5

DOUGLAS R. HOFSTADTER

The Turing Test: A Coffeehouse Conversation

PARTICIPANTS

Chris, a physics student; Pat, a biology student; and 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 it 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!

This selection appeared previously as "Metamagical Themas: A coffeehouse conversation on the Turing test to determine if a machine can think," in *Scientific American,* May 1981, pp. 15-36.

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

sandy: NO, he died back in 1954, at just forty-one. He'd only be sixty- seven this year, although he is now such a legendary figure it seems strange to imagine him still alive today.

chris: HOW did he die?

sandy: Almost certainly suicide. He was homosexual and had to deal with a lot of harsh treatment and stupidity from the outside world. In the end it apparendy got to be too much, and he killed himself.

chris: That's a sad story.

sandy: Yes, it certainly is. What saddens me is that he never got to see the amazing progress in computing machinery and theory that has taken place.

pat: Hey, are you 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 this 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 the question is not meaningful as it stands. 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 I think they confuse a lot of people. On the one hand we're given the refrain "Computers are really dumb, you have to spell everything out for them in complete detail," and on the other hand we're bombarded with advertising hype about "smart products."

chris: That's certainly true. Did you know that one computer terminal manufacturer has even taken to calling its products "dumb terminals" in order to stand out from the crowd?

sandy: That's cute, but it just plays along with the trend toward obfuscation. The term "electronic brain" always comes to my mind when I'm thinking about this. Many people swallow it completely, while others reject it out of hand. Few have the patience to sort out the issues and decide how much of it makes sense.

pat: Does Turing suggest some way of resolving it, some sort of IQ, test for machines?

sandy: That would be 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 it can think or not.

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. 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: SO, as is 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: SO tell me how Turing's test works.

sandy: The idea is based on what he calls the Imitation Game. In this game a man and a woman go into separate rooms and 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 discern which room the woman is in. Now the woman, by her answers, tries to aid the interrogator as much as possible. The man, however, 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 would very much like to participate in it someday. Would the interrogator know either the man or the woman before the test began? Would any of them know the others?

sandy: That would probably be a bad idea. All sorts of subliminal cueing might occur if the interrogator knew one or both of them. It

would be safest if all three'people were totally unknown to each other.

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 very sex-oriented questions? 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: It sounds plausible.

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 it would make a difference if the interrogator were a woman or a man. Don't you think that a woman could spot some telltale differences more quickly than a man could?

pat: If so, maybe ***that's*** how to tell a man from a woman!

sandy: Hmm . . . that's a new twist! In any case, 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 modem computer terminals. I have to admit, though, that I'm not 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 fundamentally believe in the Turing test, I'm not sure what the point is of the Imitation Game, on which it's founded!

chris: 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 your statements I gather that the Turing test is a kind of extension of the Imitation Game, only involving a machine and a person in separate rooms.

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

pat: Except for your loaded phrase "the machine tries," this sounds very interesting. But how do you know that this test 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 objec
tion 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 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 one couldn't conclude anything from a man's ability to win at the Imitation Game, how could one 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 physics, for instance, we simulate the behavior of particles, atoms, solids, liquids, gases, galaxies, and so on. But nobody 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 prefer, in their ***simulated,*** minds—the hurricane would not be a simulation but a genuine phenomenon complete with drenching and devastation.

chris: Oh, boy—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 this case, it takes special "computational glasses" to see the rain and the winds and so on.

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 tom apart in the process of simulating winds!

sandy: Certainly not, but you're 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 the memory. You'll see that some abstract links have been broken, some values of variables radically changed, and so forth. 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, a kind never before associated with hurri

canes. Using this idea, 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," located right in the middle of the eye! It's 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 that you'll insist on before you call something a hurricane.

pat: Well, wouldn't you say that being an atmospheric phenomenon is one vital 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 that 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 turn on each other? And a computer is just a fancy cash register, as I understand it.

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 till now, we'll have to write a very complicated program 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 more reliable than the post office clerk who adds up your total by hand. Now, I happen to believe eventually such a program could be written. Then we'd know something about how post office clerks and schoolkids work.

pat: I can't believe you could ever 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***—a rather ironic one, which is why I like it. The passage 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." Dennett 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 Dennett-inspired ***reductio ad absurdum*** would lead you to conclude that calculation doesn't exist, hurricanes don't exist, nothing at a higher level than particles and the laws of physics exists? What do you gain by saying 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 out 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*** hurricane 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, you've got a very different matter on your hands from simulated hurricanes.

pat: I don't see why. A brainstorm sounds to me like a mental hurricane. But seriously, you'll have to convince me.

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

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 that we call those storms hurricanes. But it's not possible to make a sharp distinction between hurricanes and nonhurricanes. 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: Yes, I remember that now. The idea of a "glitch" strikes me as wonderfully eerie—a surrealistic kind of quivering on a surrealistic kind of surface.

chris: Can you imagine—plate tectonics on a giant rotating 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 rationals, reals, complex numbers, and "on beyond zebra," as Dr. Seuss says.

pat: I think I can see your point here, Sandy. We have many examples in biology 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. 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 each other. So perhaps I can glimpse, at least a little, how to you a simulated hurricane could, in some 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 the genuine article, 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 think your slur about "wool-pulling" goes too far. 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 impossibly deep. I hope you would agree that it would then share all that abstract structure that defines the "essence of hurricane- hood." So what's to hold you back 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 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!

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

sandy: But 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. 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 the target in question 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: More exactly put, you can hypothesize about what kinds of internal structures might account for the behavior observed—but they may or may not in fact exist.

sandy: Hold on, now! Are you saying that atomic nuclei are merely hypothetical entities? After all, their existence—or should I say "hypothetical existence"?—was proven—or should I say "suggested"? —by the behavior of particles scattered off of atoms.

chris: Physical systems seem to me to be much simpler than the mind, and the certainty of the inferences made is correspondingly greater.

sandy: The experiments are also correspondingly harder to perform and to interpret. In the Turing test, you could perform many highly delicate experiments in the course of an hour. I maintain that people give other people credit for being conscious simply because of their continual external monitoring of them—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 fellow human beings and so we think they think.

sandy: TO me, that seems a highly anthropocentric view of what thought is. Does that mean you would sooner say a mannikin in a store thinks than a wonderfully programmed computer, simply because the mannikin 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 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 is what the Turing test is for. And I think it is the ***only*** way to test for "thinkinghood."

chris: But you could be fooled by the Turing test, just as an interrogator could think a man was 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. 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 avoided all emotional references?

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

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

sandy: You've hit upon a sensitive point. I simply can't believe that emotions and thought can be divorced. Put another way, I think that emotions are an automatic by-product of the ability to think. They are implied by the very nature of thought.

chris: Well, 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 requires understanding of human emotions! 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 it for advice about a complex romantic situation. It would need to know a lot about human motivations and their roots. Now if it failed at this kind of task, I would not be much inclined to say that it could think. As far as I am concerned, the ability to think, the ability to feel, 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, 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 such a machine conscious.

chris: How come? To mechanists, isn't a chess-playing machine rudi- mentarily conscious?

sandy: Not to this mechanist. The way I see it, consciousness has got to come from a precise pattern of organization—one that 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 self-model. And it's there that all existent 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 you move here, then I'll go there, and then if you go this way, I could go that way . . . "? Isn't that a sort of self-model?

sandy: Not really. Or, if you want, it's an extremely limited one. It's an understanding of self only in the narrowest sense. For instance, a chess-playing program has no concept of why it is playing chess, or the fact that it is a program, or is in a computer, or has a human opponent. It has no ideas 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 second 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 either 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: If you're going to join those who claim 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 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 shame 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 do 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: Well, we do the same thing with insects. We spot a lonely ant somewhere and say, "It's trying to get back 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 sure 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," would you?

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: What do you mean?

sandy: Take me right now, for instance. You might think that I'm just making some intellectual points, and I wouldn't need emotions to do that. But what makes me ***care*** about these points? 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 style, my personality. But that style arises from an immense number of tiny priorities, biases, leanings. When you add up a million of these interacting together, you get something that amounts to a lot of ***desires.*** It just all adds up! And that brings me to the other point, 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 "adopting 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. Of course, in my opinion, it's highly premature for anyone to adopt the intentional stance, in the full force of the term, toward 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: 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 think that the Turing test is the only way to test for the presence of thought? How can you have these beliefs at once?

sandy: Hmm.... All right. I guess I'm forced to admit that the Turing test works only above a certain level of consciousness. There can be thinking beings that could fail the test—but on the other hand, anything that passes it, in my opinion, would be a genuinely conscious, thinking being.

pat: How can you think of a computer as a conscious being? I apologize if this 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 that may sound.

chris: That does sound odd, coming from a biologist. Don't you deal with life in terms of chemistry and physics enough for all magic to seem to vanish?

pat: Not really. Sometimes the chemistry and physics just increase the feeling that there's something magical going on down there! Anyway, I can't always integrate my scientific knowledge with my gut- level 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 "machines" and get at the intuitive connotations that lurk there, out of sight but deeply influencing your opinions. I think that 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 viewed 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 even 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 thunderstorms with lightning dancing all over the sky in crazy patterns. But do you realize that just that kind of pattern 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 changes 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 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 do sound open-minded. And to tell the truth, part of me does sympathize 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 have 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: That's right. 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?" This is 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 metallic 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 in the case of 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. Probably they 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 C3PO in ***Star Wars.***

sandy: As a matter of fact I don't think of anything like them when I visualize intelligent machines. They're 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 is necessary, if people are going to try realistically to imagine an artificial intelligence, to go beyond the limited, hard-edged image of computers that comes from exposure to what we have today. The only thing that all machines will always have in

common is their underlying mechanicalness. That may sound cold and inflexible, but what could be more mechanical—in a wonderful way—than the operations 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: I'm not being sarcastic—I really ***do*** think it is 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 the farmer in the field gazing at the clouds who says, "I reckon we'll get a little snow tonight." We make models of things in our heads and use them to guess how the world will behave. We have to make do with our models, however inaccurate they may be. And if they're too inaccurate, evolution will prune us out—we'll fall over a cliff or something. And computers are 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 make 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: Boy—you throw me sometimes!

sandy: I guess I'm a strange sort of advocate for machine intelligence. To some degree I straddle the fence. I think that machines won't really be intelligent in a humanlike way until they have something like that biological wetness or slipperiness to them. I don't mean literally wet—the slipperiness could be in the software. But biologi- cal-seeming or not, intelligent machines will in any case be machines. We will have designed them, built them—or grown them! We will understand how they work—at least in some sense. Possibly no one person will really understand them, but collectively we will know how they work.

pat: It sounds like you want to have your cake and eat it too.

sandy: You're probably right. What I'm getting at is that when artificial intelligence comes, it will be mechanical and yet at the same time organic. It will have that same astonishing flexibility that we see in life's mechanisms. And when I say "mechanisms," I ***mean*** "mechanisms." DNA and enzymes and so on really ***are*** mechanical and rigid and reliable. Wouldn't you agree, Pat?

pat: That's true. But when they work together, a lot of unexpected things happen. There are so many complexities and rich modes of behavior that all that mechanicalness adds up to something very fluid.

sandy: For me it's an almost unimaginable transition from the mechanical level of molecules to the living level of cells. But it's what convinces me that people are machines. That thought makes me uncomfortable in some ways, but in other ways it is an exhilarating thought.

chris: If people are machines, how come it's so hard to convince them of the fact? Surely if we are machines, we ought to be able to recognize our own machinehood.

sandy: You have to allow for emotional factors here. To be told you're a machine is, in a way, to be told that you're nothing more than your physical parts, and it brings you face to face with your own mortality. That's something nobody finds easy to face. But beyond the emotional objection, to see yourself as a machine you have to jump all the way from the bottommost mechanical level to the level where the complex lifelike activities take place. If there are many intermediate layers, they act as a shield, and the mechanical quality becomes

almost invisible. I think that's how intelligent machines will seem to us—and to themselves!—when they come around.

pat: I once heard a funny idea about what will happen when we eventually have intelligent machines. When we try to implant that intelligence into devices we'd like to control, their behavior won't be so predictable.

sandy: They'll have a quirky little "flame" inside, maybe?

pat: Maybe.

chris: SO what's so funny about that?

pat: Well, think of military missiles. The more sophisticated their target-tracking computers get, according to this idea, the less predictably they will function. Eventually you'll have missiles that will decide they are pacifists and will turn around and go home and land quietly without blowing up. We could even have "smart bullets" that turn around in midflight because they don't want to commit suicide!

sandy: That's a lovely thought.

chris: I'm very skeptical about these ideas. Still, Sandy, I'd like to hear your predictions about when intelligent machines will come to be.

sandy: It won't be for a long time, probably, that we'll see anything remotely resembling the level of human intelligence. It just rests on too awesomely complicated a substrate—the brain—for us to be able to duplicate it in the foreseeable future. Anyway, that's my opinion.

pat: DO you think a program will ever pass the Turing test?

sandy: That's a pretty hard question. I guess there are various degrees of passing such a test, when you come down to it. It's not black and white. First of all, it depends on who the interrogator is. A simpleton might be totally taken in by some programs today. But secondly, it depends on how deeply you are allowed to probe.

pat: Then you could have a scale of Turing tests—one-minute versions, five-minute versions, hour-long versions, and so forth. Wouldn't it be interesting if some official organization sponsored a periodic competition, like the annual computer-chess championships, for programs to try to pass the Turing test?

chris: The program that lasted the longest against some panel of distinguished judges would be the winner. Perhaps there could be a big prize for the first program that fools a famous judge for, say, ten minutes.

pat: What would a program do with a prize?

chris: Come now, Pat. If a program's good enough to fool the judges, don't you think it's good enough to enjoy the prize?

pat: Sure, especially if the prize is an evening out on the town, dancing with all the interrogators!

sandy: I'd certainly like to see something like that established. I think it could be hilarious to watch the first programs flop pathetically!

pat: You're pretty skeptical, aren't you? Well, do you think any computer program today could pass a five-minute Turing test, given a sophisticated interrogator?

sandy: I seriously doubt it. It's partly because no one is really working at it explicitly. However, there is one program called "Parry" which its inventors claim has already passed a rudimentary version of the Turing test. In a series of remotely conducted interviews, Parry fooled several psychiatrists who were told they were talking to either a computer or a paranoid patient. This was an improvement over an earlier version, in which psychiatrists were simply handed transcripts of short interviews and asked to determine which ones were with a genuine paranoid and which ones with a computer simulation.

pat: You mean they didn't have the chance to ask any questions? That's a severe handicap—and it doesn't seem in the spirit of the Turing test. Imagine someone trying to tell which sex I belong to just by reading a transcript of a few remarks by me. It might be very hard! So I'm glad the procedure has been improved.

chris: HOW do you get a computer to act like a paranoid?

sandy: I'm not saying it ***does*** act like a paranoid, only that some psychiatrists, under unusual circumstances, thought so. One of the things that bothered me about this pseudo-Turing test is the way Parry works. "He"—as they call him—acts like a paranoid in that he gets abruptly defensive, veers away from undesirable topics in the conversation, and, in essence, maintains control so that no one can truly probe "him." In this way, a simulation of a paranoid is a lot easier than a simulation of a normal person.

pat: No kidding! It reminds me of the joke about the easiest kind of human for a computer program to simulate.

chris: What is that?

pat: A catatonic patient—they just sit and do nothing at all for days on end. Even I could write a computer program to do that!

sandy: An interesting thing about Parry is that it creates no sentences on its own—it merely selects from a huge repertoire of canned sentences the one that best responds to the input sentence.

pat: Amazing! But that would probably be impossible on a larger scale, wouldn't it?

sandy: Yes. The number of sentences you'd need to store to be able to respond in a normal way to all possible sentences in a conversation is astronomical, really unimaginable. And they would have to be so intricately indexed for retrieval. . . . Anybody who thinks that somehow a program could be rigged up just to pull sentences out of storage like records in a jukebox, and that this program could pass the Turing test, has not thought very hard about it. The funny part about it is that it is just this kind of unrealizable program that some enemies of artificial intelligence cite when arguing against the concept of the Turing test. Instead of a truly intelligent machine, they want you to imagine a gigantic, lumbering robot that intones canned sentences in a dull monotone. It's assumed that you could see through to its mechanical level with ease, even if it were simultaneously performing tasks that we think of as fluid, intelligent processes. Then the critics say, "You see! It would still be just a machine—a mechanical device, not intelligent at all!" I see things almost the opposite way. If I were shown a machine that can do things that I can do—I mean pass the Turing test—then, instead of feeling insulted or threatened, I'd chime in with the philosopher Raymond Smullyan and say, "How wonderful machines are!"

chris: If you could ask a computer just one question in the Turing test, what would it be?

sandy: Uhmm. . . .

pat: How about "If you could ask a computer just one question in the Turing test, what would it be?"?

Reflections

Many people are put off by the provision in the Turing test requiring the contestants in the Imitation Game to be in another room from the judge, so only their verbal responses can be observed. As an element in a parlor game the rule makes sense, but how could a legitimate scientific proposal

include a deliberate attempt to ***hide facts*** from the judges? By placing the candidates for intelligence in "black boxes" and leaving nothing as evidence but a restricted range of "external behavior" (in this case, verbal output by typing), the Turing test seems to settle dogmatically on some form of behaviorism, or (worse) operationalism, or (worse still) verifica- tionism. (These three cousins are horrible monster ***isms*** of the recent past, reputed to have been roundly refuted by philosophers of science and interred—but what is that sickening sound? Can they be stirring in their graves? We should have driven stakes through their hearts!) Is the Turing test just a case of whatjohn Searle calls "operationalist sleight-of- hand"?

The Turing test certainly does make a strong claim about what matters about minds. What matters, Turing proposes, is not what kind of gray matter (if any) the candidate has between its ears, and not what it looks like or smells like, but whether it can ***act***—or behave, if you like— intelligently. The particular game proposed in the Turing test, the Imitation Game, is not sacred, but just a cannily chosen test of more general intelligence. The assumption Turing was prepared to make was that nothing could possibly pass the Turing test by winning the Imitation Game without being able to perform indefinitely many other clearly intelligent actions. Had he chosen checkmating the world chess champion as his litmus test of intelligence, there would have been powerful reasons for objecting; it now seems quite probable that one could make a machine that can do that ***but nothing else.*** Had he chosen stealing the British Crown Jewels without using force or accomplices, or solving the Arab-Israeli conflict without bloodshed, there would be few who would make the objection that intelligence was being "reduced to" behavior or "operationally defined" in terms of behavior. (Well, no doubt ***some*** philosopher somewhere would set about diligently constructing an elaborate but entirely outlandish scenario in which some utter dolt stumbled into possession of the British Crown Jewels, "passing" the test and thereby "refuting" it as a good general test of intelligence. The true operationalist, of course, would then have to admit that such a lucky moron was, by operationalist lights, truly intelligent since he passed the defining test—which is no doubt why true operationalists are hard to find.)

What makes Turing's chosen test better than stealing the British Crown Jewels or solving the Arab-Israeli conflict is that the latter tests are unrepeatable (if successfully passed once!), too difficult (many manifestly intelligent people would fail them utterly) and too hard to judge objectively. Like a well-composed wager, Turing's test invites trying; it seems fair, demanding but possible, and crisply objective in the judging. The Turing test reminds one of a wager in another way, too. Its motivation

is to stop an interminable, sterile debate by saying "Put up or shut up!" Turing says in effect: "Instead of arguing about the ultimate nature and essence of mind or intelligence, why don't we all agree that anything that could pass this test is ***surely*** intelligent, and then turn to asking how something could be designed that might pass the test fair and square?" Ironically, Turing failed to shut off the debate but simply managed to get it redirected.

Is the Turing test vulnerable to criticism because of its "black box" ideology? First, as Hofstadter notes in his dialogue, we treat ***each other*** as black boxes, relying on our observation of apparently intelligent behavior to ground our belief in other minds. Second, the black box ideology is in any event the ideology of all scientific investigation. We learn about the DNA molecule by probing it in various ways and seeing how it behaves in response; we learn about cancer and earthquakes and inflation in the same way. "Looking inside" the black box is often useful when macroscopic objects are our concern; we do it by bouncing "opening" probes (such as a scalpel) off the object and then scattering photons off the exposed surfaces into our eyes. Just one more black box experiment. The question must be, as Hofstadter says: Which probes will be most directly relevant to the question we want to answer? If our question is about whether some entity is intelligent, we will find no more direct, telling probes than the everyday questions we often ask each other. The extent of Turing's "behaviorism" is simply to incorporate that near truism into a handy, laboratory-style experimental test.

Another problem raised but not settled in Hofstadter's dialogue concerns representation. A computer simulation of something is typically a detailed, "automated," multi-dimensional representation of that thing, but of course there's a world of difference between representation and reality, isn't there? As John Searle says, "No one would suppose that we could produce milk and sugar by running a computer simulation of the formal sequences in lactation and photosynthesis. .. If we devised a program that simulated a cow on a digital computer, our simulation, being a mere representation of a cow, would not, if "milked," produce milk, but at best a representation of milk. You can't drink that, no matter how good a representation it is, and no matter how thirsty you are.

But now suppose we made a computer simulation of a mathematician, and suppose it worked well. Would we complain that what we had hoped for was ***proofs,*** but alas, all we got instead was mere ***representations*** of proofs? But representations of proofs ***are*** proofs, aren't they? It depends on how good the proofs represented are. When cartoonists repre-

*(See selection 22, "Minds, Brains, and Programs," p. 37 2)

sent scientists pondering blackboards, what they typically represent as proofs or formulae on the blackboard is pure gibberish, however "realistic" these figures appear to the layman. If the simulation of the mathematician produced phony proofs like those in the cartoons, it might still simulate ***something*** of theoretical interest about mathematicians—their verbal mannerisms, perhaps, or their absentmindedness. On the other hand, if the simulation were designed to produce representations of the proofs a good mathematician would produce, it would be as valuable a "colleague"—in the proof-producing department—as the mathematician. That is the difference, it seems, between abstract, formal products like proofs or songs (see the next selection "The Princess Ineffabelle") and concrete, material products like milk. On which side of this divide does the mind fall? Is mentality like milk or like a song?

If we think of the mind's product as something like ***control of the body,*** it seems its product is quite abstract. If we think of the mind's product as a sort of special substance or even a variety of substances—lots 'n lots of ***love,*** a smidgin or two of ***pain,*** some ***ecstasy,*** and a few ounces of that ***desire*** that all good ballplayers have in abundance—it seems its product is quite concrete.

Before leaping into debate on this issue we might pause to ask if the principle that creates the divide is all that clear-cut at the limits to which we would have to push it, were we to confront a truly detailed, superb simulation of ***any*** concrete object or phenomenon. Any actual, running simulation is concretely "realized" in some hardware or other, and the vehicles of representation must themselves produce some effects in the world. If the representation of an event produces just about the same effects in the world as the event itself would, to insist that it is merely a representation begins to sound willful. This idea, playfully developed in the next selection, is a recurrent theme throughout the rest of the book.

D.C.D.