4

A. M. TURING

Computing Machinery and Intelligence

The Imitation Game

I propose to consider the question "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is

Excerpt from "Computing Machinery and Intelligence," *Mind,* Vol. LIX, No. 236 (1950). Reprinted by permission.

the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try to cause C to make the wrong identification. His answer might therefore be

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

Critique of the New Problem

As well as asking "What is the answer to this new form of the question," one may ask, "Is this new question a worthy one to investigate?" This latter question we investigate without further ado, thereby cutting short an infinite regress.

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to make a "thinking machine" more human by dressing it up in such artificial flesh. The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing their voices.

Some other advantages of the proposed criterion may be shown up by specimen questions and answers. Thus:

Q: Please write me a sonnet on the subject of the Forth Bridge.
A: Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764.
A: (Pause about 30 seconds and then give as answer) 105621.
Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.

The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavor that we wish to include. We do not wish to penalize the machine for its inability to shine in beauty competitions, nor to penalize a man for losing in a race against an airplane. The conditions of our game make these disabilities irrelevant. The "witnesses" can brag, if they consider it advisable, as much as they please about their charms, strength or heroism, but the interrogator cannot demand practical demonstrations.

The game may perhaps be criticized on the ground that the odds are weighted too heavily against the machine. If the man were to try and pretend to be the machine he would clearly make a very poor showing. He would be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

It might be urged that when playing the "imitation game" the best strategy for the machine may possibly be something other than imitation of the behavior of a man. This may be, but I think it is unlikely that there is any great effect of this kind. In any case there is no intention to investigate here the theory of the game, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man.

The Machines Concerned in the Game

The question which we put earlier will not be quite definite until we have specified what we mean by the word "machine." It is natural that we should wish to permit every kind of engineering technique to be used in our machines. We also wish to allow the possibility that an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental. Finally, we wish to exclude from the machines men born in the usual manner. It is difficult to frame the definitions so as to satisfy these three conditions. One might for instance insist that the team of engineers should be all of one sex, but this would not really be satisfactory, for it is probably possible to rear a complete individual from a single cell of the skin (say) of a man. To do so would be a feat of biological technique deserving of the very highest praise, but we would not be inclined to regard it as a case of "constructing a thinking machine." This prompts us to abandon the requirement that every kind of technique should be permitted. We are the more ready to do so in view of the fact that the present interest in "thinking machines" has been aroused by a particular kind of machine, usually called an "electronic computer" or "digital computer." Following this suggestion we only permit digital computers to take part in our game. . . .

This special property of digital computers, that they can mimic any discrete machine, is described by saying that they are *universal* machines. The existence of machines with this property has the important consequence that, considerations of speed apart, it is unnecessary to design various new machines to do various computing processes. They can all be done with one digital computer, suitably programmed for each case. It will be seen that as a consequence of this all digital computers are in a sense equivalent.

Contrary Views on the Main Question

We may now consider the ground to have been cleared and we are ready to proceed to the debate on our question "Can machines think?" . . . We cannot altogether abandon the original form of the problem, for opinions will differ as to the appropriateness of the substitution and we must at least listen to what has to be said in this connection.

It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question. I believe that in about fifty years' time it will be possible to program computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. I believe further that no useful purpose is served by concealing these beliefs. The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any unproved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research.

I now proceed to consider opinions opposed to my own.

1. *The Theological Objection.* Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think.[1]

I am unable to accept any part of this, but will attempt to reply in theological terms. I should find the argument more convincing if animals were classed with men, for there is a greater difference, to my mind, between the typical animate and the inanimate than there is between man and the other animals. The arbitrary character of the orthodox view becomes clearer if we consider how it might appear to a member of some other religious community. How do Christians regard the Moslem view that women have no souls? But let us leave this point aside and return to the main argument. It appears to me that the argument quoted above implies a serious restriction of the omnipotence of the Almighty. It is admitted that there are certain things that He cannot do such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit? We might expect that He would only exercise this power in conjunction with a mutation which provided the elephant with an appropriately improved brain to minister to the needs of this soul. An argument of exactly similar form may be made for the case

[1] Possibly this view is heretical. St. Thomas Aquinas (*Summa Theologica,* quoted by Bertrand Russell, *A History of Western Philosophy* [New York: Simon and Schuster, 1945], p. 458) states that God cannot make a man to have no soul. But this may not be a real restriction on His powers, but only a result of the fact that men's souls are immortal, and therefore indestructible.

of machines. It may seem different because it is more difficult to "swallow." But this really only means that we think it would be less likely that He would consider the circumstances suitable for conferring a soul. The circumstances in question are discussed in the rest of this paper. In attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates.

However, this is mere speculation. I am not very impressed with theological arguments whatever they may be used to support. Such arguments have often been found unsatisfactory in the past. In the time of Galileo it was argued that the texts, "And the sun stood still . . . and hasted not to go down about a whole day" (Joshua x. 13) and "He laid the foundations of the earth, that it should not move at any time" (Psalm cv. 5) were an adequate refutation of the Copernican theory. With our present knowledge such an argument appears futile. When that knowledge was not available it made a quite different impression.

2. *The "Heads in the Sand" Objection.* "The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so."

This argument is seldom expressed quite so openly as in the form above. But it affects most of us who think about it at all. We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be *necessarily* superior, for then there is no danger of him losing his commanding position. The popularity of the theological argument is clearly connected with this feeling. It is likely to be quite strong in intellectual people, since they value the power of thinking more highly than others, and are more inclined to base their belief in the superiority of Man on this power.

I do not think that this argument is sufficiently substantial to require refutation. Consolation would be more appropriate: perhaps this should be sought in the transmigration of souls.

3. *The Mathematical Objection.* There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete state machines. The best known of these results is known as Gödel's theorem, and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent. There are other, in some respects similar, results due to Church, Kleene, Rosser, and Turing. The latter result is the most convenient to consider, since it refers directly to machines, whereas the others can only be used in a comparatively indirect argument: for instance if Gödel's

theorem is to be used we need in addition to have some means of describing logical systems in terms of machines, and machines in terms of logical systems. The result in question refers to a type of machine which is essentially a digital computer with an infinite capacity. It states that there are certain things that such a machine cannot do. If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply. There may, of course, be many such questions, and questions which cannot be answered by one machine may be satisfactorily answered by another. We are of course supposing for the present that the questions are of the kind to which an answer "Yes" or "No" is appropriate, rather than questions such as "What do you think of Picasso?" The questions that we know the machines must fail on are of this type, "Consider the machine specified as follows. . . . Will this machine ever answer 'Yes' to any question?" The dots are to be replaced by a description of some machine in a standard form. . . . When the machine described bears a certain comparatively simple relation to the machine which is under interrogation, it can be shown that the answer is either wrong or not forthcoming. This is the mathematical result: it is argued that it proves a disability of machines to which the human intellect is not subject.

The short answer to this argument is that although it is established that there are limitations to the powers of any particular machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect. But I do not think this view can be dismissed quite so lightly. Whenever one of these machines is asked the appropriate critical question, and gives a definite answer, we know that this answer must be wrong, and this gives us a certain feeling of superiority. Is this feeling illusory? It is no doubt quite genuine, but I do not think too much importance should be attached to it. We too often give wrong answers to questions ourselves to be justified in being very pleased at such evidence of fallibility on the part of the machines. Further, our superiority can only be felt on such an occasion in relation to the one machine over which we have scored our petty triumph. There would be no question of triumphing simultaneously over *all* machines. In short, then, there might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on.

Those who hold to the mathematical argument would, I think, mostly be willing to accept the imitation game as a basis for discussion. Those who believe in the two previous objections would probably not be interested in any criteria.

4. *The Argument from Consciousness.* This argument is very well ex-

pressed in Professor Jefferson's Lister Oration for 1949, from which I quote. "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants."

This argument appears to be a denial of the validity of our test. According to the most extreme form of this view the only way by which one could be sure that a machine thinks is to *be* the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course no one would be justified in taking any notice. Likewise according to this view the only way to know that a *man* thinks is to be that particular man. It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe "A thinks but B does not" while B believes "B thinks but A does not." Instead of arguing continually over this point it is usual to have the polite convention that everyone thinks.

I am sure that Professor Jefferson does not wish to adopt the extreme and solipsist point of view. Probably he would be quite willing to accept the imitation game as a test. The game (with the player B omitted) is frequently used in practice under the name of *viva voce* to discover whether someone really understands something or has "learned it parrot fashion." Let us listen in to a part of such a *viva voce*:

INTERROGATOR: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

WITNESS: It wouldn't scan.

INTERROGATOR: How about "a winter's day"? That would scan all right.

WITNESS: Yes, but nobody wants to be compared to a winter's day.

INTERROGATOR: Would you say Mr. Pickwick reminded you of Christmas?

WITNESS: In a way.

INTERROGATOR: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.

WITNESS: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

And so on. What would Professor Jefferson say if the sonnet-writing machine was able to answer like this in the *viva voce*? I do not know whether he would regard the machine as "merely artificially signaling" these answers, but if the answers were as satisfactory and sustained as in the above passage I do not think he would describe it as "an easy contrivance." This phrase is, I think, intended to cover such devices as the inclusion in the machine of a record of someone reading a sonnet, with appropriate switching to turn it on from time to time.

In short, then, I think that most of those who support the argument from consciousness could be persuaded to abandon it rather than be forced into the solipsist position. They will then probably be willing to accept our test.

I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localize it. But I do not think these mysteries necessarily need to be solved before we can answer the question with which we are concerned in this paper.

5. *Arguments from Various Disabilities.* These arguments take the form "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X." Numerous features X are suggested in this connection. I offer a selection:

> Be kind, resourceful, beautiful, friendly . . . have initiative, have a sense of humor, tell right from wrong, make mistakes . . . fall in love, enjoy strawberries and cream . . . make someone fall in love with it, learn from experience . . . use words properly, be the subject of its own thought . . . have as much diversity of behavior as a man, do something really new. . . .

No support is usually offered for these statements. I believe they are mostly founded on the principle of scientific induction. A man has seen thousands of machines in his lifetime. From what he sees of them he draws a number of general conclusions. They are ugly, each is designed for a very limited purpose, when required for a minutely different purpose they are useless, the variety of behavior of any one of them is very small, etc., etc. Naturally he concludes that these are necessary properties of machines in general. Many of these limitations are associated with the very small storage capacity of most machines. (I am assuming that the idea of storage capacity is extended in some way to cover machines other than discrete state machines. The exact definition does not matter as no mathematical accuracy is claimed in the present discussion.) A few years ago, when very little had been heard of digital computers, it was possible to elicit much incredulity concerning them, if one mentioned their properties without describing their construction. That was presumably due to a similar application of the principle of scientific induction. These applications of the principle are of course largely unconscious. When a burned child fears the fire and shows that he fears it by avoiding it, I should say that he was applying scientific induction. (I could of course also describe his behavior in many other ways.) The works and customs of mankind do not seem to be very suitable material to which to apply scientific induction. A very large part of space-time must be investigated if reliable results are to be obtained. Otherwise we may (as most English children do) decide that everybody speaks English, and that it is silly to learn French.

There are, however, special remarks to be made about many of the disabilities that have been mentioned. The inability to enjoy strawberries and cream may have struck the reader as frivolous. Possibly a machine might be made to enjoy this delicious dish, but any attempt to make one do so would be idiotic. What is important about this disability is that it contributes to some of the other disabilities, e.g., to the difficulty of the same kind of friendliness occurring between man and machine as between white man and white man, or between black man and black man.

The claim that "machines cannot make mistakes" seems a curious one. One is tempted to retort, "Are they any the worse for that?" But let us adopt a more sympathetic attitude, and try to see what is really meant. I think this criticism can be explained in terms of the imitation game. It is claimed that the interrogator could distinguish the machine from the man simply by setting them a number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy. The reply to this is simple. The machine (programmed for playing the game) would not attempt to give the *right* answers to the arithmetic problems. It would deliberately introduce mistakes in a manner calculated to confuse the interrogator. A mechanical fault would probably show itself through an unsuitable decision as to what sort of a mistake to make in the arithmetic. Even this interpretation of the criticism is not sufficiently sympathetic. But we cannot afford the space to go into it much further. It seems to me that this criticism depends on a confusion between two kinds of mistakes. We may call them "errors of functioning" and "errors of conclusion." Errors of functioning are due to some mechanical or electrical fault which causes the machine to behave otherwise than it was designed to do. In philosophical discussions one likes to ignore the possibility of such errors; one is therefore discussing "abstract machines." These abstract machines are mathematical fictions rather than physical objects. By definition they are incapable of errors of functioning. In this sense we can truly say that "machines can never make mistakes." Errors of conclusion can only arise when some meaning is attached to the output signals from the machine. The machine might, for instance, type out mathematical equations, or sentences in English. When a false proposition is typed we say that the machine has committed an error of conclusion. There is clearly no reason at all for saying that a machine cannot make this kind of mistake. It might do nothing but type out repeatedly "0=1." To take a less perverse example, it might have some method for drawing conclusions by scientific induction. We must expect such a method to lead occasionally to erroneous results.

The claim that a machine cannot be the subject of its own thought can of course only be answered if it can be shown that the machine has *some* thought with *some* subject matter. Nevertheless, "the subject matter of a machine's operations" does seem to mean something, at least to the people who deal with it. If, for instance, the machine was trying to find a solution of the equation $x^2-40x-11=0$, one would be tempted to describe this equation as part of the machine's subject matter at that moment. In this sort of sense a machine undoubtedly can be its own subject matter. It may be used to help in making up its own programs, or to predict the effect of alterations in its own structure. By observing the results of its own behavior it can modify its own programs so as to achieve some purpose more effectively. These are possibilities of the near future, rather than Utopian dreams.

The criticism that a machine cannot have much diversity of behavior is just a way of saying that it cannot have much storage capacity. Until fairly recently a storage capacity of even a thousand digits was very rare.

The criticisms that we are considering here are often disguised forms of the argument from consciousness. Usually if one maintains that a machine *can* do one of these things, and describes the kind of method that the machine could use, one will not make much of an impression. It is thought that the method (whatever it may be, for it must be mechanical) is really rather base. Compare the parenthesis in Jefferson's statement quoted above.

6. *Lady Lovelace's Objection.* Our most detailed information of Babbage's Analytical Engine comes from a memoir by Lady Lovelace. In it she states, "The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform" (her italics). This statement is quoted by Hartree who adds: "This does not imply that it may not be possible to construct electronic equipment which will 'think for itself,' or in which, in biological terms, one could set up a conditioned reflex, which would serve as a basis for 'learning.' Whether this is possible in principle or not is a stimulating and exciting question, suggested by some of these recent developments. But it did not seem that the machines constructed or projected at the time had this property."

I am in thorough agreement with Hartree over this. It will be noticed that he does not assert that the machines in question had not got the property, but rather that the evidence available to Lady Lovelace did not encourage her to believe that they had it. It is quite possible that the machines in question had in a sense got this property. For suppose that some discrete state machine has the property. The Analytical Engine was a universal digital computer, so that, if its storage capacity and speed were adequate, it could by suitable programing be made to mimic the machine in question. Probably this argument did not occur to the Countess or to Babbage. In any case there was no obligation on them to claim all that could be claimed.

This whole question will be considered again under the heading of learning machines.

A variant of Lady Lovelace's objection states that a machine can "never do anything really new." This may be parried for a moment with the saw, "There is nothing new under the sun." Who can be certain that "original work" that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles? A better variant of the objection says that a machine can never "take us by surprise." This statement is a more direct challenge and can be met directly. Machines take me by surprise with great frequency. This is largely because I do not do sufficient calculation to decide what to expect them to do, or rather because, although I do a calculation, I do it in a hurried, slipshod fashion, taking risks. Perhaps I say to myself, "I suppose the voltage here ought to be the same as there; anyway let's assume it is." Naturally I am often wrong, and the result is a surprise for me, for by the time the experiment is done these assumptions have been forgotten. These admissions lay me open to lectures on the subject of my vicious ways, but do not throw any doubt on my credibility when I testify to the surprises I experience.

I do not expect this reply to silence my critic. He will probably say that such surprises are due to some creative mental act on my part, and reflect no credit on the machine. This leads us back to the argument from consciousness, and far from the idea of surprise. It is a line of argument we must consider closed, but it is perhaps worth remarking that the appreciation of something as surprising requires as much of a "creative mental act" whether the surprising event originates from a man, a book, a machine or anything else.

The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false. A natural consequence of doing so is that one then assumes that there is no virtue in the mere working out of consequences from data and general principles.

7. *Argument from Continuity in the Nervous System.* The nervous system is certainly not a discrete state machine. A small error in the information about the size of a nervous impulse impinging on a neuron may make a large difference to the size of the outgoing impulse. It may be argued that, this being so, one cannot expect to be able to mimic the behavior of the nervous system with a discrete state system.

It is true that a discrete state machine must be different from a continuous machine. But if we adhere to the conditions of the imitation game, the interrogator will not be able to take any advantage of this difference. The situation can be made clearer if we consider some other simpler continuous machine. A differential analyzer will do very well. (A differential analyzer is a certain kind of machine not of the discrete state type used for some kinds of calculation.) Some of these provide their answers in a typed form, and so are suitable for taking part in the game. It would not be possible for a digital computer to predict exactly what answers the differential analyzer would give to a problem, but it would be quite capable of giving the right sort of answer. For instance, if asked to give the value of π (actually about 3.1416) it would be reasonable to choose at random between the values 3.12, 3.13, 3.14, 3.15, 3.16 with the probabilities of 0.05, 0.15, 0.55, 0.19, 0.06 (say). Under these circumstances it would be very difficult for the interrogator to distinguish the differential analyzer from the digital computer.

8. *The Argument from Informality of Behavior.* It is not possible to produce a set of rules purporting to describe what a man should do in every conceivable set of circumstances. One might for instance have a rule that one is to stop when one sees a red traffic light, and to go if one sees a green one, but what if by some fault both appear together? One may perhaps decide that it is safest to stop. But some further difficulty may well arise from this decision later. To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible. With all this I agree.

From this it is argued that we cannot be machines. I shall try to reproduce the argument, but I fear I shall hardly do it justice. It seems to run something like this. "If each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines." The undistributed middle is glaring. I do not think the argument is ever put quite like this, but I believe this is the argument used nevertheless. There may however be a certain confusion between "rules of conduct" and "laws of behavior" to cloud the issue. By "rules of conduct" I mean precepts such as "Stop if you see red lights," on which one can act, and of which one can be conscious. By "laws of behavior" I mean laws of nature as applied to a man's body such as "if you pinch him he will squeak." If we substitute "laws of behavior which regulate his life" for "laws of conduct by which he regulates his life" in the argument quoted the undistributed middle is no longer insuperable. For we believe that it is not only true that being regulated by laws of behavior implies being some sort of machine (though not necessarily a discrete state machine), but that conversely being such a machine implies being regulated by such laws. However, we cannot so easily convince ourselves of the absence of complete laws of behavior as of complete rules of conduct. The only way we know of for finding such laws is scientific observation, and we certainly know of no circumstances under which we could say, "We have searched enough. There are no such laws."

We can demonstrate more forcibly that any such statement would be unjustified. For suppose we could be sure of finding such laws if they existed. Then given a discrete state machine it should certainly be possible to discover by observation sufficient about it to predict its future behavior, and this within a reasonable time, say a thousand years. But this does not seem to be the case. I have set up on the Manchester computer a small program using only 1000 units of storage, whereby the machine supplied with one sixteen-figure number replies with another within two seconds. I would defy anyone to learn from these replies sufficient about the program to be able to predict any replies to untried values.

(9) *The Argument from Extrasensory Perception.* I assume that the reader is familiar with the idea of extrasensory perception, and the meaning of the four items of it, viz., telepathy, clairvoyance, precognition, and psychokinesis. These disturbing phenomena seem to deny all our usual scientific ideas. How we should like to discredit them! Unfortunately the statistical evidence, at least for telepathy, is overwhelming. It is very difficult to rearrange one's ideas so as to fit these new facts in. Once one has accepted them it does not seem a very big step to believe in ghosts and bogies. The idea that our bodies move simply according to the known laws of physics, together with some others not yet discovered but somewhat similar, would be one of the first to go.

This argument is to my mind quite a strong one. One can say in reply that many scientific theories seem to remain workable in practice, in spite of clashing with E.S.P.; that in fact one can get along very nicely if one forgets about it. This is rather cold comfort, and one fears that thinking is just the kind of phenomenon where E.S.P. may be especially relevant.

A more specific argument based on E.S.P. might run as follows: "Let us play the imitation game, using as witnesses a man who is good as a telepathic receiver, and a digital computer. The interrogator can ask such questions as 'What suit does the card in my right hand belong to?' The man by telepathy or clairvoyance gives the right answer 130 times out of 400 cards. The machine can only guess at random, and perhaps get 104 right, so the interrogator makes the right identification." There is an interesting possibility which opens here. Suppose the digital computer contains a random number generator. Then it will be natural to use this to decide what answer to give. But then the random number generator will be subject to the psychokinetic powers of the interrogator. Perhaps this psychokinesis might cause the machine to guess right more often than would be expected on a probability calculation, so that the interrogator might still be unable to make the right identification. On the other hand, he might be able to guess right without any questioning, by clairvoyance. With E.S.P. anything may happen.

If telepathy is admitted it will be necessary to tighten our test. The situation could be regarded as analogous to that which would occur if the interrogator were talking to himself and one of the competitors was listening with his ear to the wall. To put the competitors into a "telepathy-proof room" would satisfy all requirements.

Reflections

Most of our response to this remarkable and lucid article is contained in the following dialogue. However, we wish to make a short comment about Turing's apparent willingness to believe that extrasensory perception might turn out to be the ultimate difference between humans and the machines they create. If this comment is taken at face value (and not as some sort of discreet joke), one has to wonder what motivated it. Apparently Turing was convinced that the evidence for telepathy was quite strong. However, if it was strong in 1950, it is no stronger now, thirty years later—in fact, it is probably weaker. Since 1950 there have been many notorious cases of claims of psychic ability of one sort or another, often vouched for by physicists of some renown. Some of those physicists have later felt they had been made fools of and have taken back their public pro-ESP pronouncements, only to jump on some new paranormal bandwagon the next month. But it is safe to say that the majority of physicists—and certainly the majority of psychologists, who specialize in understanding the mind—doubt the existence of extrasensory perception in any form.

Turing took "cold comfort" in the idea that paranormal phenomena might be reconcilable in some way with well-established scientific theories. We differ with him. We suspect that if such phenomena as telepathy, precognition, and telekinesis turned out to exist (and turned out to have the remarkable properties typically claimed for them), the laws of physics would not be simply *amendable* to accommodate them; only a major revolution in our scientific world view could do them justice. One might look forward to such a revolution with eager excitement—but it should be tinged with sadness and perplexity. How could the science that had worked so well for so many things turn out to be so wrong? The challenge of rethinking all of science from its most basic assumptions on up would be a great intellectual adventure, but the evidence that we will need to do this has simply failed to accumulate over the years.

D.R.H.
D.C.D.

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 apparently 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 modern 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 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 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 torn 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 hurricanes. 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 hurricanehood." 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 rudimentarily 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 biological-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) verificationism. (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 what John 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. 372)

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.