Opportunities and Challenges in Cognitive Systems Research

Pat Langley

Silicon Valley Campus
Carnegie Mellon University

Department of Computer Science
University of Auckland

http://www.cogsys.org/

Thanks to Paul Bello, Ron Brachman, Nicholas Cassimattis, Ken Forbus, John Laird, and others for discussions that helped refine the ideas in this talk.

Introductory Remarks

The Vision of Artificial Intelligence

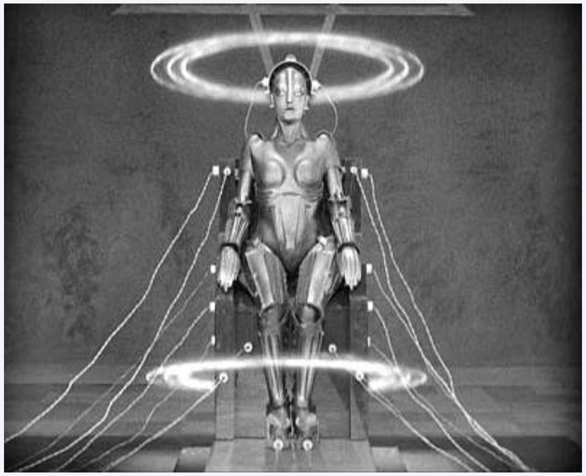
The field of artificial intelligence was launched in the summer of 1956 at the Dartmouth meeting.

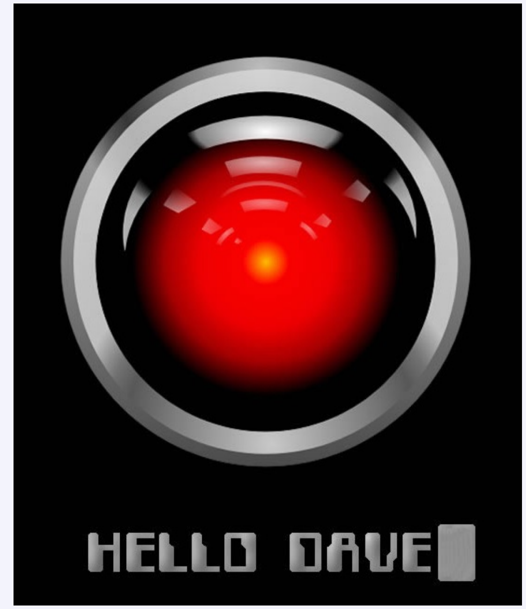
The audacious aim was to understand the mind in computational terms and reproduce all its abilities in computational artifacts.

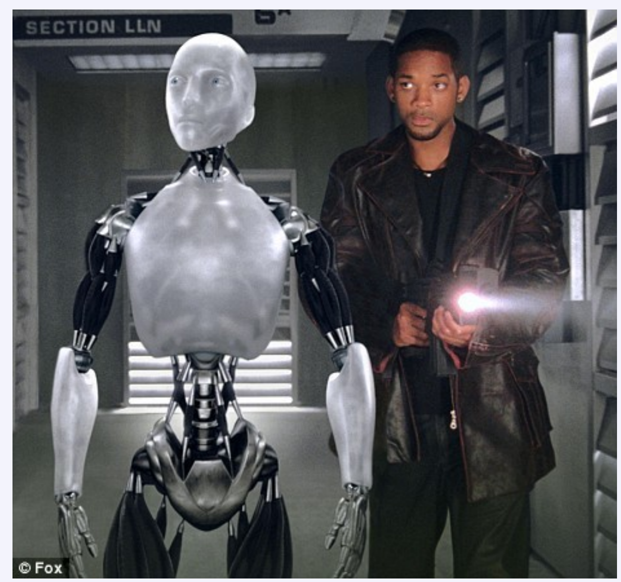
- Early researchers hoped to create systems with broad, general skills for reasoning, problem solving, and language use.
- This view continued through the mid-1980s, but recent years have seen a very different goals for AI emerge.

Why have most researchers and practitioners stepped back from the field's original aspirations? Can we remedy this situation?

AI in the Media







Historical Periods in AI

We can divide the history of AI into a number of periods with different concerns and styles:

- Audacious period / general methods (1956–late 1970s)
- Early applications / expert systems (late 1970s–late 1980s)
- AI winter / doubts about potential (late 1980s–late 1990s)
- Narrowed research and applications (late 1990s–present)

These labels do not describe all AI activities, which have been diverse and productive during all periods.

But they do reflect broad trends and attitudes about the field and its proper pursuits.

Some Narrow Successes

Expert Systems

The *expert systems* movement built upon insights about the role of knowledge in human expertise.

Work in this paradigm encoded knowledge as rules of thumb that it matched and chained to draw conclusions.

- Hundreds, if not thousands, of such systems have been deployed since the 1980s, saving large amounts. [See also *TurboTax*]
- However, they can be costly to maintain as conditions change, and their reasoning is often routine and shallow.
- And expert systems were oversold by some in the community, which led to a later backlash.

Thus, although producing the first successful applications of AI, in many ways they limited the field's scope.

Deep Blue

Playing chess, long viewed as a height of intellectual ability, was one of the original challenge problems for AI.

Early research on chess led to many insights about representation and search, two cornerstones of the field.

- *Deep Blue* was a hardware-supported chess player that searched deeper than humans or previous programs.
- In 1997, the system won a match against Gary Kasparov, the current world champion, taking 3.5 to 2.5 games.
- But Deep Blue was highly tuned, in both hardware and software, to playing chess and it lacked more general abilities.

Similar advances, with the same limitations, have occurred for checkers, backgammon, and other common games.

Spam Filters

Junk email has been an annoying problem on the Internet users for well over a decade.

Spam filters can greatly reduce this annoyance by detecting and redirecting likely candidates.

- Early spam filters were specified manually by users in terms of a constrained syntax.
- More recent filters collects user decisions as training data for supervised learning of classifiers.
- Modern systems now shelter users from the great majority of junk messages.

But these filters rely on shallow representations and statistical classification, not on the ability to understand text.

Recommender Systems

In the 1990s, the increasing availability of Web content, including products on the Web, led to *recommender systems*.

These propose, rank, or otherwise present selected items that they predict will interest the user.

- Most frameworks for recommender systems learn user profiles from explicit or implicit feedback.
- Collaborative approaches focus on similarities among user choices; content-based methods focus on item attributes.
- Amazon, Tivo, and many other companies use techniques of this sort to increase sales or customer satisfaction.

But recommender systems adopt shallow encodings of user tastes and view their task as simple classification or regression.

Other “Success” Stories

Other technologies that have been successful in their areas and have achieved wide recognition include:

- Web search engines
- Targeted advertising
- Self-driving cars
- The Watson system
- The Siri interface

Most of these systems exhibit the same narrowness and reliance on shallow encodings and methods.

The current excitement about ‘big data’ is likely to reinforce the popularity of such simple-minded approaches.

Summary

To summarize, the most visible products of AI over the past three decades have involved:

- Impressive, well-engineered systems that are
- Useful and have saved / produced substantial sums

But these computational artifacts have also been:

- Highly specialized for particular tasks and
- Often rely on shallow representations and methods

They are *idiot savants* that excel in their narrow areas but have no other competencies.

Thus, they tell us little about what makes us distinctively human or how to achieve the breadth of human intelligence.

Recent Trends in Academic AI

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 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

Commercial Success of AI

One reason for this shift has been AI's commercial successes, which have:

- led many academics to study narrowly defined tasks
- produced a bias toward near-term applications
- caused an explosion of work on "niche AI"

Moreover, component algorithms are much easier to evaluate experimentally, especially given available repositories.

Such focused efforts are appropriate for corporate AI labs, but academic researchers should aim for higher goals.

Hardware Advances

Two additional factors are faster computer processors and larger memories, which have made possible new methods for:

- playing games by carrying out far more search than humans
- finding complicated schedules that trade off many factors
- retrieving relevant items from large document repositories
- inducing opaque predictive models from large data sets

These are genuine advances, but AI might fare even better by incorporating more insights from human cognition.

Obsession with Metrics

A third influence has been increased emphasis on quantitative performance evaluations, which:

- has encouraged experiments on standardized problems
- with most studies taking the form of mindless ‘bake offs’
- that aim for ‘significant’ but not substantial improvements
- leading in turn to incremental progress but few insights

Worse, this emphasis has produced a bias against research on new functionalities and on novel but immature approaches.

Formalist Trends

Yet another factor arises from AI's typical home in departments of computer science:

- which often grew out of mathematics departments
- where analytical tractability is a primary concern
- where guaranteed optimality outranks heuristic methods
- even when this restricts work to narrow problem classes

Many AI faculty in such organizations view the field's original goals with intellectual suspicion.

This trend and others have transformed AI into a field that has adopted greatly restricted goals.

Another Perspective

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 and anxiety-controlling mechanisms … for detoxifying a chaotic and frightening 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 of still remain and provide many opportunities for research.

However, because "AI" has become associated with such limited aspirations, we need a new label.

We will use *cognitive systems*, a term coined by Brachman and Lemnios (2002), to refer to the discipline that:

- *designs, constructs, and studies computational artifacts that exhibit the full range of human intelligence*.

We can further distinguish this paradigm from what has become mainstream AI by describing its key characteristics.

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*.

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 *physical symbol systems* (Newell & Simon, 1976) of this sort are 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.

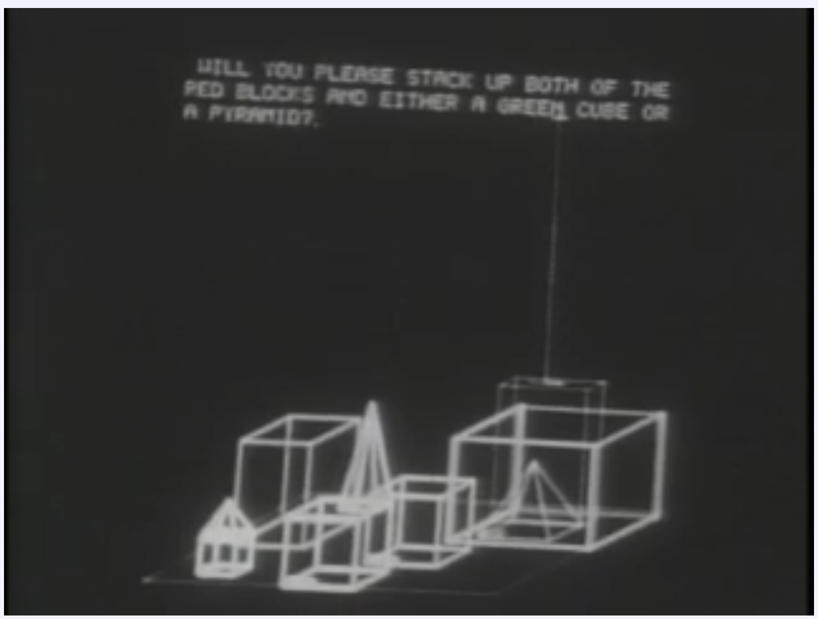
Examples of Cognitive Systems

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.

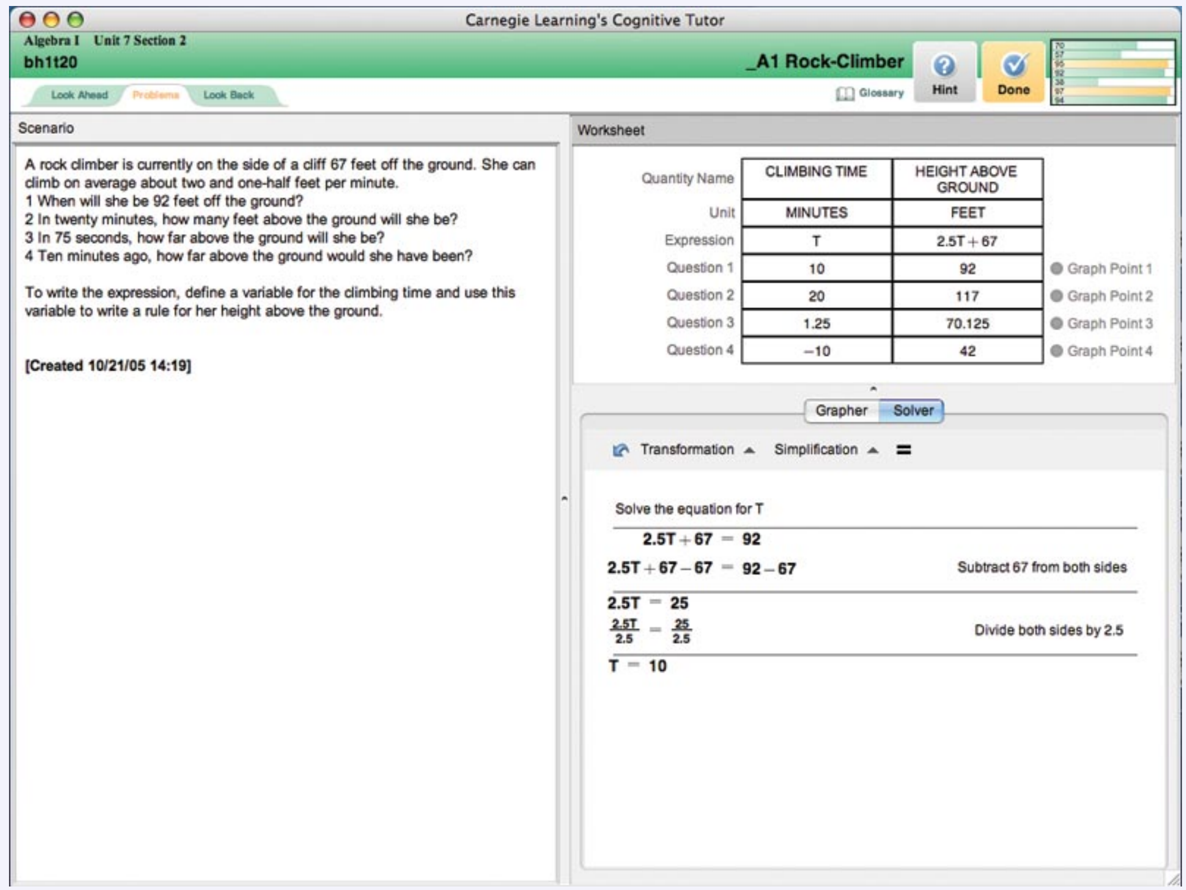
But it integrates many different facets of artistic composition and incorporates a robot arm to implement its designs.

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.

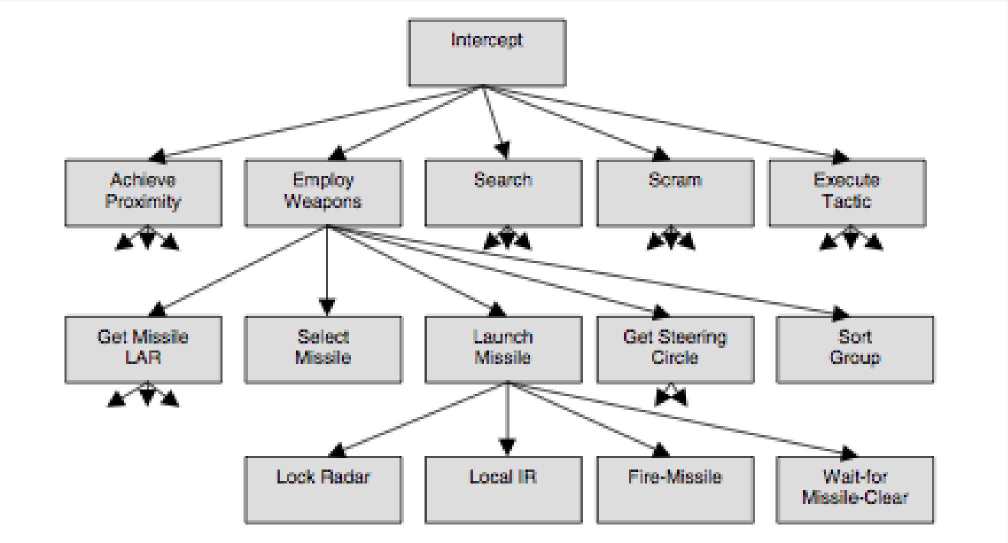


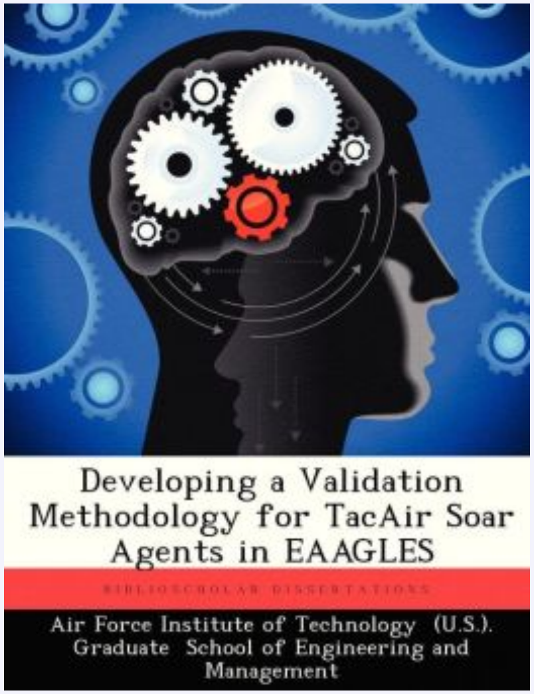
TacAir-Soar (1997)

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.





Some Recent Examples

Other efforts have also developed integrated systems that exhibit higher levels of cognition:

- The *Halo project* aims to acquire knowledge from scientific textbooks and answer questions in natural language.
- The *CALO project* developed an integrated office assistant that helps with meetings, purchase orders, and other tasks.
- The *Virtual Human project* creates synthetic characters that produce plans, have emotions, and communicate in language.
- The *Robot Scientist project* combines experiment design and execution with model revision in cell biology.

Although focused to enable progress, each has audacious goals that illustrate the cognitive systems agenda.

Research Challenges in Cognitive Systems

Some Research Priorities

We must identify challenges that can drive research on cognitive systems; some natural capabilities to study include:

- Mechanisms for flexible and scalable inference
- Flexible methods for problem solving / formulation
- Deep processing of language and dialogue
- Models of emotion and moral cognition
- Reasoning about others' mental states
- Metacognitive reasoning systems

However, we must also embed work on these topics in projects that move us toward useful software artifacts.

Deep Conversational Assistants

People carry out many tasks during a day, from cooking to driving to shopping to meeting with others.

Spoken-language dialogue is the only practical mode for helping with these tasks; an effective conversational assistant should:

- Infer the human user's goals and activities;
- Answer user questions and provide advice;
- Take into account the surrounding context;
- Store and recall previous interactions with user.

The resulting system would be similar to Siri, but it would carry out much deeper processing over more extended periods.

This would expand our understanding of task-oriented dialogue and its relation to other mental activities.

Domain-Limited Multi-Functional Systems

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

E.g., we might build 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 approach should demonstrate breadth of intellectual ability while avoiding the knowledge acquisition bottleneck.

Rich Nonplayer Game Characters

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

We should develop novel 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.

They would also advance our understanding of *social cognition*, which seems a key facet of human intelligence.

A Synthetic Entertainer

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

Imagine a synthetic character with a distinctive personality, the competencies for its profession, and memory for previous events.

The varied capabilities that it would support might include:

- Writing the music and words for new songs;
- Singing songs on a virtual stage with a backup band;
- Performing its songs in music videos directed by humans;
- Carrying out interviews with reporters and talk show hosts.

Such a system could not only clarify how different aspects of cognition interact; it could even be entertaining.

The Road Ahead

Although cognitive systems adopts the original aims of AI, its modern incarnation is relatively new.

To ensure its success as an innovative discipline, we must:

- Clarify and defend its distinctive characteristics
- Create a community of broad-minded researchers
- Identify research challenges and make progress on them
- Establish venues for communication and publication
- Recruit, train, and place promising new researchers
- Never abandon the audacious goals we have set ourselves

Understanding the mind will not happen overnight, but it is an important task that is well worth pursuing.

End of Presentation