65 The Social Implications of Artificial Intelligence

by I. J. GOOD

When the first really flexible electronic computers were still under construction, Turing[1] wrote an article on whether a computer could 'think'. He was not concerned with metaphysical questions since he *defined* 'thinking', for the purposes of his paper, as the ability to answer questions in a manner resembling a human being, though on a teleprinter. His discussion disposed of many philosophical objections and was probably influential in provoking some of the programming effort on learning machines and problem-solving machines.[2] At the time, the majority of scientists, tending to rely on primitive scientific induction, 'biological induction' as it were, denied the feasibility of intelligent machinery, and a frequently heard cliché at symposia on computing machines was that 'a computer will do only what it is programmed to do, except when it goes wrong'. They used to go wrong very frequently at that time, but even when they did they did not do anything intelligent. The cliché served a purpose in explaining the nature of current computers to the uninitiated, but it has begun to wear thin now that computers have been programmed to play draughts (checkers) and to prove elementary mathematical theorems.[3] Some of the proofs have been better than those that the programmers themselves would have thought of. A good example is the proof that the base angles of an isosceles triangle, AOB, are equal; namely that the triangles AOB and BOA are congruent since $OA = OB$, $OB = OA$, and the angles AOB and BOA are equal!

It is true that the programs so far have not produced much really original 'thought', but the work is being greatly accelerated both by improvements in computers, and in progamming techniques, especially the latter. The elementary instructions in these programs are being built up into larger and more intuitively appealing units, and they enable the human to communicate with the machine with greater and greater flexibility. Programs can be quickly modified, in minutes rather than weeks, and consequently the work on artificial intelligence can be expected to expand exponentially during say the next eight years. The variety of applications will likewise increase rapidly and it is not easy to see where the saturation point will be.

A very important question when trying to judge where this point will be is whether a computer could be effectively programmed by means of ordinary typed language. This will, I think, become possible if and only if nearly perfect mechanical translation becomes possible, for I agree with Miss Masterman[4] that the latter will require a deep analysis of semantic problems. These problems involve the logic of analogy whereas existing programming languages for computers barely do so.[5]

It is possible that programming techniques alone will not be sufficient for this goal to be reached, and it may be necessary to build machines having something of a biological structure resembling that of the brain. That is, it may be necessary to incorporate random-looking, if not actually partly random, networks in the machine. Methods of reinforcement, already used in programs,[6] may be much more effective when applied to a functionally more natural structure containing a great deal of 'parallel working'.[7] The training of such a machine would resemble the training of a baby, the primary methods being *demonstration* together with positive and negative reinforcement, followed by *talking* together with reinforcement. A machine that combined these 'bionic' features with those of an ordinary computer might well be the solution.[8]

A baby is a very complicated 'device', a product of billions of years of evolution, but only a million of those years were spent in human form. Consequently our main problem is perhaps to program or build a machine with the latent intelligence of a small lizard, totally unable to play draughts. A small lizard is handicapped by having a small brain. If we could build a machine with the latent intelligence of a small lizard, then, at many times the expense, we could probably build one with that of a baby. With a further small percentage increase in cost we could reach the level of the baby Newton and better.[9] We could then educate it and teach it its own construction and ask it to design a far more economical and larger machine.

At this stage there would unquestionably be an explosive development in science,[10] and it would be possible to let the machines tackle all the most difficult problems of science. Many of the most pressing problems, such as those of medicine and of information retrieval, would make giant strides every month, and human scientists might have to take a back seat.

Such machines, properly 'motivated', could even make useful political and economic suggestions; and would *need* to do so in order to compensate for the problems created by their own existence. There would be problems of overpopulation, owing to the elimination of disease, and of unemployment, owing to the efficiency of low-grade robots that the main machines had designed. (These robots would take their orders from people, but their 'brains' would probably be in the main machines which would communicate with their 'bodies' continually.)

Provided that the suggestions made by the machines were sensible and provided that they were clothed in impressive status-acquiring boxes, we should at last have an effective *deus ex machina.*

It may be argued that machines cannot be properly motivated because they do not feel pleasure or pain. The answer is twofold: (i) homing missiles, for example, behave *as if* they were motivated; (ii) perhaps machines *could* feel pleasure and pain after all.[11]

It would be economical for the machines to be able to communicate with one another, so as to be able to act as a single machine. Here they would have an advantage over human committees who have to communicate via the slow medium of speech. We do not need to work out how the machines would communicate with one another, since the method would itself be suggested by them. Presumably the organisation would be hierarchical. If the communication were tight enough it would be a matter of definition whether we had one machine or several.

If a machine were selfish it might not wish to replace itself by designing a much better one. But we could overcome this difficulty by permitting the machine to improve itself, for example, by becoming larger.

In fact it is desirable that, sooner or later, there should be only one of these futuristic oracles, under the control of the United Nations. Otherwise there would be a danger that the machines would come into conflict.

If the first two machines are built in America and Russia they do not need to come into conflict. For if they meet at a summit conference, possibly via satellite communication, they will presumably decide to connect themselves together into a single machine. This is a possible route to world government. Oracles of the world unite!

For what it is worth, my guess of when all this will come to pass is 1978, and the cost $\$10^{8\cdot 7 \pm 1\cdot 0}$. It would be cheap at the price.

My reason for estimating so high a cost is that the human cerebral cortex has some 10,000,000,000 neurons, each of which has perhaps a hundred dendrites. But since the electronic components of the future will operate some million times as fast as a neuron, I may have overestimated the cost.[11]

I need hardly say that my expectations may be wrong, and that we may have to be satisfied with machines that are *merely fantastically useful* for information retrieval and for other activities that require only a modicum of intelligence combined with an immense store ('memory'). Such machines will almost certainly transform both scientific research and trading methods. They will also be able to translate languages and write music and poetry, and will do these things at great speed, but most of this work will have the appearance of being uninspired by gifted human standards. If there is such a ceiling to the intelligence of machines, then the main hope of an 'explosive' development in all sciences and arts will come from specially bred men and women. (The possibility of breeding men of greater genius than has ever been known before is especially well discussed by R. A. McConnell[13] in an article that should have been submitted to this volume.) But I regard it as 'odds on' that the 'ceiling' does not exist. It is possible that the ceiling for machines will exist until the artificially-bred human geniuses attack the problem!

[Some problems in science might still be so intractable that even a very high-quality artificial intelligence would take a very long time to make appreciable progress on them. This may well be true of the really hard problems faced by psychical research; in such problems, no decisive progress has yet been made, in spite of the efforts made by many of the best scientists for many decades. Furthermore, in some of these very difficult fields of investigation, human scientists might have an advantage over an artificial intelligence, by being able to obtain clues of the right sort from experiences that are not translatable into machine language. *A. Ed.*]

REFERENCES AND NOTES

1. TURING, A. M.: 'Computing machinery and intelligence': *Mind*, volume 59, pages 433–60, 1950. Reprinted in *The World of Mathematics*, pages 2099–2123, Simon and Schuster, New York, 1956.
2. A suggestion that a problem-*putting* machine may be more economical has been made by Mieczyslaw Choynowski, 'Zalezenia cybernetyki a zagadnienia biologii ('Assumptions of cybernetics and problems of biology'), Pañstwowy Zaklad Wydawnictw Lekarskich, Warsaw, 1957.
3. For an excellent survey of such work *see* MINSKY, M. L.: 'Steps

towards artificial intelligence': *Proc. I.R.E.*, volume 49, pages 8–30, 1961.

4. Miss Margaret Masterman in a Symposium, 1961, published by the Aristotelian Society, 169–216.
5. 'Analogue' computers are stupidly named: they should be called 'continuous computers'.
6. *See*, for example, FRIEDBERG, R. M.: *IBM J. Res. Dev.*, volume 2, pages 2–13, 1958; volume 3, pages 282–7, 1959; SAMUEL, A. L.: *IBM J. Res. Dev.*, volume 3, pages 210–29, 1959; MICHIE, D.: in *Penguin Science Survey*, No. 2, 1961. GEORGE, F. H., has also done some unpublished work.
7. *See*, for example, Frank Rosenblatt's papers on the Perceptron, put out by the Cornell Aeronautical Laboratory, 1958–60; and my paper 'Speculations on Perceptrons and other Automata', *IBM Res. Rep.*, *RC* 115, June 1959.
8. 'Could a machine make probability judgments?': *Computers and Automation*, volume 8, pages 14–16 and 24–6, 1959. One of the points in this paper was that we should tend to say that a machine had made a judgement when it had made a decision whose logic was not fully understood by the designers and programmers. Such judgements are liable to occur, in suitable machines or programs, as a consequence of *reinforcement*. An important question is under what circumstances a machine should reinforce itself. One criterion, when it is trying to classify objects, is whether the objects put in each class are correlated with each other in properties *not already used* in the process. The process should then be reinforced, but also the new properties should be incorporated, at first with small weights. Other criteria occur in the 'theory of clumps'. *See* pbi No. 33. For a further note on reinforcement, *see Information Theory: Fourth London Symp.*, pages 248–9 (ed. by E. C. Cherry), London, 1961.
9. Perhaps a real or artificial brain containing more than a certain number of neurons would involve insuperable problems of organisation, and might be 'mentally unbalanced', just as business organisations and countries can be too large. This would explain why genius is allied to madness, and why people as a whole are not much more intelligent. But both these things are susceptible of more acceptable explanations. For example, geniuses and madmen are often maltreated in their youth. Also the human cerebral neuron may be as small as it physically can be without too much communicational 'noise', and the head may be about

as large as it can be without the human becoming top-heavy, and the human body may be about as tall as it can be without undue strain on the bones.

10. *J. Roy. Stat. Soc. A*, volume 114, page 107, 1951; and the IBM paper in Note 7 above. By 'explosive' I mean far faster than has ever occurred before in any science.
11. *See* 'The mind-body problem, or could an android feel pain?': in *Theories of the Mind* (ed. by J. M. Scher, Glencoe Free Press, Urbana) forthcoming. In this paper I overlooked Michael Scriven's, 'The compleat robot: a prolegomena to androidology', in *Dimensions of Mind* (ed. by S. Hook) 1960.
12. It has been emphasised by R. A. Fairthorne in a private communication that experiments on the stimulation of organisations, such as artificial neural nets, may be very misleading when the number of components is not large enough, and that there may be an analogy with the 'scale effect' in mechanical engineering. He refers to *Computer J.*, volume 1, 1958, No. 1, and to R. M. Fano, in *Proc. Int. Conf. on Sci. Information*, Washington, D.C., volume 2, page 1407, 1958.
13. McConnell, R. A.: 'The absolute weapon': *Amer. Inst. of Biol. Sc. Bull.*, volume 11, pages 14–16, 1961.

66 What to do about Automation

by OLIVER G. SELFRIDGE

One of the pressing problems of this century is Automation, and its social implications, especially unemployment. We suggest that being unemployed is a harder job than most and should be rewarded accordingly. If the pay were high enough, people would willingly give up half their pay in order to have a job.