CHAPTER 17

Solvable and Unsolvable Problems (*1954*)

Alan Turing

Introduction

Jack Copeland

Unsolvable Problems

In Chapter 1 Turing proves the existence of mathematical problems that cannot be solved by the universal Turing machine. There he also advances the thesis, now called the Church–Turing thesis, that any systematic method for solving mathematical problems can be carried out by the universal Turing machine. Combining these two propositions yields the result that there are mathematical problems which cannot be solved by any systematic method—cannot, in other words, be solved by any algorithm.

Substitution Puzzles

In 'Solvable and Unsolvable Problems' Turing sets out to explain this result to a lay audience. The article first appeared in *Science News*, a popular science journal of the time. Starting from concrete examples of problems that do admit of algorithmic solution, Turing works his way towards an example of a problem that is not solvable by any systematic method. Loosely put, this is the problem of sorting puzzles into those that will 'come out' and those that will not. Turing gives an elegant argument showing that a sharpened form of this problem is not solvable by means of a systematic method (pp. 591–2).

The sharpened form of the problem involves what Turing calls 'the substitution type of puzzle'. An typical example of a substitution puzzle is this. Starting with the word BOB, is it possible to produce BOOOB by replacing selected occurrences of the pair OB by BOOB and selected occurences of the triple BOB by O? The answer is yes:

$$\text{BOB} \rightarrow \text{BBOOB} \rightarrow \text{BBOBOOB} \rightarrow \text{BOOOB}.$$

Turing suggests that any puzzle can be re-expressed as a substitution puzzle. Some row of letters can always be used to represent the 'starting position' envisaged in a particular puzzle, e.g. in the case of a chess problem, the pieces on the board and their positions. Desired outcomes, for example board positions that count as wins, can be described by further rows of letters, and the rules of the puzzle, whatever they are, are to be represented in terms of permissible substitutions of groups of letters for other groups of letters.

As Turing points out, it is not only 'toy' puzzles that can be re-expressed as substitution puzzles, but also mathematical problems, for instance the problem of finding a proof of a given mathematical theorem within an axiom system (which Turing describes as 'a very good example of a puzzle'). The axioms—which are simply strings of mathematical symbols—form the starting position. The theorem—another string of symbols—is the winning position. The rules of the puzzle are substitutions that enable strings of mathematical symbols to be transformed into other strings, much as in the case of the transition from BOB to BBOOB in the earlier example.

Turing calls the substitution formulation of any puzzle its 'normal form' and states the following *normal form principle* (p. 588):

> Given any puzzle, we can find a corresponding substitution puzzle which is equivalent to it in the sense that given a solution of the one we can easily use it to find a solution of the other.

Normal Forms and the Church–Turing Thesis

The normal form principle for puzzles closely parallels the Church–Turing thesis, which says that given any systematic method, we can find a corresponding Turing machine that is equivalent to it.

Neither the normal form principle for puzzles nor the Church–Turing thesis is susceptible to definite proof (see 'Computable Numbers: A Guide'). While few doubt that the Church–Turing thesis is in fact true, the very nature of the thesis has always been a matter for debate. Church, for example, described the thesis as a *definition*.[1] Post, on the other hand, described it as a 'working hypothesis' that is in need of 'continual verification', and he criticized Church for masking this hypothesis as a definition.[2] Turing's remarks in 'Solvable and Unsolvable Problems' about the status of the normal form principle for puzzles are of outstanding interest for the light that they may cast on his view concerning

[1] A. Church, 'An Unsolvable Problem of Elementary Number Theory', *American Journal of Mathematics*, 58 (1936), 345–63 (356).

[2] E. L. Post, 'Finite Combinatory Processes - Formulation 1', *Journal of Symbolic Logic*, 1 (1936), 103–5 (105).

the status of the Church–Turing thesis. In this connection, see also the material from Turing's draft typescript quoted in n. 9 on p. 590.

Turing says of the normal form principle (pp. 588–9):

> The statement is . . . one which one does not attempt to prove. Propaganda is more appropriate to it than proof, for its status is something between a theorem and a definition. In so far as we know *a priori* what is a puzzle and what is not, the statement is a theorem. In so far as we do not know what puzzles are, the statement is a definition which tells us something about what they are. One can of course define a puzzle by some phrase beginning, for instance, 'A set of definite rules . . . ', but this just throws us back on the definition of 'definite rules'. Equally one can reduce it to the definition of 'computable function' or 'systematic procedure'. A definition of any one of these would define all the rest.

Turing would perhaps have said much the same concerning not only the Church–Turing thesis but also the thesis introduced in Chapter 13:

> A digital computer will replace any rival design of calculating machine.

In so far as we do not know what calculating machines are, the statement is a definition which tells us something about what they are.

Proof of Unsolvability

Having introduced the normal form principle for puzzles, Turing turns to his central project of establishing that 'there cannot be any systematic procedure for determining whether a puzzle be solvable or not' (p. 590). In particular, there cannot be a systematic procedure for determining whether substitution puzzles are or are not solvable. Turing argues by *reductio ad absurdum*. He shows that the supposition that there is a systematic procedure for determining whether substitution puzzles are or are not solvable leads to an outright contradiction, and on that basis concludes that there can be no such procedure. The argument turns on the impossibility of applying a certain procedure to itself.

Any systematic procedure is in effect a puzzle, since in following the procedure one applies rules to some 'starting position' until one or another result is achieved. So if there were a systematic procedure for determining whether each puzzle is or is not solvable, then by the normal form principle, there is a substitution puzzle—call it K—that is equivalent to this procedure. When applied to any substitution puzzle, K—if it exists—must 'come out' either with the result SOLVABLE or with the result NOT SOLVABLE. Since K is applicable to any substitution puzzle, K can be applied to itself in order to determine whether it itself is or is not solvable. Turing shows (p. 592) that this supposed ability of K to pronounce on its own solvability leads to outright contradiction, and so concludes that K cannot exist.

The Meaning of 'Unsolvable'

Turing points out that the result he has established, namely that there is no systematic method for deciding whether or not substitution puzzles come out, is often expressed by saying that there is no *decision procedure* for puzzles of this type, and that the *decision problem* for this type of puzzle is *unsolvable*. He continues (p. 592): 'so one comes to speak (as in the title of this article) about "unsolvable problems" meaning in effect puzzles for which there is no decision procedure. This is the technical meaning which the words are now given by mathematical logicians.'

As Turing says, this terminology is potentially confusing. It is natural to use the words 'unsolvable problem' to mean a problem for which no solution can possibly be found. It would be a confusion to think that Turing has shown that the problem of deciding whether or not substitution puzzles come out is an unsolvable problem in this natural sense. Indeed, with sufficient time, inventiveness, and patience, mathematicians may always be able to establish whether or not any given substitution puzzle comes out. If that is so, then the problem of deciding whether or not substitution puzzles come out is solvable, in the natural sense of the word.

What Turing has shown is that there is no systematic method for deciding whether or not substitution puzzles come out, i.e. there is no general procedure, applicable by rote, that one can employ in order to decide whether or not each substitution puzzle comes out. The 'decision problem' for substitution puzzles is the problem of finding such a rote procedure (a 'decision procedure'); in showing that there is no such procedure, Turing has shown that the decision problem for substitution puzzles is unsolvable in the natural sense.

Turing therefore recommends that, in order to 'minimize confusion', one should 'always speak of "unsolvable decision problems", rather than just "unsolvable problems"' (p. 592).

Significance of Turing's Result

Turing ends the chapter with a comment on the significance of what he has shown. His result concerning the decision problem for substitution puzzles 'may be regarded as going some way towards a demonstration, within mathematics itself, of the inadequacy of "reason" unsupported by common sense'. For he has, he says, set 'certain bounds to what we can hope to achieve purely by reasoning'.

The phrase 'purely by reasoning' here presumably means 'purely by algorithmic methods'. Some mathematical problems require for their solution not only 'reason', in this sense, but also what Turing refers to in Chapter 3 as 'intuition' (see also Chapter 4). There he says (pp. 192–3):

The activity of the intuition consists in making spontaneous judgements which are not the result of conscious trains of reasoning. ... Often it is possible to find some other way of verifying the correctness of an intuitive judgement. We may, for instance, judge that all positive integers are uniquely factorizable into primes; a detailed mathematical argument leads to the same result. This argument will also involve intuitive judgements, but they will be less open to criticism than the original judgement about factorization. ... The necessity for using the intuition is ... greatly reduced by setting down formal rules for carrying out inferences which are always intuitively valid. ... In pre-Gödel times it was thought by some that it would probably be possible to carry this programme to such a point that all the intuitive judgements of mathematics could be replaced by a finite number of these rules. The necessity for intuition would then be entirely eliminated.

The argument of 'Solvable and Unsolvable Problems' illustrates why it is that the need for intuition cannot always be eliminated in favour of formal rules.

Gödel's Theorem

Turing notes that the unsolvability of the decision problem for substitution puzzles affords an elegant proof of the following rather general statement (p. 593):

> no systematic method of proving mathematical theorems is sufficiently complete to settle every mathematical question, yes or no.

The proof Turing gives is as follows. Each statement of the form 'such-and-such substitution puzzle comes out' can be expressed in the form of a mathematical statement. So if there were a systematic method of settling every question that can be posed in mathematical form, this method would serve as a decision procedure for substitution puzzles. Given that there is no such decision procedure, it follows that no systematic method is able to settle every mathematical question.

Turing remarks that the above statement follows 'by a famous theorem of Gödel' and describes himself as providing 'an independent proof' of the statement (p. 593). Turing might also have pointed out that his own 'On Computable Numbers' yields a proof of this statement.

Gödel's famous incompleteness theorem of 1931 is, however, importantly less general than the above statement, since it concerns only one particular systematic method of proving mathematical theorems, the system set out by Whitehead and Russell in *Principia Mathematica*[3] (as explained in 'Computable Numbers: A Guide'). Gödel did later generalize his result of 1931 to all formal systems (containing a certain amount of arithmetic), but emphasized the importance that Turing's work played in this generalization. Gödel said in 1964:

[3] A. N. Whitehead and B. Russell, *Principia Mathematica*, vols. i–iii (Cambridge: Cambridge University Press, 1910–13).

[D]ue to A. M. Turing's work, a precise and unquestionably adequate definition of the general concept of formal system can now be given... Turing's work gives an analysis of the concept of 'mechanical procedure' (alias 'algorithm' or 'computation procedure' or 'finite combinatorial procedure'). ... A formal system can simply be defined to be any mechanical procedure for producing formulas, called provable formulas.[4]

In his references to Gödel's work, Turing hides his own light under a bushel.

Further reading

Boone, W. W., review of Turing's 'The Word Problem in Semi-Groups with Cancellation', *Journal of Symbolic Logic*, 17 (1952), 74–6.

Turing, A. M., 'The Word Problem in Semi-Groups with Cancellation', *Annals of Mathematics*, 52 (1950), 491–505. Reprinted in *Pure Mathematics: Collected Works of A. M. Turing*, ed. J. L. Britton (Amsterdam: North-Holland, 1992).

Provenance

What follows is the text of the original printing of 'Solvable and Unsolvable Problems' in *Science News*.[5] Unfortunately Turing's own typescript appears to have been lost. However, a sizeable fragment of a draft typescript, with additions in Turing's handwriting, has been preserved.[6] (Turing recycled the draft pages, covering the reverse sides with handwritten notes concerning morphogenesis.) The fragment corresponds to pp. 584–9. For the most part the published version follows the draft pages closely (except for punctuation and occasional changes of word and word-order). Significant differences between the draft and the published version are mentioned in footnotes.

[4] K. Gödel, 'Postscriptum', in M. Davis (ed.), *The Undecidable* (New York: Raven, 1965), 71–3 (71–2); the Postscriptum, dated 1964, is to Gödel's 1934 paper 'On Undecidable Propositions of Formal Mathematical Systems' (ibid. 41–71).

[5] Footnotes have been renumbered consecutively. Footnotes not marked 'Editor's note' appeared in *Science News*. A page reference to *Science News* been replaced by the number (in square brackets) of the corresponding page of this volume.

[6] The fragment is among the Turing Papers in the Modern Archive Centre, King's College, Cambridge; at the time of writing it is uncatalogued.

Solvable and Unsolvable Problems

If one is given a puzzle to solve one will usually, if it proves to be difficult, ask the owner whether it can be done. Such a question should have a quite definite answer, yes or no, at any rate provided the rules describing what you are allowed to do are perfectly clear. Of course the owner of the puzzle may not know the answer. One might equally ask, ‘How can one tell whether a puzzle is solvable?’, but this cannot be answered so straightforwardly. The fact of the matter is that there is *no* systematic method of testing puzzles to see whether they are solvable or not. If by this one meant merely that nobody had ever yet found a test which could be applied to any puzzle, there would be nothing at all remarkable in the statement. It would have been a great achievement to have invented such a test, so we can hardly be surprised that it has never been done. But it is not merely that the test has never been found. It has been proved that no such test ever can be found.

Let us get away from generalities a little and consider a particular puzzle. One which has been on sale during the last few years and has probably been seen by most of the readers of this article illustrates a number of the points involved quite well. The puzzle consists of a large square within which are some smaller movable squares numbered 1 to 15, and one empty space, into which any of the neighbouring squares can be slid leaving a new empty space behind it. One may be asked to transform a given arrangement of the squares into another by a succession of such movements of a square into an empty space. For this puzzle there is a fairly simple and quite practicable rule by which one can tell whether the transformation required is possible or not. One first imagines the transformation carried out according to a different set of rules. As well as sliding the squares into the empty space one is allowed to make moves each consisting of two interchanges, each of one pair of squares. One would, for instance, be allowed as one move to interchange the squares numbered 4 and 7, and also the squares numbered 3 and 5. One is permitted to use the same number in both pairs. Thus one may replace 1 by 2, 2 by 3, and 3 by 1 as a move because this is the same as interchanging first (1, 2) and then (1, 3). The original puzzle is solvable by sliding if it is solvable according to the new rules. It is not solvable by sliding if the required position can be reached by the new rules, together with a ‘cheat’ consisting of *one single* interchange of a pair of squares.[1] Suppose, for instance, that one is asked to get back to the standard position—

This article first appeared in *Science News*, 31 (1954), 7–23, published by the Penguin Press, and is printed by permission of the Estate of Alan Turing.

[1] It would take us too far from our main purpose to give the proof of this rule: the reader should have little difficulty in proving it by making use of the fact that an odd number of interchanges can never bring a set of objects back to the position it started from.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

from the position

10	1	4	5
9	2	6	8
11	3		15
13	14	7	12

One may, according to the modified rules, first get the empty square into the correct position by moving the squares 15 and 12, and then get the squares 1, 2, 3, ... successively into their correct positions by the interchanges (1, 10), (2, 10), (3, 4), (4, 5), (5, 9), (6, 10), (7, 10), (9, 11), (10, 11), (11, 15). The squares 8, 12, 13, 14, 15 are found to be already in their correct positions when their turns are reached. Since the number of interchanges required is even, this transformation is possible by sliding.[2] If one were required after this to interchange say square 14 and 15 it could not be done.

This explanation of the theory of the puzzle can be regarded as entirely satisfactory. It gives one a simple rule for determining for any two positions whether one can get from one to the other or not. That the rule is so satisfactory depends very largely on the fact that it does not take very long to apply. No mathematical method can be useful for any problem if it involves much calculation. It is nevertheless sometimes interesting to consider whether something is possible at all or not, without worrying whether, in case it *is* possible, the amount of labour or calculation is economically prohibitive. These investigations that are not concerned with the amount of work involved are in some ways easier to carry out, and they certainly have a greater aesthetic appeal. The results are not altogether without value, for if one has proved that there is no method of doing something it follows *a fortiori* that there is no practicable method. On the other hand, if one method has been proved to exist by which the decision can be made, it gives some encouragement to anyone who wishes to find a workable method.

From this point of view, in which one is only interested in the question, 'Is there a systematic way of deciding whether puzzles of this kind are solvable?', the rules which have been described for the sliding-squares puzzle are much more special and detailed than is really necessary. It would be quite enough to say: 'Certainly one can find out whether one position can be reached from another by

[2] It can in fact be done by sliding successively the squares numbered 7, 14, 13, 11, 9, 10, 1, 2, 3, 7, 15, 8, 5, 4, 6, 3, 10, 1, 2, 6, 3, 10, 6, 2, 1, 6, 7, 15, 8, 5, 10, 8, 5, 10, 8, 7, 6, 9, 15, 5, 10, 8, 7, 6, 5, 15, 9, 5, 6, 7, 8, 12, 14, 13, 15, 10, 13, 15, 11, 9, 10, 11, 15, 13, 12, 14, 13, 15, 9, 10, 11, 12, 14, 13, 15, 14, 13, 15, 14, 13, 12, 11, 10, 9, 13, 14, 15, 12, 11, 10, 9, 13, 14, 15.

a systematic procedure. There are only a finite number of positions in which the numbered squares can be arranged (viz. 20922789888000) and only a finite number (2, 3, or 4) of moves in each position. By making a list of all the positions and working through all the moves, one can divide the positions into classes, such that sliding the squares allows one to get to any position which is in the same class as the one started from. By looking up which classes the two positions belong to one can tell whether one can get from one to the other or not.' This is all, of course, perfectly true, but one would hardly find such remarks helpful if they were made in reply to a request for an explanation of how the puzzle should be done. In fact they are so obvious that under such circumstances one might find them somehow rather insulting. But the fact of the matter is, that if one is interested in the question as put, 'Can one tell by a systematic method in which cases the puzzle is solvable?', this answer is entirely appropriate, because one wants to know if there is a systematic method, rather than to know of a good one.

The same kind of argument will apply for any puzzle where one is allowed to move certain 'pieces' around in a specified manner, provided that the total number of essentially different positions which the pieces can take up is finite. A slight variation on the argument is necessary in general to allow for the fact that in many puzzles some moves are allowed which one is not permitted to reverse. But one can still make a list of the positions, and list against these first the positions which can be reached from them in one move. One then adds the positions which are reached by two moves and so on until an increase in the number of moves does not give rise to any further entries. For instance, we can say at once that there is a method of deciding whether a patience can be got out with a given order of the cards in the pack: it is to be understood that there is only a finite number of places in which a card is ever to be placed on the table. It may be argued that one is permitted to put the cards down in a manner which is not perfectly regular, but one can still say that there is only a finite number of 'essentially different' positions. A more interesting example is provided by those puzzles made (apparently at least) of two or more pieces of very thick twisted wire which one is required to separate. It is understood that one is not allowed to bend the wires at all, and when one makes the right movement there is always plenty of room to get the pieces apart without them ever touching, if one wishes to do so. One may describe the positions of the pieces by saying where some three definite points of each piece are. Because of the spare space it is not necessary to give these positions quite exactly. It would be enough to give them to, say, a tenth of a millimetre. One does not need to take any notice of movements of the puzzle as a whole: in fact one could suppose one of the pieces quite fixed. The second piece can be supposed to be not very far away, for, if it is, the puzzle is already solved. These considerations enable us to reduce the number of 'essentially different' positions to a finite number, probably a few hundred millions, and the usual argument will then apply. There are some further complications, which we will not consider in detail, if we do not know how much clearance to allow for. It is necessary to repeat the process again and again allowing successively smaller and smaller clearances. Eventually one will find that either it can be solved, allowing a small clearance margin, or else it cannot be solved even allowing a small margin of 'cheating' (i.e. of 'forcing', or having the pieces slightly overlapping in space). It will, of course, be understood that this process of trying out the possible positions is not to be done with the physical puzzle itself, but on paper, with mathematical descriptions of the positions, and mathematical criteria for deciding whether in a given position the pieces overlap, etc.

These puzzles where one is asked to separate rigid bodies are in a way like the 'puzzle' of trying to undo a tangle, or more generally of trying to turn one knot into another without cutting the string. The difference is that one is allowed to bend the string, but not the wire forming the rigid bodies. In either case, if one wants to treat the problem seriously and systematically one has to replace the physical puzzle by a mathematical equivalent. The knot puzzle lends itself quite conveniently to this. A knot is just a closed curve in three dimensions nowhere crossing itself; but, for the purpose we are interested in, any knot can be given accurately enough as a series of segments in the directions of the three coordinate axes. Thus, for instance, the trefoil knot (Figure 1*a*) may be regarded as consisting of a number of segments joining the points given, in the usual (x, y, z) system of coordinates, as (1, 1, 1), (4, 1, 1,), (4, 2, 1), (4, 2, −1), (2, 2, −1), (2, 2, 2), (2, 0, 2), (3, 0, 2), (3, 0, 0), (3, 3, 0), (1, 3, 0), (1, 3, 1), and returning again with a twelfth segment to the starting point (1, 1, 1).[3] This representation of the knot is shown in perspective in Figure 1*b*. There is no special virtue in the representation which has been chosen. If it is desired to follow the original curve more closely a greater number of segments must be used. Now let a and d represent unit steps in the positive and negative X-directions respectively, b and e in the Y-directions, and c and f in the Z-directions: then this knot may be described as *aaabffddccceeaffbbbddcee*.[4] One can then, if one wishes, deal entirely with such sequences of letters. In order that such a sequence should represent a knot it is necessary and sufficient that the numbers of a's and d's should be equal, and likewise the number of b's equal to the number of e's and the number of c's equal to the number of f's, and it must not be possible to obtain another sequence of letters with these properties by omitting a number of consecutive letters at the beginning

[3] Editor's note. In place of this sentence Turing's draft has: 'Thus for instance the trefoil knot may be regarded as consisting of a number of segments joining the points (0, 0, 0), (0, 2, 0), (1, 2, 0), (1, 2, 2), (1, −1, 2), (1, −1, 1), (−1, −1, 1), (−1, 1, 1), (2, 1, 1), (2, 0, 1), (2, 0, 3), (0, 0, 3), (0, 0, 0).'

[4] Editor's note. Turing's draft has 'bbacceeefddbbaaaeccddfff'.

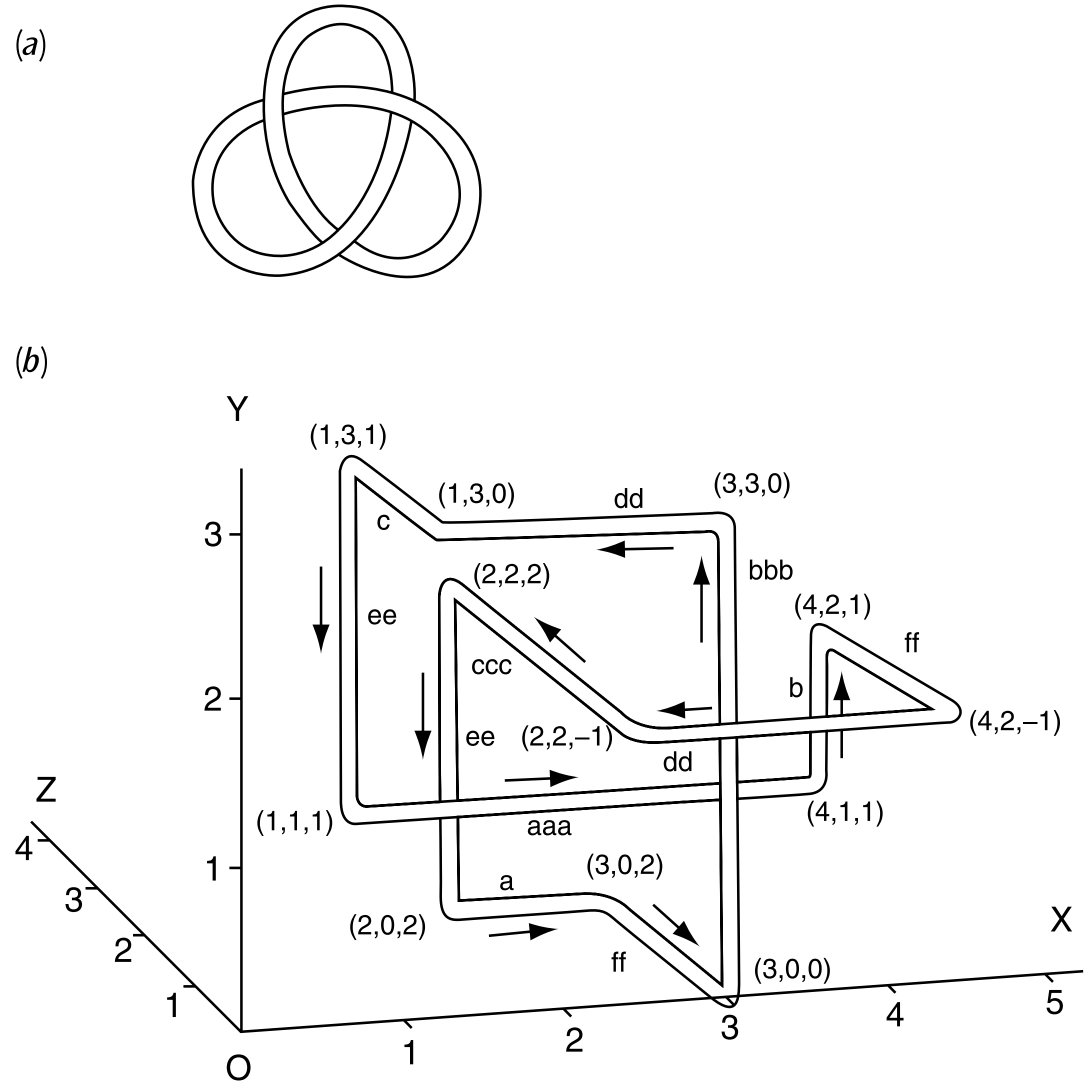

Figure 1. (*a*) The trefoil knot (*b*) a possible representation of this knot as a number of segments joining points.

or the end or both. One can turn a knot into an equivalent one by operations of the following kinds—

(i) One may move a letter from one end of the row to the other.
(ii) One may interchange two consecutive letters provided this still gives a knot.
(iii) One may introduce a letter *a* in one place in the row, and *d* somewhere else, or *b* and *e*, or *c* and *f*, or take such pairs out, provided it still gives a knot.
(iv) One may replace *a* everywhere by *aa* and *d* by *dd* or replace each *b* and *e* by *bb* and *ee* or each *c* and *f* by *cc* and *ff*. One may also reverse any such operation.

—and these are all the moves that are necessary.

It is also possible to give a similar symbolic equivalent for the problem of separating rigid bodies, but it is less straightforward than in the case of knots.

These knots provide an example of a puzzle where one cannot tell in advance how many arrangements of pieces may be involved (in this case the pieces are the letters *a, b, c, d, e, f*), so that the usual method of determining whether the puzzle is solvable cannot be applied. Because of rules (iii) and (iv) the lengths of the sequences describing the knots may become indefinitely great. No systematic method is yet known by which one can tell whether two knots are the same.

Another type of puzzle which we shall find very important is the 'substitution puzzle'. In such a puzzle one is supposed to be supplied with a finite number of different kinds of counters, perhaps just black (*B*) and white (*W*). Each kind is in unlimited supply. Initially a number of counters are arranged in a row and one is asked to transform it into another pattern by substitutions. A finite list of the substitutions allowed is given. Thus, for instance, one might be allowed the substitutions[5]

$$\text{(i)}\ WBW \rightarrow B$$
$$\text{(ii)}\ BW \rightarrow WBBW$$

and be asked to transform *WBW* into *WBBBW*, which could be done as follows

$$W\underline{BW} \xrightarrow[\text{(ii)}]{} W\underline{WB\underline{\underline{BW}}} \xrightarrow[\text{(ii)}]{} W\underline{WB}\underline{WBBW} \xrightarrow[\text{(i)}]{} W\underline{B}BBW$$

Here the substitutions used are indicated by the numbers below the arrows, and their effects by underlinings. On the other hand if one were asked to transform *WBB* into *BW* it could not be done, for there are no admissible steps which reduce the number of *B*'s.

It will be seen that with this puzzle, and with the majority of substitution puzzles, one cannot set any bound to the number of positions that the original position might give rise to.

It will have been realized by now that a puzzle can be something rather more important than just a toy. For instance the task of proving a given mathematical theorem within an axiomatic system is a very good example of a puzzle.

It would be helpful if one had some kind of 'normal form' or 'standard form' for describing puzzles. There is, in fact, quite a reasonably simple one which I

[5] Editor's note. Turing's draft has: 'Thus for instance one might be allowed the substitutions

$$WBW \rightarrow B$$
$$BWWW \rightarrow WB$$
$$BWB \rightarrow WWWB$$
$$WWB \rightarrow W$$

and be asked to transform WBWWBWBBB into WBB, and this one could do first by substituting W for WWB and getting WBWBBBB and then successively WWWWBBBB, WWBBB, WBB.'

shall attempt to describe. It will be necessary for reasons of space to take a good deal for granted, but this need not obscure the main ideas. First of all we may suppose that the puzzle is somehow reduced to a mathematical form in the sort of way that was used in the case of the knots. The position[6] of the puzzle may be described, as was done in that case, by sequences of symbols in a row. There is usually very little difficulty in reducing other arrangements of symbols (e.g. the squares in the sliding squares puzzle) to this form. The question which remains to be answered is, 'What sort of rules should one be allowed to have for rearranging the symbols or counters?' In order to answer this one needs to think about what kinds of processes ever do occur in such rules, and, in order to reduce their number, to break them up into simpler processes. Typical of such processes are counting, copying, comparing, substituting. When one is doing such processes, it is necessary, especially if there are many symbols involved, and if one wishes to avoid carrying too much information in one's head, either to make a number of jottings elsewhere or to use a number of marker objects as well as the pieces of the puzzle itself. For instance, if one were making a copy of a row of counters concerned in the puzzle it would be as well to have a marker which divided the pieces which have been copied from those which have not and another showing the end of the portion to be copied. Now there is no reason why the rules of the puzzle itself should not be expressed in such a way as to take account of these markers. If one does express the rules in this way they can be made to be just substitutions. This means to say that the *normal form for puzzles is the substitution type of puzzle*. More definitely we can say:

Given any puzzle we can find a corresponding substitution puzzle *which is equivalent to it in the sense that given a solution of the one we can easily use it to find a solution of the other. If the original puzzle is concerned with rows of pieces of a finite number of different kinds, then the substitutions may be applied as an alternative set of rules to the pieces of the original puzzle. A transformation can be carried out by the rules of the original puzzle if and only if it can be carried out by the substitutions and leads to a final position from which all marker symbols have disappeared.*

This statement is still somewhat lacking in definiteness, and will remain so. I do not propose, for instance, to enter here into the question as to what I mean by the word 'easily'. The statement is moreover one which one does not attempt to prove. Propaganda is more appropriate to it than proof, for its status is something between a theorem and a definition. In so far as we know *a priori* what is a puzzle and what is not, the statement is a theorem. In so far as we do not know what puzzles are, the statement is a definition which tells us something about what they are. One can of course define a puzzle by some phrase beginning, for instance, 'A set of definite rules . . .', but this just throws us back on the definition

[6] Editor's note. Turing's draft has 'positions'.

of 'definite rules'. Equally one can reduce it to the definition of 'computable function' or 'systematic procedure'. A definition of any one of these would define all the rest. Since 1935 a number of definitions have been given, explaining in detail the meaning of one or other of these terms, and these have all been proved equivalent to one another and also equivalent to the above statement. In effect there is no opposition to the view that every puzzle is equivalent to a substitution puzzle.[7]

After these preliminaries let us think again about puzzles as a whole. First let us recapitulate. There are a number of questions to which a puzzle may give rise. When given a particular task one may ask quite simply

(a) *Can this be done?*

Such a straightforward question admits only the straightforward answers, 'Yes' or 'No', or perhaps 'I don't know'. In the case that the answer is 'Yes' the answerer need only have done the puzzle himself beforehand to be sure. If the answer is to be 'No', some rather more subtle kind of argument, more or less mathematical, is necessary. For instance, in the case of the sliding squares one can state that the impossible cases *are* impossible because of the mathematical fact that an odd number of simple interchanges of a number of objects can never bring one back to where one started. One may also be asked

(b) *What is the best way of doing this?*

Such a question does not admit of a straightforward answer. It depends partly on individual differences in people's ideas as to what they find easy. If it is put in the form, 'What is the solution which involves the smallest number of steps?', we again have a straightforward question, but now it is one which is somehow of remarkably little interest. In any particular case where the answer to (a) is 'Yes' one can find the smallest possible number of steps by a tedious and usually impracticable process of enumeration, but the result hardly justifies the labour.

When one has been asked a number of times whether a number of different puzzles of similar nature can be solved one is naturally led to ask oneself

(c) *Is there a systematic procedure*[8] *by which I can answer these questions, for puzzles of this type?*

If one were feeling rather more ambitious one might even ask

(d) *Is there a systematic procedure*[8] *by which one can tell whether a puzzle is solvable?*

I hope to show that the answer to this last question is 'No'.

There are in fact certain types of puzzle for which the answer to (c) is 'No'.

[7] Editor's note. At this point Turing's draft contains the following: 'Some of these other definitions will be found in Refs (1), (3), (11), (13), and (16) vol II. Some equivalence theorems are proved in (4) and (14), and some propaganda on the matter will be found in (13). A very satisfactory account of all these problems will be found in (5).' Tantalizingly, the list of references is omitted.

[8] Editor's note. Turing's draft has 'systematic method'.

Before we can consider this question properly we shall need to be quite clear what we mean by a 'systematic procedure'[8] for deciding a question.[9] But this need not now give us any particular difficulty. A 'systematic procedure' was one of the phrases which we mentioned as being equivalent to the idea of a puzzle, because either could be reduced to the other. If we are now clear as to what a puzzle is, then we should be equally clear about 'systematic procedures'. In fact a systematic procedure is just a puzzle *in which there is never more than one possible move in any of the positions which arise and in which some significance is attached to the final result.*

Now that we have explained the meaning both of the term 'puzzle' and of 'systematic procedure', we are in a position to prove the assertion made in the first paragraph of this article, that there cannot be any systematic procedure for determining whether a puzzle be solvable or not. The proof does not really require the detailed definition of either of the terms, but only the relation between them which we have just explained. Any systematic procedure for deciding whether a puzzle were solvable could certainly be put in the form of a puzzle, with unambiguous moves (i.e. only one move from any one position), and having for its starting position a combination of the rules, the starting position and the final position of the puzzle under investigation.

The puzzle under investigation is also to be described by its rules and starting position. Each of these is to be just a row of symbols. As we are only considering substitution puzzles, the rules need only be a list of all the substitution pairs appropriately punctuated. One possible form of punctuation would be to separate the first member of a pair from the second by an arrow, and to separate the different substitution pairs with colons. In this case the rules

B may be replaced by *BC*
WBW may be deleted

would be represented by ': $B \rightarrow BC$: $WBW \rightarrow$:'. For the purposes of the argument which follows, however, these arrows and colons are an embarrassment. We

[9] Editor's note. At this point Turing's draft contains the following material, which is crossed out. 'It is a phrase which, like many others e.g. "vegetable" one understands well enough in the ordinary way. But one can have difficulties when speaking to greengrocers or microbiologists or when playing "twenty questions". Are rhubarb and tomatoes vegetables or fruits? Is coal vegetable or mineral? What about coal gas, marrow, fossilised trees, streptococci, viruses? Has the lettuce I ate at lunch yet become animal? The fact of the matter is that when one is applying a word, say an adjective, to something definite, one chooses the word itself so that it describes what one wants to describe fairly and squarely. If it doesn't one had better look for another word. But if one is playing twenty questions this just can't be done. The questions are about "the object", and one doesn't know what it is. The same sort of difficulty arises about question c) above. An ordinary sort of acquaintance with the meaning of the phrase "systematic method" won't do, because one has got to be able to say quite clearly about any kind of method that might be proposed whether it is allowable or not. Fortunately a number of satisfactory definitions were found in the late thirties, and they have...' [the fragment ends at this point].

shall need the rules to be expressed without the use of any symbols which are barred from appearing in the starting positions. This can be achieved by the following simple, though slightly artificial trick. We first double all the symbols other than the punctuation symbols, thus ': $BB \to BBCC$: $WWBBWW \to$:'. We then replace each arrow by a single symbol, which must be different from those on either side of it, and each colon by three similar symbols, also chosen to avoid clashes. This can always be done if we have at least three symbols available, and the rules above could then be represented as, for instance, '*CCCBBWBBCC BBBWWBBWWBWWW*'. Of course according to these conventions a great variety of different rows of symbols will describe essentially the same puzzle. Quite apart from the arbitrary choice of the punctuating symbols the substitution pairs can be given in any order, and the same pair can be repeated again and again.

Now let $P(R,S)$ stand for 'the puzzle whose rules are described by the row of symbols R and whose starting position is described by S'. Owing to the special form in which we have chosen to describe the rules of puzzles, there is no reason why we should not consider $P(R,R)$ for which the 'rules' also serve as starting position: in fact the success of the argument which follows depends on our doing so. The argument will also be mainly concerned with puzzles in which there is at most one possible move in any position; these may be called 'puzzles with unambiguous moves'. Such a puzzle may be said to have 'come out' if one reaches either the position B or the position W, and the rules do not permit any further moves. Clearly if a puzzle has unambiguous moves it cannot both come out with the end result B and with the end result W.

We now consider the problem of classifying rules R of puzzles into two classes, I and II, as follows:

Class I is to consist of sets R of rules, which represent puzzles with unambiguous moves, and such that $P(R,R)$ comes out with the end result W.

Class II is to include all other cases, i.e. either $P(R,R)$ does not come out, or comes out with the end result B, or else R does not represent a puzzle with unambiguous moves. We may also, if we wish, include in this class sequences of symbols such as *BBBBB* which do not represent a set of rules at all.

Now suppose that, contrary to the theorem that we wish to prove, we have a systematic procedure for deciding whether puzzles come out or not. Then with the aid of this procedure we shall be able to distinguish rules of class I from those of class II. There is no difficulty in deciding whether R really represents a set of rules, and whether they are unambiguous. If there is any difficulty it lies in finding the end result in the cases where the puzzle is known to come out: but this can be decided by actually working the puzzle through. By a principle which has already been explained, this systematic procedure for distinguishing the two classes can itself be put into the form of a substitution puzzle (with rules K, say). When applying these rules K, the rules R of the puzzle under investigation form the starting position, and the end result of the puzzle gives the result of the test.

Since the procedure always gives an answer, the puzzle $P(K,R)$ always comes out. The puzzle K might be made to announce its results in a variety of ways, and we may be permitted to suppose that the end result is B for rules R of class I, and W for rules of class II. The opposite choice would be equally possible, and would hold for a slightly different set of rules K', which however we do not choose to favour with our attention. The puzzle with rules K may without difficulty be made to have unambiguous moves. Its essential properties are therefore:

K has unambiguous moves.
$P(K,R)$ always comes out whatever R.
If R is in class I, then $P(K,R)$ has end result B.
If R is in class II, then $P(K,R)$ has end result W.

These properties are however inconsistent with the definitions of the two classes. If we ask ourselves which class K belongs to, we find that neither will do. The puzzle $P(K,K)$ is bound to come out, but the properties of K tell us that we must get end result B if K is in class I and W if it is in class II, whereas the definitions of the classes tell us that the end results must be the other way round. The assumption that there was a systematic procedure for telling whether puzzles come out has thus been reduced to an absurdity.

Thus in connexion with question (c) above we can say that there are some types of puzzle for which no systematic method of deciding the question exists. This is often expressed in the form, 'There is no *decision procedure* for this type of puzzle', or again, 'The decision problem for this type of puzzle is unsolvable', and so one comes to speak (as in the title of this article) about 'unsolvable problems' meaning in effect puzzles for which there is no decision procedure. This is the technical meaning which the words are now given by mathematical logicians. It would seem more natural to use the phrase 'unsolvable problem' to mean just an unsolvable puzzle, as for example 'to transform 1, 2, 3 into 2, 1, 3 by cyclic permutation of the symbols', but this is not the meaning it now has. However, to minimize confusion I shall here always speak of 'unsolvable decision problems', rather than just 'unsolvable problems', and also speak of puzzles rather than problems where it is puzzles and not decision problems that are concerned.

It should be noticed that a decision problem only arises when one has an infinity of questions to ask. If you ask, 'Is this apple good to eat?', or 'Is this number prime?', or 'Is this puzzle solvable?' the question can be settled with a single 'Yes' or 'No'. A finite number of answers will deal with a question about a finite number of objects, such as the apples in a basket. When the number is infinite, or in some way not yet completed concerning say all the apples one may ever be offered, or all whole numbers or puzzles, a list of answers will not suffice. Some kind of rule or systematic procedure must be given. Even if the number concerned is finite one may still prefer to have a rule rather than a list: it may be easier to remember. But there certainly cannot be an unsolvable decision problem in such cases, because of the possibility of using finite list.

Regarding decision problems as being concerned with classes of puzzles, we see that if we have a decision method for one class it will apply also for any subclass. Likewise, if we have proved that there is no decision procedure for the subclass, it follows that there is none for the whole class. The most interesting and valuable results about unsolvable decision problems concern the smaller classes of puzzle.

Another point which is worth noticing is quite well illustrated by the puzzle which we considered first of all in which the pieces were sliding squares. If one wants to know whether the puzzle is solvable with a given starting position, one can try moving the pieces about in the hope of reaching the required end-position. If one succeeds, then one will have solved the puzzle and consequently will be able to answer the question, 'Is it solvable?' In the case that the puzzle is solvable one will eventually come on the right set of moves. If one has also a procedure by which, if the puzzle is unsolvable, one would eventually establish the fact that it was so, then one would have a solution of the decision problem for the puzzle. For it is only necessary to apply both processes, a bit of one alternating with a bit of the other, in order eventually to reach a conclusion by one or the other. Actually, in the case of the sliding squares problem, we have got such a procedure, for we know that if, by sliding, one ever reaches the required final position, with squares 14 and 15 interchanged, then the puzzle is impossible.

It is clear then that the difficulty in finding decision procedures for types of puzzle lies in establishing that the puzzle is unsolvable in those cases where it *is* unsolvable. This, as was mentioned on page [589], requires some sort of mathematical argument. This suggests that we might try expressing the statement that the puzzle comes out in a mathematical form and then try and prove it by some systematic process. There is no particular difficulty in the first part of this project, the mathematical expression of the statement about the puzzle. But the second half of the project is bound to fail, because by a famous theorem of Gödel no systematic method of proving mathematical theorems is sufficiently complete to settle every mathematical question, yes or no. In any case we are now in a position to give an independent proof of this. If there were such a systematic method of proving mathematical theorems we could apply it to our puzzles and for each one eventually either prove that it was solvable or unsolvable; this would provide a systematic method of determining whether the puzzle was solvable or not, contrary to what we have already proved.

This result about the decision problem for puzzles, or, more accurately speaking, a number of others very similar to it, was proved in 1936–7. Since then a considerable number of further decision problems have been shown to be unsolvable. They are all proved to be unsolvable by showing that if they were solvable one could use the solution to provide a solution of the original one. They could all without difficulty be reduced to the same unsolvable problem. A number of these results are mentioned very shortly below. No attempt is made to explain the technical terms used, as most readers will be familiar with some of them, and the space required for the explanation would be quite out of proportion to its usefulness in this context.

(1) It is not possible to solve the decision problem even for substitution processes applied to rows of black and white counters only.
(2) There are certain particular puzzles for which there is no decision procedure, the rules being fixed and the only variable element being the starting position.
(3) There is no procedure for deciding whether a given set of axioms leads to a contradiction or not.
(4) The 'word problem in semi-groups with cancellation' is not solvable.
(5) It has recently been announced from Russia that the 'word problem in groups' is not solvable. This is a decision problem not unlike the 'word problem in semi-groups', but very much more important, having applications in topology: attempts were being made to solve this decision problem before any such problems had been proved unsolvable. No adequately complete proof is yet available, but if it is correct this is a considerable step forward.
(6) There is a set of 102 matrices of order 4, with integral coefficients such that there is no decision method for determining whether another given matrix is or is not expressible as a product of matrices from the given set.

These are, of course, only a selection from the results. Although quite a number of decision problems are now known to be unsolvable, we are still very far from being in a position to say of a given decision problem, whether it is solvable or not. Indeed, we shall never be quite in that position, for the question whether a given decision problem is solvable is itself one of the undecidable decision problems. The results which have been found are on the whole ones which have fallen into our laps rather than ones which have positively been searched for. Considerable efforts have however been made over the word problem in groups (see (5) above). Another problem which mathematicians are very anxious to settle is known as 'the decision problem of the equivalence of manifolds'. This is something like one of the problems we have already mentioned, that concerning the twisted wire puzzles. But whereas with the twisted wire puzzles the pieces are quite rigid, the 'equivalence of manifolds' problem concerns pieces which one is allowed to bend, stretch, twist, or compress as much as one likes, without ever actually breaking them or making new junctions or filling in holes. Given a number of interlacing pieces of plasticine one may be asked to transform them in this way into another given form. The decision problem for this class of problem is the 'decision problem for the equivalence of manifolds'. It is probably unsolvable, but has never been proved to be so. A similar decision problem which might well be unsolvable is the one concerning knots which has already been mentioned.

The results which have been described in this article are mainly of a negative character, setting certain bounds to what we can hope to achieve purely by reasoning. These, and some other results of mathematical logic may be regarded as going some way towards a demonstration, within mathematics itself, of the inadequacy of 'reason' unsupported by common sense.

Further reading

Kleene, S. C. *Introduction to Metamathematics*, Amsterdam, 1952.