Mr. David Gunning
Program Manager
Information Processing Technology Office

Learning and Reasoning:
The True Heart of the Mind

If Captain Kirk and Mr. Spock depended on today's computers, they and the crew of the Enterprise would have died a horrible death on some distant planet long, long ago. They would not have lived to see the third episode.

For example, here is a typical dialog with the deceptively simple but incredibly intelligent StarTrek computer:

Mr. Spock: "Computer, analyze the planet."

Computer: "Planet Alpha 11. Class M. Atmosphere, oxygen-nitrogen. Categorization of life form readings recorded."

This imagined computer effortlessly understands the question, fuses a complex array of sensor data, classifies the planet, and categorizes its life forms—all in a few seconds.

Now let's look at an example of some of today's smartest computer programs, such as they are. There is an annual competition called the Loebner Prize, in which computer programs try to pass the Turing Test. Human judges question computer programs, along with human entrants. The judges try to identify which subjects are human and which are computers.

So far, no computer has come close to fooling anyone, but the results illustrate just how far we have to go in our quest to create intelligent machines. Here is a transcript from the winner of the 2004 Loebner Prize, a program named Alice:

Judge 1: "What was the weather like today?"

Alice: "Ah. I can think of more than one weather like."

Judge 1: "Is it raining?"

Alice: "Light rain or heavy rain?"

Judge 1: "I'm asking you?"

Alice: "OK you can ask me."

Like much of today's software, Alice lacks the reasoning ability of the even simplest life form, on any planet.

To be fair, Alice performs so poorly because the test requires it to converse on any topic. If we narrowed the domain to only weather, we could build, and have built, dialog systems that could do much better. But, we continue to hit a barrier when it comes to achieving general purpose, human-like, machine intelligence. We have hit the barrier of machine stupidity.

This barrier prevents us from creating the kind of software we need for future, high-tech, military operations. Increasingly our military depends on a complex array of sensors, information systems, smart weapons, autonomous machines, and high-precision tactics. We couldn't imagine the Starship Enterprise going into battle without the intelligent software it has embedded in its onboard computer, tricorders, life-form scanners, and transporters. Increasingly, we won't be able to imagine our real-world military going to war without the same kind of intelligent software.

Our vision, in IPTO, is to break through this barrier and create cognitive systems that can genuinely reason and learn. In short, we want to create systems with human-like intelligence.

This is a certainly a very hard and very long-term problem. Humankind has strived for centuries to

Learning and Reasoning: The True Heart of the Mind

understand human reasoning and cognition. In the 3rd century BC, Aristotle, in an effort to make human thought more precise, invented logic. He posited a set of basic categories and rules for drawing inferences about those categories. In the 17th century, Leibniz was one of the first to imagine that this logic could be mechanized into a thinking machine. In the 19th century, Frege formalized the rules of logic into Predicate Calculus. In the 1950s, Robinson devised the basic techniques for Resolution-based Theorem Proving that turned Frege's Predicate Calculus into computer-executable algorithms.

Today, we can build large, complex knowledge bases and reasoning systems. We have systems like Cyc that turned Aristotle's categories into ontologies of 100,000 concepts and Frege's Predicate Calculus into efficient reasoning procedures over millions of axioms. Cyc could easily encode and reason about all the knowledge the Enterprise might need regarding classes of planets, life forms, alien cultures, and their weapons.

Logic-based systems of today are quite powerful and have achieved Aristotle's goal of making reasoning both mechanical and precise. Too much so, in fact. The principal weakness of these systems is that they break when faced with uncertainty or ambiguity. Cyc would not do well at interpreting the sensor data from Planet Alpha 11. Its black and white logic does not lend itself to handling the noise inherent in the sensor signal. Logic-based systems break when reasoning requires imprecision.

However, these are not the only reasoning techniques we have available. On a slightly different track, in the 17th century, Pascal and Fermat began to develop the mathematics of probability to better understand games of dice. In the 1980s, Judah Pearl developed a formal calculus for reasoning with conditional probabilities in the form of Bayesian Belief Networks.

This has resulted in a rich technology of probabilistic reasoning systems that can handle uncertainty and ambiguity. Today, we can build 50,000 nodes Bayesian Belief Nets. If we knew enough about what senor signals to expect, we could build probabilistic systems to interpret the Enterprise's planetary scans. But, they would lack the expressiveness and reasoning power of the richer logic-based systems. They would not be able to draw inferences about a possible new life form.

So, here is where we are today. We have logic-based systems that can reason. We have probabilistic systems that can handle uncertainty. But none that do both.

Even more critical, none of these systems that reason are able to learn. For both of these reasoning techniques, experts often take months or years to acquire, encode, and input the necessary knowledge to make these systems work. If the assumptions change, the systems break and need to be reprogrammed. To make these reasoning systems practical, they need to learn on their own.

We have developed an impressive collection of machine learning techniques—support vector machines, neural networks, and many others—but they do not reason. The systems that are able to reason cannot learn. The systems that learn cannot reason. Finding a way to combine these capabilities is one of our greatest challenges.

IPTO is working on these challenges. We have a current program, PAL, which stands for Personalized Assistant that Learns. The PAL project is building the first intelligent personal assistant by integrating multiple learning and reasoning techniques.

PAL is in the second year, of a 5-year program. Every year, the PAL developers are creating and testing a new version of their intelligent personal assistant.

PAL will sit on your computer, observe what you do, and learn from your actions. PAL will learn your preferences, learn what tasks you do, and—

Learning and Reasoning: The True Heart of the Mind

over time—learn to perform those tasks for you. PAL does so by combining logical and probabilistic reasoning, by combining learning by observation and learning by being told. PAL is one of first major integrations of learning and reasoning technology we have to date.

PAL can learn a new task by observing a single example and listening to a few high-level instructions, without being programmed. No matter how stupid PAL may be when it begins, it can learn. As long as it continues to learn, we have a chance of breaking the barrier of machine stupidity and creating a genuinely useful, general-purpose assistant.

To do this, the PAL developers are trying extensions, combinations, and new architectural designs of many of our current logic-based and probability reasoning systems.

IPTO is pursuing other research paths as well. In general, we are exploring new hybrid combinations of the existing learning and reasoning techniques. We don't believe the answer lies in discovering one new magic reasoning/learning algorithm, but in finding the right architectural combination of the techniques we already have.

Ted Senator has started a new program called "Transfer Learning." Humans have a very useful cognitive ability to transfer learning from one domain to another. Once we know about water— how it flows, and how water pressure behaves—it is much easier for us to learn about electricity. That's the focus of the Transfer Learning program: build artificial learning systems that have the capability to transfer learning across domains.

Tom Wagner is starting a new program called "Integrated Learning." Humans learn from only one or two examples by integrating and leveraging knowledge from diverse sources about the problem domain. That is the focus of the Integrated Learning program.

These hybrid combinations promise to yield significant results, but they may not be enough.

We may still need something completely new, if we hope to break the barrier of machine stupidity. We need to explore radically new reasoning methods, perhaps methods that are more like the human brain, and less like 18th century machines.

We have just started what we expect to be a very exciting and long-term, new initiative, called "BICA," which stands for Biologically-Inspired Cognitive Architectures. Through FMRI and brain imaging techniques, we now have a much clearer idea of how the brain works than we did in the 1960s and 70s, when most of the current artificial reasoning technologies were created.

We now have models of the neural circuitry in the neocortex. We have models of memory formation in the hippocampus. We have models of the basic interactions among the frontal and posterior cortex, hippocampus, thalamus, and basil ganglia.

Recent research on human memory shows that we may have been wrong when we assumed concepts in the human brain are symbolic. Our concepts are more like perceptual simulations or reenactments of the original experience and less like symbols. When I ask you to think about a dog, you conjure up an image of a generic dog, a composite mental simulation of the dogs you've seen in your life and not an internal symbol representing the concept DOG. When Aristotle thought about his new concepts of animals, vegetables, and minerals, his head was not full of Greek symbols as we might expect, but perceptual images of those general categories.

The basic representation of knowledge in the brain is very different from the symbolic representations we have pursued in computer science.

Under the BICA program, we solicited ideas like these from psychologists and neuroscientists to design new architectures of cognition based on the latest discoveries about how the brain actually works. A year from now, we will be ready to begin Phase 2, at which time we will solicit proposals to build computational versions of these new

Learning and Reasoning: The True Heart of the Mind

architectures. We may discover a new basis for building intelligent machines.

We need your help to break the barrier of machine stupidity. Help us to make machines learn on their own, reason flexibly, and achieve the science fiction dream of computers that truly amplify human productivity and creativity.