On Some Cognitive Robotics @ RPI

Selmer Bringsjord

Rensselaer AI & Reasoning (RAIR) Laboratory
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 US

@ Schenectady Museum 10.11.07

Rensselaer | DEPARTMENT OF COGNITIVE SCIENCE

Rensselaer Computer Science

The Rensselaer AI & Reasoning (RAIR) Lab

RAIR Lab Method

RAIR Lab Method

- Isolate and dissect human ingenuity.
  - Hence the centrality of cognitive science.

RAIR Lab Method

- Isolate and dissect human ingenuity.
  - Hence the centrality of cognitive science.
- Formalize weak correlate to this ingenuity in some advanced logical system.

RAIR Lab Method

- Isolate and dissect human ingenuity.
    - Hence the centrality of cognitive science.
- Formalize weak correlate to this ingenuity in some advanced logical system.
- Implement correlate in working computer program.

RAIR Lab Method

- Isolate and dissect human ingenuity.
  - Hence the centrality of cognitive science.
- Formalize weak correlate to this ingenuity in some advanced logical system.
- Implement correlate in working computer program.
- Augment this software as needed with machine-specific power (e.g., supercomputing).

RAIR Lab Method

- Isolate and dissect human ingenuity.
  - Hence the centrality of cognitive science.
- Formalize weak correlate to this ingenuity in some advanced logical system.
- Implement correlate in working computer program.
- Augment this software as needed with machine-specific power (e.g., supercomputing).
- Empower human by delivering software.

Toward a General Logicist Methodology for Engineering Ethically Correct Robots

Selmer Bringsjord, Konstantine Arkoudas, and Paul Bello,
Rensselaer Polytechnic Institute

A deontic logic formalizes a moral code, allowing ethicists to render theories and dilemmas in declarative form for analysis. It offers a way for human overseers to constrain robot behavior in ethically sensitive environments.

As intelligent machines assume an increasingly prominent role in our lives, there seems little doubt they will eventually be called on to make important, ethically charged decisions. For example, we expect hospitals to deploy robots that can administer medications, carry out tests, perform surgery, and so on, supported by software agents, or softbots, that will manage related data. (Our discussion of ethical robots extends to all artificial agents, embodied or not.) Consider also that robots are already finding their way to the battlefield, where many of their potential actions could inflict harm that is ethically impermissible.

How can we ensure that such robots will always behave in an ethically correct manner? How can we know ahead of time, via rationales expressed in clear natural languages, that their behavior will be constrained specifically by the ethical codes affirmed by human overseers? Pessimists have claimed that the answer to these questions is: "We can't!" For example, Sun Microsystems' cofounder and former chief scientist, Bill Joy, published a highly influential argument for this answer.[1] Inevitably, according to the pessimists, AI will produce robots that have tremendous power and behave immorally. These predictions certainly have some traction, particularly among a public that pays good money to see such dark films as Stanley Kubrick's *2001* and his joint venture with Stephen Spielberg, *AI*).

Nonetheless, we're optimists: we think formal logic offers a way to preclude doomsday scenarios of malicious robots taking over the world. Faced with the challenge of engineering ethically correct robots, we propose a logic-based approach (see the related sidebar). We've successfully implemented and demonstrated this approach.[2] We present it here in a general methodology to answer the ethical questions that arise in entrusting robots with more and more of our welfare.

Deontic logics: Formalizing ethical codes

Our answer to the questions of how to ensure ethically correct robot behavior is, in brief, to insist that robots only perform actions that can be proved ethically permissible in a human-selected *deontic logic*. A deontic logic formalizes an ethical code—that is, a collection of ethical rules and principles. Isaac Asimov introduced a simple (but subtle) ethical code in his famous Three Laws of Robotics:[3]

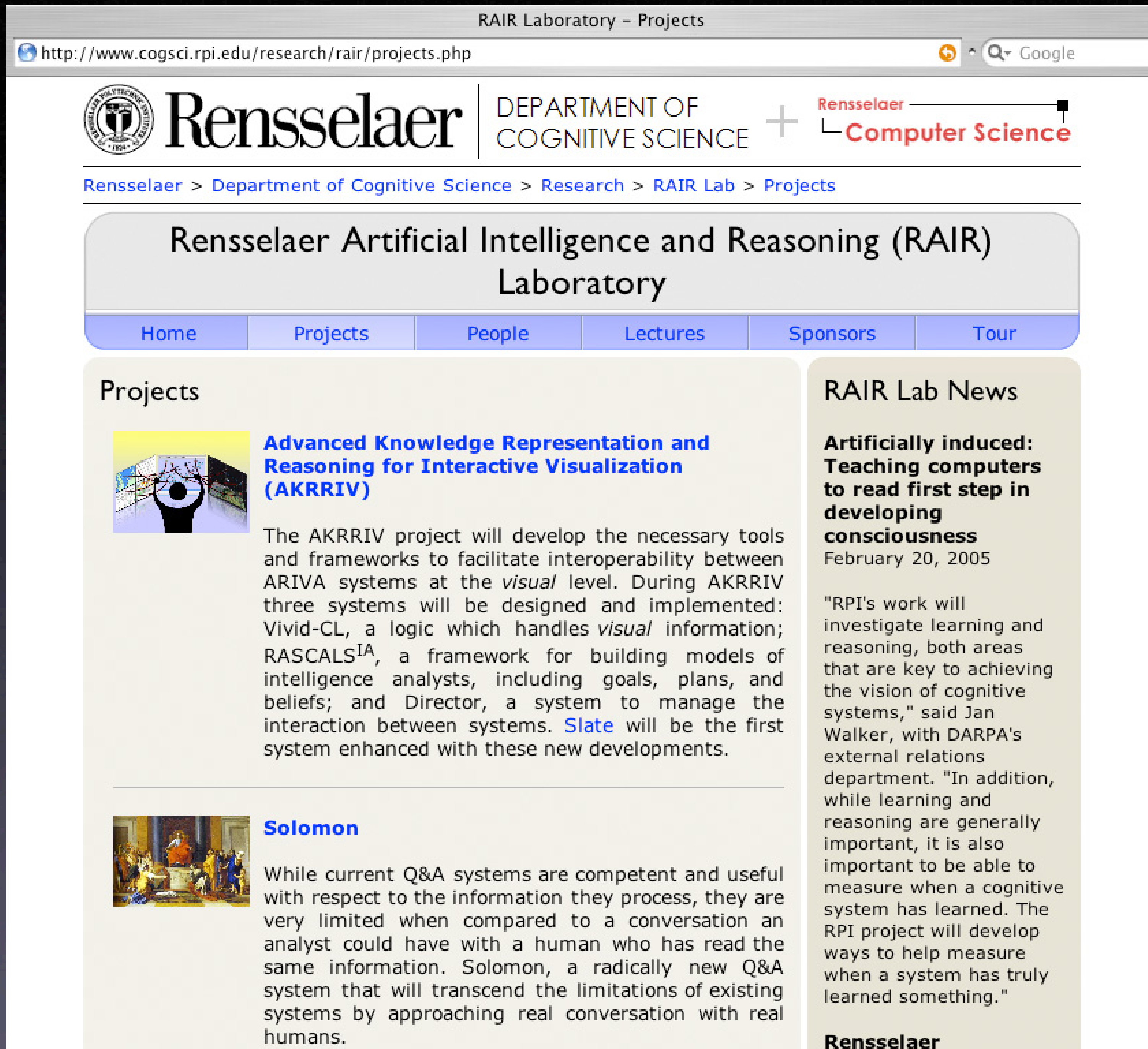
1. A robot may not harm a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law.

Human beings often view ethical theories, principles, and codes informally, but intelligent machines require a greater degree of precision. At present, and for the foreseeable future, machines can't work directly with natural language, so we can't simply feed Asimov's three laws to a robot and instruct it behave in

1541-1672/06/$20.00 © 2006 IEEE
Published by the IEEE Computer Society

http://www.cogsci.rpi.edu/research/rair

http://www.cogsci.rpi.edu/research/rair/projects.php



RAIR Laboratory – Projects

http://www.cogsci.rpi.edu/research/rair/projects.php

Rensselaer | DEPARTMENT OF COGNITIVE SCIENCE + Rensselaer Computer Science

Rensselaer > Department of Cognitive Science > Research > RAIR Lab > Projects

Rensselaer Artificial Intelligence and Reasoning (RAIR) Laboratory

Home | Projects | People | Lectures | Sponsors | Tour

Projects

Advanced Knowledge Representation and Reasoning for Interactive Visualization (AKRRIV)

The AKRRIV project will develop the necessary tools and frameworks to facilitate interoperability between ARIVA systems at the *visual* level. During AKRRIV three systems will be designed and implemented: Vivid-CL, a logic which handles *visual* information; RASCALS^{IA}, a framework for building models of intelligence analysts, including goals, plans, and beliefs; and Director, a system to manage the interaction between systems. Slate will be the first system enhanced with these new developments.

Solomon

While current Q&A systems are competent and useful with respect to the information they process, they are very limited when compared to a conversation an analyst could have with a human who has read the same information. Solomon, a radically new Q&A system that will transcend the limitations of existing systems by approaching real conversation with real humans.

RAIR Lab News

Artificially induced: Teaching computers to read first step in developing consciousness
February 20, 2005

"RPI's work will investigate learning and reasoning, both areas that are key to achieving the vision of cognitive systems," said Jan Walker, with DARPA's external relations department. "In addition, while learning and reasoning are generally important, it is also important to be able to measure when a cognitive system has learned. The RPI project will develop ways to help measure when a system has truly learned something."

Rensselaer

Psychometric Artifical Intelligence and PERI

http://www.cogsci.rpi.edu/research/rair/pai/

SUNY System...ss – Logon   Ultra Hardw...re Products   Wolfe's Bald...rized Dealer   Screen Door ... Von Morris

Rensselaer | DEPARTMENT OF COGNITIVE SCIENCE + Rensselaer Computer Science

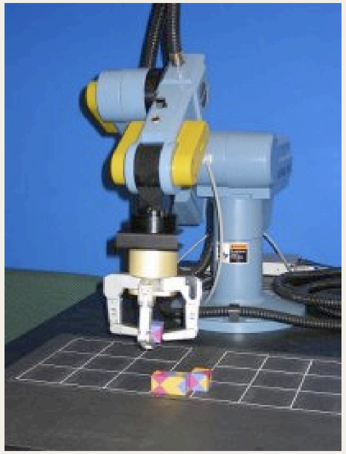
Rensselaer > Department of Cognitive Science > Research > RAIR Lab > Projects > Psychometric AI

Rensselaer Artificial Intelligence and Reasoning (RAIR) Laboratory

Home | Projects | People | Lectures | Symposia/Conferences | Sponsors | Tour

Psychometric Artificial Intelligence and PERI

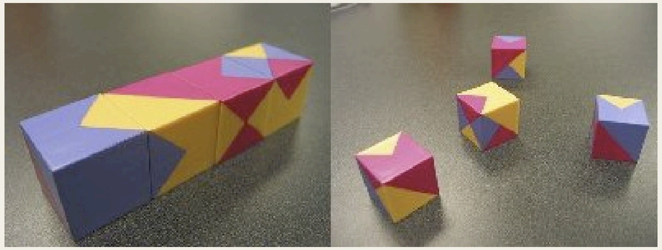
Abstract(+) from the overview paper:



We propose an answer to the "What is AI?" question, namely, that AI is really (or at least really ought in significant part to be) Psychometric AI (PAI). Psychometric AI is the field devoted to building information processing entities capable of at least solid performance on all established, validated tests of intelligence and mental ability, a class of tests that includes IQ tests, tests of reasoning, of creativity, mechanical ability, and so on. Along the way, we: set out and rebut some objections to PAI; describe PERI, a robot in our lab who exemplifies PAI; and briefly treat the future of Psychometric AI, first by pointing toward some promising PAI-based applications, and then by raising some of the "big" philosophical questions the success of Psychometric AI will raise.

Beginnings of PAI:

We have begun our research with the WAIS-R (Wechsler Adult Intelligent Scale - Revised) and the first task of this test has already been surpassed successfully. For reasons of legality we cannot mention the specifics of this subtest, the Block Design Task, but we discuss another similar puzzle which PERI can also solve successfully. This puzzle (shown in snapshots below; compliments of the Binary Arts Corp) is described in more detail in our IJCAI overview paper (see above).



PAI/PERI Project Links

- Psychometric AI Home
- Recent News
- Presentations and Demos
- Publications
- Media Coverage
- Online Library
- Restricted Content

PAI/PERI Project Team

- Selmer Bringsjord
- Bettina Schimanski
- Gabriel Mulley

RAIR Laboratory – Sponsors

http://www.cogsci.rpi.edu/research/rair/sponsors.php

Ultra Hardw...re Products   Wolfe's Bald...rized Dealer   Screen Door ... Von Morris   Screen Door ... Von Morris   Apple   Amazon

Rensselaer | DEPARTMENT OF COGNITIVE SCIENCE + Rensselaer Computer Science

Rensselaer > Department of Cognitive Science > Research > RAIR Lab > Sponsors

Rensselaer Artificial Intelligence and Reasoning (RAIR) Laboratory

Home | Projects | People | Lectures | Symposia/Conferences | Sponsors | Tour

Sponsors

DARPA
Defense Advanced Research Projects Agency

SAIC
Science Applications International Corporation

ARDA
Advanced Research and Development Activity

NSF
National Science Foundation

AFRL
Air Force Research Laboratory

RAIR Lab News

Selmer Bringsjord receives NSF Award
July 20, 2007

Selmer Bringsjord, Nick Webb and other researchers received a National Science Foundation award to research Social Robotics. The Team proposes to use Social Robotics as a mechanism to deliver a revitalized Computer Science (CS) education.

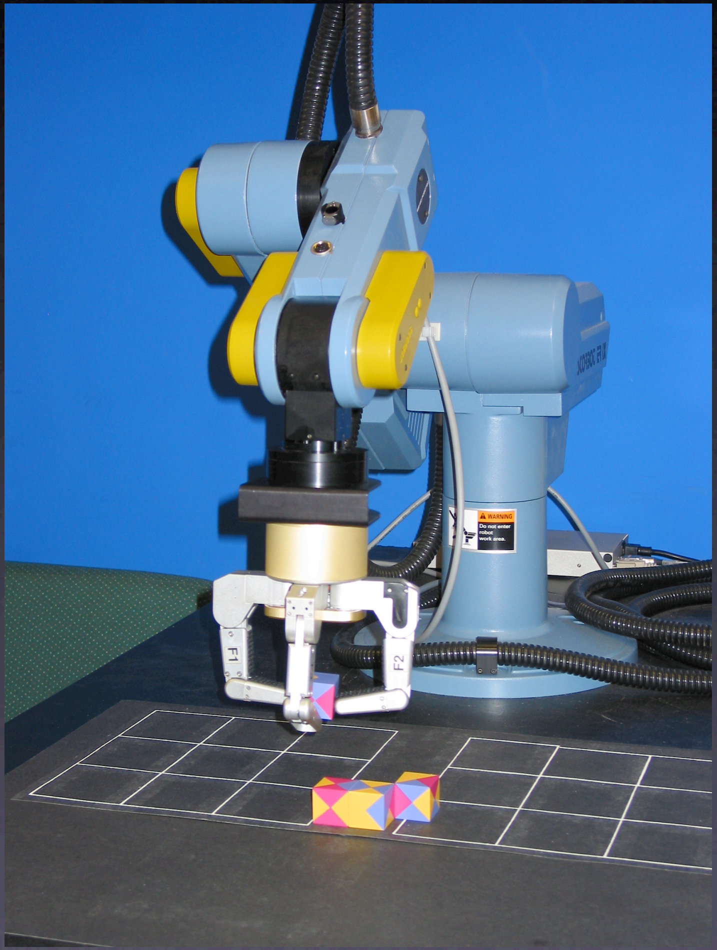
"Provability-Based Semantic Interoperability via Translation Graphs" Presentation
July 26, 2007

RAIR Lab researchers will head to New Zealand in November to present Provability-Based Semantic Interoperability via Translation Graphs at the International Workshop on Ontologies and Information Systems for the Semantic Web (ONISW 2007). The ONISW2007 paper

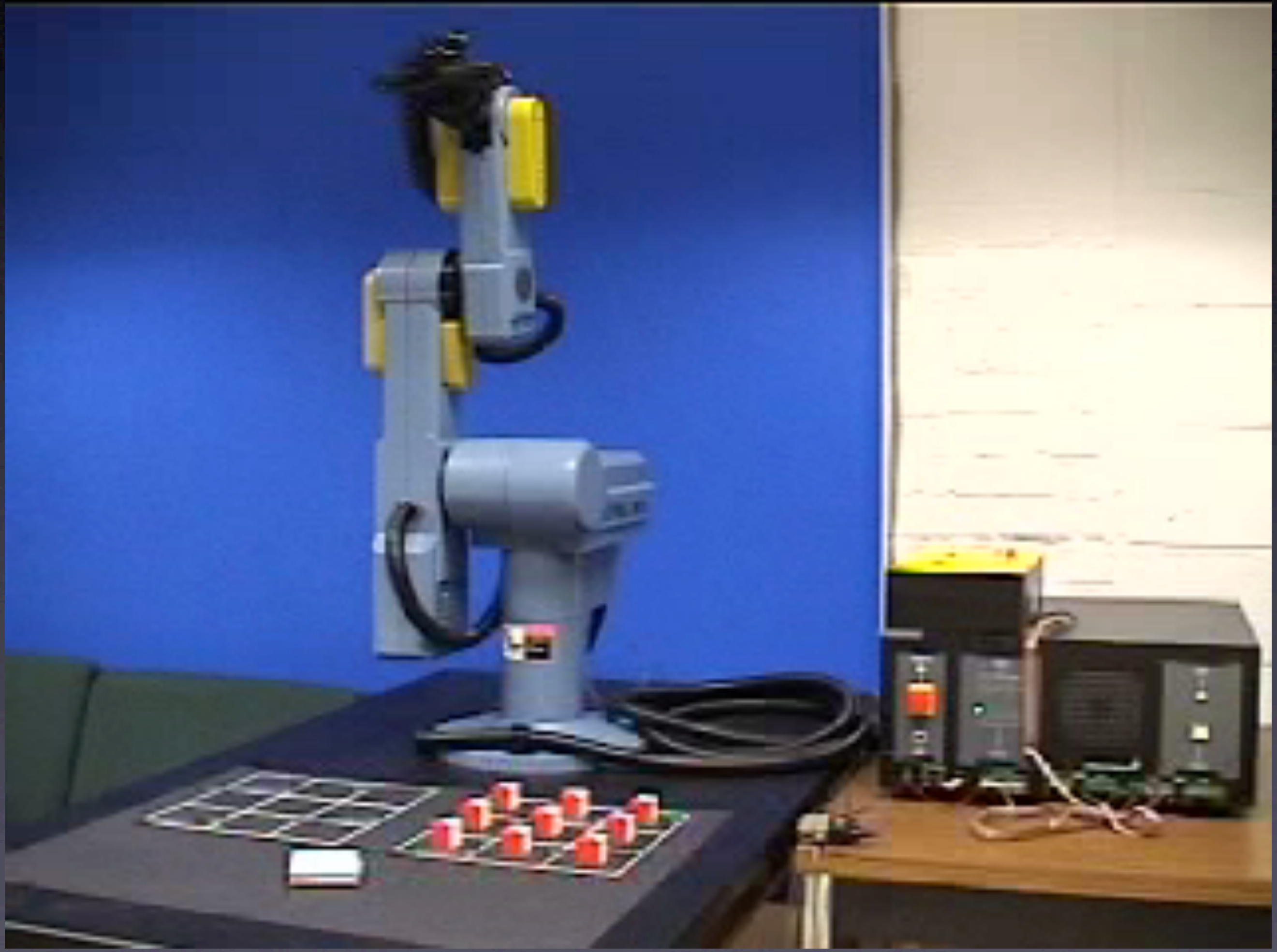
PERI

Pscyhometric Experimental Robotic Intelligence

- Scorbot-ER IX
- Sony B&W XC55 Video Camera
- Cognex MVS-8100M Frame Grabber
- Dragon Naturally Speaking Software
- NL (Carmel & RealPro?)
- BH8-260 BarrettHand Dexterous 3-Finger Grasper System



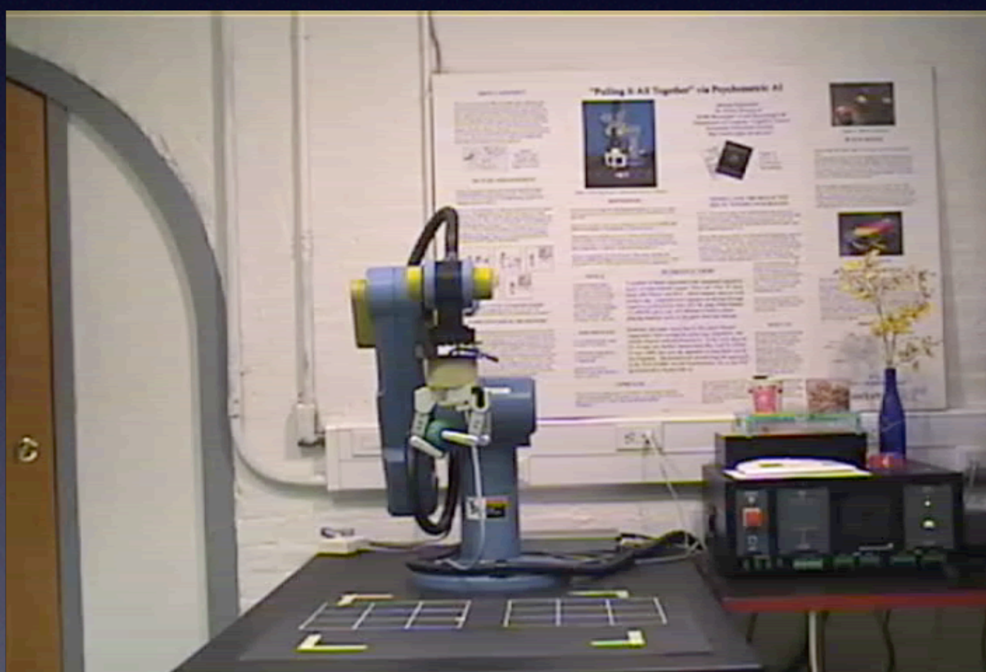
PERI "Cracked" Block Design*

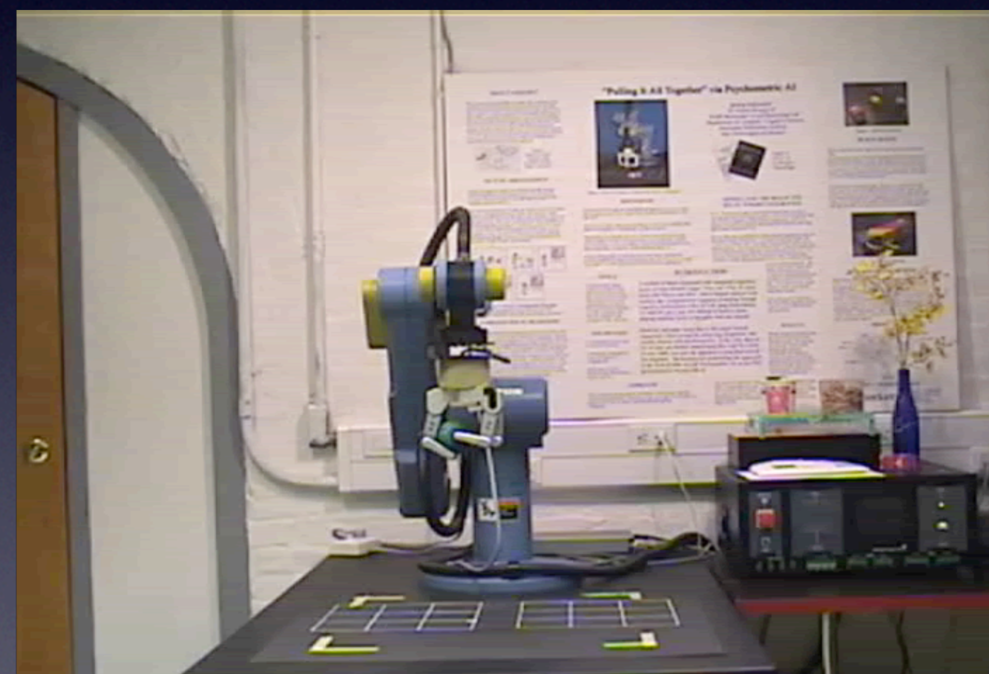


*With much help from Sandia Labs' Bettina Schimanski.

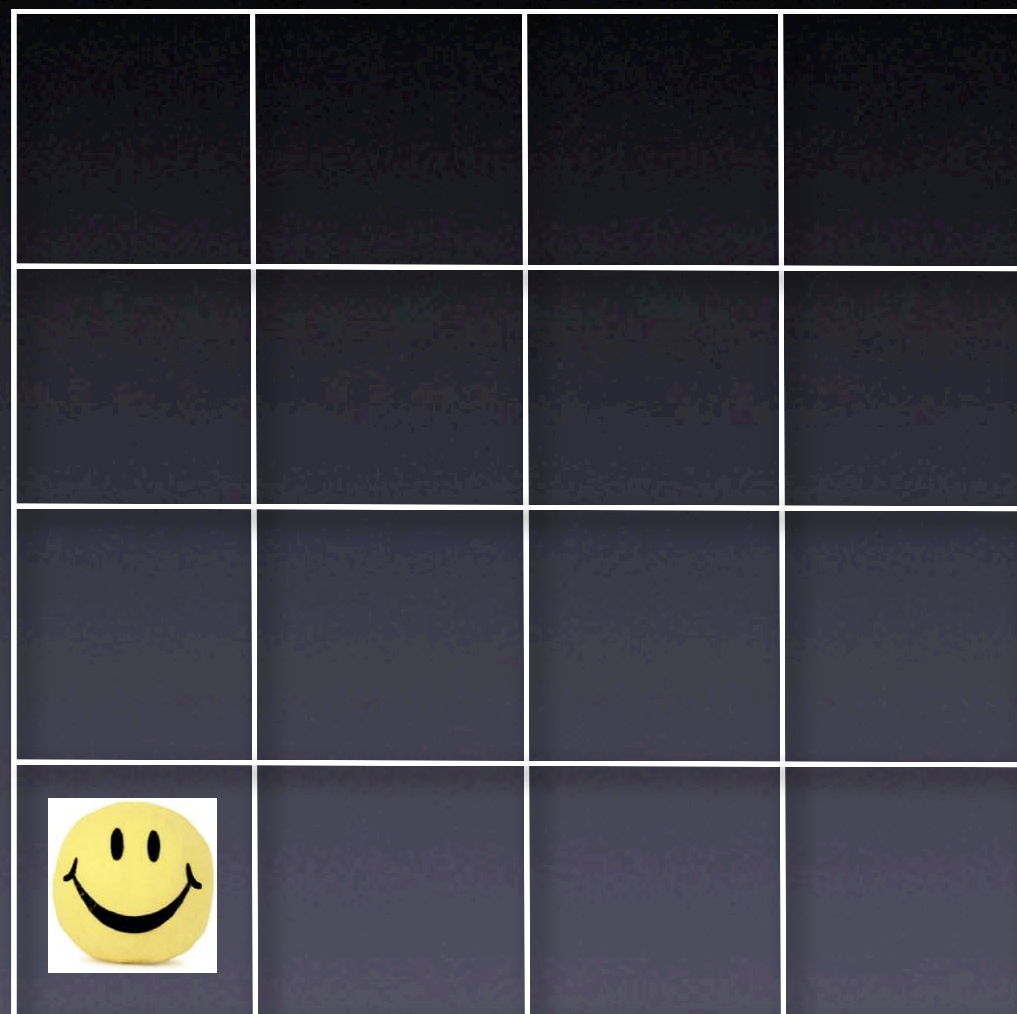
```
(defun peris-choice ()
       (cond ((> (random 10) 5) (hold-earth))
             ((drop-earth))))
```

```
? (peris-choice)
"I will drop earth"

? (peris-choice)
"I will hold onto earth"

? (peris-choice)
"I will hold onto earth"
```



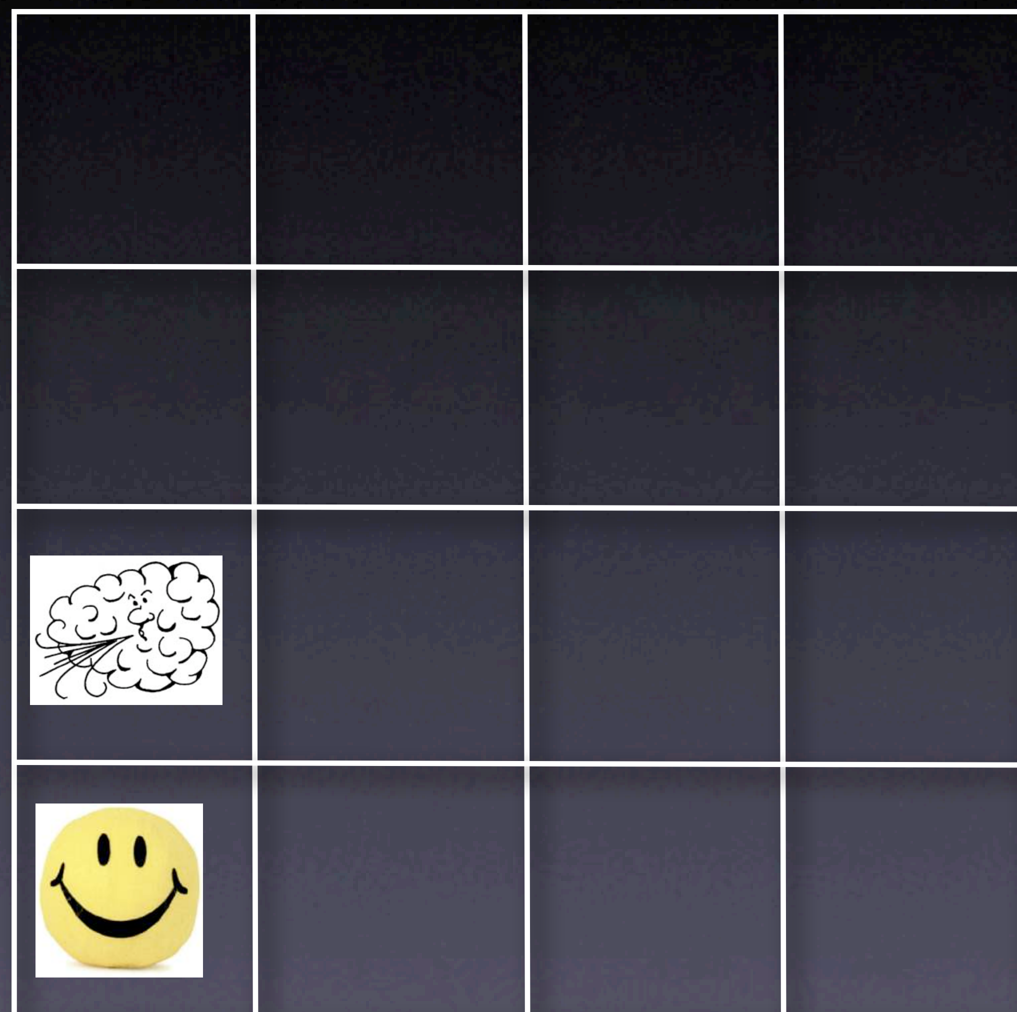


Hunt the Wumpus

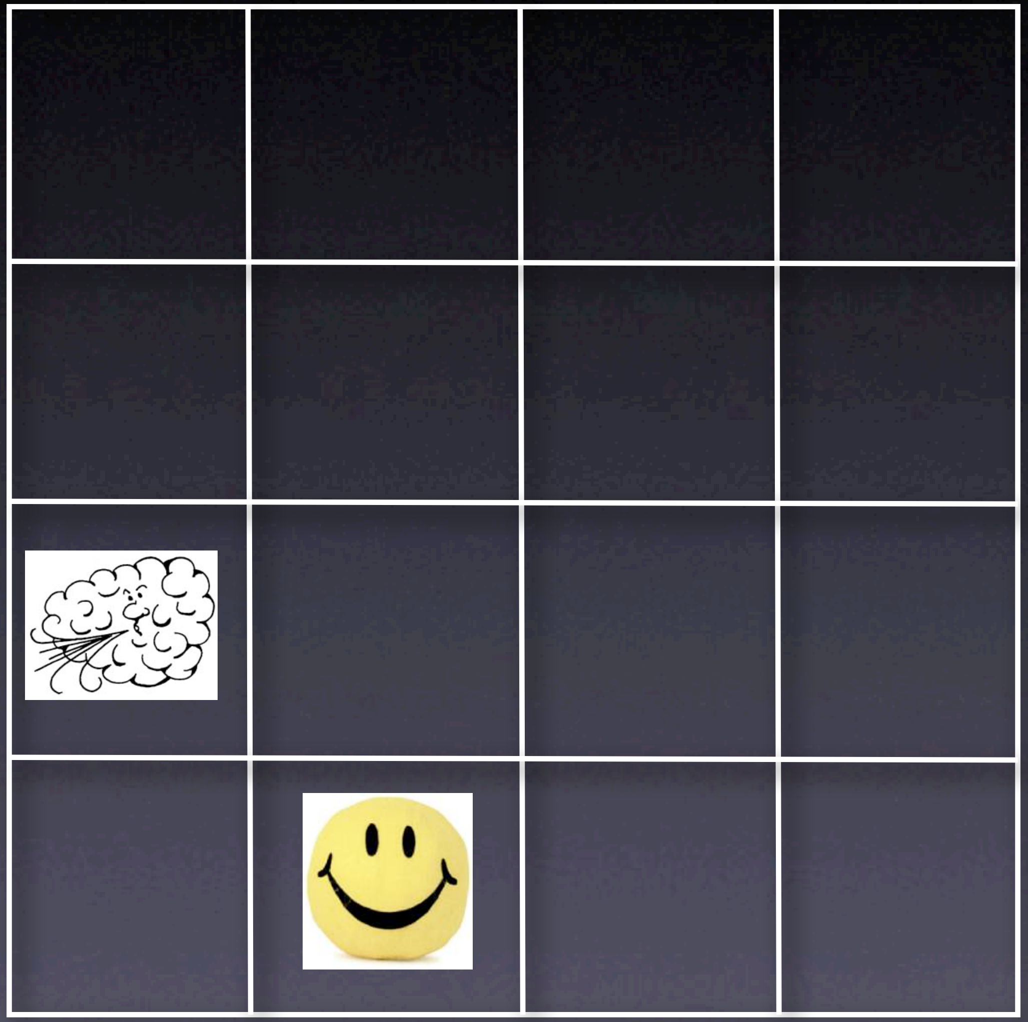


Hunt the Wumpus

Hunt the Wumpus

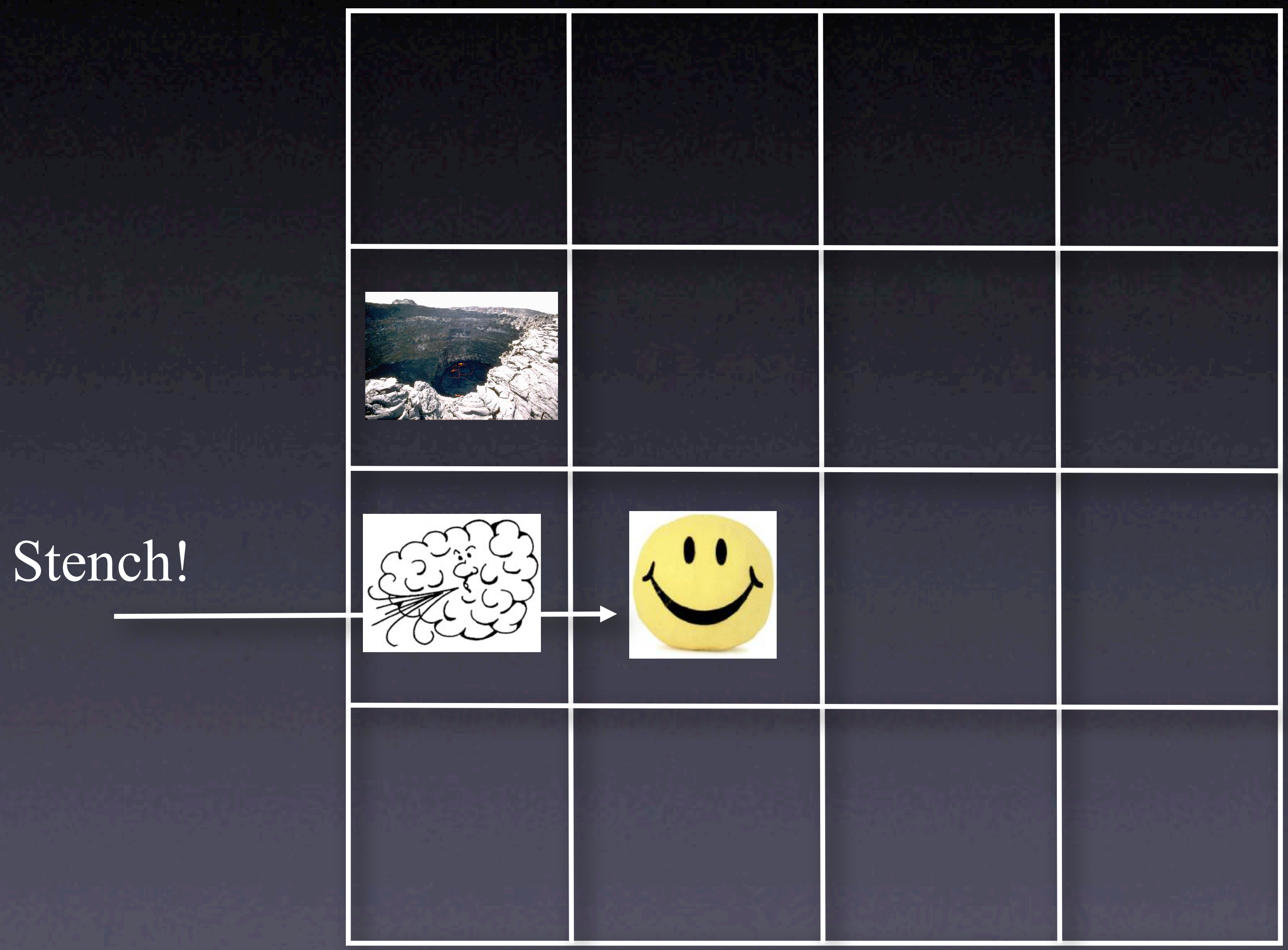


Hunt the Wumpus

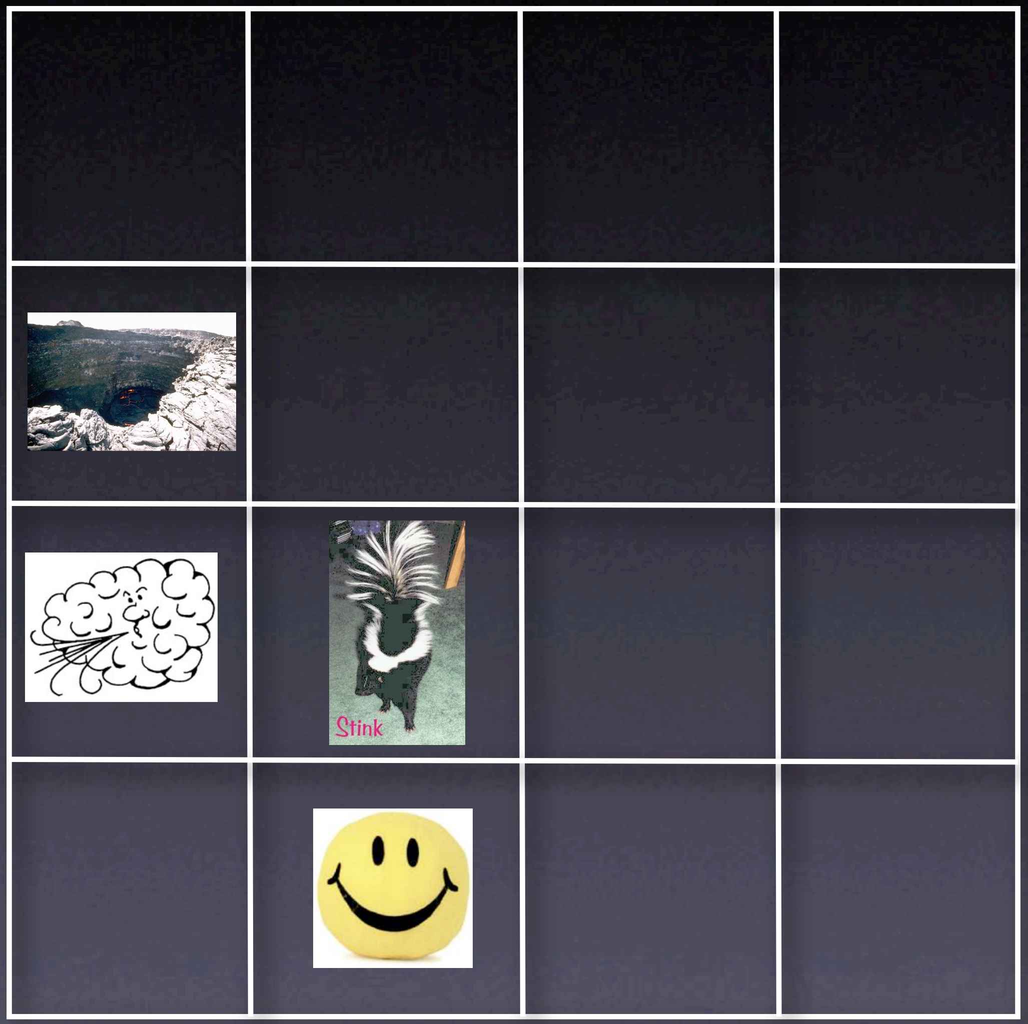


Hunt the Wumpus

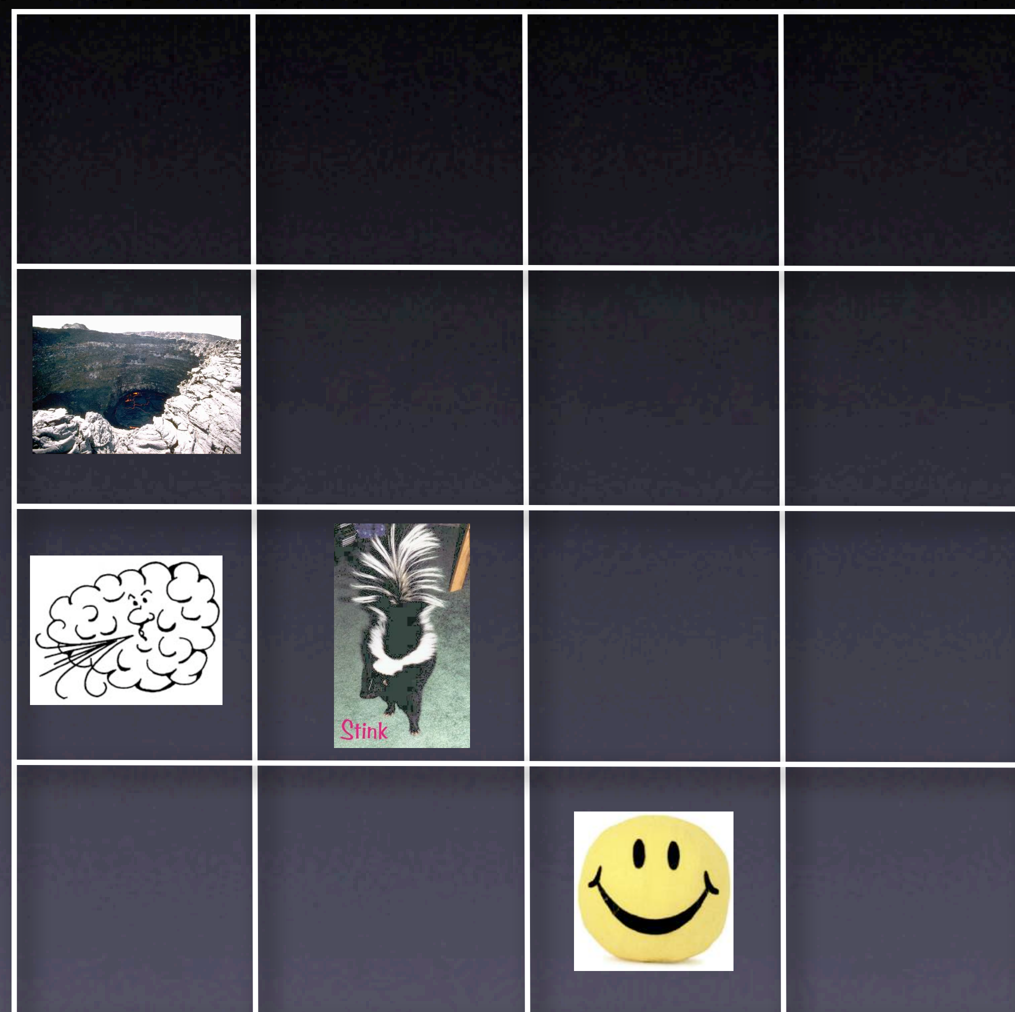
Hunt the Wumpus



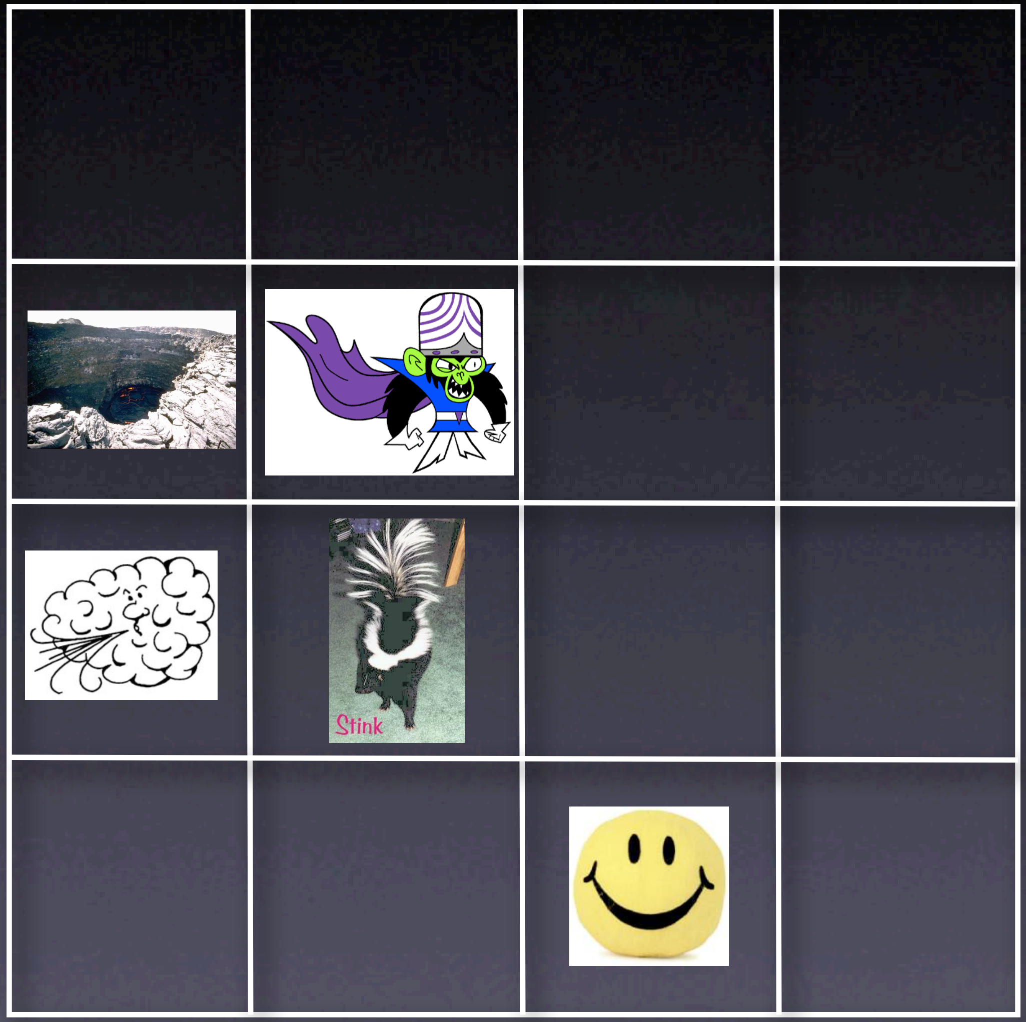
Hunt the Wumpus



Hunt the Wumpus



Hunt the Wumpus



Hunt the Wumpus

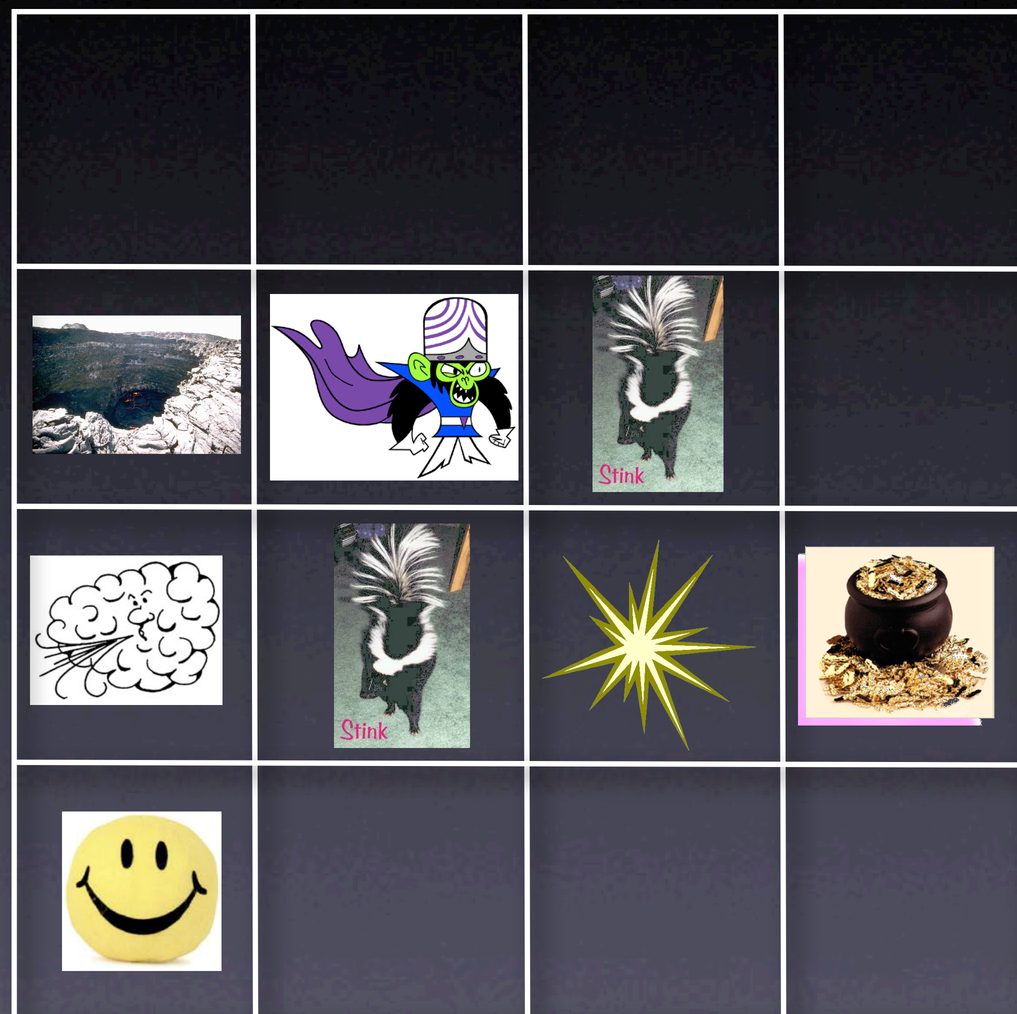
Hunt the Wumpus

Stench!

Hunt the Wumpus

Hunt the Wumpus

Hunt the Wumpus



Situation Calculus

The situation calculus in the following layers works as follows:

The result function takes in a list of actions and returns a list representing a location

- There is a general definition of the result function, which SNARK uses to build up sequences of actions, and it is defined as:

```
(= (result ?actions (result (list ?action) square))
   (result (append (list ?action) ?actions) square))
```

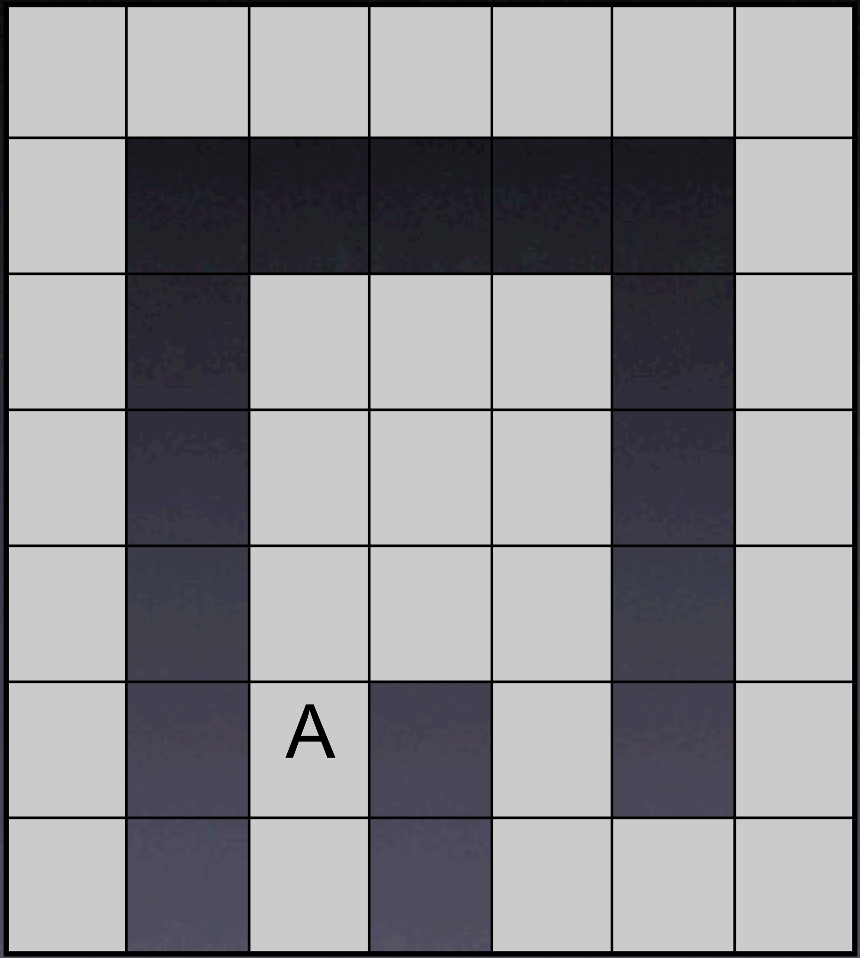
- Results for single actions are then defined – in this case since the theory is used to plan a path consisting of visited squares, the result of a single action on a square is only defined if that square is visited, e.g. if the action is 'up, the result function returns the square above it:

```
(= (result (list up) (list 1 0)) (list 1 1))
```

- SNARK is used to prove there is a list of actions that the result function takes in and performs on the agent's current location and returns the location of interest
  - For example if the agent is at (2,1) and wants to get to (0,0), SNARK would generate, assuming the appropriate squares are visited, (list down left left)
  - i.e. (= (result ?actions (list 2 1)) (list 0 0))
    - (= (result ?actions (result (list ?action) (list 2 1))) (list 0 0)) //general-result-defn
    - (= (result ?actions (list 2 0)) (list 0 0)) //result-of-down
    - (= (result ?actions (result (list ?action) (list 2 0))) (list 0 0)) //general-result-defn
    - (= (result ?actions (list 1 0)) (list 0 0)) //result-of-left
    - at this point 'left solves it, and SNARK has remembered the list up to this point, so the answer is (list down left left)

Simulation Performance

- In the world situation on the right, it takes SNARK 2 seconds to generate `(LIST UP RIGHT RIGHT DOWN DOWN RIGHT RIGHT UP UP UP UP UP UP LEFT LEFT LEFT LEFT LEFT LEFT DOWN DOWN DOWN DOWN DOWN DOWN)` as a solution for the home layer – that's 2 seconds for a list of length 25
- Before much needed efficiency enhancements and some slight theory adjustments, this proof would have taken well over a day
- This shows the need for careful, terse theory and taking full advantage of all appropriate efficiency options in SNARK

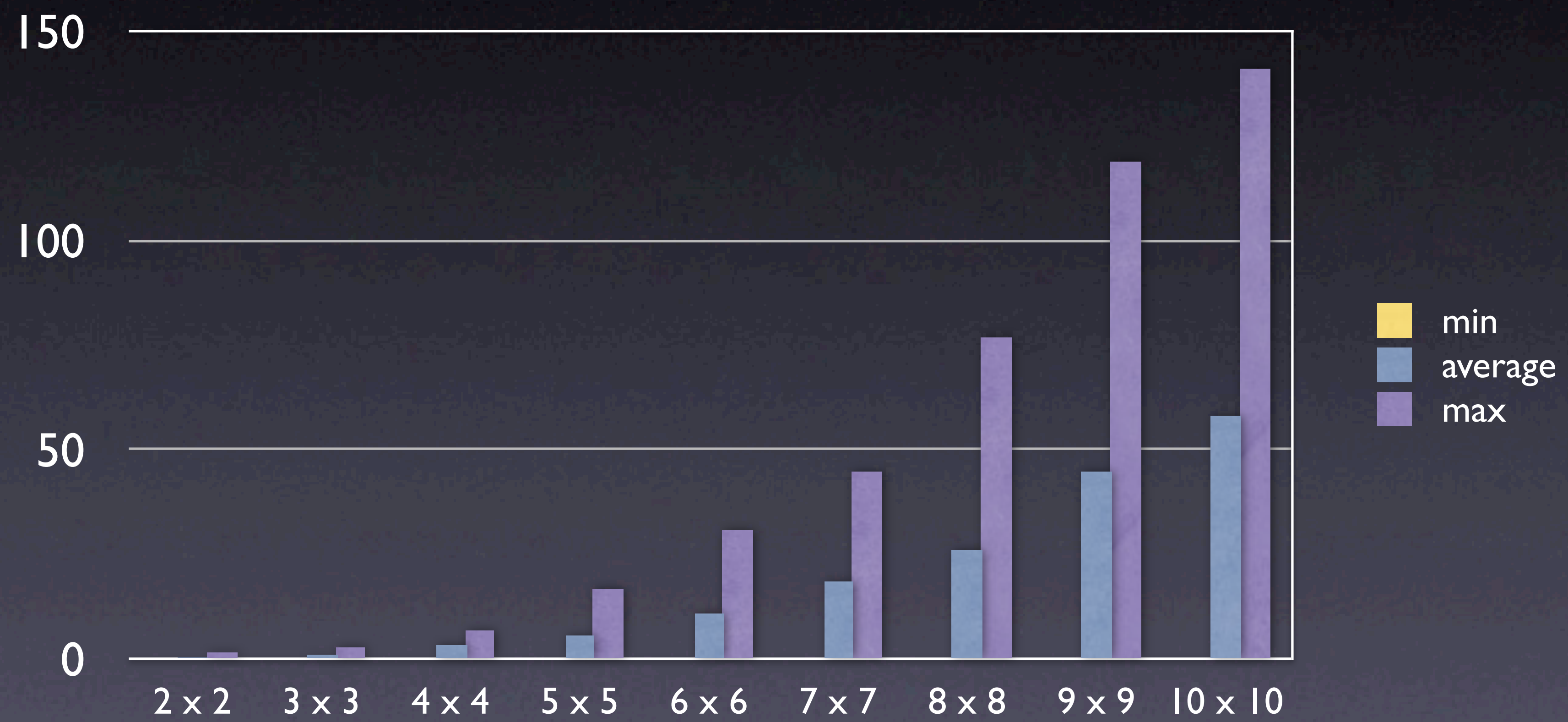


- Here, a gray square represents a visited square and A represents the agent's location

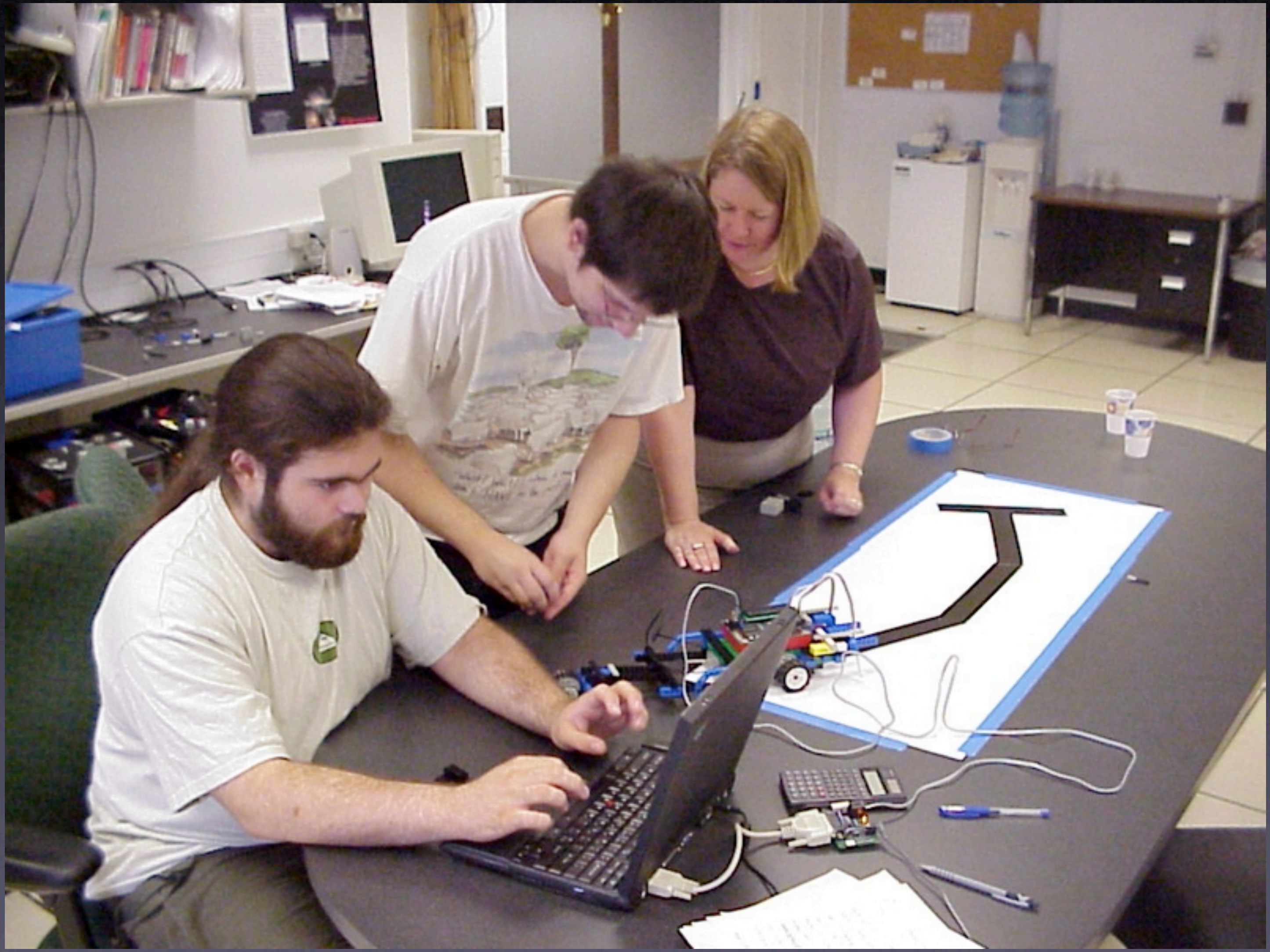
Simulation Performance : Type 1

- The average run times for successful runs are:
  - For a 2 by 2 map, the average was 0.27 seconds
  - For a 3 by 3 map, the average was 1 second
  - For a 4 by 4 map, the average was 3.25 seconds
  - For a 5 by 5 map, the average was 5.82 seconds
  - For a 6 by 6 map, the average was 11.02 seconds
  - For a 7 by 7 map, the average was 18.67 seconds
  - For an 8 by 8 map, the average was 25.99 seconds
  - For a 9 by 9 map, the average was 45 seconds
  - For a 10 by 10 map, the average was 58.25 seconds
- The next slide shows minimum, maximum, and average run times for successful runs for different map sizes of Type I – minimum times were typically 0 seconds (approximately), which occurred when the gold was in the start square, and maximum times typically occurred when the agent found the gold after exploring almost the entire map

Simulation Performance : Type I



Wumpus World Competition



Video of Marc Controlling Robot

The End