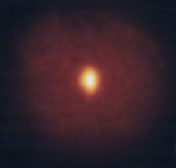
HAL'S LEGACY

2001'S COMPUTER AS DREAM AND REALITY



EDITED BY DAVID G. STORK

FOREWORD BY ARTHUR C. CLARKE

HAL's Legacy

2001*'s Computer as Dream and Reality*

edited by David G. Stork

The MIT Press
Cambridge, Massachusetts
London, England

Third printing, 2000

Library of Congress Cataloging-in-Publication Data

HAL's legacy : 2001's computer as dream and reality / edited by David G. Stork.
p. cm.
Includes index.
ISBN 0-262-19378-7 (hc : alk. paper)
1. Computer science. 2. Supercomputers. I. Stork, David G.
QA76.H265 1997
791.43'72—dc20 96-31375
CIP

Contents

Foreword: The Birth of HAL

Arthur C. Clarke

"Is it true, Dr. Chandra, that you chose the name Hal to be one step ahead of IBM?"

"Utter nonsense! Half of us come from IBM and we've been trying to stamp out that story for years. I thought that by now every intelligent person knew that H-A-L is derived from Heuristic ALgorithmic."

Afterwards, Max swore that he could distinctly hear the capital letters.
—2010: Odyssey Two

Sometimes you just can't win. I deliberately inserted that passage into *Odyssey Two* because for decades I too had been "trying to stamp out that story." I don't know when or how it originated, but believe me it's pure coincidence, even though the odds against it are 26^3 to 1. (Much less, of course, if you eliminate ridiculous—or vulgar—combinations.)

I was embarrassed by the whole affair, and I felt that IBM, which was very helpful to Stanley Kubrick during the making of *2001*, would be annoyed. (Bell Telephone and PanAm—remember them?—also provided useful services. How difficult it is to foresee the future! At least Hilton and Howard Johnson are still with us, though they are not yet in space.)

Well, recently I gave a satellite address to an IBM conference in Europe and was pleasantly surprised to discover that all had been forgiven. In fact, because Big Blue now seems quite proud of the link and no longer fears guilt by association, I've happily abandoned my attempt to set the record straight.

Of course, I'm pleased that the distinguished contributors of this book are also trying, so to speak, to put the record straight—even though the hindsight of specialists is bound to be more accurate than the foresight of a writer and filmmaker—especially in a field as rapidly changing as computer science. A great deal has happened since Stanley and I worked so hard to learn all we could about that nascent field (and about space travel too, for that matter).

Although I've never considered *2001* as a strict prediction—but as more of a vision, a way thinks could work—I have long kept track, informally, of how our vision compares with computer science reality. Some things we got right—even righter than we ever had a reason to suspect. Others, well, who could have known?

In analyzing these issues, the contributors to *HAL's Legacy* have done us all a great service. They've given us so much more than a scorecard for the film and novel. These creators of the real technology and science have shown the reasons for the way things developed—and may continue to develop—to 2001 and beyond. Their clear analysis of the details of *2001*—a single chess move, say, or the click of an (analog!) camera, the use of the single word *take*—illuminates both science and science fantasy. They also present informed and exciting speculations about the future of computing and artificial intelligence. I have learned a lot from this book and have been especially happy to know how its contributors were affected by *2001*.

Still, there are many things about HAL no one else could explain. For instance, why was he 'born' at Urbana, Illinois? Though I have long forgotten most of the reasons for decisions we made a third of a century ago, I remember this one clearly.

My applied mathematics tutor at King's College, Cambridge, when I took my degree in 1947–1948 was the distinguished cosmologist George McVittie. He taught me the elements of perturbation theory, which I have used in several of my stories—as I acknowledged in *Reach for Tomorrow.* During the 1950s George moved to the United States, where he took up a post at the University of Illinois, Urbana.

gramophone wound down and the words began
to blur and slow, "... leave ... my ... soul ... alone ..."
to cease at last when something other died.
And silence matched the silence under snow.
—Dannie Abse

I will always remember my collaboration with Stanley Kubrick as some of the most intellectually stimulating (and demanding!) work of my career. Anecdotes such as those of Dr. Minsky's recounts in chapter 2 bring back fond memories of the set. Speaking of the set, I do not know how many actors Stanley interviewed before he settled on Douglas Rain as the voice of HAL; but I am almost certain that one of them was Martin Balsam, who comes to a memorably sticky end in *Psycho*. Apparently Martin made some recordings but decided the role wasn't for him. So here is another piece of unknown Kubricana—like the custard-pie fight in the war room of *Dr. Strangelove* that never made it into the final version. (Did you ever wonder what those tables of goodies were doing at the back of the room?) I still hear Douglas Rain's silky voice every time I tell my computer to do something stupid and it says reproachfully, "I'm sorry, Dave, I can't do that."

Further Readings

Dannie Abse. *Collected Poems 1948–76*. London: Hutchinson, 1977.

Jerome Agel, ed. *The Making of Kubrick's 2001*. New York: Signet, 1970. Responses to the film, as well as some discussion of the filming.

Piers Bizony. *2001: Filming the Future. London, UK:* Aurum Press, 1994. This beautiful large-format book is the product of many years of devoted research. It gives the entire history of the film and is full of original art work, engineering drawings, and stills taken during production, most of which have never appeared before.

Arthur C. Clarke. *2001: A Space Odyssey*. London: Hutchinson/Star, 1968. A novel based on the original screenplay by Stanley Kubrick and Arthur C. Clarke.

Arthur C. Clarke. *The Lost Worlds of 2001*. New York: New American Library, 1972. Clarke's view of the writing of the script, the later novel, and alternative chapters, plots, and so on.

Arthur C. Clarke, *2010: Odyssey Two*. New York: Ballantine Books, 1982. The sequel to *2001*, explains several issues in the first novel and the film.

Andrew Hodges. *Alan Turing: The Enigma*. New York: Simon and Schuster, 1983. The definitive biography of one of the deepest thinkers at the dawn of the modern computer age.

Neil McAleer. *Odyssey: The Authorized Biography of Arthur C. Clarke*. London: Victor Gollancz, 1992.

Operating Manual for the HAL 9000 Computer: Revised Edition. Oakland, Calif.: Miskatonic University Press, 2010. This edition, essential for all surviving users of this versatile machine, advises the fitting of small explosive charges at key points in the mainframe.

Preface

I am a HAL Nine Thousand computer, Production Number 3. I became operational at the HAL Plant in Urbana, Illinois, on January 12, 1997.
—*HAL*, 2001: A Space Odyssey (the novel)

At a dinner party some time ago, an acquaintance, a nonscientist, asked me in a casual way about my duties as chief scientist at a research lab. I said that one of my great joys was overseeing a wide range of projects, to varying extents, and I mentioned a few of them: pattern recognition, machine learning, neural networks, computer-chip design, supercomputer design, image compression, expert systems, handwriting recognition, document analysis, uses of global networks such as the World Wide Web, novel human-machine interfaces, and so on. Then I turned to one of the areas of my particular expertise: lipreading by computer.

"Oh," she said, "Like HAL." Ah, a kindred soul, I thought. We spent quite some time discussing the state of the art and the challenges of computer lipreading, its possible applications, and so on. Later our discussion turned to other topics suggested by the movie—language understanding, chess, computer vision, artificial intelligence. It was clear that she was interested in the current state of the art and that many years before the film had both caught her imagination and helped her identify crucial issues in today's computer science. One of the questions she asked was, "How realistic was HAL?"

This book is for people like her. And because no one is an expert in all the topics covered in the film, even scientists are sure to learn from the accounts of other areas. The book is much more than an answer to her question, though. It has four major goals, which it addresses in varying proportions in the sixteen chapters.

Analysis

It is a testament to Clarke and Kubrick's achievement that *2001* still holds up to close scrutiny in the late 1990s. Under the expert eyes of the contributors, the most innocuous aspects of scenes—a line of computer code on a screen, a chess move, the use of a word, the form of a button—reveal a great deal. Even though I've seen the film several dozen times, I have learned an immense amount from the contributors. *HAL's Legacy* seeks to do for *2001* what good art history does for a major painting; namely, make the viewer see it in a new light—a tall order, to be sure!

Teaching

The film illustrates key ideas in several disciplines of computer science, and thus provides a springboard for discussions of the field in greater depth, including our own research. Descriptions of the world computer chess champion Deep Blue system, the commercially successful VOICE recognition system, the massive CYC artificial-intelligence project, the award-winning Mathematica software system, and much more are here discussed by their creators at a level accessible to the general reader.

Prognostication

It is natural, too, to look to the future. Several contributors make informed and fascinating predictions based on developments in the field. What are the most promising approaches toward artificial intelligence? Will we ever be able to "reverse engineer" a human brain and represent it in a computer?

Reflection

2001 transcends the label of "science fiction movie" and captures many of the central metaphors of *our* time, telling us much about society and its aspirations. The film has even been praised by the pope! Many people have been deeply affected by the film, among them several contributors who reflect here about its influence on their own careers and on computer science in general.

Clearly, *HAL's Legacy* differs from books on the making of the film or its cinematography. It differs, too, from books that analyze the science shown in movies or on television—science that is incidental and just "goes along for the ride." To an extent unprecedented and never duplicated in a feature film, the makers of *2001* were as careful as possible to get things right; when they did make errors, they often did so in illuminating ways.

Now seems like the perfect time for *HAL's Legacy.* Birthdays are an important theme in the film (there are at least five of them), and in the novel, HAL "becomes operational . . . on January 12, 1997." Kubrick changed the year to 1992 for the film version—perhaps to give HAL a longer lifetime and so make his death more poignant. On the 1992 date, I—along with colleagues, faculty, and assorted Silicon Valley friends—held a birthday party for HAL. I was interviewed by several papers, and an Associated Press photo of me cutting the HAL cake (shaped like his console, complete with red LED under a clear plastic hemisphere) appeared worldwide. I was pleasantly surprised to learn that much of the general public was interested in HAL too.

It has been particularly rewarding for me to work with this group of contributors—all of whom were chosen because of their preeminence in their respective subfields. I have known a few of them personally for many years; Azriel Rosenfeld was on my dissertation committee. Others I met serving on panel discussions. I'll never forget the time I came dressed in a suit while fellow panelist Marvin Minsky showed up in a Pac Man T-shirt. Yet others I knew primarily through their books—Dan Dennett and Don Norman, for example—and still others are inventors of products I use regularly (e.g., Steve Wolfram's Mathematica). At our meetings and dinners in Stanford, Urbana, and Cambridge, and through frequent written messages, we passed

1 *"The Best-Informed Dream": HAL and the Vision of 2001*

David G. Stork

With most science fiction films, the more science you understand the *less* you admire the film or respect its makers. An evil interstellar spaceship careens across the screen. The hero's ship fires off a laser blast, demolishing the enemy ship—the audience cheers at the explosion. But why is the laser beam visible? There is nothing in space to scatter the light back to the viewer. And what slowed the beam a billionfold to render its advance toward the enemy ship perceptible? Why, after the moment of the explosion, does the debris remain centered in the screen instead of continuing forward as dictated by the laws of inertia? What could possibly drag and slow down the expanding debris (and cause the smoke to billow) in the vacuum of outer space? Note too the graceful, falling curve of the debris. Have the cinematographers forgotten that there is no gravity—no "downward"—in outer space? Of course, the scene is accompanied by the obligatory deafening boom. But isn't outer space eternally silent? And even if there were some magical way to hear the explosion, doesn't light travel faster than sound? Shouldn't we *see* the explosion long before we *hear* it, just as we do with lighting and thunder? Finally, isn't all this moot? Shouldn't the enemy ship be invisible anyway, as there are no nearby stars to provide illumination?

But with other, less numerous films, the more science you know the more you appreciate a film and esteem its makers. *2001* is, of course, the premier example of this phenomenon. Director Stanley Kubrick and author Arthur

C. Clarke consulted scientists in universities and industry and at NASA in their effort to portray correctly the technology of future space travel. They tried to be plausible as well as visionary. Every detail—from the design of the space ship, the timing of the mission, and the technical lingo to the typography on the computer screens and the space stewardesses' hats (bubble-shaped and padded to cushion bumps in the zero gravity of space travel)—was carefully considered in light of the then-current technology and informed predictions. The film, which has been used in the training of NASA astronauts, doesn't look dated even though thirty years have passed since its release.

We acknowledge, of course, that science fantasy is a literary or cinematic genre and need not get the science right to succeed as art. Indeed, *2001* succeeds on the strength and boldness of its vision, the profundity of its central thesis, and the clarity and unsurpassed mastery of its cinematic technique. Nevertheless, the incorporation of science and technology—*real* science and technology—is a vital part of its success, a part that has not, until now, received adequate consideration.

It is widely believed that architects design their best buildings when they are confronted by obstacles and challenges. Think of Frank Lloyd Wright's Fallingwater house, nestled among rock outcrops and perched over a stream. So too, faithfulness to scientific constraints led Kubrick and Clarke to create especially brilliant cinematic solutions. Would a lesser director have obeyed the laws of physics and portrayed Frank's murder or Dave's reentry through the emergency airlock in *silence*?

Virtually all the film's rare departures from scientific veracity were deliberate compromises by Kubrick and Clarke. For instance, even though *Discovery*'s speed is extremely high by terrestrial standards, the ship would not appear to move relative to the stars. The filmmakers were well aware of this. But when test sequences showing the stars motionless in the background made the ship look too static, they compromised and introduced a slow drift of the stars. (This solution was immensely challenging technically and required meticulous microalignment of many separate sequences.) Similarly, although in reality half of *Discovery* would have been invisible to any cosmic viewer—because it would receive no light from the sun—Clarke and Kubrick, realizing that such a half-visible ship would distract the audience, reluctantly illuminated all of it. In short, the filmmakers knew and cared about getting the science right and made as few artistic exceptions to accuracy as possible. Their care extended, too, to HAL, the central and surely the most memorable character in the film.

But before we turn to HAL, we might ask: Why analyze the science in the film at all? Why not just consider the film as art, with its own conventions and logic? We believe that, just as an art history book can deepen our understanding of a painting or sculpture, scientific analyses can lead to a richer aesthetic experience of the film. We seek to see the film from an additional, fresh perspective—not to diminish its art, but to appreciate it more fully. Such a scientific analysis also provides those of us who are not working scientists an opportunity to learn more about the research going on in real laboratories and about the recent history of science—computer science in particular.

Consider, for example, how such a perspective augments a traditional cinematic view in relation to the issue of software. In Clarke's novel (adapted from the screenplay), HAL's birthday is January 12, 1997, whereas in Kubrick's passage in the screenplay it is 1992. Why the difference? A traditional analysis (centered on character development, plot devices, and so forth) might suggest that Kubrick wanted to make HAL's life somewhat longer in order to make his death more poignant. But from our—and Clarke's—technological perspective, the 1992 date is implausible; there is simply no need for such a long history. Who would use a nine-year-old computer on the most technologically sophisticated and challenging adventure in the history of mankind? In fact, HAL's software could be downloaded onto a 1997-vintage computer in a few days (and wouldn't require inserting millions of floppy disks). Consequently, there is no scientific reason for the early birthday.

Taking our scientific and technological perspective a step farther, we can then ask: What else does this lack of appreciation of software imply, and how was it manifested in the film? Kubrick and Clarke—and indeed all but a few computer visionaries in the 1960s—failed to understand the important and unique nature of software: that it is general purpose, infinitely malleable, and can be divorced from hardware. This lack of understanding helps explain the excessive number of control buttons we see in the film, especially in the pods. Currently, jetliners and fighter jets are equipped with numerous computer screens that display different types of information and replace mechanical buttons. One good computer screen with windows and software buttons would have sufficed for *Discovery*. But *2001* was made before the Macintosh computer interface was developed, so we can't blame Kubrick and Clarke for overlooking this important trend, nor for the click of the (analog) shutter of a bulky camera where software (and smaller, digital cameras) would be used today. And, as Stephen Wolfram points out, the snippets of computer code visible on HAL's screens are reminiscent of BASIC or Fortran, popular computer languages at the time the film was made. Presumably, the software written for HAL's hardware is much less intuitive (i.e., easy for humans to understand) than languages developed since then.

In short, our technical analysis of the software issues suggested by the film draws together such details as HAL's birthday, the click of a camera, and the plethora of buttons, as well as snippets of code on the screens. It's hard to imagine how a traditional cinematic or literary analysis could shed such light or deepen our understanding of the film in this way.

In fact, *2001* is suffused—explicitly and implicitly—with the motif of birthdays: everything from the "dawn of humanity" and the explicit date when HAL "became operational," to the birthday of Heywood Floyd's daughter "Squirt" (played by Kubrick's daughter Vivian), to Frank's birthday (complete with his parents singing "Happy Birthday" via a radio transmission delayed by the immense distances of inter-planetary space), to the final scene—the birth of a star child. All this, of course, occurs at the birth of the millennium.

As the chapters in this book illustrate, the film ranges widely over issues that are still salient in computer science (and in space travel), and presents a rich array of "predictions"—though Clarke prefers to consider them "visions." It is a testament to the thoroughness of its makers that the film re-surpassed the vision of a HAL. As David Kuck (chapter 3) points out, recent advances in computer design and the spectacular and continuous rate of improvement in hardware summarized by Moore's law tell us that we could soon build a computer the size and power of HAL. Ravishankar Iyer, reviewing the progress in computer reliability, shows how we could make the hardware of a massive computer like HAL reliable, even for a prolonged space mission. Alas, the techniques for insuring reliable hardware are more effective than those for software, and this is surely an imposing, but not insurmountable, barrier. In principle, there seems to be nothing to prevent us from making a large computer that is fault-tolerant. Does this sound too much like the hubris of HAL and his designers? As Frank points out in that context, making a computer with no errors "sounds a little like famous last words."

Still, in limited-application domains, we *have* made steady improvements. We have computer chess systems that beat all but a few dozen human grandmasters, and they are improving every year. It seems all but certain that the world's best chess player will soon be a computer—perhaps even by the year 2001. In chapter 5, Murray Campbell, a member of IBM's Deep Blue team that challenged Garry Kasparov in February of 1996, analyzes *2001*'s chess scene in fascinating detail. Campbell also reflects on the changing public perception of chess machines, as illustrated in the film.

We have made several important strides toward automated speech recognition, especially in the initial stages of transcribing raw sound into phonemes. Kurzweil summarizes that progress and applies his own commercial VOICE speech-recognition system to the *2001* soundtrack. He finds that the system's simple phonological transcription is good, and can be particularly accurate when restricted to just two talkers (e.g., Frank and Dave in the quiet spaceship). General speech recognition, however, relies very heavily on semantics, common sense, context and world knowledge. (For example, how do we give a computer the ability to distinguish among such homonyms as *their, they're,* and *there*?) We are still far from solving problems like this one.

Similarly, in my own research (chapter 11), I have developed speechreading (lipreading) systems that use both sight and sound. These systems outperform purely acoustical speech recognizers in noisy rooms. Alas, no current system even remotely approaches HAL's proficiency at speechreading in silence. In fact, the lipreading scene in the pod is the only one Clarke thought was technologically implausible for the year 2001. Automatic speech recognition and speechreading will always be limited by the problems of representing semantics, common sense, and world knowledge—profoundly difficult issues that will occupy science for many decades to come.

As humans, we take our faculties of vision for granted, but making computers see has proven to be extremely difficult. In chapter 10, Azriel Rosenfeld discusses some of the successes of research in "early" vision, such as edge and motion detection, face tracking, and the recognition of emotions (which he illustrates with images from *2001*). Full vision, however, would include the ability to analyze scenes. Imagine, for example, a computer that could look at an arbitrary scene—anything from a sunset over a fishing village to Grand Central Station at rush hour—and produce a verbal description. This is a problem of overwhelming difficulty, relying as it does on finding solutions to both vision and language and then integrating them. I suspect that scene analysis will be one of the last cognitive tasks to be performed well by computers.

Other capabilities have proven remarkably difficult to develop, including one that Clarke wasn't sure we could solve by 2001: making a computer produce natural-sounding speech. In chapter 6, Joe Olive reviews the development of computer speech generation, such as the systems needed to convert text to speech for the blind. Although these experimental systems work adequately for short utterances or single words, with sentences (let alone entire speeches) they cannot convey the human subtleties of stress and intonation. A convincing artificial speaking system would require the system to understand what it is saying—again, an extremely hard and unsolved problem.

Reading Roger Schank's discussion of language in chapter 8 is like seeing the Wizard of Oz behind the curtain. With a few simple tricks, we could duplicate some of HAL's linguistic performance. For instance, it would be a fairly straightforward task to record a large number of canned stories and have HAL play them back when he hears appropriate "trigger" questions. Such a program might even persuade a casual or unsuspecting person that HAL understood language. Indeed, a noted computer program (ELIZA) around the time of *2001*'s release, sought to mimic a Rogerian therapist; in limited dialogues it convinced naive users that they were conversing with a real person. Such demonstrations tell us more about the novelty of computer discourse of the time and the gullibility of users than they do about true machine intelligence. ELIZA and its descendants are a far, far cry from true language understanding and general intelligence.

In the film, HAL makes several plans: to test whether the AE35 unit was in fact faulty, to navigate the ship, to search for extraterrestrial life around Jupiter, and to kill the crew. Such planning, Dave Wilkins points out in chapter 14, requires the identification of subgoals, anticipating obstacles, retracing steps, and so on. His review of the progress in planning by computer demonstrates the difficulty of solving even apparently simple problems in a rarefied and idealized world of stacking blocks. The bottom line is that planning is *hard*! HAL was not particularly good at it, and neither are current computer systems.

As the characters in the film admit, it is hard to tell whether HAL's emotions are programmed in to make him easy to talk to, or whether they are genuine. Rosalind Picard (chapter 13) discusses HAL's emotions—his pride, anger, fear, paranoia, concern—as well as his ability to recognize emotion in crew members—and how current research approaches similar problems. If emotions are essential for computer cognition, as she and Minsky and Norman argue, how will we deal with such "affective" computers? Would you trust an affective computer with your spreadsheets?

Don Norman (in chapter 12) underscores the importance of emotion, and points to a notion of machine (and human) intelligence much broader than the kind of logical intellect that preoccupies the field. He looks at what it might be like to live in space and work with computers for extended periods. His discussion stresses the need to take into consideration the "softer" aspects of cognition, those related to emotions, making mistakes, and so forth.

It's interesting that Kubrick and Clarke seem to want us to have stronger feelings toward HAL than we do toward the crew. HAL is the only one in the film to show emotions: "I'm afraid. I'm afraid. I'm losing my mind. I can feel it. I can feel it." In contrast, the dull, robotlike astronauts sleepwalk through boring meetings and chat about ham sandwiches. (Three other characters are known solely through the trace of their biological functions during hibernation.) When the BBC announcer, Mr. Amer, greets the crew, they mumble indistinctly in response. HAL, however, answers with clarity, animation, and interest: "Everything is going extremely well." By the middle of the movie, we have grown accustomed to HAL and accept the fact that he has more personality than the crew. While we may be shocked and surprised at the death of Frank, HAL's plaintive "I'm afraid . . . I'm afraid" evokes our sympathy.

In fact, the audience's sympathy (or anger or both) toward HAL makes more poignant a number of ethical questions. Does HAL *murder* Frank and the three hibernating crewmen? If so, who (or what) should be punished? Is it immoral to disconnect HAL (without a trial!)? Daniel Dennett argues that higher-order intentionality and responsibility are necessary conditions for moral responsibility—and hence blame—and that HAL exhibits such intentionality. After all, he tells Dave "I want to help you" (though, of course he says this to try to stop Dave from dismantling him).

A broad question touched on throughout this book is whether we should try to make computers intelligent by mimicking a human brain or, instead, exploit their particular strengths—such as rapid search and large memories. So far, different domains have tried different approaches. Researchers in computer chess, for instance, began by trying to reproduce the methods of human grandmasters—in particular recognizing key configurations of pieces on the board—but found that this approach quickly led to such difficulties as determining and representing the key properties of the arrangements. Massive and rapid searches of possible sequences of moves have proven more successful. This approach is used for Deep Blue, which has rapid search capacities that are distinctly unlike those of human grandmasters (although grandmaster Garry Kasparov has commented that Deep Blue plays like an intelligent human). Likewise, the best computer speechreading systems operate not at all like humans. Both of these applications are limited and—as valuable and interesting as they may be—do not lead us toward true AI.

Kurzweil's brain-scan approach is the most extreme of the duplicate-a-human-brain approaches. Certain cognitive abilities—in particular, language and our species-specific common sense—are distinctive to human beings. There are strong arguments for trying to copy these capacities in computers. While this would not require duplicating a brain to the level of nerve cells, it would, presumably, involve identifying the specific methods we use to carry out myriad tasks.

Another theme suggested by *2001* is society's perception of computers, contemporaneous and future. Were public fears about computers in the 1960s borne out? The fact that Frank takes his loss to HAL at chess without the slightest surprise reflects an attitude that is radically different from the public's perception in that decade when the thought of a computer becoming the world chess champion evoked anger and hostility.

As we approach 2001, we might ask why we have not matched the dream of making a HAL. The reasons are instructive. In broad overview, we have met, and surpassed, the vision of HAL in those domains—speech, hardware, planning, chess—that can be narrowly defined and easily specified. But in domains such as language understanding and common sense, which are basically limitless in their possibilities and hard to specify, we fall far short. Perhaps too we need to ask whether, as a culture, we are willing to support the undertaking of producing artificial intelligence.

The vision that produced HAL was clearly that "bigger is better." In fact, one of the major trends that the filmmakers did not foresee was that there was "room at the bottom"; that is, that computers would become smaller and smaller and evolve into doorknobs and pocket pagers. There are no personal computers or personal digital assistants in *2001*; Frank and Dave take notes with pen on paper tablets on clipboards. The role of networks is not explored, although it is unclear how they would have been incorporated into the story even if the creators had been fantastically prescient enough to predict them.

2 Scientist on the Set: An Interview with Marvin Minsky

David G. Stork

David G. Stork: You—along with John McCarthy, Claude Shannon, Nathaniel Rochester, and others—are credited with founding the field of artificial intelligence (AI) at the famous Dartmouth conference in 1956. A decade later, in the mid-sixties, when Clarke and Kubrick began work on *2001*, where was the field of AI? What were you trying to do?

Marvin Minsky: Well, we were trying to make intelligent machines. There were lots of good students working on interesting and important problems. In 1964 Tom Evans's program ANALOGY had excellent results on automated analogies—you know, figure A is to figure B as figure C is to what . . .

Jim Slagel wrote a program that could get an *A* on an MIT calculus exam. This is a tricky domain because, unlike simple arithmetic, to solve a calculus problem—and in particular to perform integration—you have to be smart about which integration technique should be used: integration by partial fractions, integration by parts, and so on. Around 1967 Dan Bobrow wrote a program to do algebra problems based on symbols rather than numbers.

It was somewhat later, 1974, that Eugene Charniak wrote a program to try to do word problems in algebra—of the sort, "Mary bought two pounds of flour at fifty cents a pound, and bought three days later a pound of sugar at two dollars. How much did she spend altogether?" He found this was extremely difficult, and basically his program didn't work.

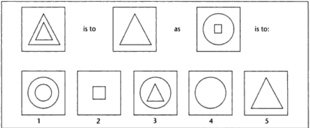


Figure 2.1
A Sample Problem of the General Sort Solved by Evans's ANALOGY Program

Stork: Was this because of insufficiently sophisticated language understanding, or instead lack of common sense or world knowledge?

Minsky: World knowledge. For instance, in the flour problem, how would the computer know that Mary didn't buy three *days*? In fact, somewhat later, around 1970, Terry Winograd addressed a simple Blocks World environment made up of simple objects and actions—"put the big block next to the smallest block," and so on—and showed that one could have a fairly seamless transition from syntax to semantics. So language, as difficult as a domain as it is, was not the obstacle to Charniak's program.

Stork: In the late sixties did you really think that toy world domains such as Blocks World captured all the essential aspects of intelligence?

Minsky: I did then, and I *still* do! In fact, I think it was the move away from such problems that is the main reason for lack of progress in AI. The problem is that in working on specific problems (such as chess, character recognition, and so on), there is no depth.

I think a key to AI is the need for several representations of the knowledge, such that when the system is stuck (using one representation) it can jump to use another. When David Marr at MIT moved into computer vision, he