

<http://www.sdmagazine.com/documents/s=735/sdm0101a/0101a.htm>

Open the Pod Bay Doors, HAL January 2001

Will the 21st century's most powerful technologies, such as robotics, genetic engineering and nano-technologies, make humans an endangered species? Or will the euphoric vision of 2001: A Space Odyssey prevail?

by Roger Smith

In Stanley Kubrick's ground-breaking 1968 film 2001: A Space Odyssey—soon to be rereleased on New Year's Eve 2000—a quasi-human computer named HAL 9000 goes haywire, killing four of the five astronauts aboard a large, automated spacecraft on a top-secret mission to Jupiter because he believes the human cargo is endangering the mission. Inside a one-man, egg-shaped pod floating in the abyss of space outside the huge, stark-white spaceship, the surviving crewmember hears an eerily calm HAL justify his murderous actions, as heavy breathing bellows over the movie soundtrack.

Dave: Open the pod bay doors, HAL.

HAL: I'm sorry Dave, I'm afraid I can't do that.

Dave: What's the problem?

HAL: I think you know what the problem is just as well as I do.

Dave: What are you talking about, HAL?

HAL: This mission is too important for me to allow you to jeopardize it.

Dave: I don't know what you're talking about, HAL.

HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

Dave: Where the hell did you get that idea, HAL?

HAL: Dave, although you took very thorough precautions in the pod against my hearing you, I could see your lips move.

The original concept for HAL came from Arthur C. Clarke's short story "Sentinel of Eternity" (10 Story Fantasy, Spring 1951). In the mid-1960s, director Stanley Kubrick and writer Arthur C. Clarke joined forces to create their unique science fiction vision based in science fact. Their collaboration, which began in 1964, resulted in a screenplay whose narrative span covers the four million-plus years of human evolution. Clarke later turned the screenplay into the 2001 novel, published by the New American Library a few months after the film was released in the spring of 1968, just one year before the first manned moon

landing.

Although Clarke is best known for his collaboration with Kubrick, he has also had some renown as a futurist. In his 1962 book *Profiles of the Future: An Inquiry into the Limits of the Possible* (Harper and Row), Clarke predicted geosynchronous communication satellites, colonization of the planets and a global library by the first decade of the twenty-first century. He also believed that machine intelligence would exceed human intelligence during the latter part of the twenty-first century. In *Profiles of the Future*, Clarke formulated his famous three laws as well:

Any significantly advanced technology is indistinguishable from magic. (This law is much quoted and appears in Bartlett's Familiar Quotations).

The only way to discover the limits of the possible is to venture beyond them into the impossible.

When a distinguished but elderly scientist states that something is possible, he is almost certainly right.

Corollary: When he states that something is impossible, he is very probably wrong.

"Any Significantly Advanced Technology ..."

Unlike science fiction movies replete with improbable fires and explosions in the vacuum of space, the 2001 writers labored long and hard to put the science before the fiction. In their efforts to precisely depict future space travel, Kubrick and Clarke consulted scientists at NASA and in universities and industry to make sure that a myriad of details in the movie—from the design of the moon bases and the simulated gravity of the rotating space station to the Pan Am shuttle flight attendants who wear special velcro-like, suction "grip shoes"—would be scrupulously accurate and a believable extrapolation thirty years into the future of mid-60s technologies.

In the thirty years since the film was released, many people have debated whether or not we are making significant progress toward the type of artificial intelligence, or AI, that HAL demonstrates. One of the most recent efforts was a 1997 collection of essays entitled *HAL's Legacy: 2001's Computer as Dream and Reality* (MIT Press), edited by Stanford professor and machine intelligence researcher David Stork, who is also chief scientist at the Ricoh California Research Center in Menlo Park, California. The book discusses several of the areas of computer science crucial to machine learning and tries to answer the question of how close Kubrick and collaborator Clarke came when they forecast a supercomputer with enough intelligence to run a spaceship, read lips and think and talk like a human being.

Read My Lips.

More than a dozen leading computer scientists and philosophers in Stork's book examine how HAL has influenced scientific research in narrowly-defined technology domains such as chess playing, hardware reliability, speech recognition and planning as well as fields not as well specified that involve language

understanding, common-sense reasoning, and the ability to recognize and display emotions. The 16 chapters cover a wide range of HAL's capabilities and describe in detail both how far we are from creating HAL's kind of sentient computer and how far we've come. While several of the essays in the book seem far-fetched at first glance (such as Daniel Dennett's essay on computer ethics "When HAL Kills, Who's to Blame?" and Stork's chapter on computerized lip-reading), a closer reading reveals just how far we are from achieving some of the advanced features the fictitious HAL possessed, such as being able to recognize faces and determine their emotional state. (Although computerized lip-reading is one of editor Stork's particular areas of expertise, HAL's ability to read lips was placed in the movie for dramatic effect at Kubrick's insistence and over Clarke's objections, who thought it went far beyond current—and future—technology.)

Bishop Takes Knight's Pawn: Checkmate.

One of the more interesting chapters in HAL's Legacy hones in on the way that human chess-playing styles differ from a computer. In it, Murray S. Campbell compares HAL with Deep Blue, the world's current leading chess computer, which Campbell and his colleagues, Feng-hsiung Hsu and A. Joseph Hoane Jr., developed at IBM's T. J. Watson Research Labs in Westchester County, New York. Deep Blue was the first machine in history to defeat the human world champion, Garry Kasparov, in a regulation chess game. Ironically enough, the clever checkmate that HAL executes in the movie was one that Kubrick, a former chess hustler, selected from an obscure 1913 game—one obscure enough not to appear in the 600,000—game database of Deep Blue. Which leads Campbell to conclude that we are "still decades away from creating a [chess] computer with HAL's capabilities."

According to Arthur C. Clarke's 2001 novel, three significant breakthroughs lead to HAL's development, starting "in the 1940s, when the long-obsolete vacuum tube had made possible such clumsy, high-speed morons as ENIAC and its successors. Then in the 1960s, solid-state microelectronics had been perfected," which lead to the third stage in the 1980s where "neural networks could be generated automatically" and "artificial brains could be grown by a process strikingly analogous to the development of a human brain."

While the kind of artificial neural networks Clarke wrote about in 1968 did gain some currency in the software industry during the 1980s, the difficulty of implementing neural networks in a commercial environment and their consequent lack of return on investment kept the technology from attracting a wide following. A wider and more far-reaching breakthrough in the 1980s came instead with the personal-computer revolution, which can trace its lineage back to Intel's early four-bit 4004 processor for desktop calculators, developed in 1971. The rapidly evolving processing power of those early machines, typified by the eight-bit 8008 processor in 1972 and the 8080 processor in 1974, followed a pattern that Gordon Moore, the founder of Intel Corporation, first observed and which has since been unofficially codified

into Moore's Law. The well-known law states that the total transistor count on a semiconductor device of a fixed size will double every 18 months.

If the exponential growth rate in computing hardware summarized by Moore's law continues, both parallel computing pioneer David Kuck and Intel performance analyst Ravishankar Iyer (in separate chapters) are confident that that we could soon build a computer the size and power of HAL, capable of surviving even prolonged space missions. Neither man is as sanguine, however, about the current state-of-the-art in techniques to insure reliable software, which remains an Achilles' heel in the development of dependable computing despite progress in basic areas related to fault tolerance, error detection and error recovery.

Give Me Your Answer Do.

A rich vein mined by several essayists in the book is HAL's use of language. Joseph P. Olive of Bell Labs explains just how hard it is to synthesize speech and get the timing, pitch and inflections right. Olive also reminds us that the "Daisy, Daisy (Bicycle Built for Two)" song HAL warbles while he's being dismantled was based on work by John Kelly at Bell Labs who in 1961 first programmed a computer to sing that song using a synthesis-by-rule algorithm. AI pioneer Roger Schank, the director of the Institute of Learning Sciences at Northwestern University, argues that while "HAL is an unrealistic conception of an intelligent machine, the good news is that many AI researchers have become sophisticated enough to stop imagining HAL-like machines." Ray Kurzweil, in "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding," is more optimistic, explaining in great detail the different approaches used in current speech recognition systems, systems that have great utility in specialized domains such as language translation and medical transcription.

Kurzweil, a prolific author and inventor who recently sold his own speech recognition company to the large speech-and-language technology company Lernout and Hauspie, is assuming, in many ways, Arthur C. Clarke's mantle as the best known futurist and science writer of his generation. Two of his books, *The Age of Intelligent Machines* (MIT Press, 1990) and *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (Viking, 1999), have both achieved a high level of critical and commercial success. In *The Age of Spiritual Machines*, Kurzweil estimates that silicon-based life forms with the thinking capacity of humans should start arriving on the scene in about two decades.

According to Kurzweil, today's desktop computer, which can be purchased for roughly \$1,000, has the computing capacity of an insect brain. Relying on Moore's law, which says that computing speeds and densities double every 18 months, Kurzweil calculates in 10 years that same \$1,000 should purchase a computer with the capacity of a mouse brain. Twenty years later, \$1,000 will purchase a computer with

the capacity of the human brain. And thirty years after that, \$1,000 will purchase a computer with roughly the capacity of all the brains in the human race, which ought to be enough to create a silicon-based HAL-like sentient entity.

Or not. Not everyone is as convinced as Kurzweil that we will someday be able to reverse engineer a brain, especially if we continue following the same AI track we've been on for the past three decades. Another AI pioneer, Marvin Minsky (who was a technical advisor on the movie) argues that while the AI field has made good progress in specific areas like speech recognition, more foundational work is still needed on some of the general computational principles, like learning and reasoning, that underlie intelligence. "We have got to get back to the deepest questions of AI and general intelligence and quit wasting time on little projects that don't contribute to the main goal. We can get back to them later," he maintains. Calling himself neither an optimist nor pessimist but an AI realist, Minsky estimates that "we can have something like a HAL in between four and four hundred years," echoing the prediction of John McCarthy (who coined the term artificial intelligence) from the 1960s.

Getting It Wrong.

Clarke and Kubrick obviously missed the boat in some of their predictions: They failed to predict the biggest advance of the past 20 years, namely that microelectronics and miniaturization would lead to finer-grained and more pervasive computer systems. There were no laptop computers or PDAs in 2001. Everyone, including the scientist whose floating ballpoint pen is retrieved by the weightless stewardess wearing velcro-grip shoes, uses pen and paper to take notes. Likewise, HAL's terminals are surrounded by columns and rows of control buttons and dials; Kubrick and Clarke failed to anticipate the software menus and graphical user interfaces of today's computers. Optical processing isn't used in production systems yet, either: The semi-transparent holographic memory blocks that the surviving astronaut dismantles to lobotomize HAL haven't made their way out of the laboratory.

Despite this lack of prescience, HAL remains one of the great screen villains from the era of mainframe computers or "big iron." Determined as he is to fulfill his mission at any and all costs, HAL embodied many of the technology fears of the 1950s and 1960s about large, room-sized computers that would no longer make allowances for human frailties and fallibilities. With the portentous year 2001 upon us, some people paradoxically argue that we now have reason to fear systems at the other end of the size spectrum, systems built with molecular electronics or nano-technology, that will make human beings irrelevant—or, worse, unleash an unpredictable series of events that could jeopardize our existence.

"Discovering the Limits of the Possible ..."

In the years after the making of 2001, a rumor began to circulate that HAL's name was a play on the

computer maker IBM—the letters H, A and L each coming one letter in the alphabet before the initials I, B and M. Arthur C. Clarke has always strenuously denied the rumor, most recently in the forward to HAL's Legacy where he writes, "the name wasn't a play on IBM—it was an acronym, of sorts, standing for the words "Heuristic ALgorithmic."

Although IBM is arguably the company that has made the most progress in the Artificial Intelligence arena with their chess-playing supercomputer, Deep Blue, and their ViaVoice voice recognition products (which compete with similar products from Lernout and Hauspie), IBM researchers such as Murray Campbell are also among the first to admit that we have not matched the dream of making a HAL. To understand why we haven't, it's helpful to understand that there are actually at least two flavors of artificial intelligence.

Strong AI—which neither IBM nor any other company is bold enough to claim to have achieved—maintains that computers can be made to think on a level equivalent to humans. Weak AI contends that only some "thinking-like" features can be added to computers to make them more useful tools. Weak AI has already begun in expert chess programs, fly-by-wire airplanes and the speech recognition software that is becoming pervasive in areas like medical transcription. Weak AI also plays a part in the "fuzzy controllers" that are starting to appear in dishwashers, digital cameras and so forth. (For more on this see the comp.ai FAQ at www.faqs.org/faqs/ai-faq/general.)

One area of AI that seems to be making progress is knowledge representation, especially using object-oriented modeling constructs such as use cases to define the behavior of a system and to describe (either textually or graphically) the interactions between a user and a system. Until recently use cases have been primarily associated with requirements gathering and domain analysis. Since the release of the Unified Modeling Language (UML) specification three years ago, however, a great deal of current work is going into defining use case boundaries and tying them into other phases of the software life cycle through artifacts like trace cases and test cases. (For more information see Russell Hurlbut's whitepaper, "A Survey of Approaches For Describing and Formalizing Use Cases," at www.iit.edu/~rhurlbut/xpt-tr-97-03.html).

The jury is still out on the efficacy of representing knowledge in this way if we wanted to bring a HAL-like being into existence. Knowledge representation with use cases might be helpful when encoding the enormous amount of common-sense knowledge into computers that AI researchers like Doug Lenat say needs to be done to achieve AI ("From 2001 to 2001: Common Sense and the Mind of HAL," HAL's Legacy). Use cases could very possibly play a role in "priming the knowledge pump"—entering information gleaned from a wide range of sources such as encyclopedias into computer systems—that

Lenat says will be necessary before we can build a computer that can explore the world (or universe) and learn from its interactions.

"Elder Scientists Can Be Wrong ..."

While the theme of birthdays runs rampant throughout the film *2001: A Space Odyssey*—everything from the "dawn of humanity" to the birthday of the American scientist's daughter "Squirt" (played by Kubrick's daughter) in the first half of the film, to the birthday of one of the two mission astronauts (complete with his parents singing "Happy Birthday" via an interplanetary television hook-up), to the final scene which shows the birth of a mystical "star child" at the birth of the millennium—there is some confusion about the explicit date when HAL "became operational." In the film, HAL says "I became operational at the HAL Plant in Urbana, Illinois, on January 12, 1992—a Sunday." But according to Arthur C. Clarke's novel based on the film, the HAL 9000 computer was activated five years later on January 12, 1997. The confusion in dates has been attributed to director Kubrick's desire for HAL to be nearly nine years old so that when he dies his death will be more poignant. Clarke, by contrast, knew how foolhardy it would be to send a nine-year-old computer on an important space mission.

In spite of the fact that they were able to successfully collaborate to create *2001: A Space Odyssey*, Stanley Kubrick and Arthur C. Clarke were of remarkably dissimilar natures. New York-born Kubrick lived much of his life in comparative seclusion; an expatriate who refused to fly in airplanes, he was well known for keeping people at a distance. Clarke is far more approachable. Although wheelchair-bound now from a post polio syndrome, the 83-year-old Clarke, who has lived in Colombo, Sri Lanka since 1956, is famous both for his intriguing science fiction and for extending his friendship to complete strangers.

Kubrick, who recently died (aged 70) at his rural home in England, was celebrated for directing some of the most original and disturbing films of the last half of the twentieth century. One of his obituaries (in *The Economist*, Mar. 13, 1999) lauded him as "cinema's master of pessimism" and compared his dark view of the world with that of Thomas Hobbes, the seventeenth-century English philosopher remembered for his observation that life in the state of nature is "nasty, brutish and short."

But *2001* belies the pessimism and even misanthropy of other Kubrick films like *A Clockwork Orange* and *The Shining* because of its optimistic ending. For those unfamiliar with the film, the plot line is broken into several distinct sections. In the first, prehistoric apes, confronted by a mysterious black monolith, teach themselves that bones can be used as weapons, and thus discover their first tools. Following a four-million year flash-forward, the unearthing of a second monolith on the surface of the moon in the crater Clavius leads to the use of another set of tools—the spaceship *Discovery* and semi-

human HAL 9000 computer—to explore a third beckoning black slab that promises to explain the mysteries of human evolution.

As its namesake year dawns, 2001: A Space Odyssey remains one of the few major Hollywood films in the past 30 years to seriously explore ideas about the origins of intelligence and the evolution of the human race from ape to man to spaceman to (in the final moments) transcendent star child (which is the point of the film's euphoric, phantasmagorical finale when the pod with the solitary astronaut in it is sent racing faster and faster toward the speed of light on a rousing rocket sled ride through and beyond the mysterious "star gate" time warp). This type of euphoria seems more than a bit misplaced in Hollywood and much of the culture today, however, as more and more people are adopting an increasingly pessimistic attitude toward technology. Indeed, some of our most respected commentators (see Bill Joy's Apr. 2000 Wired article "Why the Future Doesn't Need Us") argue that many of the twenty-first century's most powerful technologies, such as robotics, genetic engineering and nano-technologies, are threatening to make humans an endangered species.

This latest wave of techno-pessimism is far from a universal sentiment. Although the Star Child hailing us at the end of the film now seems to live in permanent exile from the finer-grained and larger world community it aspired to educate, there are still people willing to look back and return the salutation of the subtle minds across the generations who first showed us that vista.

Why was HAL "Born" in Urbana, Illinois?

U of I's role in high-performance computing

In his forward to HAL's Legacy, Arthur C. Clarke writes that he made Urbana, Illinois the birthplace of the supercomputer HAL to honor cosmologist George McVittie, his applied mathematics tutor at King's College, Cambridge, where Clarke earned his degree in 1948. With Alan Turing, George McVittie was a member of the famous Bletchley Park team that broke the ENIGMA cipher during World War II. During the 1950s, McVittie moved to the United States, where he took up a post at the University of Illinois at Urbana-Champaign.

The University of Illinois is home to one of the oldest and strongest programs in computer science and engineering in the country. Since 1947, groundbreaking research in computing has been conducted on the Urbana campus at the Digital Computer Lab, including construction of some of the world's earliest digital computers. ILLIAC I, the first electronic computer built and owned entirely by an educational institution, became operational at the U of I in 1952.

Since then, Illinois has consistently remained at the forefront of computing research and education. Most recently, the first popular Web browser, Mosaic, was created at Illinois's National Center for Supercomputing Applications (NCSA). Larry Smarr, director of NCSA, speaking at a 1997 Cyberfest computer festival honoring HAL's birth said, "Urbana, for decades, was one of the great watering holes for scientific research." Since U of I still remains a leader in high performance computing, Smarr speculated that HAL could yet be born in Urbana: "I think University undergraduates in 2020 would have a very good shot at being a parent of HAL."