Intelligent Machinery, A Heretical Theory

A. M.TURING

‘You cannotmakea machinetothink foryou.’Thisisacommonplacethat is usually acceptedwithout question. Itwill bethe purposeof thispaper to question it.

Mostmachinery developedfor commercialpurposes isintended tocarry outsomeveryspecificjob,andtocarryitoutwithcertainty andconsiderablespeed. Very oftenit does thesame series ofoperations overandover again without any variety.Thisfact about the actualmachinery available isapowerful argumenttomany infavourofthesloganquoted above. To amathematical logician thisargument isnotavailable,forithas been shownthat therearemachinestheoretically possiblewhichwilldosomethingvery closetothinking. Theywill,forinstance, testthevalidity ofaformalproof in thesystem of Principia Mathematica,or eventell ofa formulaof thatsystemwhether itisprovableordisprovable.Inthecase thattheformulais neitherprovable nor disprovablesuchamachine certainlydoes not behavein averysatisfactory manner, for itcontinues towork indefinitelywithout producing any result atall, butthis cannot be regarded as very different fromthereaction ofthemathematicians, whohave forinstance worked for hundredsof yearsonthequestion as to whether Fermant’s last theorem is true or not.For the case of machines ofthiskindamoresubtlekindofagreement isnecessary. By Gödel’sfamous theorem, or somesimilarargument, one canshowthat however the machine isconstructedtherearebound to be cases where the machinefails to give an answer, but a mathematician would beableto. On theotherhand, the machine has certainadvantagesover the mathematician.Whatever it does can be reliedupon, assuming no mechanical ‘breakdown’，whereas themathematician makes a certainproportion of mistakes.I believe thatthis dangerof themathematician makingmistakesisanunavoidable corollaryofhispowerofsometimes hitting uponanentirelynewmethod. This seemstobeconfirmed by the well known factthat themost reliablepeople will not usually hitupon reallynew methods.

My contention is that machines can be constructed which will simulate the behaviour of thehuman mind very closely. They will make mistakes at times, and at times they may make new and very interesting statements, and on the whole the output of them willbe worth attention to the same sort of extent as the output of a human mind.The content of this statement lies in the greater frequencyexpected for the true statements,and it cannot, I think, be given an exact statement. It would not, forinstance, be sufficient to say simply that the machine will make any truestatement sooner or later, for an example of such a machine would be one whichmakes all possible statementssooner or later. We know how to construct these, and as they would (probably) produce true and falsestatements about equally frequently, their verdicts would be quite worthless.It would be the actual reaction of the machine to circumstances that wouldprove my contention, if indeed it can be proved at all.

Let us go rather more carefully into the nature of this ‘proof'. It is clearly possible to produce amachine which would give a very good account of itself for any range of tests, if the machine were made sufficiently elaborate. However,this again would hardly be considered an adequate proof. Such a machine wouldgive itself away by making the same sort of mistake over and over again, andbeing quite unable to correct itself, or to be corrected by argument fromoutside. If the machine were able in some way to ‘ learn by experience ’ itwould be much more impressive. If this were the case there seems to be no real reason why one should not start from a comparatively simplemachine, and, by subjecting it to a suitable range of ‘experience’ transform itinto one which was much more elaborate, and was able to deal with a far greaterrange of contingencies. This process could propably be hastened by a suitable selection of the experiences to which it was subjected. This might becalled ‘education ’. But here we have to be careful. It would be quite easy toarrange the experiences in such a way that theyautomatically caused the structure of the machine to build up into a previously intendedform, and this would obviously be a gross form of cheating, almost on a parwith having a man inside the machine. Here again the criterion as to what wouldbe considered reasonable in the way of ‘education’ cannot be put intomathematical terms, but I suggest that the following would be adequate inpractice. Let us suppose that it is intended that the machine shall understandEnglish, and that owing to its having no hands or feet, and not needing to eat,not desiring to smoke, it will occupy its time mostly in playing games such as Chess and GO,and possibly Bridge. The machine isprovided with a typewriter keyboard on which any remarks to it are typed, andit also types out any remarks that it wishes to make. I suggest that theeducation of the machine should be entrusted to some highly competent schoolmasterwho is interested in the project but who is forbidden any detailed knowledge ofthe inner workings of the machine. The mechanic who has constructed themachine, however, is permitted to keep the machine in running order, and if hesuspects that the machine has been operating incorrectly may put it back to oneof its previous positions and ask the schoolmaster to repeat his lessons fromthat point on, but he may not take any part in the teaching. Since this procedurewould only serve to test the bona fides of the mechanic, I need hardly saythat it would not be adopted in the experimental stages. As I see it, thiseducation process would in practice be an essential to the production of areasonably intelligent machine within a reasonably short space of time. Thehuman analogy alone suggests this.

I maynow give some indication of the way in which such a machine might be expectedto function. The machine would incorporate a memory. This does not need verymuch explanation. It would simply be a list of all the statements that had beenmade to it or by it, and all the moves it had made and the cards it had playedin its games. These would be listed in chronological order. Besides thisstraightforward memory there would be a number of ‘indexes of experience’. Toexplain this idea I will suggest the form which one such index might possiblytake. It might be an alphabetical index of the words that had been used givingthe 'times’ at which they had been used, so that they could be looked up in thememory. Another such index might contain patterns of men or parts of a GO boardthat had occurred. At comparatively late stages of education the memory mightbe extended to include important parts of the configuration of the machine ateach moment, or in other words it would begin to remember what its thoughts hadbeen. This would give rise to fruitful new forms of indexing. New forms ofindex might be introduced on account of special features observed in the indexesalready used. The indexes would be used in this sort of way. Whenever a choicehas to be made as to what to do next features of the present situation arelooked up in the indexes available, and the previous choice in the similarsituations, and the outcome, good or bad, is discovered. The new choice is madeaccordingly. This raises a number of problems. If some of the indications arefavourable and some are unfavourable what is one to do? The answer to this willprobably differ from machine to machine and will also vary with its degree ofeducation. At first probably some quite crude rule will suffice, e.g., to dowhichever has the greatest number of votes in its favour. At a very late stageof education the whole question of procedure in such cases will probably havebeen investigated by the machine itself, by means of some kind of index, andthis may result in some highly sophisticated, and, one hopes, highlysatisfactory, form of rule. It seems probable however that the comparativelycrude forms of rule will themselves be reasonably satisfactory, so that progresscan on the whole be made in spite of the crudeness of the choice rules. Thisseems to be verified by the fact that Engineering problems are sometimes solvedby the crudes t rule of thumb procedure which only deals with the mostsuperficial aspects of the problem, e.g., whether a function increases or decreaseswith one of its variables. Another problem raised by this picture of the waybehaviour is determined is the idea of 'favourable outcome’. Without some such idea,corresponding to the ‘pleasure principle' of the psychologists, itis very difficult to see how to proceed. Certainly it would be most natural tointroduce some such thing into the machine. I suggest that there should be twokeys which can be manipulated by the schoolmaster, and which represent theideas of pleasure and pain. At later stages in education the machine wouldrecognise certain other conditions as desirable owing to their having beenconstantly associated in the past with pleasure, and likewise certain others asundesirable. Certain expressions of anger on the part of the schoolmastermight, for instance, be recognised as so ominous that they could never beoverlooked, so that the schoolmaster would find that it became unnecessary to ‘apply the cane’any more.

To makefurther suggestions along these lines would perhaps be unfruitful at thisstage, as they are likely to consi.st of nothing more than an analysis ofactual methods of education applied to human children. There is, however, onefeature that I would like suggest should be incorporated in the machines, andthat is a ‘randomelement'. Each machine should be supplied with a tape bearing arandom series of figures, e.g., 0 and 1 in equal quantities, and this series offigures should be used in the choices made by the machine. This would result inthe behaviour of the machine not being by any means completely determined bythe experiences to which it was subjected, and would have some valuable uses whenone was experimenting with it. By faking the choices made one would be able tocontrol the development of the machine to some extent. One might, for instance,insist on the choice made being a particular one at, say, 10 particular places,and this would mean that about one machine in 1024 or more would develop to ashigh a degree as the one which had been faked. This cannot very well be givenan accurate statement because of the subjective nature of the idea of ‘degreeof development’ to say nothing of the fact that the machine that had been fakedmight have been also fortunate in its unfaked choices.

Let us nowassume, for the sake of argument, that these machines are a genuine possibility,and look at the consequences of constructing them. To do so would of course meetwith great opposition, unless we have advanced greatly in religious tolerationfrom thedays of Galileo. There would be great opposition from theintellectuals who were afraid of being put out of a job. It is probable thoughthat the intellectuals would be mistaken about this. There would be plenty todo in trying, say, to keep one's intelligence up to the standard set by themachines, for it seems probable that once the machine thinking method hadstarted, it would not take long to outstrip our feeble powers. There would beno question of the machines dying, and they would be able to converse with eachother to sharpen their wits. At some stage therefore we should have to expectthe machines to take control, in the way that is mentioned in Samuel Butler’s Erewhon.

ABSTRACT.In this posthumous essay, Turing contends that it may be possible to constructa machine in which there would be an element of randomness and an analogue ofthe pleasure principle of psychology, that could be taught, and that couldeventually be more intelligent than humans.