FURTHER READING

Hammond, K., *Case-based Planning: Viewing Planning as a Memory Task.* San Diego: Academic Press 1989.
Lakatos, I., *Proofs and Refutations.* Cambridge: Cambridge University Press 1972.
Riesbeck, C. and Schank, R., *Inside Case-based Reasoning.* Hillsdale, NJ: Lawrence Erlbaum 1989.
Schank, R., *Dynamic Memory: A Theory of Reminding and Learning in Computers and People.* Cambridge: Cambridge University Press 1982.
Schanck, R., *Tell Me a Story: A New Look at Real and Artificial Memory.* New York: Scribner 1990.
Schank, R. and Abelson, R., *Scripts, Plans, Goals, and Understanding.* Hillsdale, NJ: Lawrence Erlbaum 1977.
Sussman, G. *A Computer Model of Skill Acquisition.* New York: American Elsevier 1975.
Waterman, D. and Hayes-Roth, F. (eds.), *Pattern-directed Inference Systems.* New York: Academic Press 1978.

5

Mathematical intelligence

ROGER PENROSE

What is mathematical intelligence? Is there anything that is essentially different about the way that we reason mathematically, from the way in which we think generally? Is mathematical intelligence different from any other kind of intelligence?

I feel certain that there is no fundamental difference between mathematical and other kinds of thinking. It is true that many people find it difficult to cope with the abstract type of thinking that is needed for mathematics, whilst finding comparatively little difficulty with the equally convoluted judgements that are involved in day-to-day relationships with other human beings. Some kinds of thinking come easily to certain people, whereas other kinds come more easily to others. But I do not think that there is any essential difference – or that there is more difference between mathematical thinking and, say, planning a holiday, than there is between the latter activity and understanding a music-hall joke. Human mathematical intelligence is just one particular form of human intelligence and understanding. It is more extreme than most of these other forms in the abstract, impersonal, and universal nature of the concepts that are involved, and in the rigour of its criteria for establishing truth. But mathematical thinking is in no way removed from other qualities that are important ingredients in our general ability for intelligent comprehension, such as intuition, common-sense judgement, and the appreciation of beauty.

What, after all, is intelligence? What is thinking? There is a prevalent viewpoint in current philosophising that holds that whatever it is that in detail constitutes the physical activity underlying our thought processes, it cannot, in effect, be other than the carrying out of some vastly complicated calculation. The relevant actions of our brains, so it is argued, are simply to provide our bodies with a very effective control system – a control system that could in principle be effected by a computer, if only one knew enough of the details of those computational procedures that the brain actually carries out. One might well imagine that, in accordance with this view, such an underlying computational basis to our thinking ought to be most manifest with mathematical thinking. For is not mathematics a computational activity *par excellence*? Indeed, it is not! It is one of my purposes here to emphasise that there is a great deal of what is essential in mathematical thinking that is not of a computational character. Indeed, it turns out that it is possible actually to demonstrate that there is something in our mathematical understanding – in our insights as to mathematical truth – that eludes any computational description whatever. It is the very precision and the universal character of mathematical argument that allows such a demonstration to be possible. But the conclusion is in no way restricted to an intelligence that relates merely to mathematical thinking. As I have argued, there is nothing essential that separates mathematical from other types of thinking, so our demonstration that mathematical understanding is something that cannot be simulated in a computational way can be thought of, also, as a demonstration that understanding itself – one of the most essential ingredients of genuine intelligence – is something that lies beyond any kind of purely computational activity.

MATHEMATICAL VISUALISATION

What are we doing when we conjure up in our minds the image of some mathematical structure? Are we performing some internal calculation, like those that lead to the impressive computer graphic displays with which we are now so familiar? Perhaps our brains are

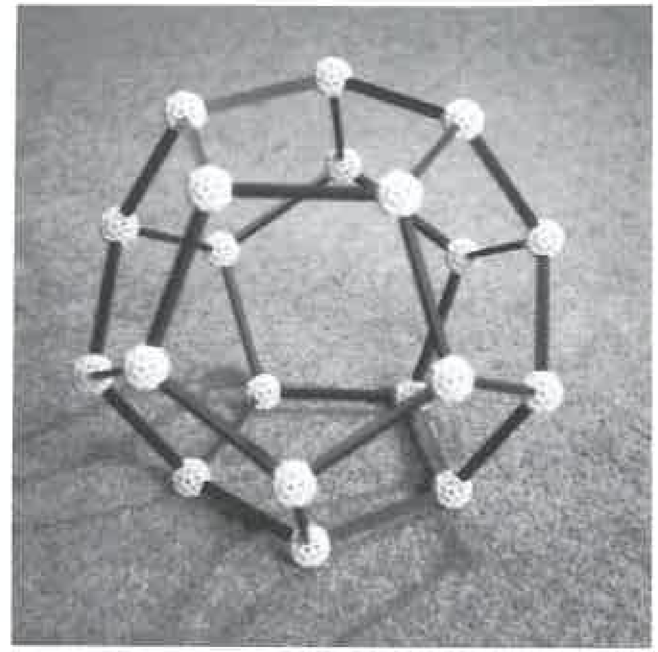


Figure 1 A regular dodecahedron

acting out something like the computational procedures that give rise to what is called 'virtual reality', whereby an entire three-dimensional seemingly consistent structure, such as a non-existent building, can be visually presented to a human subject, through the agency of a pair of special stereoscopic goggles. The detailed scene that each eye perceives is the result of a complicated calculation performed in 'real time' so that the structure appears to remain consistent no matter how the subject turns his head or moves his body. Are we doing something similar when, in our 'mind's eyes' we conjure up some consistent mental image of a three-dimensional object, whether real or entirely imagined?

I shall argue that we are actually doing something very different from this. Let us consider an example. Figure 1 is a photograph of a regular dodecahedron. With some effort, it may be possible for us to rotate this image to a different orientation. In fact, we may feel that we have some conception of the object as an actual three-dimensional structure rather than as something that needs a particular vantage

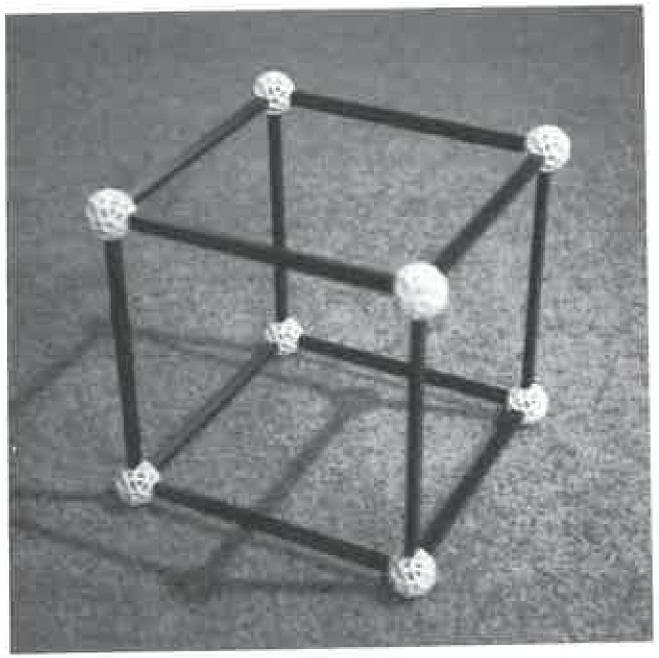


Figure 2 A cube

point from which to view it. Many people would find considerable difficulty in visualising an entire dodecahedron, but the cube depicted in the photograph of Figure 2 is a good deal easier. It is not that hard to transfer the flat image on the page to a 'solid' three-dimensional imagined structure. This may seem to be similar to what is involved with computer graphic displays. In Figure 3, I have provided a sequence of computer images of a regular dodecahedron viewed from successively slightly different vantage points, so that the

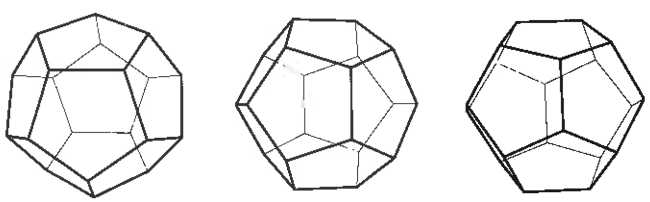


Figure 3 Computer pictures of a dodecahedron, from a gradually moving vantage point

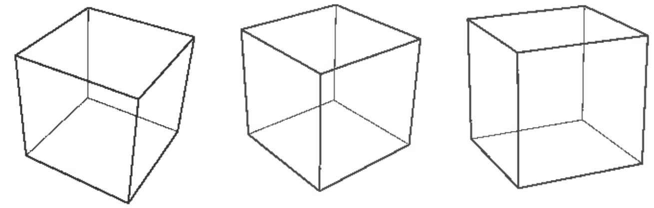


Figure 4 Computer pictures of a cube, from a gradually moving vantage point

dodecahedron appears to rotate just as it would if we slowly move around it. In Figure 4, I have done the same thing with a cube. In each case, there is, inside the computer, a stored representation of the dodecahedron, or cube, that does not change, but the chosen vantage point is gradually altered. Might not our own visual images be something like such stored computer representations?

I think it is unlikely – and to support this contention, let me indicate some fairly clear-cut distinctions. In the first place, the computer displays are far more accurate than anything that can be at all easily achieved by human imagination. Of course, it might well be argued that we are simply being very inefficient and inaccurate in our visualisations, as compared with a modern computer. Indeed, it would not be hard to introduce inaccuracy into our computer simulations, so that they fall to the level of accuracy that would be relevant to any particular human individual. If it were just our inaccuracy that distinguishes our own acts of visualisation from the outputs of computers then my argument would certainly be a very weak one. But visualisation carries with it strong elements of understanding, and it is actual understanding that the computer simulations lack.

To illustrate this point, consider Figure 5. Here I have added some lines to the photograph of Figure 1 to show that a cube can be found inside the dodecahedron, its eight vertices coinciding exactly with a selection of eight from the twenty vertices of the dodecahedron. It is not hard to see that this selection of eight vertices indeed gives us an exact cube. Symmetry considerations alone will tell us this; each face must clearly be an exact rectangle, at least, and the rectangle's sides must indeed be equal since each is a 'diagonal' of one of the equal

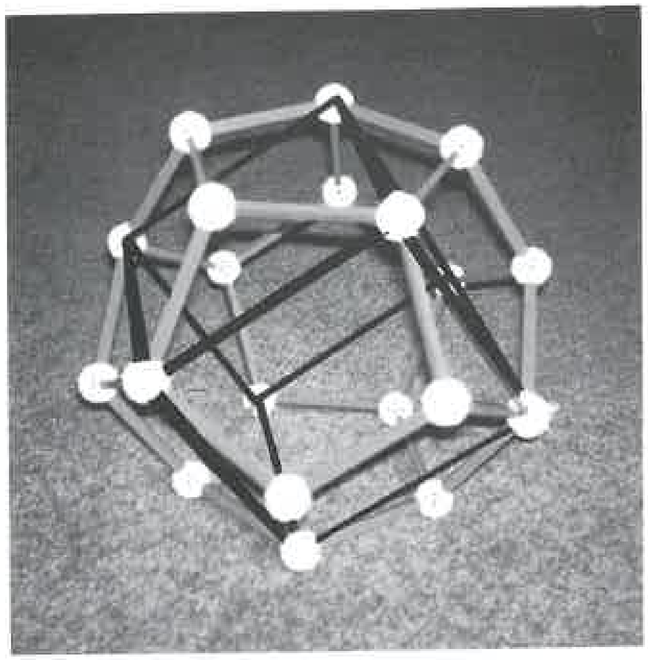


Figure 5 Dodecahedron with additional lines to show that a cube sits within it, sharing eight of its vertices

regular-pentagonal faces of the dodecahedron. Can the computer simulation 'see' this fact? This position would be hard to maintain, I feel, but we can at least test it by asking whether or not two of the edges of the computer's proposed 'cube' are equal. (This, in itself, is far from sufficient, of course, but at least it is a start.) When I tried this on two edges, the computer came out with two ten-digit decimal expressions that differed in the final place. When I asked it whether these two numbers were equal, it asserted that indeed they were. (Apparently, the program allows for some round-off error.)

This is hardly a convincing demonstration that the computer in any sense 'knew' that there is an absolutely exact cube in the regular dodecahedron. Rather, it seems to me, it establishes the contrary conclusion, that all this particular computer 'visualisation' can do is come up with approximations, albeit approximations with nine or ten figures of accuracy. It has no way of reaching the exact conclusion that our own visualisations – and accompanying understandings – are capable of: that indeed our proposed cube is geometrically precise.

In fact, being wise after the event, a computer programmer might replace the particular way in which the dodecahedron is stored in the machine by another one whereby exact information about distances and angles could be retrieved on demand. It would then be possible for the machine to give a correctly affirmative answer to our question about whether the suggested cube is indeed an exact one. Actually, being wise after the event, there would be an even easier way: the computer could be instructed simply to answer 'yes' to this particular question about the cube! The trouble with this, of course, is that the computer itself could in no way be said to possess any mathematical understanding of the exactness of the cube. It would simply be parrotting the information that its programmer had provided it with, and no-one would argue that any understanding of the exactness lay with the computer rather than with its human programmer.

One might try to do better than this, of course, and perhaps equip the computer with the complete system of axioms for three-dimensional Euclidean geometry. It could then try to ascertain whether a given statement, such as the exactness of the cube referred to above, could be deduced from these axioms. In this way it could, in principle at least, provide the correct answers to many geometrical problems. Of course, it might still be questioned whether what the computer does bears any relation to what a human mathematician does when understanding that a geometrical statement is actually true. That human understanding has to do with a belief in the validity of those intuitions – based to a good extent on symmetry considerations – underlying the very choice of the axioms themselves. The issue is a somewhat delicate one for there are valid geometric axiom systems that are distinct from those of Euclid. Indeed, when I later present powerful arguments in support of the thesis that our insights and understandings are not things that can be reduced to computation, it will be necessary to turn away from geometry, and to address the issue of computation directly, where it will be our intuitions concerning the natural numbers 0, 1, 2, 3, 4, 5, . . . , rather than geometrical forms, that will be the subject of our deliberations.

But before turning to such matters, let me give credit to what might be called one of the early 'success stories' of Artificial Intelligence. In the early 1960s, H. L. Gelernter programmed a computer to derive propositions in Euclidean plane geometry from the axioms with which it had been initially provided. When the computer came up with its proof that if a triangle has two equal sides, then the angles opposite to those sides are also equal, Gelernter was startled. For the computer's proof had been unknown to him, and it was considerably simpler than that given by Euclid. The computer's argument was this (see Figure 6): since AB=AC, the triangles ABC and ACB are congruent (side-side-side); therefore <ABC equals <ACB, QED! In fact this argument was not new. (It was given in the fourth century AD by Pappus.) But it was undoubtedly a striking fact that a computer could come up with something so elegant and unexpected.

In this example, the computer's success arose because its 'blind' rule-following stopped it from being distracted from the seeming absurdity of its own argument. No doubt there are many other situations in which human mathematicians have been similarly distracted from seeing arguments that they should have seen. However, this particular example is one for which the chain of reasoning is very short, and it is not hard to find it by means of a mindless search. When the derivations from axioms are long and complicated, as they tend to be with mathematical arguments of considerable sophistication, then the 'mindless search' method becomes hopelessly inefficient.

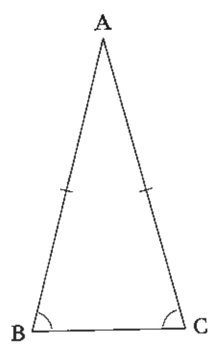


Figure 6 An isosceles triangle; the angles at B and C are equal – as proved by Gelernter's computer program, since ABC and ACB are congruent

Insight and understanding are necessary ingredients, if the search is to become manageable.

PROOF BY GEOMETRICAL INSIGHT 1: FAREY FRACTIONS

Choose a reasonable-sized natural number n (say n=9) and write down, in order of size, all the fractions from 0 to 1, expressed in their lowest terms, whose denominators do not exceed n (here with n=9)

0/1, 1/9, 1/8, 1/7, 1/6, 1/5, 2/9, 1/4, 2/7, 1/3, 3/8, 2/5, 3/7, 4/9, 1/2, 5/9, 4/7, 3/5, 5/8, 2/3, 5/7, 3/4, 7/9, 4/5, 5/6, 6/7, 7/8, 8/9, 1/1

Such an array is referred to as a sequence of *Farey fractions*. Let me point out one remarkable property that such a sequence possesses. If we take the difference between any pair of consecutive fractions in the list, then we find that the numerator is always 1. For example

$$2/5-3/8 = (16-15)/(5\times8) = 1/40$$
$$4/7-5/9 = (36-35)/(7\times9) = 1/63$$

How do we know that this must always be the case? We must find a way of convincing ourselves that if a/b and c/d are consecutive fractions in the list, then

$$ad - bc = 1$$

(since $a/b-c/d=(ad-bc)/bd$). To do this, we can use a geometrical argument. Let us imagine each fraction a/b to be plotted as the point (a,b) on a graph, where we also take the points representing successive fractions to be joined by straight-line segments. Figure 7 gives an accurate representation of all this, for n=9, but the figure is rather crowded and it will be much clearer if we use the inaccurate representation of part of the sequence that is illustrated in Figure 8. (It is a striking fact about mathematical diagrams, and the attendant visualisation of abstract concepts – which is indeed what is going on here – that an inaccurate image, if it is inaccurate in an appropriate way, may often give more of an insight as to what is going on than

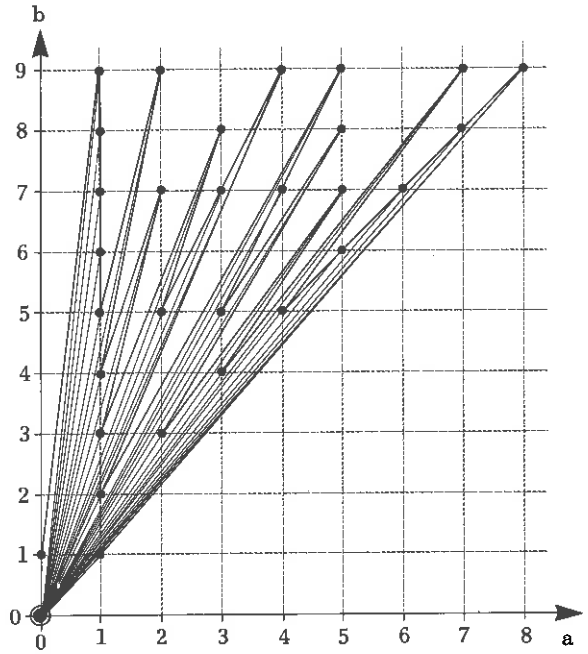


Figure 7 Farey fractions illustrated geometrically

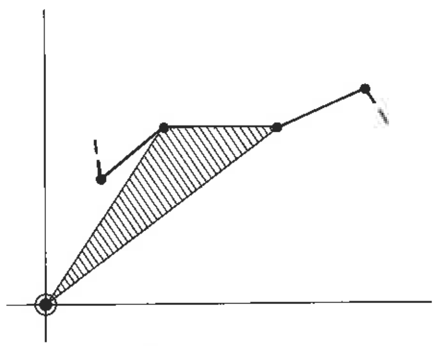


Figure 8 The area of the shaded triangle is ½ unit

an accurate one!) We wish to show that $ad-bc=1$. It is a well-known formula of coordinate geometry that the *area* Δ of the triangle whose vertices are (0,0), (a,b) and (c,d) is given by

$$\Delta = (ad - bc)/2$$

so what we must establish is that this area is precisely 1/2 (Figure 9).

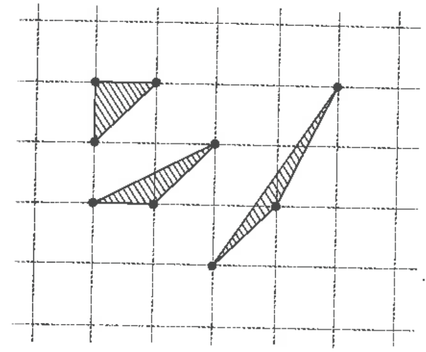


Figure 9 Triangles whose vertices are on lattice points and which contain no other lattice points, either internally or on an edge, are each of area = ½ unit

In fact, it is a theorem (essentially a special case of one due to Minkowski) that if a triangle's vertices all lie at integer lattice points (the points (x,y) where x and y are integers) and if it has the property that it contains no other lattice point either in its interior or on its edges, then it must have an area equal to 1/2. Thus, what we need to show is that our triangle indeed has this property. This is actually not hard to see from the fact that the two fractions a/b and c/d are both in their lowest terms (so that there are no further lattice points on the sides of the triangle out from the origin (0,0)), and from the fact that all fractions with denominators no greater than n have been included – a fact that would be contradicted if our triangle constructed from two successive fractions a/b and c/d contained any other lattice point.

It remains to establish the aforementioned theorem. One way of doing this (Figure 10) is to consider a sequence of transformations whereby the triangle is successively moved without changing its area until it becomes just half of a lattice square. Each transformation consists of taking one of the vertices and moving it parallel to the opposite side, until it reaches another lattice point that lies closer to the other two vertices than before. I shall not bother with the full details of this here, but I hope that the rough idea is clear.

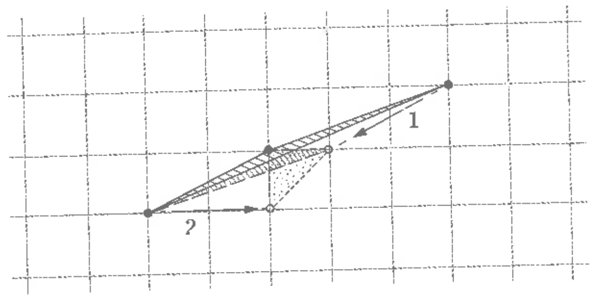


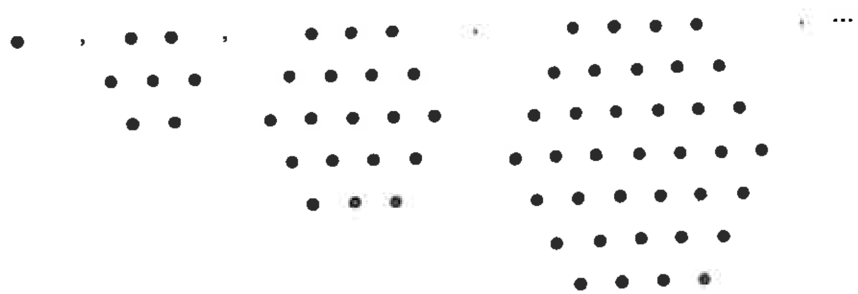
Figure 10 The fact that the area is $\frac{1}{2}$ unit can be proved by successively displacing a vertex to another lattice point closer to the opposite side, in a direction parallel to that side, until a diagonally bisected square is obtained

PROOF BY GEOMETRICAL INSIGHT 2: HEXAGONAL NUMBERS

My next example involves what are called hexagonal numbers

1, 7, 19, 37, 61, 91, 127, . . .

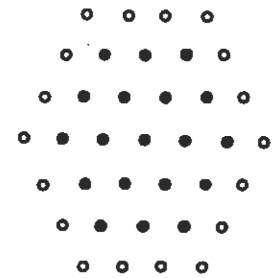
namely the numbers that can be arranged as regular hexagonal arrays (excluding the vacuous array):



These numbers are obtained, starting from 1, by adding successive multiples of 6

6, 12, 18, 24, 30, 36, . . .

as we see from the fact that each hexagonal number can be obtained from the one before it by adding a hexagonal ring around its border:



For the number of spots in this ring is a multiple of 6, the multiplier increasing by 1 each time, as the hexagon gets larger.

Now let us add together the hexagonal numbers successively, up to a certain point, starting with 1. What do we find?

$$1=1,\ 1+7=8,\ 1+7+19=27,\ 1+7+19+37=64,\ 1+7+19+37+61=125$$

The numbers 1, 8, 27, 64, 125 are all *cubes*. A cube is a number multiplied by itself three times

$$1=1^3=1\times1\times1,\ 8=2^3=2\times2\times2,\ 27=3^3=3\times3\times3,\ 64=4^3=4\times4\times4,\ 125=5^3=5\times5\times5,\ \ldots$$

Is this a general property of hexagonal numbers? Let's try the next case. We indeed find

$$1+7+19+37+61+91 = 216 = 6\times6\times6 = 6^3$$

I am going to try to convince you that this is always true. First of all, a cube is called a cube because it is a number that can be represented as a cubic array of spots

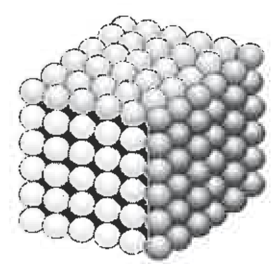


Figure 11 A cubical array of spheres

I want you to try to think of such an array as built up successively, starting at one corner and then adding a succession of three-faced arrangements each consisting of back wall, side wall and ceiling, as depicted thus:

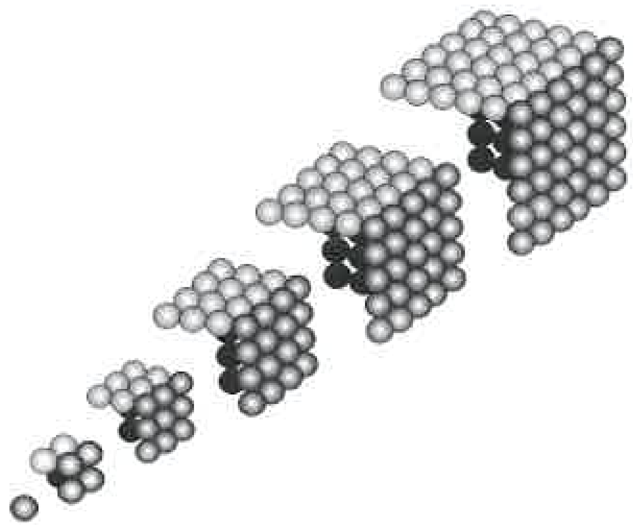


Figure 12 The cubic array is separated into a succession of layers consisting of side wall, back wall and ceiling, each of which is viewed from a long distance off

Now view this three-faced arrangement from a long way out, along the direction of the corner common to all three faces. What do we see? A hexagon:

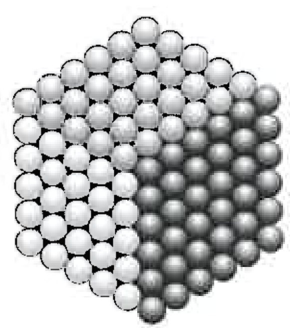


Figure 13 Each layers appear as a hexagonal arrangement – a hexagonal number

The spots that constitute these hexagons, successively increasing in size, when taken together, correspond to the spots that constitute the entire cube. This establishes the fact that adding together successive hexagonal numbers, starting with 1, will always give a perfect cube.

PROOF BY GEOMETRICAL INSIGHT 3: RULES OF ARITHMETIC

This last example again shows the power of mathematical (geometrical) visualisation. Let us try something else; this time a good deal more elementary. How do we know that, for two natural numbers a and b, we always have

$$a + b = b + a?$$

For this, all we need to do is visualise a collection of things (imagined to be 'a' in number) to which we add another collection of things (imagined to be 'b' in number). The total number of things altogether is clearly the same whichever order we add them in, so the required result follows.

This example is very trivial, and it is perhaps not quite 'geometrical' in the ordinary sense, but I think that it represents a genuine mathematical insight from an act of imagination. One might try to argue that this insight is really just an aspect of our 'experience' of the persistence of objects in the world. For example, apples and oranges do not just disappear when placed in a box, nor do they magically appear within the box. But there is more to it than this. None of us has ever directly experienced precisely 88990012345 objects or 60606999931 objects, yet we would have no doubt that indeed

$$88990012345 + 60606999931 = 60606999931 + 88990012345.$$

Our experiences with apples and oranges merely act as a guide towards our mathematical insights. Those insights are genuine abstractions that are valid methods of reasoning about abstract mathematical objects. (As always, in mathematics, one must be

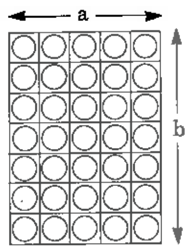


Figure 14 A rectangular array of a columns of b objects gives the same total number as would b columns of a objects – by rotation of the figure

extremely careful, however. There are kinds of *infinite* numbers for which the rule $a+b=b+a$ fails – but that's another story!)

Let's try another rule of elementary arithmetic:

$$a \times b = b \times a$$

This tells us that if we are to imagine taking 'a' collections of objects, where each collection contains 'b' objects, then the total number would be unaltered if we did the same thing but with 'a' and 'b' interchanged. Stated this way, it is not really obvious; but if we imagine our collections to be arranged as a succession of 'a' columns, where each column contains 'b' objects, then the symmetry becomes obvious (Figure 14). We can imagine rotating the resulting rectangular array through a right angle to achieve this. Even easier than this (if we have mental difficulty in visualising the rotation) is simply to read our array off the other way around rather than mentally rotating it, that is as row-by-row rather than column-by-column. This works just as well for, say,

$$97666000011 \times 777708999 = 777708999 \times 97666000011$$

as for $5\times7=7\times5$, even though we cannot precisely visualise collections of things that represent these actual individual numbers.

How about the associative law

$$(a \times b) \times c = a \times (b \times c)$$

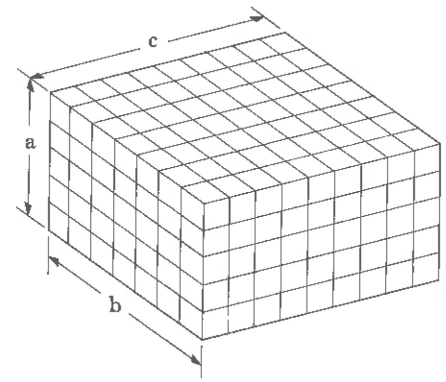


Figure 15 A three-dimensional rectangular array shows that $(a \times b) \times c = a \times (b \times c)$

We can see this one by imagining a three-dimensional rectangular array which we read off in two different ways: see Figure 15. Again this works just as well for, say

$$(97666000011\times777708999)\times83383302222$$
$$= 97666000011\times(777708999\times83383302222)$$

as for $(5\times7)\times9=5\times(7\times9)$, even though these large numbers cannot be individually visualised precisely.

An important thing to note about these 'visualisations' is that they are not really pictures of actual things in space, but of something much more abstract. It really does not matter whether 'actual' space has the necessary accurately Euclidean structure, or that it extends outwards far enough so the particular numbers that we wish to represent can be realised in terms of actual objects within that space. (For example, if we wished to represent the number $10^{10^{1000}}$, which is a perfectly good number, still subject to the same algebraic laws as the numbers that we can directly visualise, then we could not do so within the 'observable universe' with actual physical objects.) Nevertheless, our simple visualisations are indeed sufficient to provide the necessary insights that can convince us that the algebraic properties that we have been considering are actually true of all natural numbers, no matter how large they may be.

Now, let us consider the relation

$$((a\times b)\times c)\times d = a\times(b\times(c\times d))$$

Can we be sure that this one is universally true? Now we use a different method. It is not much good to use a four-dimensional array, which is what would be needed if we were to try to visualise the relation directly. We have no direct experience of four spatial dimensions, so our geometrical intuitions are not of much immediate use in this case. Instead, we can turn to algebra and deduce this last relation from what we have established previously. First

$$((a\times b)\times c)\times d = (a\times b)\times(c\times d)$$

by the earlier relation (but with $a \times b$, c, d, in place of a, b, c, respectively); then

$$(a\times b)\times(c\times d) = a\times(b\times(c\times d))$$

again by this earlier relation (but now just with $c\times d$ in place of c). The result follows by combining these two.

Algebra provides a very useful means of replacing our direct insights by calculational procedures. We do not now have to think of what our expressions actually 'mean'. We can now just calculate! In fact this is somewhat overstating the case I want to make. The effective use of algebra often requires a good deal of understanding, subtlety, and even artistry. But in many ways, it is the power of a good calculus (like ordinary algebra or, indeed, ordinary arithmetical notation) that, to a considerable degree, it enables understanding to be temporarily suspended and replaced by blind calculation. It is in this facility for blind calculation, rather than understanding, that computers can far exceed the capabilities of even the most effective of human experts.

CAN UNDERSTANDING ALSO BE REDUCED TO BLIND CALCULATION?

What about the human faculty of understanding? Can we be sure that it is not itself some kind of calculational activity? Presumably anyone who believes that genuine Artificial Intelligence is possible must also believe that the quality of understanding can be artificially simulated – for understanding is surely an essential part of genuine intelligence. The present meaning of the term Artificial Intelligence (or 'AI') is that it is something according to which the artificial simulation is indeed performed calculationally, by which I mean by the use of electronic computers. Although it is possible to imagine that, in the future, some means of 'artificial simulation' which is different from anything that can be achieved calculationally might be introduced, I shall stick to the standard 'calculational' meaning of the term here. Accordingly, those who believe that genuine AI is possible must believe that the quality of understanding can indeed be simulated calculationally.

Before giving my main reason for disbelieving this possibility, I should be a little clearer about what I mean by the term 'calculational'. In effect, as I indicated earlier, I mean anything that can be performed by a modern general-purpose computer. This is perhaps not very precise as a mathematical definition. What I really mean, technically, is anything that can be performed by a Turing machine. However, I appreciate that many readers will not know precisely what the term 'Turing machine' means, so it is easier just to refer to a modern general-purpose computer, where we require this 'computer' to be a mathematically idealised concept. The idealisations that we require are that the computer never makes mistakes, that it can continue indefinitely without ever wearing out, and that it has an unlimited storage capacity. If you think of this last idealisation as being a little unreasonable, just imagine that it is always possible to add more storage capacity (that is 'memory') to the computer whenever it runs out.

An important point to make here is that actions of things like 'parallel computers' and (artificial) 'neural networks' (or 'connection machines'), that we hear quite a lot about these days, are all included in what I mean by 'calculations'. For some types of problem, a computer designed according to what is called a 'parallel architecture' may be much more efficient or much faster than an ordinary serial computer, but there is no difference between the two in principle. Likewise, systems like neural networks – referred to as 'bottom-up' systems – which improve upon their performance by a 'learning' process aimed at optimising the quality of their output, are also calculational. A bottom-up action contrasts with the standard 'top-down' calculational procedures in that the latter operate according to an algorithm that is known to work correctly for the class of problem that it is concerned with, whereas with bottom-up action no such algorithm is given beforehand and, instead, a means is initially provided whereby the system is to improve its performance as it gains experience. This is still a calculational procedure (and therefore still an 'algorithm'), however, because the very means whereby the system is to improve its performance is itself given by a set of calculational rules. From the point of view of the present discussion, the essential difference between a bottom-up and a top-down system is that the former is only an approximate means of obtaining the required answers even though it may sometimes be a very effective one.

A simple test for deciding whether a system is calculational is to ask: can it be run on an ordinary computer? If it can be then the system is indeed a calculational one. In fact, as far as I am aware, most of the (artificial) neural network systems that have been constructed to date are actually, as they stand, simulations run on ordinary computers – so that there is no question but that they must be calculational! (Perhaps it is intended eventually to construct some special electronic hardware, on which the neural network system would be run much more efficiently, but that makes no difference to the fact that such systems are always calculational in nature.)

I now wish to present an argument that effectively demonstrates that mathematical understanding is not a calculational activity. The argument is based on a form of the famous theorem of Kurt Gödel, that he proved in 1930, but where I also call upon some later ideas

introduced mainly by Alan Turing in about 1935. I shall not need any of the technical details of Gödel's argument, and although the argument may be found to be somewhat confusing, it does not use any difficult mathematics.

The type of calculation that we shall be concerned with will be an operation that can be performed on a natural number. The action of some calculation C on a natural number n is written C(n). We may think of C as being given by a computer program, where after feeding the program into our computer, we then supply the computer with the number n, which the computer operates on to produce the answer. (Technically, C may be thought of as a Turing machine, and C(n) is the action of that Turing machine on the natural number n.) I shall not be very concerned here with the actual result of the calculation, but mainly whether or not it ever eventually stops.

Let us consider some examples. One possible calculation would be to form the square n^2 of the natural number n. This particular calculation encounters no problem about eventually stopping, since the square of any given natural number can certainly be formed in a finite time. (Recall that there is to be no limit on the computer's storage space.) More subtle is the following example of a calculation, which depends on the given natural number n.

> *Find the smallest natural number that is not the sum of* n *squares.*

Our calculation would proceed, trying the natural numbers in turn: 0, 1, 2, 3, 4, 5, . . . ; until it finds one that is not the sum of n squares. To get the idea, let us first see how this works for $n=2$. We start with 0, and find that 0 is indeed the sum of two squares, namely $0=0^2+0^2$. We must move on to 1, and find that although 0^2+0^2 doesn't work, we do find that indeed $1=0^2+1^2$. Thus, we must move on to 2, finding that although 0^2+0^2, or 0^2+1^2, or 0^2+2^2 do not work, we indeed have $2=1^2+1^2$. Moving on to 3, we find that none of 0^2+0^2, 0^2+1^2, 0^2+2^2, 0^2+3^2, 1^2+1^2, 1^2+2^2, 1^2+3^2, 2^2+2^2, 2^2+3^2, or 3^2+3^2, will work (cutting the calculation off when the number to be squared reaches the number to be summed to – though we could be more efficient, cutting things off earlier). Thus we find 3 as the smallest number that is not the sum of two squares. We could now try this all over again with $n=3$, finding, in this case, that the number 7 is the smallest that is not the sum of three squares. We can also go back and test the case $n=1$, finding that 2 is the smallest number not the sum of one square (and examining the logic involved in the case $n=0$, we find that 1 is the smallest number not the sum of zero squares).

Now let us consider $n=4$. Our calculation proceeds, finding

$$0=0^2+0^2+0^2+0^2,\ 1=0^2+0^2+0^2+1^2,\ 2=0^2+0^2+1^2+1^2, \ldots$$
$$6=0^2+1^2+1^2+2^2,\ 7=1^2+1^2+1^2+2^2,\ 8=0^2+0^2+2^2+2^2, \ldots$$
$$23=1^2+2^2+3^2+3^2,\ 24=0^2+2^2+2^2+4^2, \ldots$$
$$71=2^2+3^2+3^2+7^2,$$

and so on.

It seems never to stop at all! In fact it never does stop. According to a famous theorem first proved in the eighteenth century by the great French mathematician Joseph L. Lagrange, *every* number is, indeed, the sum of four squares. It is not such an easy theorem. Even Lagrange's contemporary, the great Swiss mathematician Leonhard Euler, a man of astounding mathematical insight, originality and productiveness, had tried but failed to find a proof, so I am certainly not going to trouble the reader with the details of Lagrange's argument here.

Instead, let us try a calculation that is very much easier to see never to stop.

> *Find the smallest odd number that is the sum of* n *even numbers.*

The poor computer that is set mindlessly upon this task will certainly never complete its work, no matter what n is – because even numbers always add to even numbers.

I have given some examples of calculations, some of which will eventually terminate to produce an answer and some of which continue for ever. How are we to decide which of these two possibilities will occur in any particular case? When a calculation does not ever

stop, by what means can we ascertain this fact? We have seen that this may be hard, as was the case with Lagrange's theorem, but sometimes it is easy, as in the last example. Are mathematicians themselves using some calculational procedure in order to ascertain that non-stopping calculations actually do not stop?

Let us imagine that they do use such a procedure, and, moreover, they are aware of the nature of this procedure and of the fact that the procedure that they use is sound – that is to say, that it does not erroneously come to the conclusion that a calculation does not stop when in fact it does. It will not be necessary to assume that this procedure can, in every case, ascertain that a non-stopping calculation does not in fact stop.

Let us call our putative procedure A. Then when A is presented with a calculation C and with the number n on which C acts, it will be set into action. If the calculation A itself successfully comes to a halt, then it will have decided that C(n) does not in fact terminate. (Note that A is not a procedure for deciding that calculations *do* terminate. We might have some other procedure B for that kind of decision. If we want to incorporate B into A, we can do so by employing the device of putting A into a 'loop' whenever B successfully comes to its conclusion, thus making sure that the 'A' used in the argument will actually not terminate when the calculation does. This is just a technical point. I mention it only because sometimes people are disturbed if A ignores arguments that show that a calculation will stop.)

In order to be a little clearer about how a calculation (A) can act on another calculation (C), let us specify the various calculations C by giving each one a separate number. Thus the different calculations will be

$$C_0, C_1, C_2, C_3, C_4, C_5, \ldots$$

where we can think of this ordering as being provided by the numerical ordering of the computer programs that specify these calculations in turn. Technically, C_r could be the 'r^{th} Turing machine' in some standard system of numbering. We can now think of A as a calculation acting on the two numbers r and n, and conclude

If $A(r,n)$ *stops, then* $C_r(n)$ *does not stop*

The calculation A is still not quite of the form of the other calculations that we have been considering since it acts on two natural numbers, not one. Let us remedy this by considering only the cases for which $r=n$. (This perhaps seems an odd thing to do, but it is the crucial step in Gödel's and Turing's argument, itself taken from the famous 'diagonal slash' of the highly original nineteenth-century mathematician Georg Cantor.) We obtain

If $A(n,n)$ *stops, then* $C_n(n)$ *does not stop*

Now, $A(n,n)$ is of the form of the calculations that we have been considering, so it must be one of them, say the k[th] one, and we have

$$A(n,n) = C_k(n)$$

and therefore, putting this in the displayed statement above

If $C_k(n)$ *stops, then* $C_n(n)$ *does not stop*

Taking the particular case $n=k$, we obtain

If $C_k(k)$ *stops, then* $C_k(k)$ *does not stop*

From this we deduce that $C_k(k)$ certainly will not stop (because if it did, then it doesn't!).

The remarkable thing, here, is that although we have ourselves just seen that $C_k(k)$ does not stop, the calculation A is incapable of ascertaining this fact. For $A(k,k)$ is the same as $C_k(k)$, so if the latter does not stop, the former cannot stop either. Thus, A cannot successfully come to the conclusion that $C_k(k)$ does not stop! Since we have actually just established that $C_k(k)$ does not stop, it follows that the mathematical procedures that we use in order to establish that calculations do not stop are not accessible to the calculational procedure A. We note this would apply whatever A is, provided that we know what A is, and we know A to be sound. The inescapable conclusion seems to be:

Mathematicians are not using a knowably sound calculational procedure in order to ascertain mathematical truth

We deduce that mathematical understanding – the means whereby mathematicians arrive at their conclusions with respect to mathematical truth – cannot be reduced to blind calculation!

DISCUSSION OF THE IMPLICATIONS OF THE GÖDEL ARGUMENT

I should address some of the possible loopholes and objections that various people have made to arguments of the type that I have just given.

A common reaction to the Gödel argument is simply not to take it seriously, for 'how could an argument of that kind possibly have anything to say about the mind?' But although this may be a natural reaction, it is no answer to the argument. If one believes that the conclusion is wrong, then one must find a flaw in the argument.

A worry that people often have is that I have given the argument in terms of a single A, whereas there might be a whole host of calculational procedures that mathematicians use. However, this is not a real objection. There is no difficulty about combining together many different such procedures (even an infinite number of them) into a single 'A', provided that it is a calculational matter to decide which procedure to use. It is only for convenience that I have phrased things as I have, and there is no loss of generality involved.

One common objection is to point out that the Gödel(–Turing) argument is itself something that one could envisage putting on a computer. There is nothing non-computable about generating all the steps of the argument as I have given them and, if we wish, we could include some version of the Gödel argument into our rules for deciding that calculations will not stop. But if we simply adjoin this new argument to the 'A' that we had before, we are really cheating, because that 'A' was already supposed to represent the totality of the means that are available to mathematicians for ascertaining that calculations do not stop. If we accept the Gödel argument as a new means for ascertaining that certain calculations do not stop, then our previous 'A' did not represent that totality. Instead, we should be using some new 'A', say A*, that includes this version of the Gödel argument. But if A* is supposed to represent that totality then we can apply our argument instead to A*, and again we obtain a contradiction. The point is that we cannot put the entire idea of the Gödel argument into calculational form even though we can incorporate certain instances of it into a calculation.

There is a closely related objection that people sometimes try to make against the version of the Gödel argument that I have given. The claim is sometimes made that the argument applies only to the one particular A that has been singled out, and that it is not a general objection against all As. This is a misconception about how the argument is being used, however. The argument has the form, familiar to mathematicians, of a *reductio ad absurdum*, whereby a hypothesis is put forward (here that there is some knowably sound calculational procedure that we use – and that we are calling it A) from which a contradiction is obtained, thereby showing that the hypothesis was false. The argument indeed rules out all such As, not just a particular one.

The Gödel argument is more often phrased in terms of some axiom system F, and in terms of the provability of mathematical results from F. Gödel's most familiar theorem shows that provided F is (and is believed to be) consistent – so it cannot be used to prove that a statement is true and false at the same time – then there are mathematical propositions that are (and are seen to be) true but which cannot be derived from F. The argument is often made that we cannot actually see that these propositions are true unless we can show that the axioms are consistent. I have not phrased my own argument in this way, but have referred to the soundness of the procedure A. If we trust A not to make mistakes, then we see, by the Gödel argument that A cannot represent the totality of our mathematical insights, whatever A might be. Likewise, if we trust F – and this implies that

we believe F to be consistent, for otherwise we could use it to prove a nonsense like 1=2 – then we see that F does not represent the totality of our mathematical insights. If we do not trust F (or A) then it certainly cannot represent our insights!

Another objection I have seen made is that we cannot be sure that the numbers that we are talking about are actually the natural numbers 0, 1, 2, 3, 4, 5, . . . , but they might be some funny kind of 'surreal' numbers, where some of the things that would be true for the natural numbers turn out to be false for the surreal numbers. Although I cannot really take this argument seriously – because the numbers that we are talking about *are* the natural numbers, and not anything else – this objection contains, in a sense, the 'nub' of the mystery. How do we know that it is the natural numbers that we are indeed talking about? We cannot, merely by specifying a finite system of axioms or rules, completely distinguish the natural numbers from all the various kinds of 'surreal' numbers. Yet every child knows what the things 0, 1, 2, 3, 4, 5, . . . mean, despite this fact. Somehow we have a direct intuition which tells us what the 'natural number' concept is, given only very inadequate hints in terms of 'two bananas, five oranges, zero socks' and so on.

Let us then accept the apparently inescapable implication of the Gödel(–Turing) argument: *mathematicians do not simply ascertain mathematical truth by means of knowably sound calculational procedures*. There remain the possibilities that they might use unknowable or unsound calculational procedures – or, as is my own belief, that they simply do not just use calculational procedures when they ascertain truth. With regard to the calculational possibilities, I should point out that mathematicians certainly don't *think* that they are using unknowable or unsound procedures in order to ascertain mathematical truth! They are of the opinion that they are perfectly aware of what they are doing when they use whatever methods they use and, moreover, that these methods are perfectly sound. It is undoubtedly true that mathematicians make mistakes from time to time, but these mistakes are recognisable as such. Another mathematician might point out the mistake, or the very mathematician who made the mistake might notice it later. It is not that there are inbuilt errors that mathematicians are completely incapable of seeing as errors. Consciously, the methods that mathematicians use are neither unknowable nor unsound. If they are indeed using a horrendously complicated unknowable calculational procedure X, or an unsound calculational procedure Y, then these things would have to be completely unconscious.

Is it plausible that they are actually using such an X or Y without knowing it? One point should be emphasised here, and that is the apparently universal nature of the criteria that mathematicians use to establish the truth of their results. Suppose that each mathematician used a different X or Y, personal to that particular mathematician, then they would not be able to convince one another of their arguments. We require a universal X or Y that would have to be built into their brains in a way that would be common to all. How could such an X or Y have arisen? It would have to have been by means of the powerful processes of natural selection that Darwin himself revealed to us. But anyone who has glanced at any respectable modern mathematics research journal will realise how far-removed from the activities of the outside world are the deliberations of mathematicians. If it were the horrendously complicated unknowable X, or the complicatedly erroneous Y, that somehow got implanted in our brains, via our genes, it is very hard to see how this could have been by the process of natural selection, a process geared to promote the survival of our primitive remote ancestors. Much more likely is that there is no such X or Y but, instead, it is a non-calculational quality – the ability to *understand* – that natural selection has favoured. This quality is in no way specific to mathematics, but would have been immensely valuable to our ancestors in many different ways, providing a powerful selective advantage. Only incidentally does it turn out that this same quality is what is needed for mathematics.

WHAT UNDERLIES NON-COMPUTATIONAL BRAIN ACTION?

If we accept that we do something beyond computation when we understand, how can this be reconciled with the view that our brains are just physical objects governed by precise physical laws? One way out might be to adopt a mystical viewpoint according to which the behaviour of the mind could not be accounted for simply in terms of the physical brain. Apparently Gödel himself felt driven to this kind of solution.

For myself, I reject mysticism in favour of a scientific explanation. There are various possibilities to consider. For example, perhaps we use a calculational procedure that is continually improving itself. If so, then, we must ask, how is this improvement coming about. If the improvement is itself governed by some preassigned mechanism, then, as was the case with a neural network, it is still calculational. Might the improvements come about via some continual interaction with the environment? But if this is to give us something beyond calculation, it would imply that there is an essential feature of our environment that cannot even be *simulated* computationally. Of course, it may not be feasible to simulate the particular environment of a specific individual, but to suggest that it is in principle impossible to simulate any appropriate plausible environment is to suggest that there is something essential in the physical action of the world that lies beyond calculation. Once that possibility is accepted, then the possibility that our very brains might act according to some non-calculational action must also be allowed.

What about random ingredients? Would they count as 'non-calculational'? In the sense that a strict Turing machine does not allow for such ingredients, their inclusion would, indeed, take us out of calculational activity. However, in practice, purely random ingredients would add nothing useful to pure calculation. In fact, there are many calculational procedures that call for the inclusion of random ingredients, but these are usually implemented in practice by incorporating what are called 'pseudo-random numbers', these being numbers that are generated by some suitably complicated process that gives them the appearance of randomness even though they are not strictly random. For practical purposes randomness gives us nothing useful that cannot be achieved purely calculationally. (Closely related is the behaviour of what are called 'chaotic systems' which have the appearance of randomness even though they are entirely calculational.)

Finally, there is the possibility that, in appropriate circumstances, the actual behaviour of physical systems might be essentially non-calculational, a conscious brain being one example. My personal belief is that this is indeed the case, but there are several speculative elements that are involved in such a belief. First, one must ask where in physics non-calculational action might be found. I believe that such action must be in an area where present-day physics is in need of radical improvement – what is referred to as 'the measurement problem' in quantum theory. Roughly speaking, such an improved theory would supply a more satisfactory link between the micro-level of atoms and molecules (the 'quantum level') and the macro-level of discernable phenomena (the 'classical level'). I believe that brain action will never be properly understood without such a theory. At least something of this nature will be needed in order to explain the non-calculational aspects of mathematical and other kinds of understanding.

FURTHER READING

Boden, Margaret A., *The Creative Mind: Myths and Mechanisms*, London: Weidenfeld 1990.

Broadbent, D. (ed.), *The Simulation of Human Intelligence*, Oxford: Blackwell 1993.

Chou, Shang-Ching, *Mechanical Geometry Theorem Proving*, Dordrecht: Reidel 1987.

Nagel, E., and Newman, J.R., *Gödel's Proof*, London: Routledge and Kegan Paul 1958.

Penrose, Roger, *The Emperor's New Mind*, Oxford: Oxford University Press 1989. Vintage paperback, 1990.

6

Intelligence in Traditional Music

SIMHA AROM

Tradition might be seen as the process of transmission, within a community that identifies itself as such, of a specific knowledge and of relatively stable forms of behaviour. These define 'symbolic communities' which, as the French anthropologist Jean Molino puts it, are 'groups of individuals sharing some relatively stable features of language and culture, i.e., relatively stable features of their symbolic organisation systems'.

Traditional music is a symbolic production which, like language for a given community, is transmitted from mouth to ear, from generation to generation, and represents a major constituent of the group's cultural identity. Almost all traditional musics share this character: they are transmitted orally. Memory thus plays an essential role. Even in societies which have systems of notation, such as China, India, Tibet, and others, writing only fulfils a mnemotechnic role, as a memory support. It never assumes a prescriptive function. Since orally transmitted musics are not fixed once and for all in writing, there obviously is large scope for improvisation and variations.

Traditional music can either be art music (the French call it *musique savante*), or popular music. As art music, it may be the subject of abstract speculation, of deductions based on acoustic rules, of a constituted body of codes – sometimes written. In such cases, one can properly speak of a theory, since it is presented in explicit form. Such