

JAMES H. MOOR

## AN ANALYSIS OF THE TURING TEST

(Received 11 September, 1975)

### I. THE TURING TEST

In his classic article, 'Computing Machinery and Intelligence', A. M. Turing suggests that the ambiguous question, 'Can machines think?' should be replaced with a new set of questions involving a game which he calls 'the imitation game'.<sup>1</sup> In this paper I wish to argue that the proponents and critics of the imitation game have misunderstood its significance. The real value of the imitation game lies not in treating it as the basis for an operational definition but in considering it as a potential source of good inductive evidence for the hypothesis that machines think. With this understanding the four standard criticisms against the game do not apply.

A standard version of the imitation game involves a man, a computer, and a human interrogator. The interrogator stays in a room apart from the man and the computer and must on the basis of answers to questions that he puts to each via a teletypewriter decide which respondent is the man and which respondent is the computer.<sup>2</sup> In this version of the game, which is often called 'the Turing test', the basic question which replaces the question 'Can machines think?' might be put, 'On the average after  $n$  minutes or  $m$  questions is an interrogator's probability of correctly identifying which respondent is a machine significantly greater than 50 percent?'

If the number of minutes and questions were kept very small, then playing the imitation game would be little more than an entertaining pastime. But, in order to make the imitation game less of a game and more of a test with interesting results let us assume that the following situation occurs. The imitation game is played by many different interrogators each of whom has ample opportunity to ask many questions (each taking a week to make thousands of inquiries if you wish) and the results are such that the probability of the average interrogator correctly identifying the machine is

not significantly greater than 50 percent. If such a situation did occur, then one would have little doubt that the imitation game was played well by the machine, but one might have a lingering doubt about the significance of this result. One is tempted to say, "Sure, the machine plays the imitation game very well but so what?"

## II. THE SIGNIFICANCE OF THE TURING TEST

Unfortunately, Turing does not help us very much in understanding the significance of the imitation game. At one point he claims that the question, 'Can machines think?' is "too meaningless to deserve discussion".<sup>3</sup> But, if Turing intends that the question of the success of the machine at the imitation game replace the question about machines thinking, then it is difficult to understand how we are to judge the propriety and adequacy of the replacement if the question being replaced is too meaningless to deserve discussion. Our potential interest in the imitation game is aroused not by the fact that a computer might learn to play yet another game, but that in some way this test reveals a connection between possible computer activities and our ordinary concept of thinking.

The usual understanding of the Turing test put forth by its proponents (and critics) is that it provides an operational definition. For instance, P. H. Millar describes it as a virtue of the Turing test that "it constitutes an operational definition which, given a computer terminal system, can be used as a criterion."<sup>4</sup> It is usually not specified whether the Turing test is intended as an operational definition of 'computer thinking', "the equivalence of human and computer thinking", or what. But in any case the benefits of this type of interpretation seem to me to be illusory. Either one is giving an arbitrary definition of a new term in which case the operational definition needs no justification but the Turing test becomes very uninteresting, or one is attempting to capture a notion which is at least related to our ordinary concept of thinking in which case the Turing test is very interesting but the operational definition requires some justification. In short, one cannot assess the significance of the Turing test without attending to the ordinary concept of thinking as unclear as that concept might be.

To think is to process information in ways which involve recognition, imagination, evaluation and decision. To a moderate extent computers perform these processes already. The question is whether a computer could

eventually process information in these ways with the sophistication of a normal, living, adult human being; and if it could, how would we know it? The clue lies in considering the basis for our knowledge that other humans think. I believe that another human being thinks because his ability to think is part of a theory I have to explain his actions. The theory postulates a number of inner information processes, but the evidence for the theory comes from the outward behavior of the person. On the basis of his behavior I can confirm, disconfirm, and modify my theory. Furthermore, there is no reason why knowledge of computer thinking can not arise in the same way. I can use the computer's behavior as evidence in assessing my theory about its information processing. In neither the human case nor the computer case must I consider the thinking to be on a close analogy with my own, for the evidence might dictate that the human or computer discriminates and evaluates quite differently than I do.

If passing the Turing test is not an operational definition, then is it at least a necessary condition for making an inductive inference about computer thinking? Again, I believe the answer is negative. One might have solid inductive evidence based on the computer's behavior that it was thinking as well as a human but the computer could not pass the Turing test simply because it was a poor actor. Thus, the Turing test is not essential to our knowledge about computer thinking. One could certainly construct and test a theory about computer thinking in much the same way one constructs and tests a theory about human thinking without becoming engaged in the Turing test.

Then what is the value of the Turing test? I believe that the significance of the Turing test is that it provides one good format for gathering inductive evidence such that if the Turing test was passed, then one would certainly have very adequate grounds for inductively inferring that the computer could think on the level of a normal, living, adult human being. Beyond the obvious reason that the Turing test eliminates prejudice due to the appearance of the computer I believe there are two strong arguments why the Turing test is a good format for gathering inductive evidence. First, the Turing test permits direct or indirect testing of virtually all of the activities one would count as evidence for thinking. Foremost, the Turing test permits (even demands) evaluation of linguistic behavior which is central to our inductive inferences about how others think. In the Turing test format the nonverbal behavior of the respondents cannot be directly observed by the interrogator, but this

limitation is not as severe as is usually supposed. Since nonverbal behavior can be described, probing questions can be put to the respondents about how they would perform various activities involving thinking, e.g., designing a house, balancing a bank account, playing bridge, etc. Secondly, the Turing test encourages severe testing. It is a familiar point that the confirmation of a theory can come all too easily. If a computer simply repeated one cognitive activity many times and did it well, even if it is a complex activity like playing chess, it does not follow that the computer's thinking capacity has been critically tested. In the Turing test, however, the computer would be tested in detail over a wide range of subjects. Moreover, the interrogator's goal is to find a refuting instance which gives the computer away.

### III. REPLIES TO OBJECTIONS

#### *1. The Objection Concerning Behaviorism*

One of the most common objections to the Turing test is that the test must be based upon a behavioristic construal of the concept of thinking and any such behavioristic analysis is absurd. Probably the most imaginative form of this argument is given by Keith Gunderson. Gunderson creates a parody of the imitation game by asking whether rocks could imitate.<sup>5</sup> He imagines an elaborate apparatus consisting of a rock box, electric eye, releasing mechanism, etc. which could land on someone's toe with about the same effect as a person's foot. It would be the interrogator's job to determine if it was a rock box or a person stepping on his toe. If the interrogator could not reliably determine whether it was a person or a rock box that was stepping on his toe, the rock box would pass the test. Gunderson concludes:

The parody comparison can be pushed too far. But I think it lays bare the reason why there is no contradiction involved in saying, "Yes, a machine can play the imitation game, but it can't think". It is for the same reason that there is no contradiction in saying, "Of course a rock box of such-and-such a sort can be set up, but rocks surely can't imitate". For thinking (or imitating) cannot be fully described simply by pointing to net results such as those illustrated above. For if this were not the case it would be correct to say that a phonograph could sing, and that an electric eye could see people coming.<sup>6</sup>

I have tried to argue that an understanding of the Turing test is not necessarily dependent upon a behavioristic analysis of thinking, i.e., providing an operational definition of thinking. On the contrary, since our knowledge

of thinking by others has an inductive basis, it certainly is neither a surprise nor a criticism to point out that the statement "Yes, a machine can play the imitation game, but it can't think" is not a contradiction. It is doubtful whether the toe-stepping game provides a good format for generating adequate evidence to establish that the rock box apparatus (let along the rocks themselves as Gunderson misleadingly suggests) imitates. It surely does not follow from this that the Turing test fails to provide a good format for generating adequate behavioral evidence to justify the induction that a computer thinks.

## *2. The Objection Concerning Mechanism*

Michael Apter suggests that Turing's position is question-begging since it assumes that the brain is a machine.<sup>7</sup> As a factual matter Turing probably did believe that the brain is a machine in some sense although not a discrete state machine. Turing says, "In considering the functions of the mind or the brain we find certain operations which we can explain in purely mechanical terms".<sup>8</sup> But he is careful to qualify his view stating that he does not claim to be giving convincing arguments, rather they should be described as "recitations tending to produce belief". What is important for our purposes, however, is that the Turing test in no way depends upon the assumption that the brain is a machine. My claim is that if the test was passed in the sense discussed above, then one should conclude that both men and machines can think. One is not forced to assume or conclude from this that brains are machines (or machines are brains) any more than one would assume or conclude that a human who printed very well must be a typewriter.

I do not doubt that 'machines' might be construed broadly enough to include brains, but the Turing test is far more interesting if 'machines' is taken in a more ordinary and narrow sense which would include digital computers, even if made of another generation of electronic parts, but which would exclude brains. For if brains were machines, then the question 'Can machines think?' seems to be answered very easily in the affirmative.

## *3. The Objection Concerning Internal Operation*

If a digital computer could pass the Turing test, then there would surely be a natural curiosity and legitimate interest in how the computer could

accomplish this feat. It is sometimes argued that if the computer was to accomplish this feat by using very unorthodox methods, then there would be grounds for believing that the computer did not think; and therefore, the Turing test is inadequate.

In general, I believe there is something valuable in this criticism. It does seem possible that a computer might accomplish its behavioral repertoire on a much different basis than a human being; and there would indeed be benefit, as Gunderson points out, in making comparisons between the cognitive processes of human beings and computers and between the operations of the human nervous system and the computer's circuitry. What is not clear is how this is a criticism of the Turing test. The underlying danger is that two very similar claims may be confused:

- (i) Evidence about the internal operation of a computer *might alter* a justified inductive inference that the computer can think on the level of a normal, living, adult human being.
- (ii) Evidence about the internal operation of a computer *is necessary to make* a justified inductive inference that the computer can think on the level of a normal, living, adult human being.

I believe that (i) is true. For example, an extreme case is that of Baron von Kempelen's very successful chess-playing machine in which a man was so cleverly hidden that although people were allowed to examine all of the interior of the machine at different times, the man inside escaped detection by moving around.<sup>9</sup> In this case it is clear that information about the internal operation of the chess-playing machine would influence one to reconsider his judgements about the machine's actual abilities. In less extreme cases it may be question of comparing different physical processes which perform similar informational functions. Just how different such physical processes can be and still be regarded as the same information process is an open and interesting question. Although even if the information processes of computers were considered different from information processes of humans, it is not clear whether one would conclude that computers did not think or just that they thought by a different means from humans.

But, the essential point is that to grant (i) is only to admit that further evidence might alter inductive inferences and this is certainly no criticism of the Turing test. In order for critics to use the objection concerning internal

operation as a criticism against the Turing test they must show that (ii) is true. Not only do the critics fail to show that (ii) is true, but I believe for the reasons given in section II that (ii) must be false. Beyond what we infer about cognitive processes from the behavioral level, we know very little about the relevant internal operations of human beings. Yet, people have legitimately inferred on the basis of behavior that others could think at least since Aristotle who, of course, believed that the brain was a cooling agent for the blood.

#### 4. *The Objection Concerning the Scope of the Test*

If one agrees that it is the behavioral evidence which is most crucial in assessing the thinking capabilities of a computer, it can still be objected that the Turing test is inadequate because it is only one test of behavior. Gunderson compares the situation to a vacuum cleaner salesman who claims that his vacuum cleaner is all-purpose but only demonstrates that the vacuum cleaner can pick up bits of dust.<sup>10</sup> Gunderson's point is that one expects other activities from a computer which is claimed to think than merely the ability to play one game. Jerry Fodor argues further, "Turing would presumably have been dissatisfied with a device that could answer questions about how to boil water if it routinely put the kettle in the icebox when told to brew the tea".<sup>11</sup>

Again it is important to distinguish two very similar claims:

- (iii) Behavioral evidence which cannot be directly obtained in the Turing test *might alter* a justified inductive inference that a computer can think on the level of a normal, living, adult human being.
- (iv) Behavioral evidence which cannot be directly obtained in the Turing test *is necessary to make* a justified inductive inference that a computer can think on the level of a normal, living, adult human being.

I believe that (iii) is true. I do not wish to deny that further testing beyond the Turing test would be valuable and that the results of such further testing might make one revise inferences based on the results of the Turing test

alone. It is interesting to note, however, that if the disconfirmation was not too severe, e.g., a situation in which the computer passed the Turing test, had reasonably good nonverbal behavior, but routinely put the kettle in the icebox when told to brew the tea, then one might attribute the problem to whimsy or to difficulties in the computer's motor apparatus rather than reject the hypothesis that the computer was capable of thinking.

Again the essential point is that to grant (iii) is only to admit a well known fact about inductive inferences. In order to attack the Turing test the critic must show that (iv) is true. And again, not only do the critics fail to show that (iv) is true, but I believe for the reasons given in Section II that (iv) must be false. It is simply a misleading numbers game to suggest that the Turing test is only one test. The Turing test provides a format for directly or indirectly examining any of a wide variety of activities which would count as evidence for thinking.

#### IV. CONCLUSION

In this paper I have argued that the Turing test is best not treated as the basis for an operational definition, and in any event acceptance of the Turing test does not allow us to avoid the question of how knowledge of computer thinking is possible. Nonetheless, the Turing test is a significant test for computer thought if it is interpreted inductively. Under such an interpretation the standard criticisms of the Turing test demonstrate not that the test is defective but only that it is subject to the canons of good scientific methodology.

There are, however, important limitations of the Turing test. As a practical matter the test is of little value in guiding research and investigators have quite rightly proposed much more limited modifications of the Turing test or alternatives to it. Another drawback of the Turing test is that it places so much emphasis upon the computer's ability to act and deceive that the computer's grasp of the world qua computer is not very well tested.<sup>12</sup> I believe it is likely that if thinking is attributed to computers, the attribution will be a gradual process as computers acquire more and more skills (in much the same way that attribution of thinking is gradually given to a child as he develops). Thus, ironically if a computer was eventually developed which could pass the Turing test, then it would probably be unnecessary to run the



test since the inductive evidence for the computer's thinking capacity would have already been gathered during the computer's development.

Finally, it is important to distinguish between Turing's imitation game, a possible empirical test which can be defended conceptually, and Turing's prophecy that "in about fifty years' time it will be possible to programme 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".<sup>13</sup> Since computer science is a young field, one can understand researchers' high hopes pinned to visions of the development of better computers with bigger memories, more parallel processing, and improved heuristic programming. In the long run researchers must have significant results as well as high hopes if Turing's promissory note is to maintain its currency. But, if the vision behind Turing's prophecy should become a reality, then attributing thought to computers might be less a matter of regressing to anthropomorphism than escaping from egocentricity.

*Dartmouth College*

#### NOTES

<sup>1</sup> A. M. Turing, 'Computing Machinery and Intelligence', *Mind* LIX (October, 1950), pp. 433-460.

<sup>2</sup> I believe this version represents the usual interpretation of the imitation game although Turing's own description is ambiguous. See Turing, pp. 433-434.

<sup>3</sup> Turing, p. 442.

<sup>4</sup> P. H. Millar, 'On the Point of the Imitation Game', *Mind* LXXXII (October, 1973), p. 595.

<sup>5</sup> Keith Gunderson, *Mentality and Machines* (Garden City, New York, Doubleday & Company, Inc., 1971), p. 41f.

<sup>6</sup> Gunderson, p. 44.

<sup>7</sup> Michael J. Apter, *The Computer Simulation of Behaviour* (New York, Harper & Row, Publishers, 1971), p. 68.

<sup>8</sup> Turing, p. 454.

<sup>9</sup> Cf. Jurg Nievergelt and J. Craig Farrar, 'What Machines Can and Cannot Do', *American Scientist* 61 (May-June, 1973), p. 309.

<sup>10</sup> Gunderson, pp. 53-55.

<sup>11</sup> Jerry Fodor, *Psychological Explanation: An Introduction to the Philosophy of Psychology* (New York, Random House, Inc., 1968), pp. 126-127.

<sup>12</sup> For example, it might turn out that the computer played the imitation game so well because it believed that it really was human! In such a case the computer would likely be regarded as a thinker but as somewhat psychotic.

<sup>13</sup> Turing, p. 442