# Symposium on Cognitive Systems and Discovery Informatics

**Pat Langley**

Silicon Valley Campus
Carnegie Mellon University

Department of Computer Science
University of Auckland

*http://www.cogsys.org/symposium/2013/*

This meeting has been funded jointly by the US National Science Foundation and by Stanford University. Thanks to Ben Meadows for administrative assistance.

# The Original AI Vision

The early days of artificial intelligence research were guided by a common vision:

- *Understanding and reproducing, in computational systems, the full range of intelligent behavior observed in humans*.

This paradigm was adopted widely from the field's founding in the 1950s through the 1980s.

Yet the past 25 years have seen a very different AI emerge that has largely abandoned these initial goals.

This has happened for many reasons, but together they have led to greatly reduced aspirations among researchers.

# The Cognitive Systems Movement

The field's original challenges still remain and provide many research opportunities, but we need a new label for their pursuit.

We will use *cognitive systems* (Brachman & Lemnios, 2002) to refer to the movement that pursues AI's original goals.

We can define cognitive systems as the research discipline that:

- *designs, constructs, and studies computational artifacts that exhibit the distinctive features 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.

# Status of the Movement

The cognitive systems movement is young, but it is engaging in a number of activities to encourage research:

- Holding an annual refereed conference
  - Arlington (11/2011), Palo Alto (12/2012), Baltimore (12/2013)
  - http://www.cogsys.org/conference/2013/
- Publishing a refereed, archival journal
  - Two volumes now published electronically
  - http://www.cogsys.org/journal/
- Organizing invited symposia on related topics
  - http://www.cogsys.org/symposium/2013/

The aim is to raise cognitive systems to an intellectual discipline that is both visible and vital.

# Science and Computation

Without doubt, scientific research is one of the most complex of human activities in that it:

- Examines some of the most complex phenomena
- Develops some of the most complex accounts
- Depends on some of the most complex social interactions

And this process is becoming ever more complicated, dealing with more data, more sophisticated models, and larger groups.

The daunting complexity of this enterprise suggests the need for computational assistance.

These features also make science a natural target for cognitive systems research.

# Historical Successes

This is not a new idea. digital computers have been used to aid the scientific process in many ways for decades, including:

- Computational encoding / simulation of models for complex phenomena
- Computer analysis of scientific data sets and discovery of new laws / relations
- Collection, storage, and management of scientific data sets and scientific knowledge
- Computational support for communication and interaction among scientists

Information technology has increasingly become a key tool for most scientific disciplines.

# Research on Computational Scientific Discovery (from 1979 to 2000)

<table>
<tr><td colspan="2">1979</td><td colspan="2">1980</td><td colspan="2">1981</td><td colspan="2">1982</td><td colspan="2">1983</td><td colspan="2">1984</td><td colspan="2">1985</td><td colspan="2">1986</td><td colspan="2">1987</td><td colspan="2">1988</td><td colspan="2">1989</td><td colspan="2">1990</td><td colspan="2">1991</td><td colspan="2">1992</td><td colspan="2">1993</td><td colspan="2">1994</td><td colspan="2">1995</td><td colspan="2">1996</td><td colspan="2">1997</td><td colspan="2">1998</td><td colspan="2">1999</td><td colspan="2">2000</td></tr>
<tr><td colspan="12">Bacon.1–Bacon.5</td><td colspan="4">Abacus, Coper</td><td colspan="8">Fahrehneit, E*, Tetrad, IDS$_N$</td><td colspan="4">Hume, ARC</td><td colspan="6">DST, GP$_N$ LaGrange</td><td colspan="4">SDS</td><td colspan="6">SSF, RF5, LaGramge</td></tr>
<tr><td colspan="5">←AM</td><td colspan="3"></td><td colspan="5">Glauber</td><td colspan="5">NGlauber</td><td colspan="4"></td><td colspan="4">IDS$_Q$, Live</td><td colspan="8"></td><td colspan="6">RL, Progol</td><td colspan="4">HR</td></tr>
<tr><td colspan="5">←Dendral</td><td colspan="3"></td><td colspan="6">Dalton, Stahl</td><td colspan="5">Stahlp, Revolver</td><td colspan="5">Gell-Mann</td><td colspan="4">BR-3, Mendel</td><td colspan="4">Pauli</td><td colspan="4">BR-4</td><td colspan="8"></td></tr>
<tr><td colspan="12"></td><td colspan="4">IE</td><td colspan="2"></td><td colspan="6">Coast, Phineas, AbE, Kekada</td><td colspan="4"></td><td colspan="12">Mechem, CDP</td><td colspan="4">Astra, GP$_M$</td></tr>
</table>

| *Legend* | Numeric laws | Qualitative laws | Structural models | Process models |
|---|---|---|---|---|

# Successes of Computational Scientific Discovery

Systems of this type have helped discover new knowledge in many scientific fields:

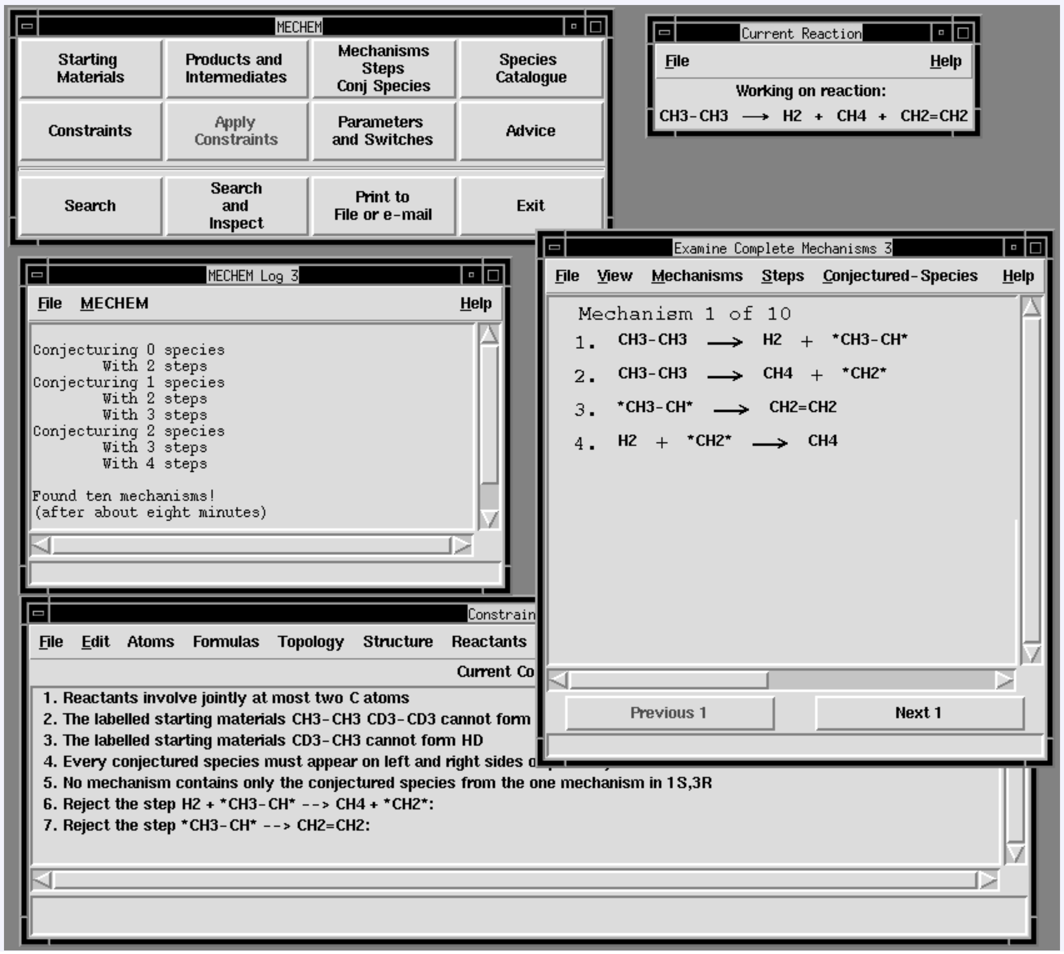
- stellar taxonomies from infrared spectra (Cheeseman et al., 1989)
- qualitative chemical factors in mutagenesis (King et al., 1996)
- quantitative laws of metallic behavior (Sleeman et al., 1997)
- quantitative conjectures in graph theory (Fajtlowicz et al., 1988)
- temporal laws of ecological behavior (Todorovski et al., 2000)
- reaction pathways in catalytic chemistry (Valdes-Perez, 1994)

Each of these has led to publications in the refereed literature of the relevant scientific field.

# The MECHEM Environment

MECHEM (Valdes-Perez, 1994) was an interactive system that generated plausible pathways to explain chemical reactions.

Users could access the software through a graphical interface.




This front end made MECHEM more accessible to chemists.

Using the system, Valdes-Perez and collaborators discovered many new chemical pathways.

A number of these led to peer-reviewed publications in the chemistry literature.

# Discovery Informatics

Despite many successes, each subarea has been isolated and has ignored aspects of the scientific enterprise.

There remains a need for broader computational research that attempts to:

- Understand, in computational terms, the representations and processes that underlie scientific research;
- Develop and study computational systems that embody these new understandings; and
- Apply these systems to specific scientific problems in order to support new research.

We will refer to this group of activities as *discovery informatics* because they address the overall context of discovery.

# What About Big Data?

Does the recent excitement about ‘data-intensive science’ and ‘big data’ make other aspects of science irrelevant?

Definitely not; science is becoming more complicated along four different dimensions:

- Larger *data sets* (although not yet in all fields)
- Larger *models* to visualize, reason over, and construct
- Larger *problem spaces* in which to search for models
- Larger *groups* of scientists in collaborative teams

A well-balanced field of discovery informatics should explore computational responses to each of these challenges.

The cognitive systems paradigm offers insights for each case.

# Questions for Discovery Informatics

We have organized this meeting to answer some key questions:

- *What structured representations can encode scientific phenomena, models, hypotheses, and models? How do representations aid processing?*

# Questions for Discovery Informatics

We have organized this meeting to answer some key questions:

- What structured representations can encode scientific phenomena, models, hypotheses, and models? How do representations aid processing?
- *What role does multi-step reasoning play in the generation of scientific predictions and explanations? Does this involve deduction or abduction?*

# Questions for Discovery Informatics

We have organized this meeting to answer some key questions:

- What structured representations can encode scientific phenomena, models, hypotheses, and models? How do representations aid processing?
- What role does multi-step reasoning play in the generation of scientific predictions and explanations? Does this involve deduction or abduction?
- *What role does problem solving play in the construction and revision of scientific hypotheses and models, and what heuristics guide it?*

# Questions for Discovery Informatics

We have organized this meeting to answer some key questions:

- What structured representations can encode scientific phenomena, models, hypotheses, and models? How do representations aid processing?
- What role does multi-step reasoning play in the generation of scientific predictions and explanations? Does this involve deduction or abduction?
- What role does problem solving play in the construction and revision of scientific hypotheses and models, and what heuristics guide it?
- *What major activities make up the scientific process and how do they interact to advance acquisition of data and generation of models?*

# Questions for Discovery Informatics

We have organized this meeting to answer some key questions:

- What structured representations can encode scientific phenomena, models, hypotheses, and models? How do representations aid processing?
- What role does multi-step reasoning play in the generation of scientific predictions and explanations? Does this involve deduction or abduction?
- What role does problem solving play in the construction and revision of scientific hypotheses and models, and what heuristics guide it?
- What major activities make up the scientific process and how do they interact to advance acquisition of data and generation of models?
- *What lessons do the history of science and cognitive psychology offer about representations and mechanisms that underlie scientific research?*

# Questions for Discovery Informatics

We have organized this meeting to answer some key questions:

- What structured representations can encode scientific phenomena, models, hypotheses, and models? How do representations aid processing?
- What role does multi-step reasoning play in the generation of scientific predictions and explanations? Does this involve deduction or abduction?
- What role does problem solving play in the construction and revision of scientific hypotheses and models, and what heuristics guide it?
- What major activities make up the scientific process and how do they interact to advance acquisition of data and generation of models?
- What lessons do the history of science and cognitive psychology offer about representations and mechanisms that underlie scientific research?
- *What computational abstractions recur across scientific disciplines despite different types of phenomena, formalisms, and content?*

The cognitive systems paradigm offers a natural framework in which to pursue these issues.

# Symposium Schedule

| | **Friday, June 21** | **Saturday, June 22** |
|---|---|---|
| 8:35 AM | Continental breakfast | Continental breakfast |
| 9:00 AM | Session 1 (two talks) | Session 1 (two talks) |
| 10:30 AM | *Morning break* | *Morning break* |
| 11:00 AM | Session 2 (two talks) | Session 2 (two talks) |
| 12:30 PM | *Lunch (provided)* | *Lunch (provided)* |
| 2:00 PM | Session 3 (two talks) | Session 3 (two talks) |
| 3:30 PM | *Afternoon break* | *Afternoon break* |
| 4:00 PM | Session 4 (two talks) | Session 4 (two talks) |
| 5:30 PM | *Open discussion* | *Closing discussion* |
| 6:00 PM | Reception / Buffet dinner | Symposium ends |

**Web site:** *http://www.cogsys.org/symposium/2013/*

**Wireless Network:** CMUSV

**Wireless Password:** None

**Talks:** 35 minutes for presentations, 10 minutes for questions

**Restrooms:** Outside doors on the left, others on second floor

**Etiquette:** Please take trash with you and please **recycle**

End of Presentation