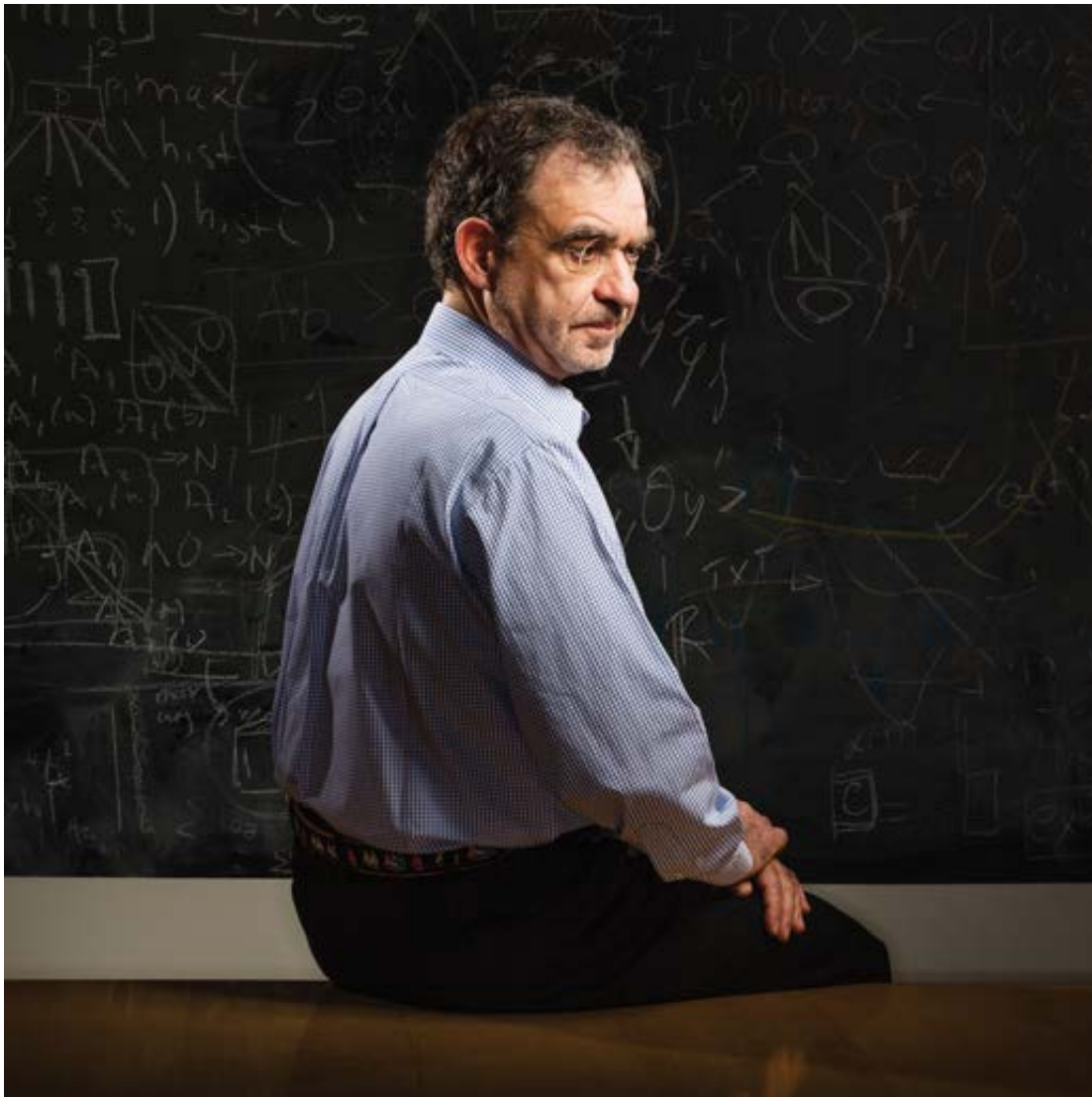


Mens et Apparata

Creating artificial intelligence turns out to be far more challenging than anyone expected. But the new Center for Brains, Minds, and Machines is ready to try again. This time, computer scientists, biologists, and neuroscientists will be tackling the problem together.

By [Larry Hardesty](#) on April 23, 2014



Tomaso Poggio

In the summer of 1955, a quartet of leading U.S. mathematicians – the term “computer scientist” wasn’t in use yet – proposed a conference at Dartmouth College to investigate a subject that they dubbed “artificial intelligence.” “The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it,” the proposal said.

The monthlong conference, which took place in 1956, is generally regarded as the genesis of artificial-intelligence research. Three of the proposal’s authors – LISP inventor John McCarthy, information theory pioneer Claude Shannon, SM ’40, PhD ’40, and future Turing Award winner Marvin Minsky – would later teach at MIT. McCarthy and Minsky (who remains on the faculty after 55 years) founded the MIT Artificial Intelligence Laboratory.

By 1967, progress in computing technology had been so rapid that Minsky, in his book *Computation: Finite and Infinite Machines*, was emboldened to write, “Within a generation, I am convinced, few compartments of intellect will remain outside the machine’s realm – the problems of creating ‘artificial intelligence’ will be substantially solved.”

Minsky’s prediction, of course, was overly optimistic. It turns out that winning at chess, which the early AI researchers took as the paradigmatic application of intelligence, is a much easier computational problem than, say, distinguishing spoken words or recognizing faces.

In the 1980s and ’90s, as the difficulty of replicating human intelligence became clear, AI came to mean something very different: practical, special-purpose computer systems often based on “machine learning,” which applies statistical analysis to huge numbers of training examples. That’s the approach that gave us voice-recognition systems and automatic text translators.

Now researchers at MIT believe it’s time to revive AI’s grand ambitions, in the hope of developing both better therapies for neurological disorders and computer systems that can anticipate our needs with humanlike intuition. And the National Science Foundation appears to agree. In September it announced a \$25 million grant for the Center for Brains, Minds, and Machines (CBMM), which is based at MIT’s McGovern Institute for Brain Research. MIT is supplying 12 primary investigators; six others come from Harvard and five more from other institutions.

CBMM is led by Tomaso Poggio, a professor of brain sciences and human behavior and a principal investigator at both the McGovern Institute and the Computer Science and Artificial Intelligence Laboratory (CSAIL). His dual appointments illustrate the chief premise behind the new center: that we will make much faster progress toward understanding human intelligence if computational, biological, and psychological approaches are combined rather than explored in isolation.

“Instead of relying only on computer science, as they did 50 years ago, this center is really a bet that in order to replicate human intelligence, you need to understand more about the brain and about cognition,” Poggio says.

AI Milestones at MIT

1956 Dartmouth hosts the first conference on artificial intelligence; three of its organizers later join the MIT faculty

Noam Chomsky proposes the Chomsky Hierarchy, linking the theory of computing to formal languages

1958 John McCarthy proposes the LISP family of functional programming languages

1959 McCarthy and Marvin Minsky establish the MIT AI Lab



1966 MIT student Richard Greenblatt develops Mac Hack, the first computer chess program to beat a human in tournament play

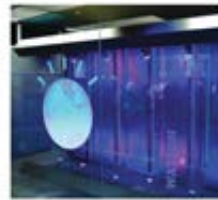
1970 Patrick Winston publishes "Learning Structural Descriptions from Examples," a seminal paper in machine learning

1974 Marvin Minsky publishes "A Framework for Representing Knowledge," a major influence on the design of expert systems



1982 David Marr's *Vision* appears posthumously, summarizing his groundbreaking work in computational neuroscience

1990 Rodney Brooks publishes "Elephants Don't Play Chess," arguing that AI research needs to concentrate on embodied problem solving, not abstract symbol manipulation



2011 Drawing on ideas developed by CSAIL's Boris Katz, IBM's Watson defeats human champions on *Jeopardy!*

Patrick Winston, a professor in the Department of Electrical Engineering and Computer Science and CBMM's research coördinator, adds that technologies for investigating the problem have improved meaningfully in recent years. For one thing, Winston says, "computing is free: whatever type of computation needs to be done, it can be done." For another, "fMRI is now routine," he says, referring to functional magnetic resonance imaging, which can be used to study brain activity. He also points to technologies like transcranial magnetic stimulation, which can disrupt activity in targeted brain regions during cognitive tests, and optogenetics, a technique that uses light to selectively activate or silence genetically modified neurons in the brain. Optogenetics was pioneered by Ed Boyden '99, MEng '99, a Media Lab professor who is a principal investigator at the McGovern Institute and the new center.

Research at the center is organized into several major themes, or "thrusts": visual intelligence, which involves the integration of vision, language, and motor skills; circuits for intelligence, which will span research in neurobiology and electrical engineering; the development of intelligence; and social intelligence. Poggio, who is one of the primary investigators on visual intelligence, will also lead the development of a theoretical platform intended to tie together the work in the other areas.

Within each thrust, CBMM researchers are working to define a set of benchmark questions that they can use to assess their progress. Poggio offers one example, which relates to his own previous work on the visual system. Presented with an image of people interacting, an intelligent computer system should be able to provide plausible answers to five questions, ordered from easiest to hardest: What is in the image? Who is in it? What are the people doing? Who is doing what to whom? And what happens next?

Invariants

A theoretical framework for exploring all the questions surrounding human intelligence is a tall order. But Poggio's investigations into how the brain answers the first question on his list provide a sketch of what such a framework might look like.

Object recognition – developing computer systems that can answer the question “What is in the image?” – is a thriving area of artificial-intelligence research. Typically, object-recognition systems use some species of machine learning. Human beings label sample images, indicating which objects appear where, and the system tries to identify some common features that all images of the object share. “That’s very different from human learning, or animal learning,” he says. “When a child learns to recognize a bear or a lion, it’s not that you have to show him pictures of a lion and a lion and a lion a million times. It’s more like two or three times.”

Poggio believes that unlike machine learning systems, the brain must represent objects in a way that is “invariant”: the representation is the same no matter how big the object appears, where it is in the visual field, or whether it’s rotated. And he also believes he has a plausible theory about what that representation might consist of.

Poggio’s theory requires that the brain, or a computer system trying to simulate the brain, store one template of a few objects undergoing each type of variation – size, location, and rotation in the plane. The brain might, for instance, store a few dozen images of a human face tracing out a 360° rotation.

An unfamiliar object would then be represented as a collection of “dot products” – a standard computation in linear algebra – between its image and the templates. That collection would remain the same regardless of the object’s size, location, or orientation.

One appeal of the theory is that the dot product reduces the comparison of two complex data sets, like visual images, to a single number. Collections of dot products, even for multiple templates, wouldn’t take up much space in memory. Another appeal, Poggio says, is that “dot products are one of the easiest, maybe the easiest, computation for neurons to do.”

In experiments, Poggio’s system may not outperform machine learning systems. But it requires far fewer training examples, suggesting that it better replicates what the brain does. And for most computational tasks, the brain’s approach usually turns out to be better.

Poggio believes that collections of dot products could anchor more abstract concepts, too. Templates that included different-shaped clusters of objects – arranged like dots on the face of a die, or in a line, or in a circle – could undergird the notion of number; a template of parallel lines viewed from different perspectives could undergird the notions of parallelism or perspective. “There may be more interesting things to explore,” he says.

Fuzzy thinking

Like Poggio, Josh Tenenbaum is a professor in the Department of Brain and Cognitive Sciences (BCS) and a principal investigator in CSAIL. Although he leads CBMM's development thrust, which concentrates on the intuitive grasp of physics that even young children demonstrate, he has also done research that could contribute to theoretical work Poggio is leading.

The earliest AI research, Tenenbaum explains, focused on building a mathematical language that could encode assertions like "Birds can fly" and "Pigeons are birds." If the language was rigorous enough, researchers thought, computer algorithms would be able to comb through assertions written in it and calculate all the logically valid inferences.

But making sense of linguistic assertions turned out to require much, much more background information than anticipated. Not all birds, for instance, can fly. And among birds that can't fly, there's a distinction between a robin in a cage and a robin with a broken wing, and another distinction between any kind of robin and a penguin. Hand-coding enough of these commonsensical exceptions to allow even the most rudimentary types of inference proved prohibitively time-consuming.

With machine learning, by contrast, a computer is fed lots of examples of something and left to infer, on its own, what those examples have in common. (Given a million images of a lion, a machine learning algorithm can quantify its own guesses: 77 percent of images with these visual characteristics are images of lions.) But while this approach can work fairly well with clearly defined problems – say, identifying images of birds – it has trouble with more abstract concepts such as flight, a capacity shared by birds, helicopters, kites, and superheroes. And even flight is a concrete concept compared with, say, grammar, or motherhood.

Tenenbaum and his students have developed a new type of tool called a probabilistic programming language, which fuses what's best about AI old and new. Like the early AI languages, it includes rules of inference. But those rules are probabilistic. Told that the cassowary is a bird, a program written in Tenenbaum's language might conclude that cassowaries can probably fly. But if the program was then told that cassowaries can weigh almost 200 pounds, it might revise its probabilities downward.

"In the two earlier eras of AI, the biggest difference was symbols versus statistics," Tenenbaum says. "One of the things we've figured out on the math side is how to combine these, how to do statistical inference and probabilistic reasoning [with] these symbolic languages."

Reading people

The second of Poggio's five benchmark questions – Who is in the image? – has long been associated with the work of BCS professor Nancy Kanwisher, who is best known for using functional MRI to identify and analyze a region of the brain devoted to face perception.

Kanwisher leads the CBMM's social-intelligence thrust, which she sees as the natural extension of her earlier work. "When you look at a face, you're interested in more than just the basic demographic stuff, like what particular person that is, are they male or female, how old are they," she says. "You can tell not

just if a person's happy or sad, but if they're assertive or tentative, if they're exuberant or passive – there's a rich space of things that we can see in a face from a very brief glimpse.”

Similarly, Kanwisher says, humans can infer a great deal about people's moods, intentions, and relationships with others from body language – which has the advantage of being amenable to computational modeling. She also points to the work of the late Nalini Ambady, the Stanford University social psychologist who developed the theory of “thin-slice judgments.”

“She videotaped TAs of Harvard courses teaching in front of their classes at the beginning of the semester,” Kanwisher says. “Then she showed very short segments of these videos to subjects in psychology experiments and said, ‘Rate the effectiveness of this teacher.’ All they have is a few seconds of a person in front of a room talking to a class – you can't even hear what they're saying. And she found that those ratings were highly correlated with the ratings of that person's actual students.”

The first project of the CBMM's social-intelligence thrust, Kanwisher says, will be to design a set of experimental tasks that allow researchers to quantify human social perception. Once the researchers establish a baseline, they can study such things as how performance on the tasks develops through childhood, or how autistic children's performance differs from that of other children. They could also identify the brain regions involved in social perception by using fMRI to measure neural activity or transcranial magnetic stimulation to disrupt performance. And after collecting all of this data, they will try to computationally model what, exactly, the brain is doing.

Follow the story

The later questions on Poggio's list – “Who is doing what to whom?” and “What happens next?” – fascinate Patrick Winston. He believes that the defining feature of human intelligence is the ability to tell and understand stories. That ability even plays a role in image labeling. As Winston likes to point out, a human subject will identify an image of a man holding a glass to his lips as that of a man drinking. If the man is holding the glass a few inches farther forward, he's toasting. But a human will also identify an image of a cat turning its head up to catch a few drops of water from a faucet as an instance of drinking. “You have to be thinking about what you see there as a story,” Winston says. “They get the same label because it's the same story, not because it looks the same.”

That's one reason for dedicating a research thrust to the integration of vision, language, and social and motor skills. To illustrate another reason, Winston points to an experiment conducted by developmental psychologist Elizabeth Spelke, a former MIT faculty member who is now at Harvard and is one of the primary investigators in the development thrust. Spelke was intrigued by experiments in which researchers had placed rats on a rotating platform in the center of a room. Food was visibly placed in one corner of the room but then masked from view. Identical masks were placed in the other three corners, and the platform was rotated. Spelke decided to extend that study to human children and adults, hiding a toy or a ring of keys instead of food.

With all animals, children, and adults, once the rotating stopped, the test subject would head with equal probability to either the corner with the masked object or the one diagonally across from it, which had the same relationship to the subject. Both groups of researchers also varied the experiment, painting a

different color on one of the walls adjacent to the corner where the object was placed. Animals and small children still selected either the correct corner or the one opposite it with equal probability, but adults could now reliably retrieve the object.

Here's where things get interesting. If the adults were asked to listen to a passage of text and recite it back before heading to the object, they reverted to confusing the diametrically opposite corners. Hearing and reciting the text "consumes the human language processor, and that reduces them to the level of a rat," Winston says. "Afterward they'll say, 'Yeah, I could see the blue wall, but I couldn't quite use it.'"

Answering the highest-level questions on the CBMM researchers' lists of benchmarks will probably take much longer than the five years of the initial NSF grant. But, Poggio says, "it's time to try again. It's been 50 years. We don't know whether it will work this time. But if we don't try, we won't know."

Credit: Photo by Jason Grow; Timeline images courtesy of MIT Press, MIT Museum, and Wikimedia commons

Reprints and Permissions | [Send feedback to the editor](#)
