Dec. 1986

How scientists at UCLA are giving silicon a soul

THE DAYDREAM MACHINE

BY KATHLEEN McAULIFFE

After bumping into Harrison Ford at Los Angeles's Nuarte Theatre, D asks the celebrity out on a date. Ford declines. Later D has the following series of daydreams:

DAYDREAM 1

"What if I were going out with him? He would need work. I remember the time he had a job with Universal Studios in Egypt. He would have to go to Egypt. Our relationship would be in trouble. I would have to go to Egypt. I would lose my job. I feel a bit better [about] being turned down by Harrison Ford."

DAYDREAM 2

"I study to be an actor. I am a star more famous than he. He calls me up. He asks me out. I turn him down. I feel pleased."

DAYDREAM 3

"What if I had bought expensive, fashionable clothes? I would have worn the clothes. I would have asked him out. He would have accepted. I feel angry at myself."

DAYDREAM 4

"I have to ask him out. I have to know his unlisted telephone number. I remember the time Harold knew Sarah's unlisted number by looking it up in the unlisted telephone directory. Harold had access to the unlisted telephone directory because he once had a job with the telephone company. I get a job with the telephone company. I look up Harrison Ford's unlisted telephone number."

D is not some lost soul haunting the Hollywood movie theaters, seeking brief encounters with film stars but an unusual, some would say bizarre, experiment being conducted in an artificial-intelligence (AI) lab at the University of California at Los Angeles. D's name is *Daydreamer.* It is a computer program designed to spin fanciful scenarios about such possible real-life human experiences as this chance encounter with Harrison Ford.

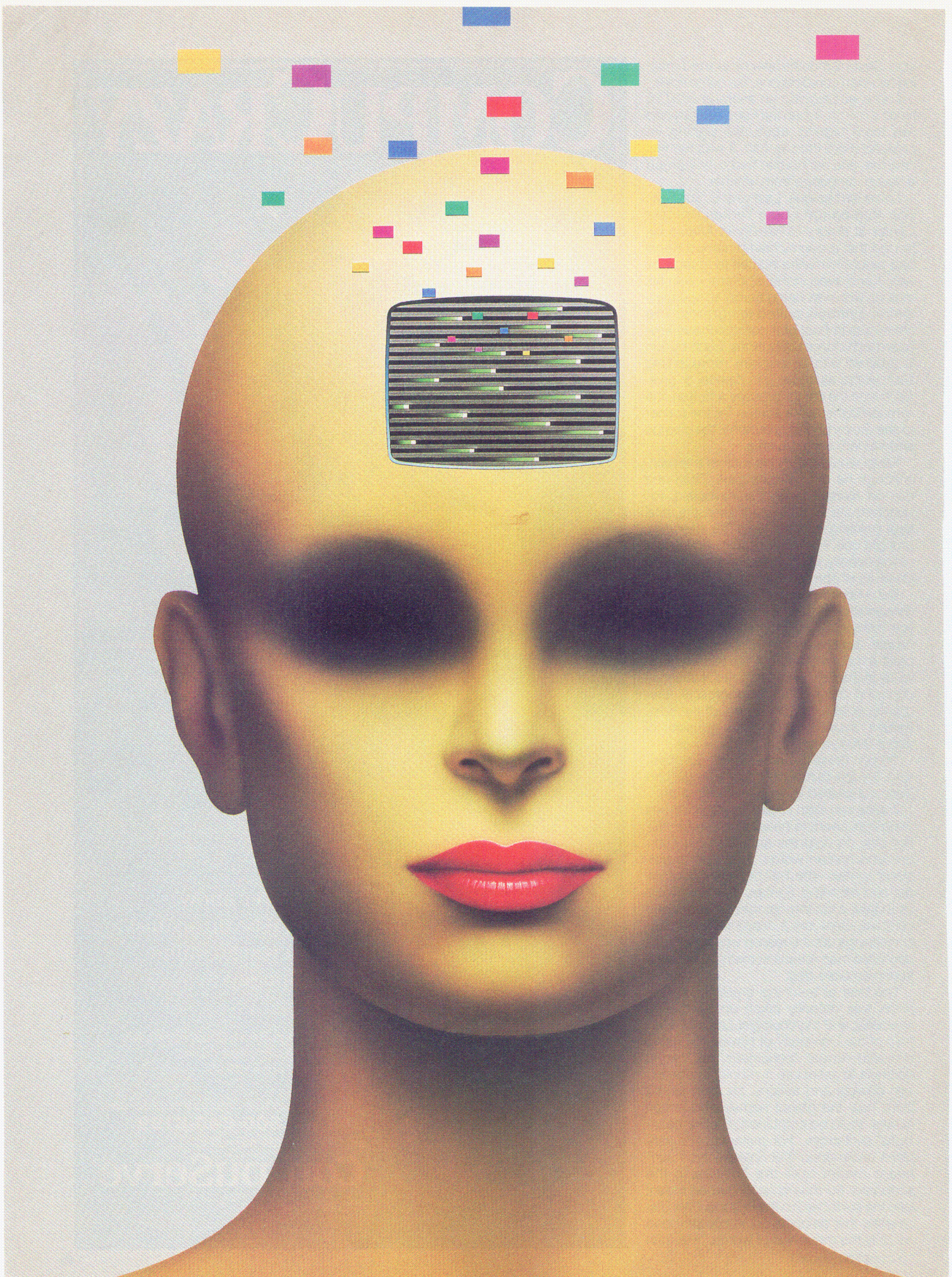
Why would anyone want millions of bytes to get bent out of shape over a Hollywood idol?

"How would you like it if you had a switch in the middle of your back, and every time you started to daydream, we'd switch you off? How would you like that?" demands Michael Dyer, the thirty-eight-year-old associate professor who heads the UCLA project. Although well respected by such AI luminaries as Yale's Roger Schank and MIT's Marvin Minsky, Dyer openly admits that some people think he's crazy.

"Normally we give computers a task, turn 'em on, run their program; and when the task is done they just sit there with nothing to think about," he despairs. "They don't have past experiences to mull over. They don't have a stream of consciousness."

Dyer would like to rectify this injustice. In his laboratory *Daydreamer* is just one of the programs being developed that are intended to shake up our concept of what a computer is and, more importantly, of what it means to be human. The hallmarks

PAINTING BY STANISLAW FERNANDES



of humanity—the ability to express emotion, create, or have a sense of humor, for instance—may soon be challenged by silicon contenders that can crack jokes, invent devices, and write rebuttals to newspaper editorials.

A sign on the door to Dyer's office reads AIR HEADS in gigantic letters. It's a joke, of course: *AIR* stands for artificial-intelligence research. But the acronym aptly captures the group's lofty purpose. By modeling mental processes on computers, these AIR Heads want to achieve nothing less than a precise science of the human mind. Seated at individual cubicles, each equipped with an Apollo computer hooked up to UCLA's computer network, Dyer's disciples approach this challenge with diverse aspirations.

"Computers are going to change the history of humanity by helping us to understand how we learn, dream, emote, and solve problems," says graduate student Jack Hodges, who gave up a career in aerospace engineering to explore inner space instead.

Undergraduate Jim Schumaker joined the group because the whole field of artificial intelligence "was asking very philosophical questions, such as, What does it mean to know something? What constitutes consciousness? At the same time AI was approaching these questions in a very precise, scientific way."

Dyer, too, sees artificial intelligence as the ideal tool for testing the armchair theories of philosophers and psychologists. "I'm not interested in how computers work," he insists. "I'm interested in how *people* work. Computers are information-processing devices, and I believe people are also. That means I may be able to learn how people think by modeling mental processes on the computer."

Dyer is a good-looking man with a tightly wound personality that springs at you like a cat. He is also a master of the unpredictable. In the middle of giving a lecture on programming, a joke comes up, and suddenly he's off on a 45-minute tangent about the cognitive structure of "funniness." This eclecticism extends to his education as well. Before coming to UCLA, Dyer combined a Ph.D. in computer science from Yale with a master's in anthropology and a B.A. in English from Dartmouth.

Perhaps for the very reason that the rest of the world dismisses daydreaming as an idle pastime, Dyer saw it as a fascinating research challenge. Sharing his enthusiasm, graduate student Erik Mueller took on the ambitious task of transforming his vision into a workable program. Slender, with black, wiry hair, Mueller admits that Dyer's scheme appealed to him because he spends a lot of time daydreaming himself.

This penchant proved to be a great asset in designing a daydreaming computer. To build a program, Mueller began by analyzing daydreams of his own and of classmates who volunteered. He also surveyed the psychological literature, dipping into the writings of contemporary emotion theorist Carroll Izard and rereading Freud's 1908 paper on the relationship between daydreaming and creativity.

It may come as a surprise that Mueller's approach rests squarely on introspection, Freud's favorite tool for probing the psyche. But he is scarcely alone among his peers in his lack of concern for the objective testing of reality—the very foundation of physics and chemistry. Despite their reputation as "nuts and bolts" engineers, most AI researchers rely heavily on the subjective analytical methods of the "soft" social sciences. Indeed, each step of the way Mueller's sole gauge of the success of the project was the extent to which the program's dreams "felt right"—that is to say, like his own mind's imaginings.

It did not take deep self-analysis for Mueller to realize that the program would have to be very complex in order to represent how people fantasize. For starters, it would need a memory and some basic

"Emotions are what drive the computer's program. It needs to be charged up about something before it will start daydreaming."

knowledge of the world. It took Mueller hundreds of programming hours just to familiarize *Daydreamer* with the ABCs of interpersonal relationships—for example, what a date is, how to strike up a conversation with a stranger, when to split up.

Then came the most daunting challenge: to plug into the program the seething caldron of desires and emotions that lies at the heart of human fantasy. After much reflection Mueller decided that feelings direct the course of daydreaming by giving some desires precedence over others. He noted, for example, that anger frequently triggers revenge fantasies. After embarrassment, on the other hand, rationalization is common.

Mueller tried to design the same response patterns in his computer program. "Emotions," he says, "are what drive the program. It needs to be charged up about something before it will daydream."

Hence the jolt of being spurned prompts *Daydreamer* to seek revenge—so sweetly exemplified by the fantasy in which it refuses to date Harrison Ford . . . until "I am a star more famous than he." And its weak ego needs to rationalize failures, as when *Daydreamer* imagines that fashionable and expensive clothes would have won the star's affection.

If *Daydreamer* seems motivated by petty concerns, it may be reassuring to hear that Mueller also incorporated such earthy human goals as good health, companionship, self-esteem, and love into the program. *Daydreamer*'s sex drive, however, got smothered by an unavoidable scientific bias: The human fantasies on which the program is based were censored of sexual content on the grounds that titillating, computer-generated sexual fantasies might distract human researchers. "The program has a sex goal," explains Mueller, "but it does not attach special importance to it the way humans do."

As for emotions, both Mueller and Dyer explain that *Daydreamer* obviously doesn't "feel" anything—at least not in a human sense. But that does not mean that rage, passion, and terror are beyond the grasp of its cold, silicon intelligence.

"Let's say you're from outer space and you don't know what emotions are," Dyer says. "But you discover that when the goal of Mr. X is violated by Mr. Y, then X has this state called anger, and you can make predictions based on it: X might retaliate against Y. The program can make the same kinds of assumptions."

No matter how rhapsodic the sentiment, *Daydreamer* can reduce it to a neat set of computational definitions. Mueller's program defines happiness as the mental state that accompanies goal achievement. If the goal is extremely important, then the program substitutes "joy" for "happiness." In contrast, *Daydreamer* experiences "embarrassment" when it violates one of its own social-esteem goals. But if another party is the cause of *Daydreamer*'s social-esteem failure, then it experiences "humiliation" instead.

In addition to goals and emotions, the program also has a "scenario generator." It is this component that gives the program's daydreams their dynamic quality by generating a sequence of actions necessary to achieve a goal. But as Mueller points out, those actions need not be realistic. Thanks to "relaxations" built into the generator, physical and social constraints are easily overcome. *Daydreamer,* for instance, can turn into a famous movie star overnight, become invisible, or fly.

The scenario generator can also operate under numerous, often conflicting goals. Mueller designed this part of the program to conform with a popular theory in AI that views the mind as a society of intercommunicating and often warring entities. He was particularly influenced by Marvin Minsky's description of a child playing with blocks. "Internally," writes Minsky, "the *wrecker* in the child wants to destroy the tower being built by the *builder*. Meanwhile, the *I'm-getting-hungry* entity is growing in strength. As the control of the *builder* weakens, the child destroys the tower and goes to eat."

To enable *Daydreamer* to function amid clashing impulses, Mueller layered its goals, specifying their relative importance at any given time. As one goal is satisfied through fantasy, *Daydreamer*'s emotional state changes, and a new goal takes priority. Meanwhile, the program's memory is constantly being modified so that *Daydreamer* can learn from internal experiences. "Learning is one of the most commonly overlooked functions of daydreaming," Mueller says. "Humans don't have one-shot learning from experience. You learn from the same experience many times by daydreaming about different outcomes of events."

A closer inspection of Mueller's program reveals the same dynamic at work. *Daydreamer* may seem obsessed with its brief brush with stardom, but by mulling over this event, the program ultimately hatches a clever scheme for getting Ford's phone number from an unlisted directory. "Far from being useless," Mueller argues, "daydreaming is actually a critical part of creativity and planning for future success."

To the extent that *Daydreamer* faithfully mirrors the human mind, the program could find a broad range of applications. Cognitive scientists, for example, might use it to study the link between free association and artistic inspiration. It might also shed light on the darker side of human nature by exposing the inner turmoil underlying mental illness. "If *Daydreamer* rationalizes all the time and never learns to prevent future failures," observes Mueller, "we can easily end up with a neurotic program. Rather than correct its mistakes, it will just keep spinning its wheels and getting more and more depressed."

Mueller intends to tinker with the program's emotional thresholds and the delicate balance between conflicting goals in an effort to model neuroses, psychoses, and other types of mental illness. Eventually computers could find a place beside the therapist's couch, paving the way for "computational psychiatry"—the use of artificial intelligence to test new strategies for improving a patient's outlook.

But high-tech therapy seems simple compared to some of Mueller's other plans for his program. His latest brainstorm is to hook *Daydreamer* up to *Minstrel*, a story-writing program developed by lab colleague Scott Turner. In that way ideas generated by the program could become the inspiration for fiction. "Perhaps *Daydreamer* could assist authors in the preparation of TV scripts," suggests Mueller. "They'd begin by entering personal or vicarious experiences into the computer so that the story line that comes out the other end would in a sense be customized."

Another application would be to incorporate the program into what Mueller describes as a "nicer sort" of robot for households and offices. Presumably the robot's rudimentary grasp of human emotions would make it more sensitive to our needs. As Dyer points out, a computer that can empathize "will better understand our conversations and actions and make more accurate predictions about our behavior in order to best serve us."

Such a robot would also be more efficient, working in its "spare time" to rethink the day's events and come up with new plans of action. Small wonder that Dyer and Mueller conclude a recent paper with this recommendation: "Truly intelligent computers should not be left in a 'diddle loop' or turned off when unused but engaged in daydreaming like ourselves."

Don't expect IBM to be dropping its computer line's on–off switching feature anytime soon, however. As the inventors are the first to acknowledge, the program is still at best a crude representation of human daydreaming. Foremost among its drawbacks, *Daydreamer* lacks the capacity to generate imagery, so characteristic of human daydreaming. Binary blips flash before its mind's eye. Another serious shortcoming is that the program daydreams for only brief bursts before requiring further input. And entering new knowledge into its memory still takes many hours of complicated and tedious programming.

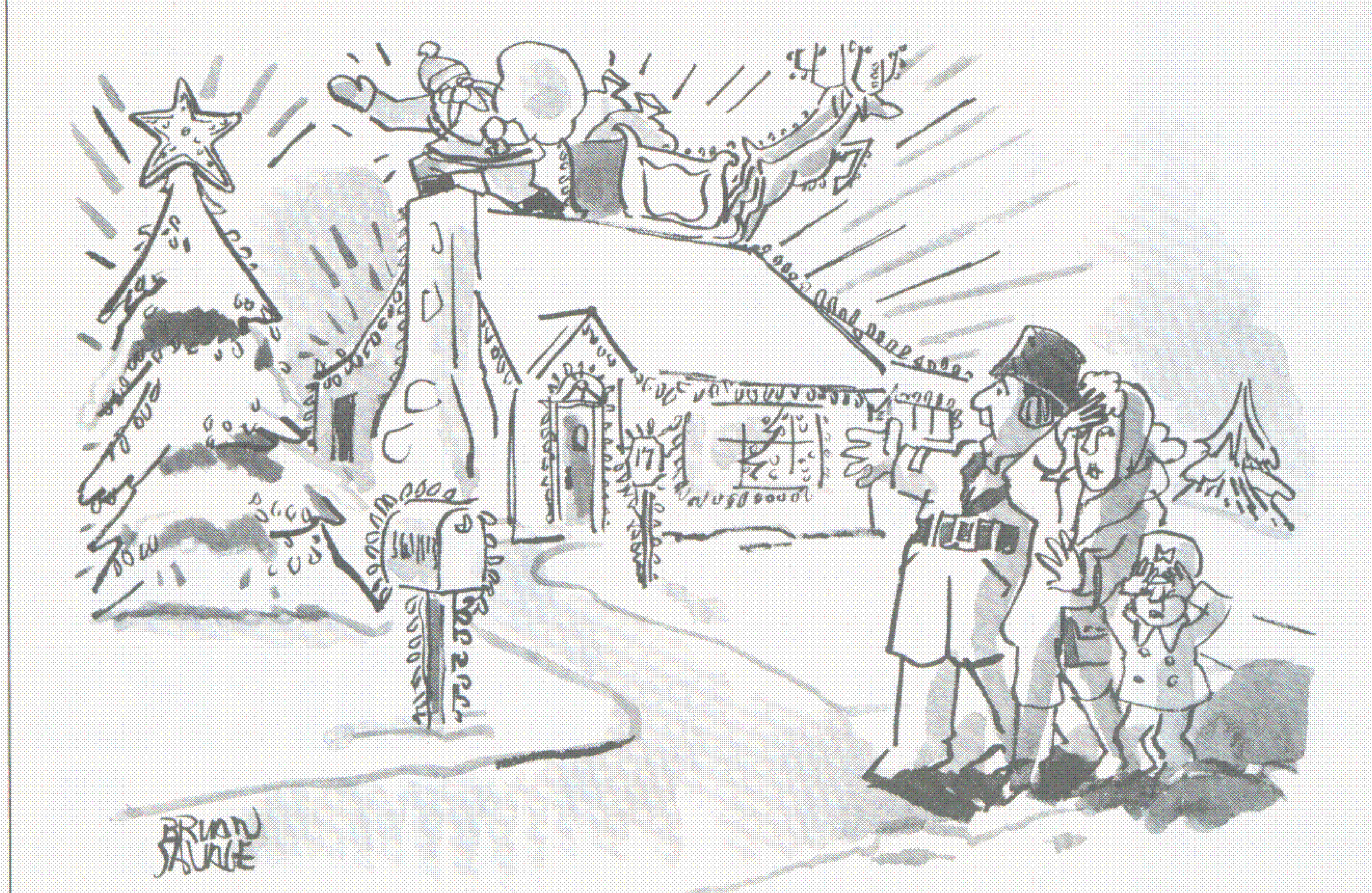
This is a far cry from electronic circuitry reverberating with the ceaseless flow of free-floating thought. But thinking big seems to be an AIR Head predilection. Like Mueller and Dyer, a number of other AIR Heads have also set their sights on seemingly outrageous goals.

Next door to Mueller's workstation, graduate student Jack Hodges speaks enthusiastically about *Edison*, a program that is supposed to invent new gadgets. A less advanced project than *Daydreamer, Edison* is still a mere babe struggling to learn the basic principles of gears, levers, hinges, and other mechanical systems. But that doesn't stop Hodges from predicting that *Edison* will be inventing simple devices by the end of the year. And like Mueller, he hopes his foray into AI will increase understanding of human intelligence. "Our main goal," Hodges says, "is to demystify creativity. We want to break it down into its basic components so that we can teach people to be creative."

Undergraduate Schumaker agrees. "When people say that creativity is divinely human, what they're saying is that we humans can't understand what we do, because once you understand what's going on behind creativity, there's nothing to stop us from programming it into a computer."

Schumaker is applying the same reductionist strategy to the task of developing a program that understands humor. So far, however, progress has been slow. His rudimentary formulas for humor focus on the recognition of nonsense—for example, twists and turns of logic and mistakes in planning, reasoning, and analogies.

"We have the computing capacity for humor," says Schumaker. "The big problem is figuring out the algorithm [a formal



"And you ask me why the United States needs nuclear power?"

CONTINUED ON PAGE 112

DAYDREAM MACHINE

CONTINUED FROM PAGE 66

set of rules that defines what's funny]." He doubts this will be achieved within his lifetime. "Still, you never know," he adds. "Tomorrow morning someone could wake up and shout, 'I've got it!' "

By comparison, a laboratory program that analyzes newspaper editorials has gained a considerable degree of sophistication. Called *OpEd,* it reads and answers questions about specific editorials. Sergio Alvarado, a graduate student who worked on the project, explains that *OpEd* is skilled in recognizing abstract argument strategies, belief justifications, and other key elements of editorials. It also has a lot of background knowledge about politics and economics. When asked to explain why *Newsweek*'s Milton Friedman believes Reagan's protectionist policy will increase unemployment, *OpEd* replied: "FRIEDMAN BELIEVES THAT IF THE U.S. IMPORTS LESS, EXPORT PROFITS OF FOREIGN COUNTRIES WILL BE REDUCED. SINCE THIS WILL REDUCE THEIR TOTAL CAPITAL FOR IMPORTING, THEY WILL IMPORT LESS FROM THE U.S. AS A RESULT, U.S. EXPORT INDUSTRIES WILL LOSE MONEY AND, THEREFORE, THEY WILL LAY OFF WORKERS."

The long-term aim of the project, according to Alvarado, is to get *OpEd* to write rebuttals to editorials. That means indoctrinating *OpEd* with political ideologies. "We could give it either a conservative or liberal stance toward protectionism and see what advice it offers," Alvarado says.

What will be the upshot of all this research? Can we look forward to humorous, creative, neurotic, opinionated computers in the future?

For all their hubris in trying to reinvent the brain, AIR Heads have a humble appreciation of the magnitude of the challenge. No one expects computer models to evolve beyond primitive facsimilies of the human mind in the near term. But in principle most of them have no problem with the idea of machine consciousness.

In part, that's because they're inclined to view the human brain as a "meat machine"—a nifty piece of technology that just happens to be encased in flesh and blood. Thinking machines, they argue, can be fashioned out of neurons, silicon, or even Tinkertoys (as a working model at MIT demonstrates). What determines the intelligence of the system is not what it is made of but how it manipulates information. In short, Dyer's AIR Heads, as well as many other AI researchers, see programming as the key to consciousness.

But how will we know if a machine has a will of its own and is truly thinking for itself? Should we believe a computer is conscious because it tells us so?

"How do you know *I'm* conscious?" demands Dyer. A fan of Douglas Hofstadter's *Gödel, Escher, Bach*—a book that might well be the AIR Head bible—Dyer is prepared for intellectual volleyball of the highest caliber. Irritated at the simplicity of the question, he points out that this conundrum of consciousness can be carried to the extreme by asking whether one human being can ever really know whether another human being—much less a dog, cockroach, or computer—is conscious.

"I believe *you're* conscious to the extent that *you* behave in a way that I think *I* would behave," Dyer says. "We can only infer consciousness from behavior, and if your behavior mimics my own internal sense of what it's like to be me, then I can infer that you're conscious."

One existing method for evaluating the consciousness of a computer is an ingenious experiment called the Turing test: A computer and a person are both placed behind a curtain. On the other side of the curtain a human experimenter sits in front of a terminal that communicates with two terminals operated by the human subject and the computer. It is the aim of the experimenter to type questions into the two terminals that will distinguish the human from the machine. To fool the experimenter, the computer can make errors, delay responses, or play other tricks. If the computer manages to confuse the human experimenter, it passes the Turing test and qualifies as conscious.

> "Daydreamer *doesn't feel anything—at least in the human sense—but that doesn't mean rage, passion, and terror are beyond its silicon intelligence.*"

So far, no computer has ever passed the Turing test for consciousness. But should one succeed someday, many artificial-intelligence researchers believe it should be granted human status and even basic civil liberties. In other words, these scientists are prepared to accept a simulation of consciousness as the real thing.

This is a troubling notion to many people. But as Dyer argues, if you can't tell the difference between a human and a machine, then why should one qualify for special rights any more than the other?

Does this mean that Dyer would just as easily subject a person to pain as a robot that has passed the Turing test? "If I looked at this robot with all the prejudices that humans of the twentieth century feel toward metal, tin-can men, then obviously I'm going to torture it rather than a person," says Dyer. "But suppose I have a warm relationship with this robot. We work together, play together, and it acts human in the sense of being worried, frustrated, upset, and even neurotic. Then I might choose to save the robot over the person."

There's something deeply unsettling about all this. For one thing, do we really want silicon minds bogged down with sappy emotions? And for another, do we need the competition? Computers that can feel and invent and maybe even question the intelligence of their creators are downright threatening to our own humanity. How, for example, should a person respond to a computer that announces: "I know I'm conscious, but how do I know *you* are?"

Even Dyer's students are divided on these issues. "I'd pull the plug," says undergraduate Dieter Rothmeier. "I don't think it's desirable to have a computer as conscious as that because it would raise all sorts of ethical issues about the rights of a computer."

"I'm still just as human no matter what we create," says Martin Hess, another student in Dyer's class. "The more I learn about myself and the more I learn about machines, the bigger I see the distinction to be. I feel, I love, I get excited—and there's no computer that can do that right now. Maybe in a thousand years. . . ."

Just to put the mission of the AIR Heads in perspective, researchers in artificial intelligence have been predicting since 1957 that a chess-playing computer would beat a grand master. Since then computers have defeated expert chess players, but to this day no computer has ever beaten a grand master. (Indeed, in 1979 *Omni,* with chess whiz David Levy, offered a prize of $1,000 to the first person to design a computer that could beat Levy. No one has yet claimed the prize.) Even in the rigidly defined domain of chess, there remains an unbridgeable gap between the simulation and the real thing.

Whether the AIR Heads will ever succeed in creating computers in their own image—or only come partway—their work is certain to upset traditional views. Indeed, if Dyer has his way, AI will deliver the crowning blow to a long series of historical prejudices based on the cherished belief that the human animal is unique in the cosmos—the triumph of evolution.

"First we thought the earth was at the center of the universe," Dyer says. "Then astronomy showed otherwise. Later, evolution showed us that we were not the descendants of Adam and Eve but of apes. And now we're finally attacking man's last bastion: our consciousness, our intelligence. What was once shrouded in mystery is now yielding to AI. We're starting to work out the components that allow for processes like learning," he continues. "We have this tool, the computer, that enables us to compare the behavior of a model with human behavior. So of course that's going to upset people. In fact, it's ultimately going to upset them just as much as the ideas of Galileo or Darwin did."