techniques that learn far more slowly than humans
  - almost exclusive focus on classification and reactive control

The Cognitive Revolution

During the 1950s and 1960s, the key breakthroughs in AI and cognitive psychology resulted from:

- Rejecting behaviorists' obsession with learning on simple tasks and information theory's focus on statistics;
- Studying problem solving, language understanding, and other tasks that involve *thinking* (i.e., *cognition*);
- Emphasizing the role of *mental structures* in supporting such complex behaviors.

Unfortunately, many modern AI researchers have abandoned the main insights of this cognitive revolution.

Reasons for the Shift

This shift in AI's focus has occurred for a number of reasons, including:

- Commercial successes of 'niche' AI
- Faster processors and larger memories
- Obsession with quantitative metrics
- Formalist trends imported from computer science

Together, these have drawn researchers' attention away from the field's original vision.

Why Has AI Gone Astray?

Maslow (1966) postulates some other reasons why a scientific field can become narrow and conservative:

> … these “good”, “nice” scientific words – prediction, control, rigor, certainty, exactness, preciseness, neatness, …, quantification, proof, … – are all capable of being pathologized when pushed to the extreme. [They] may be pressed into the service of safety needs [to] become … anxiety-avoiding … mechanisms … for detoxifying a … frightening world as well as ways of … understanding a fascinating … world.

But Maslow notes that science need not proceed in this way:

> … healthy scientists [can] enjoy not only the beauties of precision but also the pleasures of sloppiness, casualness and ambiguity…They are not afraid of hunches, intuitions, or improbable ideas…All of this is exemplified in the greater versatility of the great scientist, of the creative, courageous, and bold scientists.

The Cognitive Systems Paradigm

The Cognitive Systems Movement

The field's original challenges still remain and provide many opportunities for research.

Because "AI" is associated with limited aspirations, we adopt the term *cognitive systems* (Brachman & Lemnios, 2002).

This paradigm aims to *design, construct, and study computational artifacts that exhibit the full range of human intelligence*.

We can distinguish cognitive systems from (current) mainstream AI by six characteristics.

See *Advances in Cognitive Systems* (http://www.cogsys.org/).

Feature 1: Focus on High-Level Cognition

One distinctive feature of the cognitive systems movement lies in its emphasis on *high-level cognition*.

People share basic capabilities for categorization and empirical learning with dogs and cats, but only humans can:

- Understand and generate language
- Solve novel and complex problems
- Design and use complex artifacts
- Reason about others' mental states
- Think about their own thinking

Computational replication of these abilities is the key charge of cognitive systems research.

Feature 2: Structured Representations

Another distinctive aspect of cognitive systems research concerns its reliance on *structured representations* and *knowledge*.

The insight behind the 1950s AI revolution was that computers are not mere number crunchers.

Computers and humans are *general symbol manipulators* that:

- Encode information as list structures or similar formalisms
- Create, modify, and interpret this relational content
- Incorporate numbers only as annotations on these structures

The paradigm assumes that representing and reasoning over rich symbolic structures is key to human-level cognition.

Feature 3: Systems Perspective

Research in our paradigm is also distinguished by approaching intelligence from a *systems perspective*.

While most AI efforts idolize component algorithms, work on cognitive systems is concerned with:

- How different intellectual abilities interact and fit together
- Cognitive architectures that offer unified theories of mind
- Integrated intelligent agents that combine capabilities

Such systems-level research provides the only avenue to artifacts that exhibit the breadth and scope of human intelligence.

Otherwise, we will remain limited to the *idiot savants* that have become so popular in academia and industry.

Feature 4: Influence of Human Cognition

Research on cognitive systems also draws ideas and inspiration from findings about *human cognition*.

Many of AI's earliest insights came from studying human problem solving, reasoning, and language use, including:

- How people represent knowledge, goals, and beliefs
- How humans utilize knowledge to draw inferences
- How people acquire new knowledge from experience

We still have much to gain by following this strategy, even when an artifact's operation differs in its details.

Human capabilities also provide *challenges* for cognitive systems researchers to pursue.

Feature 5: Heuristics and Satisficing

Another assumption of cognitive systems work is that intelligence relies on *heuristic methods* that:

- Are not guaranteed to find the best or even *any* solution but
- Greatly reduce search and make problem solving tractable
- Apply to a broader range of tasks than methods with guarantees

They mimic high-level human cognition in that they *satisfice* by finding acceptable rather than optimal solutions.

Much of the flexibility in human intelligence comes from its use of heuristic methods.

Feature 6: Exploratory Research

Cognitive systems research also differs from mainstream AI in its approach to *evaluation* in that it encourages:

- Demonstrations of entirely new functionality
- Novel approaches to well-established problems
- Analyses of challenging cognitive tasks
- Architectures and frameworks for integrated intelligence

These evaluation styles encourage *exploratory research*, which is crucial given how little we understand about the mind.

Studies must still make clear claims and support them, but many forms of evidence are possible.

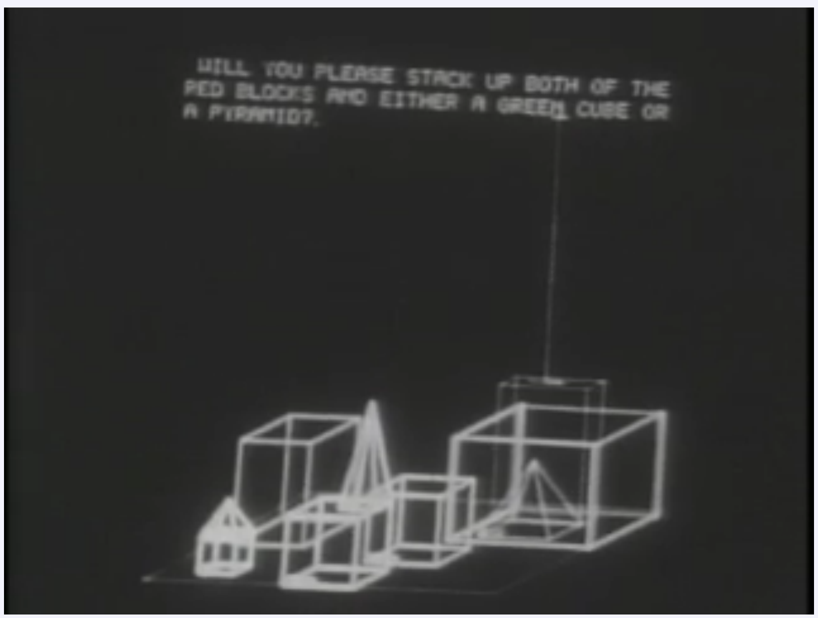
Examples of Cognitive Systems

SHRDLU (1970)

SHRDLU was an early AI system that interacted with users in natural language.

It inferred sentence meanings to answer queries and executed complex commands.

Although limited in scope, SHRDLU had many features of a cognitive system.



Person: Pick up a big red block.
Computer: *OK.*
Person: Grasp the pyramid.
Computer: *I don't understand which pyramid you mean.*
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: *By "it", I assume you mean the block which is taller than the one I am holding.*
Computer: *OK.*
Person: What does the box contain?
Computer: *The blue pyramid and the blue block.*
Person: What is the pyramid supported by?
Computer: *The box.*
Person: How many blocks are not in the box?
Computer: *Four of them.*
Person: Is at least one of them narrower than the one which I told you to pick up?
Computer: *Yes, the red cube.*

Aaron (1973–present)

The Aaron system composes and physically paints novel art work.

In some sense, it is only a rule-based expert system that operates in an area we usually associate with creativity.

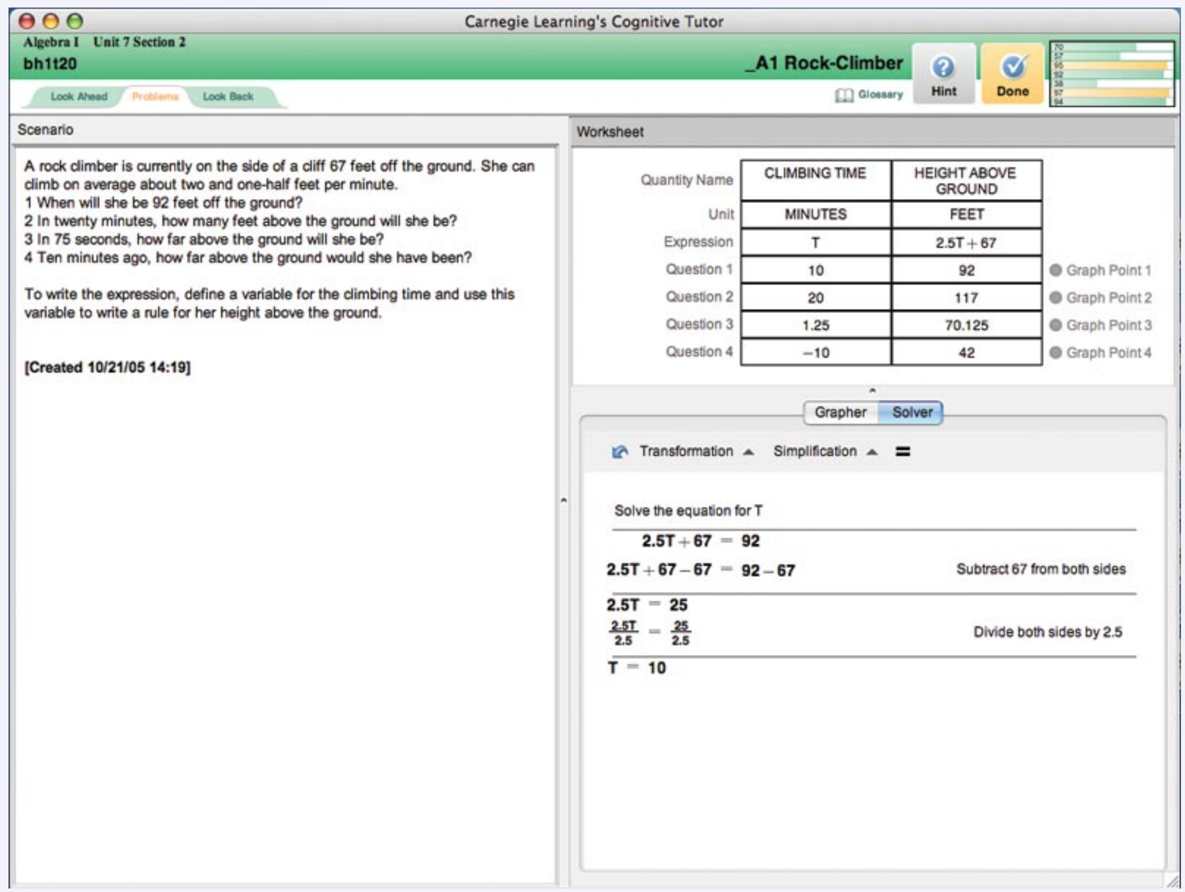
But it integrates many different facets of artistic composition and incorporates a robot arm to implement its designs.

Carnegie Learning’s Algebra Tutor (1999–present)

This tutor encodes knowledge about algebra as production rules, infers models of students’ knowledge, and provides them with personalized instruction.

The system has been adopted by hundreds of US middle schools.

Studies have shown that it improves student learning in this domain by 75 percent.

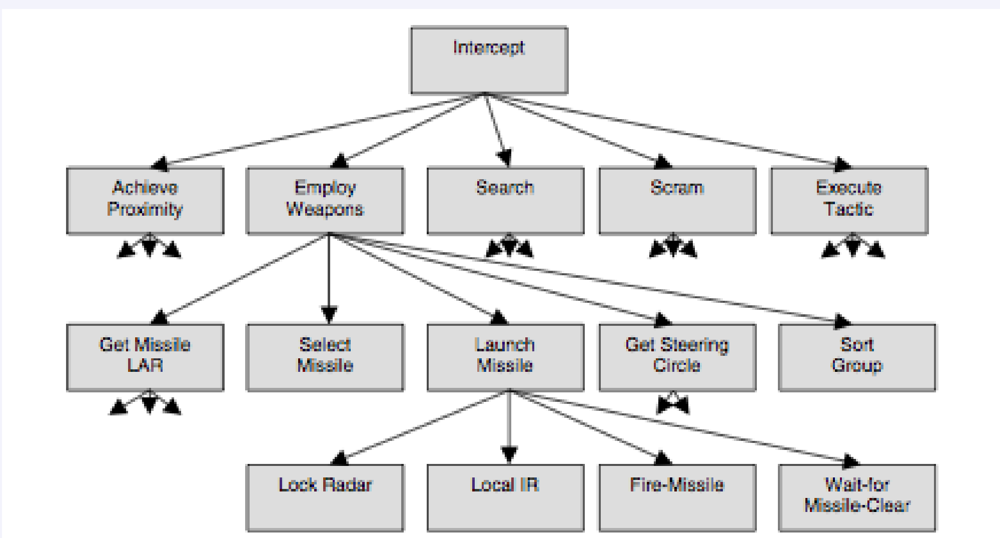


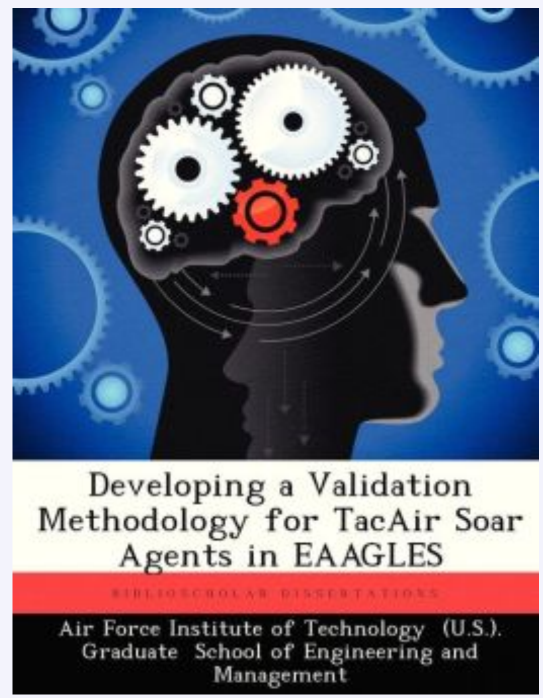
TacAir-Soar (1997)

The TacAir-Soar system reproduces pilot behavior in tactical air combat.

It combines abilities for spatio-temporal reasoning, plan generation / recognition, language, and coordination.

The system flew 722 missions during the STOW-97 simulated training exercise.





Façade (2003–2007)

Mateas and Stern's Façade is a graphical environment in which characters interact with the user and each other.

The agents understand and generate sentences, control gaze and expression, and they exhibit distinct personalities.

Façade characters use a rich knowledge base to produce inferences, carry out physical activities, and engage socially.

Three Hypotheses for Cognitive Systems Research

Laws of Qualitative Structure

Newell and Simon (1976) have argued that any scientific field depends on *laws of qualitative structure*, such as:

- The cell doctrine in biology
- Plate tectonics in geology
- The germ theory of disease
- The atomic theory of matter

They also proposed two such laws, one about mental *structures* and another and the other about mental *processes*.

Physical Symbol Systems

In their Turing Award article, Newell and Simon introduced the *physical symbol system hypothesis*, which stated that:

- *Symbols* – physical patterns that are stable unless modified – can be organized into *symbol structures*;
- A *physical symbol system* has processes for creating, modifying, and interpreting such symbol structures;
- Such a physical symbol system has the necessary and sufficient means for *general intelligent action*.

Symbolic processing of this sort is the fundamental idea behind most successes in artificial intelligence.

Heuristic Search

Newell and Simon also made another claim, the *heuristic search hypothesis*, about the nature of problem solving:

- A problem solver *represents* candidate situations, actions, and solutions as symbol structures;
- Problem solving involves a *search process* that generates, modifies, and tests these structures;
- Search is guided by *heuristics* – rules of thumb – that focus attention down promising paths.

Heuristics are needed because, in practice, one cannot search most problem spaces exhaustively.

Social Cognition

In the same spirit, we propose a third law of qualitative structure about the nature of intelligence.

The *social cognition hypothesis* states that intelligence requires the ability to:

- *Represent* models of other agents' *mental states*;
- *Generate* such mental models and *reason* over them;
- *Use* these models for *informed interaction* with others.

The Turing test's focus on extended conversation, however problematic, reflected this intuition.

This theoretical claim suggests some interesting challenges for research on cognitive systems.

Research Challenges for Interactive Cognitive Systems

Features of Challenge Problems

We should identify challenge problems that can drive research on interactive cognitive systems and that:

- Focus on tasks that require *high-level* cognition;
- Benefit from *structured* representations and knowledge;
- Require *system-level* integration of capabilities;
- Have *human* role models that offer insights;
- Be complex enough to need *heuristic* approaches; and
- Depend centrally on processing *social* structures.

They must also move beyond the Turing test by emphasizing *goal-oriented* behavior.

Rich Nonplayer Game Characters

Synthetic characters are rampant in today's computer games, but they are typically shallow.

We should develop more compelling nonplayer characters that:

- Infer other players' goals and use them toward their own ends;
- Interact with human players in constrained natural language;
- Cooperate with them on extended tasks of common interest;
- Form long-term relationships based on previous interactions.

Such agents would generate much richer and more enjoyable experiences for human players.

For this purpose, they must reason about others' mental states.

Deep Conversational Assistants

Spoken-language dialogue is the natural mode for providing aid on tasks like driving, cooking, and shopping.

But systems like Siri are primitive, and we need more effective conversational assistants that:

- Carry out extended dialogues about goal-directed activities;
- Take into account the surrounding task context;
- Infer *common ground* (Clark, 1996) for joint beliefs / goals;
- Store and utilize previous interactions with the user.

These would carry out deep language processing, reason about others' mental states, and depend crucially on *social cognition*.

A Truly General Game Player

Humans use their domain knowledge in different ways, and we need multifunctional systems with the same versatility.

One example might be a system that, given knowledge about a class of games, can:

- Play that class of game in competitions;
- Discuss previous games with other players;
- Provide commentary on games played by others;
- Analyze and discuss particular game situations;
- Teach the game to a human novice.

This should demonstrate breadth of intellectual ability but avoid the knowledge acquisition bottleneck.

A Synthetic Entertainer

Our society devotes far more attention to its pop stars than to its scientists and scholars.

Imagine a *synthetic entertainer* with a distinctive personality, a memory for previous events, and the ability to:

- Write the music and words for entirely new songs;
- Sing songs on a virtual stage with a backup band;
- Perform its songs in music videos directed by humans;
- Carry out interviews with reporters and talk show hosts.

Building such systems will not only clarify how different facets of cognition interact; they could even be, well, *entertaining*.

But, again, they would depend centrally on social cognition.

A Synthetic Defense Attorney

Another occupation with high status in our society, at least in the media, is the *criminal attorney*.

A fifth challenge involves constructing a trial lawyer with legal knowledge and ability to defend clients in mock trials by:

- Interviewing the client to gather case details;
- Planning a defense to use in court;
- Interacting with the judge during pretrial hearing;
- Examining and cross examining witnesses; and
- Preparing and presenting a closing statement.

Success in legal defense relies strongly on reasoning about, and manipulating, others' mental states, i.e., social cognition.

A Synthetic Politician

A final challenge involves yet another high-visibility profession; building a synthetic *politician* who runs for office.

This agent would have knowledge of current issues, memory for its career, and the ability to get elected by:

- Reasoning about a specified set of current issues
- Writing and delivering speeches on these topics;
- Answering questions from the press; and
- Participating in debates with other candidates.

The agent would generate plans to address current issues, defend them against critics, and, most important, reason about voters' beliefs / goals and how to sway them.

Some Necessary Components

Although cognitive systems involve integration, we also need research on core abilities for social cognition, including:

- *Representing* other agents' mental states;
- *Reasoning* flexibly about others' beliefs and goals;
- Social *plan understanding* from others' observed behavior;
- Social *plan generation* to manipulate others' actions;
- Understanding and planning in task-oriented *dialogue*;
- Cognitive accounts of *emotion*, *morals*, and *personality*.

We can study these in the context of challenge problems, but we can also use simpler domains like fables (Pearce et al., 2014).

Summary Remarks

In this talk, I discussed the cognitive systems paradigm, which pursues AI's original vision, by:

- Stating six distincive features of systems research in this area;
- Reviewing five examples of innovative cognitive systems;
- Proposing three hypotheses about intelligent behavior;
- Posing six challenge tasks for interactive cognitive systems.

Research in this emerging field retains the audacity of early AI and promises to keep us occupied for decades to come.

Readings on Cognitive Systems

- Laird, J. E. (2002). Research in human-level AI using computer games. *Communications of the ACM*, *45*, 32-35.
- Langley, P. (2012). The cognitive systems paradigm. *Advances in Cognitive Systems*, *1*, 3-13.
- Langley, P. (2012). Intelligent behavior in humans and machines. *Advances in Cognitive Systems*, *2*, 3-12.
- Langley, P. (2014). Four research challenges for cognitive systems. *Advances in Cognitive Systems*, *3*, 3-11.
- Jones, R. M., & Laird, J. E. (1997). Constraints on the design of a high-level model of cognition. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Stanford, CA.
- Swartout, W. R., Gratch, J., Hill, R. W., Hovy, E., Marsella, S., Rickel, J., & Traum, D. (2006). Toward virtual humans. *AI Magazine*, *27*, 96-108.

End of Presentation