A Coffeehouse Conversation on the Turing Test

May, 1981

Douglas R. Hofstadter

Participants in the dialog: Chris, a physics student; Pat, a biology student; Sandy, a philosophy student.

Chris: Sandy, I want to thank you for suggesting that I read Alan Turing's article "Computing Machinery and Intelligence." It's a wonderful piece and certainly made me think-and think about my thinking.

Sandy: Glad to hear it. Are you still as much of a skeptic about artificial intelligence as you used to be?

Chris: You've got me wrong. I'm not against artificial intelligence; I think it's wonderful stuff-perhaps a little crazy, but why not? I simply am convinced that you AI advocates have far underestimated the human mind, and that there are things a computer will never, ever be able to do. For instance, can you imagine a computer writing a Proust novel? The richness of imagination, the complexity of the characters-

Sandy: Rome wasn't built in a day!

Chris: In the article, Turing comes through as an interesting person. Is he still alive?

Sandy: No, he died back in 1954, at just 41. He'd be only 70 or so now, although he is such a legendary figure it seems strange to think that he could still be living today.

Chris: How did he die?

Sandy: Almost certainly suicide. He was homosexual, and had to deal with some pretty barbaric treatment and stupidity from the outside world. In the end, it got to be too much, and he killed himself.

Chris: That's horrendous, especially in this day and age.

Sandy: I know. What really saddens me is that he never got to see the amazing progress in computing machinery and theory that has taken place since 1954. Can you imagine how he'd have been wowed?

Chris: Yeah . . .

Pat: Hey, are you two going to clue me in as to what this Turing article is about?

Sandy: It is really about two things. One is the question "Can a machine think?"-or rather, "Will a machine ever think?" The way Turing answers the question-he thinks the answer is yes, by the way-is by batting down a series of objections to the idea, one after another. The other point he tries to make is that, as it stands, the question is not meaningful. It's too full of emotional connotations. Many people are upset by the suggestion that people are machines, or that machines might think. Turing tries to defuse the question by casting it in less emotional terms. For instance, what do you think, Pat, of the idea of thinking machines?

Pat: Frankly, I find the term confusing. You know what confuses me? It's those ads in the newspapers and on TV that talk about "products that think" or "intelligent ovens" or whatever. I just don't know how seriously to take them.

Sandy: I know the kind of ads you mean, and they probably confuse a lot of people. On the one hand, we're always hearing the refrain "Computers are really dumb; you have to spell everything out for them in words of one syllable"-yet on the other hand, we're constantly bombarded with advertising hype about "smart products."

Chris: That's certainly true. Do you know that one company has even taken to calling its products "dumb terminals" in order to stand out from the crowd?

Sandy: That's a pretty clever gimmick, but even so it just contributes to the trend toward obfuscation. The term "electronic brain" always comes to my mind when I'm thinking about this. Many people swallow it completely, and others reject it out of hand. It takes patience to sort out the issues and decide how much of it makes sense.

Pat: Does Turing suggest some way of resolving it, some kind of IQ test for machines?

Sandy: That would be very interesting, but no machine could yet come close to taking an IQ test. Instead, Turing proposes a test that theoretically could be applied to any machine to determine whether or not it can think.

Pat: Does the test give a clear-cut yes-or-no answer? I'd be skeptical if it claimed to.

Sandy: No, it doesn't claim to. In a way that's one of its advantages. It shows how the borderline is quite fuzzy and how subtle the whole question is.

Pat: And so, as usual in philosophy, it's all just a question of words!

Sandy: Maybe, but they're emotionally charged words, and so it's important, it seems to me, to explore the issues and try to map out the meanings of the crucial words. The issues are fundamental to our concept of ourselves. So we shouldn't just sweep them under the rug.

Pat: Okay, so tell me how Turing's test works.

Sandy: The idea is based on what he calls the *Imitation Game*. Imagine that a man and a woman go into separate rooms, and from there they can be interrogated by a third party via some sort of teletype set-up. The third party can address questions to either room, but has no idea which person is in which room. For the interrogator, the idea is to determine which room the woman is in. The woman, by her answers, tries to help the interrogator as much as she can. The man, though, is doing his best to bamboozle the interrogator, by responding as he thinks a woman might. And if he succeeds in fooling the interrogator...

Pat: The interrogator only gets to see written words, eh? And the sex of the author is supposed to shine through? That game sounds like a good challenge. I'd certainly like to take part in it someday. Would the interrogator have met either the man or the woman before the test began? Would any of them know any of the others?

Sandy: That would probably be a bad idea. All kinds of subliminal cueing might occur if the interrogator knew one or both of them. It would certainly be best if all three people were totally unknown to one another.

Pat: Could you ask any questions at all, with no holds barred?

Sandy: Absolutely. That's the whole idea!

Pat: Don't you think, then, that pretty quickly it would degenerate into sex-oriented questions? I mean, I can imagine the man, overeager to act convincing, giving away the game by answering some very blunt questions that most women would find too personal to answer, even through an anonymous computer connection.

Sandy: That's a nice observation. I wonder if it's true ....

Chris: Another possibility would be to probe for knowledge of minute aspects of traditional sex-role differences, by asking about such things as dress sizes and so on. The psychology

of the Imitation Game could get pretty subtle. I suppose whether the interrogator was a woman or a man would make a difference. Don't you think that a woman could spot some telltale differences more quickly than a man could?

Pat: If so, maybe the best way to tell a man from a woman is to let each of them play interrogator in an Imitation Game and see which of the two is better at telling a man from a woman!

Sandy: Hmm . . . that's a droll twist. Oh well, I don't know if this original version of the Imitation Game has ever been seriously tried out, despite the fact that it would be relatively easy to do with modern computer terminals. I have to admit, though, that I'm not at all sure what it would prove, whichever way it turned out.

Pat: I was wondering about that. What would it prove if the interrogator-say a woman-couldn't tell correctly which person was the woman? It certainly wouldn't prove that the man was a woman!

Sandy: Exactly! What I find funny is that although I strongly believe in the idea of the Turing Test, I'm not so sure I understand the point of its basis, the Imitation Game.

Chris: As for me, I'm not any happier with the Turing Test as a test for thinking machines than I am with the Imitation Game as a test for femininity.

Pat: From what you two are saying, I gather the Turing Test is some kind of extension of the Imitation Game, only involving a machine and a person instead of a man and a woman.

Sandy: That's the idea. The machine tries its hardest to convince the interrogator that it is the human being, and the human tries to make it clear that he or she is not the computer.

Pat: The machine *tries*? Isn't that a loaded way of putting it?

Sandy: Sorry, but that seemed the most natural way to say it.

Pat: Anyway, this test sounds pretty interesting. But how do you know that it will get at the essence of thinking? Maybe it's testing for the wrong things. Maybe, just to take a random illustration, someone would feel that a machine was able to think only if it could dance so well that you couldn't tell it was a machine. Or someone else could suggest some other characteristic. What's so sacred about being able to fool people by typing at them?

Sandy: I don't see how you can say such a thing. I've heard that objection before, but frankly, it baffles me. So what if the machine can't tap-dance or drop a rock on your toe? If it can discourse intelligently on any subject you want, then it has shown that it can think-to me, at least! As I see it, Turing has drawn, in one clean stroke, a clear division between thinking and other aspects of being human.

Pat: Now you're the baffling one. If you couldn't conclude anything from a man's ability to win at the Imitation Game, how could you conclude anything from a machine's ability to win at the Turing Game?

Chris: Good question.

Sandy: It seems to me that you could conclude something from a man's win in the Imitation Game. You wouldn't conclude he was a woman, but you could certainly say he had good insights into the feminine mentality (if there is such a thing). Now, if a computer could fool someone into thinking it was a person, I guess you'd have to say something similar about it-that it had good insights into what it's like to be human, into "the human condition" (whatever that is).

Pat: Maybe, but that isn't necessarily equivalent to thinking, is it? It seems to me that passing the Turing Test would merely prove that some machine or other could do a very good job of simulating thought.

Chris: I couldn't agree more with Pat. We all know that fancy computer programs exist today for simulating all sorts of complex phenomena. In theoretical physics, for instance, we simulate the behavior of particles, atoms, solids, liquids, gases, galaxies, and so on. But no one confuses any of those simulations with the real thing!

Sandy: In his book *Brainstorms*, the philosopher Daniel Dennett makes a similar point about simulated hurricanes.

Chris: That's a nice example, too. Obviously, what goes on inside a computer when it's simulating a hurricane is not a hurricane, for the machine's memory doesn't get torn to bits by 200 mile-an-hour winds, the floor of the machine room doesn't get flooded with rainwater, and so on.

Sandy: Oh, come on-that's not a fair argument! In the first place, the programmers don't claim the simulation really is a hurricane. It's merely a simulation of certain aspects of a hurricane. But in the second place, you're pulling a fast one when you imply that there are no downpours or 200-mile-an-hour winds in a simulated hurricane. To us there aren't any, but if the program were incredibly detailed, it could include simulated people on the ground who would experience the wind and the rain just as we do when a hurricane hits. In their minds-or, if you'd rather, in their simulated minds-the hurricane would be not a simulation, but a genuine phenomenon complete with drenching and devastation.

Chris: Oh, my-what a science-fiction scenario! Now we're talking about simulating whole populations, not just a single mind!

Sandy: Well, look-I'm *simply* trying to show you why your argument that a simulated McCoy isn't the real McCoy is fallacious. It depends on the tacit assumption that any old observer of the simulated phenomenon is equally able to assess what's going on. But in fact, it may take an observer with a special vantage point to recognize what is going on. In the hurricane case, it takes special "computational glasses" to see the rain and the winds.

Pat: "Computational glasses"? I don't know what you're talking about.

Sandy: I mean that to see the winds and the wetness of the hurricane, you have to be able to look at it in the proper way. You-

Chris: No, no, no! A simulated hurricane isn't wet! No matter how much it might seem wet to simulated people, it won't ever be genuinely wet! And no computer will ever get torn apart in the process of simulating winds.

Sandy: Certainly not, but that's irrelevant. You're just confusing levels. The laws of physics don't get torn apart by real hurricanes, either. In the case of the simulated hurricane, if you go peering at the computer's memory, expecting to find broken wires and so forth, you'll be disappointed. But look at the proper level. Look into the *structures* that are coded for in memory. You'll see that many abstract links have been broken, many values of variables radically changed, and so on. *There's* your flood, your devastation-real, only a little concealed, a little hard to detect.

Chris: I'm sorry, I just can't buy that. You're insisting that I look for a new kind of devastation, one never before associated with hurricanes. That way you could call anything a hurricane as long as its effects, seen through your special "glasses," could be called "floods and devastation."

Sandy: Right-you've got it exactly! You

recognize a hurricane by its effects. You have no way of going in and finding some ethereal "essence of hurricane," some "hurricane soul" right in the middle of the storm's eye. Nor is there any ID card to be found that certifies "hurricanehood." It's just the existence of a certain kind of pattern--a spiral storm with an eye and so forth-that makes you say it's a hurricane. Of course, there are a lot of things you'll insist on before you call something a hurricane.

Pat: Well, wouldn't you say that being an atmospheric phenomenon is one prerequisite? How can anything inside a computer be a storm? To me, a simulation is a simulation is a simulation!

Sandy: Then I suppose you would say that even the calculations computers do are simulated-that they are fake calculations. Only people can do genuine calculations, right?

Pat: Well, computers get the right answers, so their calculations are not exactly fake-but they're still just patterns. There's no *understanding* going on in there. Take a cash register. Can you honestly say that you feel it is *calculating* something when its gears mesh together? And the step from cash register to computer is very short, as I understand things.

Sandy: If you mean that a cash register doesn't feel like a schoolkid doing arithmetic problems, I'll agree. But is that what "calculation" means? Is that an integral part of it? If so, then contrary to what everybody has thought up till now, we'll have to write a very complicated program indeed to perform *genuine* calculations.

Of course, this program will sometimes get careless and make mistakes, and it will sometimes scrawl its answers illegibly, and it will occasionally doodle on its paper .... It won't be any more reliable than the store clerk who adds up your total by hand. Now, I happen to believe that eventually such a program could be written. Then we'd know something about how clerks and schoolkids work.

Pat: I can't believe you'd ever be able to do that!

Sandy: Maybe, maybe not, but that's not my point. You say a cash register can't calculate. It reminds me of another favorite passage of mine from Dennett's *Brainstorms*. It goes something like this: "Cash registers can't really calculate; they can only spin their gears. But cash registers can't really spin their gears, either: they can only follow the laws of physics." Bennett said it originally about computers; I modified it to talk about cash registers. And you could use the same line of reasoning in talking about people: "People can't really calculate; all they can do is manipulate mental symbols. But they aren't really manipulating symbols: all they are doing is firing various neurons in various patterns. But they can't really make their neurons fire; they simply have to let the laws of physics make them fire for them." Et cetera. Don't you see how this *reduction ad absurdum* would lead you to conclude that calculation doesn't exist, that hurricanes don't exist-in fact, that nothing at a level higher than particles and the laws of physic exists? What do you gain by saying that a computer only pushes symbols around and doesn't truly calculate?

Pat: The example may be extreme, but it makes my point that there is a vast difference between a real phenomenon and any simulation of it. This is so for hurricanes, and even more so for human thought.

Sandy: Look, I don't want to get too tangled up in this line of argument, but let me try one more example. If you were a radio ham listening to another ham broadcasting in Morse code and you were responding in Morse code, would it sound funny to you to refer to "the person at the other end"?

Pat: No, that would sound okay, although the existence of a person at the other end would be an assumption.

Sandy: Yes, but you wouldn't be likely to go and check it out. You're prepared to recognize personhood through those rather unusual channels. You don't have to see a human body or hear a voice. All you need is a rather abstract manifestation-a code, as it were. What I'm getting at is this. To "see" the person behind the dits and dahs, you have to be willing to do some *decoding*, some interpretation. It's not direct perception; it's indirect. You have to peel off a layer or two to find the reality hidden in there. You put on your "radio-ham's glasses" to "see" the person behind the buzzes. Just the same with the simulated hurricane! You don't see it darkening the machine room; you have to decode the machine's memory. You have to put on special "memory-decoding" glasses. *Then* what you see is a hurricane.

Pat: Oh ho ho! Talk about fast ones-wait a minute! In the case of the shortwave radio, there's a real person out there, somewhere in the Fiji Islands or wherever. My decoding act as I sit by my radio simply reveals that that person exists. It's like seeing a shadow and concluding there's an object out there, casting it. One doesn't confuse the shadow with the object, however! And with the hurricane there's no *real* storm behind the scenes, making the computer follow its patterns. No, what you have is just a shadow hurricane without any genuine hurricane. I just refuse to confuse shadows with reality.

Sandy: All right. I don't want to drive this point into the ground. I even admit it is pretty silly to say that a simulated hurricane is a hurricane. But I wanted to point out that it's not as silly as you might think at first blush. And when you turn to simulated thought then you've got a very different matter on your hands from simulated hurricanes.

Pat: I don't see why. You'll have to convince me.

Sandy: Well, to do so, I'll first have to make a couple of extra points about hurricanes.

Pat: Oh no! Well, all right, all right.

Sandy: Nobody can say just exactly what a hurricane is-that is, in totally precise terms. There's an abstract pattern that many storms share, and it's for that reason we call those storms hurricanes. But it's not possible to make a sharp distinction between hurricanes and no hurricanes. There are tornados, cyclones, typhoons, dust devils .... Is the Great Red Spot on Jupiter a hurricane? Are sunspots hurricanes? Could there be a hurricane in a wind tunnel? In a test tube? In your imagination, you can even extend the concept of "hurricane" to include a microscopic storm on the surface of a neutron star.

Chris: That's not so far-fetched, you know. The concept of "earthquake" has actually been extended to neutron stars. The astrophysicists say that the tiny changes in rate that once in a while are observed in the pulsing of a pulsar are caused by "glitches"starquakes-that have just occurred on the neutron star's surface.

Sandy: Oh, I remember that now. That "glitch" idea has always seemed eerie to me-a surrealistic kind of quivering on a surrealistic kind of surface.

Chris: Can you imagine-plate tectonics on a giant sphere of pure nuclear matter?

Sandy: That's a wild thought. So, starquakes and earthquakes can both be subsumed into a new, more abstract category. And that's how science constantly extends familiar concepts, taking them further and further from familiar experience and yet keeping some essence constant. The number system is the classic example-from positive numbers to negative numbers, then rationale, reels, complex

numbers, and "on beyond zebra," as Dr. Seuss says.

Pat: I think I can see your point, Sandy. In biology, we have many examples of close relationships that are established in rather abstract ways. Often the decision about what family some species belongs to comes down to an abstract pattern shared at some level. Even the concepts of "male" and "female" turn out to be surprisingly abstract and elusive. When you base your system of classification on very abstract patterns, I suppose that a broad variety of phenomena can fall into "the same class," even if in many superficial ways the class members are utterly unlike one another. So perhaps I can glimpse, at least a little, how to you, a simulated hurricane could, in a funny sense, *be* a hurricane.

Chris: Perhaps the word that's being extended is not "hurricane," but "be."

Pat: How so?

Chris: If Turing can extend the verb "think," can't I extend the verb "be"? All I mean is that when simulated things are deliberately confused with genuine things, somebody's doing a lot of philosophical wool pulling. It's a lot more serious than just extending a few nouns, such as "hurricane."

Sandy: I like your idea that "be" is being extended, but I sure don't agree with you about the wool pulling. Anyway, if you don't object, let me just say one more thing about simulated hurricanes and then I'll get to simulated minds. Suppose you consider a really deep simulation of a hurricane-I mean a simulation of every atom, which I admit is sort of ridiculous, but still, just consider it for the sake of argument.

Pat: Okay.

Sandy: I hope you would agree that it would then share all the abstract structure that defines the "essence of hurricanehood." So what's to keep you from calling it a hurricane?

Pat: I thought you were backing off from that claim of equality.

Sandy: So did I, but then these examples came up, and I was forced back to my claim. But let me back off, as I said I would do, and get back to thought, which is the real issue here. Thought, even more than hurricanes, is an abstract structure, a way of describing some complex events that happen in a medium called a brain. But actually, thought can take place in any one of several billion brains. There are all these physically very different brains, and yet they all support "the same thing": thinking. What's important, then, is the abstract *pattern*, not the medium. The same kind of swirling can happen inside any of them, so no person can claim to think more "genuinely" than any other. Now, if we come up with some new kind of medium in which the *same style* of swirling takes place, could you deny that thinking is taking place in it?

Pat: Probably not, but you have just shifted the question. The question now is: How can you determine whether the "same style" of swirling is really happening?

Sandy: The beauty of the Turing Test is that it *tells* you when! Don't you see?

Chris: I don't see that at all. How would you know that the same style of activity was going on inside a computer as inside my mind, simply because it answered questions as I do? All you're looking at is its outside.

Sandy: I'm sorry, I disagree entirely! How do you know that when I speak to you, anything similar to what you call thinking is going on inside me? The Turing Test is a fantastic probe, something like a particle accelerator in physics. Here, Chris-I think you'll like this analogy. Just as in physics, when you want to understand

what is going on at an atomic or subatomic level, since you can't see it directly, you scatter accelerated particles off a target and observe their behavior. From this, you infer the internal nature of the target. The Turing Test extends this idea to the mind. It treats the mind as a "target" that is not directly visible but whose structure can be deduced more abstractly. By "scattering" questions off a target mind, you learn about its internal workings, just as in physics.

Chris: Well . . . to be more exact, you can *hypothesize* about what kinds of internal structures might account for the behavior observed-but please remember that they may or may not in fact exist.

Sandy: Hold on, now! Are you suggesting that atomic nuclei are merely hypothetical entities? After all, their existence (or should I say *hypothetical* existence?) was proved (or should I say *suggested*?) by the behavior of particles scattered off atoms.

Chris: I would agree, but you know, physical systems seem to me to be much simpler than the mind, and the certainty of the inferences made is correspondingly greater. And the conclusions are confirmed over and over again by different types of experiments.

Sandy: Yes, but those experiments still are of the same sort-scattering, detecting things indirectly. You can never *handle* an electron or a quark. Physics experiments are also correspondingly harder to do and to interpret. Often they take years and years, and dozens of collaborators are involved. In the Turing Test, though, just one person could perform many highly delicate experiments in the course of no more than an hour. I maintain that people give other people credit for being conscious simply because of their continual external monitoring of other people-which is itself something like a Turing Test.

Pat: That may be roughly true, but it involves more than just conversing with people through a teletype. We see that other people have bodies, we watch their faces and expressions-we see they are human beings, and so we think they think.

Sandy: To me, that seems a narrow, anthropocentric view of what thought is. Does that mean you would sooner say a mannequin in a store thinks than a wonderfully programmed computer, simply because the mannequin looks more human?

Pat: Obviously, I would need more than just vague physical resemblance to the human form to be willing to attribute the power of thought to an entity. But that organic quality, the sameness of origin, undeniably lends a degree of credibility that is very important.

Sandy: Here we disagree. I find this simply too chauvinistic. I feel that the key thing is a similarity of *internal* structure-not bodily, organic, chemical structure but *organizational* structure software. Whether an entity can think seems to me a question of whether its organization can be described in a certain way, and I'm perfectly willing to believe that the Turing Test detects the presence or absence of that mode of organization. I would say that your depending on my physical body as evidence that I am a thinking being is rather shallow. The way I see it, the Turing Test looks far deeper than at mere external form.

Pat: Hey now-you're not giving me much credit. It's not just the shape of a body that lends weight to the idea that there's real thinking going on inside. It's also, as I said, the idea of common origin. It's the idea that you and I both sprang from DNA molecules, an idea to which I attribute much depth. Put it this way: the external form of human bodies reveals that they share a deep biological history, and it's that depth that lends a lot of credibility to the notion that the owner of such a body can

think.

Sandy: But that is all indirect evidence. Surely you want some *direct* evidence. That's what the Turing Test is for. And I think it's the *only* way to test for thinkinghood.

Chris: But you could be fooled by the Turing Test, just as an interrogator could mistake a man for a woman.

Sandy: I admit, I could be fooled if I carried out the test in too quick or too shallow a way. But I would go for the deepest things I could think of.

Chris: I would want to see if the program could understand jokes-or better yet, make them! *That* would be a real test of intelligence.

Sandy: I agree that humor probably is an acid test for a supposedly intelligent program, but equally important to me perhaps more so-would be to test its emotional responses. So I would ask it about its reactions to certain pieces of music or works of literature-especially my favorite ones.

Chris: What if it said, "I don't know that piece," or even, "I have no interest in music"? What if it tried its hardest (oops!-sorry, Pat!) .... Let me try that again. What if it did everything it could, to steer clear of emotional topics and references?

Sandy: That would certainly make me suspicious. Any consistent pattern of avoiding certain issues would raise serious doubts in my mind as to whether I was dealing with a thinking being.

Chris: Why do you say that? Why not just conclude you're dealing with a thinking but unemotional being?

Sandy: You've hit upon a sensitive point. I've thought about this for quite a long time, and I've concluded that I simply can't believe emotions and thought can be divorced. To put it another way, I think emotions are an automatic by-product of the ability to think. They are entailed by the very nature of thought.

Chris: That's an interesting conclusion, but what if you're wrong? What if I produced a machine that could think but not emote? Then its intelligence might go unrecognized because it failed to pass your kind of test.

Sandy: I'd like you to point out to me where the boundary line between emotional questions and nonemotional ones lies. You might want to ask about the meaning of a great novel. This certainly requires an understanding of human emotions! Now is that thinking, or merely cool calculation? You might want to ask about a subtle choice of words. For that, you need an understanding of their connotations. Turing uses examples like this in his article. You might want to ask for advice about a complex romantic situation. The machine would need to know a lot about human motivations and their roots. If it failed at this kind of task, I would not be much inclined to say that it could think. As far as I'm concerned, thinking, feeling, and consciousness are just different facets of one phenomenon, and no one of them can be present without the others.

Chris: Why couldn't you build a machine that could feel nothing (we all know machines don't feel anything!), but that could think and make complex decisions anyway? I don't see any contradiction there.

Sandy: Well, I do. I think that when you say that, you are visualizing a metallic, rectangular machine, probably in an air conditioned room-a hard, angular, cold object with a million colored wires inside it, a machine that sits stock still on a tiled floor, humming or buzzing or whatever, and spinning its tapes. Such a

machine can play a good game of chess, which, I freely admit, involves a lot of decision making. And yet I would never call it conscious.

Chris: How come? To mechanists, isn't a chess-playing machine rudimentarily conscious?

Sandy: Not to *this* mechanist! The way I see it, consciousness has got to come from a precise pattern of organization, one we haven't yet figured out how to describe in any detailed way. But I believe we will gradually come to understand it. In my view, consciousness requires a certain way of mirroring the external universe internally, and the ability to respond to that external reality on the basis of the internally represented model. And then in addition, what's really crucial for a conscious machine is that it should incorporate a well-developed and flexible sell-model. And it's there that all existing programs, including the best chess-playing ones, fall down.

Chris: Don't chess programs look ahead and say to themselves as they're figuring out their next move, "if my opponent moves here, then I'll go there, and then if they go this way, I could go below that way..."? Doesn't that usage of the concept "I" require a sort of self-model?

Sandy: Not really. Or, if you want, it's an extremely limited one. It's an understanding of self in only the narrowest sense. For instance, a chess-playing program has no concept of why it is playing chess, or of the fact that it is a program, or is in a computer, or has a human opponent. It has no idea about what winning and losing are, or-

Pat: How do *you* know it has no such sense? How can *you* presume to say what a chess program feels or knows?

Sandy: Oh come on! We all know that certain things don't feel anything or know anything. A thrown stone doesn't know anything about parabolas, and a whirling fan doesn't know anything about air. It's true I can't *prove* those statements-but here we are verging on questions of faith.

Pat: This reminds me of a Taoist story I read. It goes something like this. Two sages were standing on a bridge over a stream. One said to the other, "I wish I were a fish. They are so happy." The other replied, "How do you know whether fish are happy or not? You're not a fish!" The first said, "But you're not me, so how do you know whether I know how fish feel?"

Sandy: Beautiful! Talking about consciousness really does call for a certain amount of restraint. Otherwise, you might as well just jump on the solipsism bandwagon ("I am the only conscious being in the universe") or the panpsychism bandwagon ("*Everything* in the universe is conscious!").

Pat: Well, how do you know? Maybe everything is conscious.

Sandy: Oh Pat, if you're going to join the club that maintains that stones and even particles like electrons have some sort of consciousness, then I guess we part company here. That's a kind of mysticism I just can't fathom. As for chess programs, I happen to know how they work, and I can tell you for sure that they aren't conscious. No way!

Pat: Why not?

Sandy: They incorporate only the barest knowledge about the goals of chess. The notion of "playing" is turned into the mechanical act of comparing a lot of numbers and choosing the biggest one over and over again. A chess program has no sense of disappointment about losing, or pride in winning. Its self-model is very crude. It gets away with doing the least it can, just enough to play a game of chess and nothing more. Yet interestingly enough, we still tend to talk about the "desires" of a chess-playing computer. We say, "it wants to keep its

king behind a row of pawns" or "it likes to get its rooks out early" or "it thinks I don't see that hidden fork."

Pat: Yes, and we do the same thing with insects. We spot a lonely ant somewhere and say, "It's trying to get hack home" or "It wants to drag that dead bee back to the colony." In fact, with any animal we use terms that indicate emotions, but we don't know for certain how much the animal feels. I have no trouble talking about dogs and cats being happy or sad, having desires and beliefs and so on, but of course I don't think their sadness is as deep or complex as human sadness is.

Sandy: But you wouldn't call it "simulated" sadness.

Pat: No, of course not. I think it's real.

Sandy: It's hard to avoid use of such teleological or mentalistic terms. I believe they're quite justified, although they shouldn't be carried too far. They simply don't have the same richness of meaning when applied to present-day chess programs as when applied to people.

Chris: I still can't see that intelligence has to involve emotions. Why couldn't you imagine an intelligence that simply calculates and has no feelings?

Sandy: A couple of answers here. Number one, any intelligence has to have motivations. It's simply not the case, whatever many people may think, that machines could think any more "objectively" than people do. Machines, when they look at a scene, will have to focus and filter that scene down into some preconceived categories, just as a person does. And that means seeing some things and missing others. It means giving more weight to some things than to others. This happens on every level of processing.

Pat: I'm not sure I'm following you.

Sandy: Take me right now, for instance. You might think I'm just making some intellectual points, and I wouldn't need emotions to do that. But what makes me *care* about these points? Just now-why did I stress the word "care" so heavily? Because I'm emotionally involved in this conversation! People talk to each other out of conviction-not out of hollow, mechanical reflexes. Even the most intellectual conversation is driven by underlying passions. There's an emotional undercurrent to every conversation-it's the fact that the speakers want to be listened to, understood, and respected for what they are saying.

Pat: It sounds to me as if all you're saying is that people need to be interested in what they're saying. Otherwise, a conversation dies.

Sandy: Right! I wouldn't bother to talk to anyone if I weren't motivated by *interest*. And "interest" is just another name for a whole constellation of subconscious biases. When I talk, all my biases work together, and what you perceive on the surface level is my personality, my style. But that style arises from an immense number of tiny priorities, biases, leanings. When you add up a million of them interacting together, you get something that amounts to a lot of desires. It just all adds up! And that brings me to the other answer to Chris's question about feelingless calculation. Sure, that exists-in a cash register, a pocket calculator. I'd say it's even true of all today's computer programs. But eventually, when you put enough feelingless calculations together in a huge coordinated organization, you'll get something that has properties on another level. You can see it-in fact, you have to see it-not as a bunch of little calculations but as a system of tendencies and desires and beliefs and so on. When things get complicated enough, you're *forced* to change your level of description. To some extent that's already happening, which is why we use words such as

"want," "think," "try," and "hope" to describe chess programs and other attempts at mechanical thought. Dennett calls that kind of level switch by the observer "adapting the intentional stance." The really interesting things in AI will only begin to happen, I'd guess, when the program *itself* adopts the intentional stance toward itself!

Chris: That would be a very strange sort of level-crossing feedback loop.

Sandy: It certainly would. When a program looks at itself from the outside, as it were, and tries to figure out why it acted the way it did, then I'll start to think that there's someone in there, doing the looking.

Pat: You mean an "I"? A self?

Sandy: Yes, something like that. A soul, even-although not in any religious sense. Of course, it's highly premature for anyone to adopt the intentional stance (in the full force of the term) with respect to today's programs. At least that's my opinion.

Chris: For me an important related question is: To what extent is it valid to adopt the intentional stance toward beings other than humans?

Pat: I would certainly adopt the intentional stance toward mammals.

Sandy: I vote for that.

Chris: Now that's interesting. How can that be, Sandy? Surely you wouldn't claim that a dog or cat can pass the Turing Test? Yet don't you maintain the Turing Test is the only way to test for the presence of consciousness? How can you have these beliefs simultaneously?

Sandy: Hmm.... All right. I guess that my argument is really just that the Turing Test works only above a certain level of consciousness. I'm perfectly willing to grant that there can be thinking beings that could fail at the Turing Test-but the main point that I've been arguing for is that anything that passes it would be a genuinely conscious, thinking being.

Pat: How can you think of a computer as a conscious being? I apologize if what I'm going to say sounds like a stereotype, but when I think of conscious beings, I just can't connect that thought with machines. To me, consciousness is connected with soft, warm bodies, silly though it may sound.

Chris: That does sound odd, coming from a biologist. Don't you deal with life so much in terms of chemistry and physics that all magic seems to vanish?

Pat: Not really. Sometimes the chemistry and physics simply increase the feeling that there's something magical going on down there! Anyway, I can't always integrate my scientific knowledge with my gut feelings.

Chris: I guess I share that trait.

Pat: So how do you deal with rigid preconceptions like mine?

Sandy: I'd try to dig down under the surface of your concept of "machine" and get at the intuitive connotations that lurk there, out of sight but deeply influencing your opinions. I think we all have a holdover image from the Industrial Revolution that sees machines as clunky iron contraptions gawkily moving under the power of some loudly chugging engine. Possibly that's even how the computer inventor Charles Babbage saw people! After all, he called his magnificent many-geared computer the "Analytical Engine."

Pat: Well, I certainly don't think people are just fancy steam shovels or electric can openers.

There's something about people, something that-that-they've got a sort of *flame* inside them, something alive, something that flickers unpredictably, wavering, uncertain-but something *creative*!

Sandy: Great! That's just the sort of thing I wanted to hear. It's very human to think that way. Your flame image makes me think of candles, of fires, of vast thunderstorms with lightning dancing all over the sky in crazy, tumultuous patterns. But do you realize that just that kind of thing is visible on a computer's console? The flickering lights form amazing chaotic sparkling patterns. It's such a far cry from heaps of lifeless, clanking metal!
It *is* flamelike, by God! Why don't you let the word "machine" conjure up images of dancing patterns of light rather than of giant steam shovels?

Chris: That's a beautiful image, Sandy. It does tend to change my sense of mechanism from being matter-oriented to being pattern-oriented. It makes me try to visualize the thoughts in my mind-these thoughts right now, even!-as a huge spray of tiny pulses flickering in my brain.

Sandy: That's quite a poetic self-portrait for a mere spray of flickers to have come up with!

Chris: Thank you. But still, I'm not totally convinced that a machine is all that I am. I admit, my concept of machines probably does suffer from anachronistic subconscious flavors, but I'm afraid I can't change such a deeply rooted sense in a flash.

Sandy: At least you sound open-minded. And to tell the truth, part of me sympathizes with the way you and Pat view machines. Part of me balks at calling myself a machine. It is a bizarre thought that a feeling being like you or me might emerge from mere circuitry. Do I surprise you?

Chris: You certainly surprise *me*. So, tell us-do you believe in the idea of an intelligent computer, or don't you?

Sandy: It all depends on what you mean. We've all heard the question "Can computers think?" There are several possible interpretations of this (aside from the many interpretations of the word "think"). They revolve around different meanings of the words "can" and "computer."

Pat: Back to word games again ....

Sandy: I'm sorry, but that's unavoidable. First of all, the question might mean, "Does some present-day computer think, right now?" To this I would immediately answer with a loud no. Then it could be taken to mean, "Could some present-day computer, if suitably programmed, potentially think?" That would be more like it, but I would still answer, "Probably not." The real difficulty hinges on the word "computer." The way I see it, "computer" calls up an image of just what I described earlier: an air-conditioned room with cold rectangular metal boxes in it. But I suspect that with increasing public familiarity with computers and continued progress in computer architecture, that vision will eventually become outmoded.

Pat: Don't you think computers as we know them will be around for a while?

Sandy: Sure, there will have to be computers in today's image around for a long time, but advanced computers-maybe no longer called "computers"-will evolve and become quite different. Probably, as with living organisms, there will be many branchings in the evolutionary tree. There will be computers for business, computers for schoolkids, computers for scientific calculations, computers for systems research, computers for simulation, computers for rockets going into space, and so on. Finally, there will be computers for the

study of intelligence, It's really only these last that I'm thinking of-the ones with the maximum flexibility, the ones that people are deliberately attempting to make smart. I see no reason that these will stay fixed in the traditional image. They probably will soon acquire as standard features some rudimentary sensory systems mostly for vision and hearing, at first. They will need to be able to move around, to explore. They will have to be physically flexible. In short, they will have to become more animal-like, more self-reliant.

Chris: It makes me think of the robots R2D2 and C3P0 in the movie Star Wars.

Sandy: Not me! In fact, I don't think of anything remotely like them when I visualize intelligent machines. They are too silly, too much the product of a film designer's imagination. Not that I have a clear vision of my own. But I think it's necessary, if people are realistically going to try to imagine an artificial intelligence, to go beyond the limited, hard-edged picture of computers that comes from exposure to what we have today. The only thing all machines will always have in common is their underlying mechanicalness. That may sound cold and inflexible, but then-just think-what could be more mechanical, in a wonderful way, than the workings of the DNA and proteins and organelles in our cells?

Pat: To me, what goes on inside cells has a "wet," "slippery" feel to it, and what goes on inside machines is dry and rigid. It's connected with the fact that computers don't make mistakes, that computers do only what you tell them to do. Or at least that's my image of computers.

Sandy: Funny-a minute ago, your image was of a flame, and now it's of something wet and slippery. Isn't it marvelous, how contradictory we can be?

Pat: I don't need your sarcasm.

Sandy: No, no, I'm not being sarcastic-I really do think it's marvelous.

Pat: It's just an example of the human mind's slippery nature-mine, in this case.

Sandy: True. But your image of computers is stuck in a rut. Computers certainly can make mistakes-and I don't mean on the hardware level. Think of any present-day computer predicting the weather. It can make wrong predictions, even though its program runs flawlessly.

Pat: But that's only because you've fed it the wrong data.

Sandy: Not so. It's because weather prediction is too complex. Any such program has to make do with a limited amount of data-entirely correct data-and extrapolate from there. Sometimes it will make wrong predictions. It's no different from a farmer gazing at the clouds and saying, "I reckon we'll get a little snow tonight." In our heads, we make models of things and use those models to guess how the world will behave. We have to make do with our models, however inaccurate they may be, or evolution will prune us out ruthlessly-we'll fall off a cliff or something. And for intelligent computers, it'll be the same. It's just that human designers will speed up the evolutionary process by aiming explicitly at the goal of creating intelligence, which is something nature just stumbled on.

Pat: So you think computers will be making fewer mistakes as they get smarter?

Sandy: Actually, just the other way around! The smarter they get, the more they'll be in a position to tackle messy real-life domains, so they'll be more and more likely to have inaccurate models. To me, mistake making is a sign of high intelligence!

Pat: Wow-you throw me sometimes!