From: Imre Lakatos, Mathematics, Science & Epistemology, eds. J. Worrall & G. Currie, (CUP: Cambridge, 1978) 61-69.

4

What does a mathematical proof prove?*

On the face of it there should be no disagreement about mathematical proof. Everybody looks enviously at the alleged unanimity of mathematicians; but in fact there is a considerable amount of controversy in mathematics. Pure mathematicians disown the proofs of applied mathematicians, while logicians in turn disavow those of pure mathematicians. Logicists disdain the proofs of formalists and some intuitionists dismiss with contempt the proofs of logicists and formalists.

I shall begin with a rough classification of mathematical proofs; I classify all proofs accepted as such by working mathematicians or logicians under three heads:

(1) pre-formal proofs
(2) formal proofs
(3) post-formal proofs.

Of these (1) and (3) are kinds of informal proofs.

I am afraid that some ardent Popperite may already be rejecting all that I am about to say on account of my classification. He will say that these misnomers clearly prove that I really think that mathematics has some necessary, or at least standard, pattern of historical development – pre-formal, formal and post-formal stages, and that I am already showing my hand – that I want to inject a disastrous historicism into sound mathematical philosophy.

It will turn out in the course of my paper that this, in fact, is just what I should like to do; I am quite convinced that even the poverty of historicism is better than the complete absence of it – always providing of course that it is handled with the care necessary in dealing with any explosives.

As a consequence of the unhistorical conception of 'formal theory' there has been a lot of discussion as to what constitutes a respectable formal system out of the immense multitude of capriciously proposed consistent formal systems which are mostly uninteresting games. Formalists had to disentangle themselves from these difficulties. They could of course have done this by dropping their basic outlook, but they have

* This paper seems to have been written some time between 1959 and 1961 for Dr T. J. Smiley's seminar at Cambridge. Lakatos's own copy contains several handwritten corrections; some by himself and some by Dr Smiley. We have incorporated them into the text. There is no indication that Lakatos ever returned to this paper after 1961. He subsequently changed his mind on some of the points made in the paper and had no plans to publish it himself. (*Eds.*)

tended to prefer complicated *ad hoc* corrections. They look for criteria distinguishing those formal systems which are '*interesting*' or '*acceptable*' and so on, thus betraying their bad consciences in accepting the pure formalist conception according to which mathematics is the set of *all consistent* formal systems. For instance, Kneale says that a mathematical system should be 'interesting'. His definition runs as follows: 'A possible – [possible means complying with some usual concept of modern rigour – i.e. consistent] system is interesting mathematically if it is rich in theorems and has many connections with other parts of mathematics, and in particular with the arithmetic of natural numbers.'[1] Curry, who is a most extreme representative of formalism, introduces the notion of 'acceptability'. He says: 'The primary criterion of acceptability is empirical; and the most important considerations are adequacy and simplicity.'[2] I fear there is a point on which I slightly disagree with their approach: they select from a previously given set of formal systems those which are interesting or acceptable. I should like to reverse the order: we should speak of formal systems only if they are formalizations of established informal mathematical theories. No further criteria are needed. There is indeed no respectable formal theory which does not have in some way or another a respectable informal ancestor.

Now I come back to our original subject: proofs. Most of the students of the modern philosophy of mathematics will instinctively define proof according to their narrow formalist conception of mathematics. That is, they will say that a proof is a finite sequence of formulae of some given system, where each formula of the sequence is either an axiom of the system or a formula derived by a rule of the system from some of the preceding formulae. 'Pure' formalism admits any formal system, so we must always specify in which system S we operate; then we speak only about an S-proof. Logicism admits essentially only one large distinguished system, and so essentially admits a single concept of proof.

One of the most outstanding features of such a formal proof is that we can mechanically decide of any given alleged proof if it really was a proof or not.

But what about an *in*formal proof? Recently there have been some attempts by logicians to analyse features of proofs in informal theories. Thus a well known modern text-book of logic says that an 'informal proof' is a formal proof which suppresses mention of the logical rules of inference and logical axioms, and indicates only every use of the specific postulates.[3]

Now this so-called 'informal proof' is nothing other than a proof in an axiomatized mathematical theory which has already taken the shape of a hypothetico-deductive system, but which leaves its underlying logic unspecified. At the present stage of development in mathematical logic a competent logician can grasp in a very short time what the necessary underlying logic of a theory is, and can formalize any such proof without too much brain-racking.

[1] Kneale [1955], p. 106.
[2] Curry [1958], p. 62.
[3] Suppes [1957], p. 128.

But to call this sort of proof an informal proof is a misnomer and a misleading one. It may perhaps be called a quasi-formal proof or a 'formal proof with gaps' but to suggest that an informal proof is just an incomplete formal proof seems to me to be to make the same mistake as early educationalists did, when, assuming that a child was merely miniature grown-up, they neglected the direct study of child-behaviour in favour of theorizing based on simple analogy with adult behaviour.

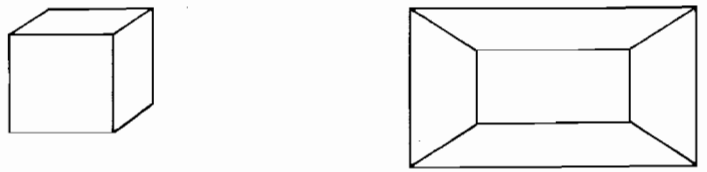
But now I should like to exhibit some truly informal, or, to be more precise, pre-formal proofs.

My first example will be a proof of Euler's well-known theorem on simple polyhedra.[1] The theorem is this: Let V denote the number of vertices, E the number of edges and F the number of faces of a simple polyhedron; then invariably

$$V-E+F=2$$

By a polyhedron is meant a solid whose surface consists of a number of polygonal faces, and a simple polyhedron is one without 'holes', so that its surface can be deformed continuously into the surface of a sphere. The proof of this theorem runs as follows:

Let us imagine a simple polyhedron to be hollow, with a surface made of thin rubber (see Figure 1 (a)). Then if we cut out one of the faces of the hollow polyhedron, we can deform the remaining surface until it stretches out flat on a plane (see Figure 1 (b)). Of course, the areas of the faces and the angles between the edges of the polyhedron will have changed in this process. But the network of vertices and edges in the plane will contain the same number of vertices and edges as did the original polyhedron, while the number of polygons will be one less than in the original polyhedron, since one face was removed. We shall now show that for the plane network, $V-E+F=1$, so that, if the removed face is counted, the result is $V-E+F=2$ for the original polyhedron.

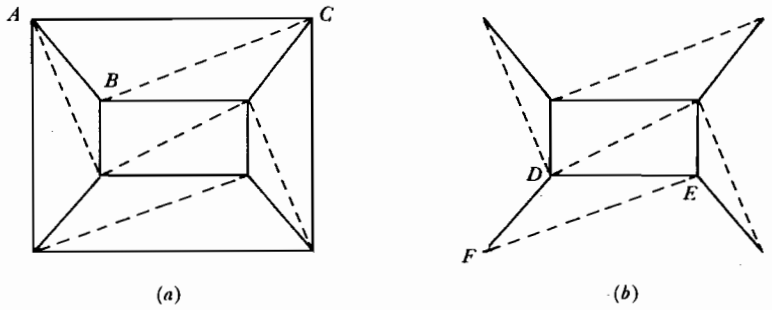


(a) (b)

Figure 1

[1] For a full discussion of the history of this theorem, see Lakatos [1976c].

We 'triangulate' the plane network in the following way: in some polygon of the network which is not already a triangle we draw a diagonal. The effect of this is to increase both E and F by 1 thus preserving the value of $V-E+F$. We continue drawing diagonals joining pairs of points until the figure consists entirely of triangles, as it must eventually (see Figure 2(*a*)). In the triangulated network, $V-E+F$ has the value that it had before the division into triangles, since the drawing of diagonals has not changed it. Some of the triangles have edges on the boundary of the plane network. Of these some, such as ABC, have only one edge on the boundary, while other triangles may have two edges on the boundary. We take any boundary triangle and remove that part of it which does not also belong to some other triangle. Thus, from ABC we remove the edge AC and the face, leaving the vertices A, B, C, and the two edges AB and BC [see Figure 2(*a*)]; while from DEF we remove the face, the two edges DF and FE, and the vertex F [see Figure 2(*b*)]. The removal of a triangle of type ABC decreases E and F by 1, while V is unaffected, so that $V-E+F$ remains the same. The removal of a triangle of type DEF decreases V by 1, E by 2 and F by 1, so that $V-E+F$ again remains the same. By a properly chosen sequence of these operations we can remove triangles with edges on the boundary (which changes with each removal) until finally only one triangle remains, with its three edges, three vertices and one face. For this simple network $V-E+F=3-3+1=1$. But we have seen that by constantly erasing triangles $V-E+F$ was not altered. Therefore in the original plane network $V-E+F$ must equal 1 also, and thus equals 1 for the polyhedron with one face missing. We conclude that $V-E+F=2$ for the complete polyhedron.



(*a*) (*b*)

Figure 2

I think that mathematicians would accept this as a proof, and some of them will even say that it is a beautiful one. It is certainly sweepingly convincing. But we did not *prove* anything in any however liberally interpreted logical sense. There are no postulates, no well-defined underlying logic, there does not seem to be any feasible way to formalize this reasoning. What we were doing was *intuitively showing*

that the theorem was true. This is a very common way of establishing mathematical *facts*, as mathematicians now say. The Greeks called this process *deikmyne* and I shall call it *thought experiment.*

Now is this a proof? Can we give a definition of proof which would allow us to decide at least *practically*, in most cases, if our proof is really a proof or not? I am afraid the answer is 'no'. In a genuine low-level pre-formal theory proof cannot be defined; theorem cannot be defined. There is no method of verification. As a strict logician like Dr Nidditch would surely say, it is – I quote – '*mere* persuasive argumentation, rhetorical appeal, reliance on intuitive insight or worse'.[1]

But if there is no method of verification, there is certainly a method of falsification. We can point out some hitherto unthought of possibilities. For instance assume that we had omitted to stipulate that the polyhedron be simple. We may not have thought of the possibility of the polyhedron having a hole in it (in which case the theorem would be subject to many counterexamples).* Actually Cauchy made this 'mistake'.[2] This is the frequently occurring phenomenon of mathematical theorems being 'stated in a false generality'.

For the sake of a better and simpler illustration let me quote another famous thought experiment with a celebrated falsification. The problem is to find the two points P and Q that are as far apart as possible on the surface or boundary of any triangle. The answer is easy to guess; P and Q are the ends of the longest side. This can easily be proved by the sort of thought experiment which we just used: no axioms, no rules, but convincing force. Let us see:

If one of the points, say P, lies on the *inside* of the triangle, then PQ obviously does not have its maximum length. For on the extension of the line PQ there is obviously a point P' that is further from Q than P is, and that is still inside the triangle. If both P and Q lie on the *boundary* of the triangle, but one of them, say P, is not a vertex, then we can obviously find a nearby point P' on the boundary that is further from Q than the distance PQ. Therefore PQ can be a maximum only if both P and Q are vertices; otherwise it certainly is not. Thus PQ is a side of the triangle and must obviously be the longest side.

It is obvious that the same thought experiment can be accomplished for polygons to 'prove' the following theorem: in order that two points on the surface of a polygon be farthest apart, they must be two of the vertices that are farthest apart.

I think this should be quite convincing. Nevertheless there is an unthought-of possibility which may spoil our pleasure. Apply the same thought-experimental procedure to this figure:

[1] Nidditch [1957], p. 5.

* One such counterexample is the 'picture frame' (Lakatos [1976*c*], p. 19) (*eds.*).

[2] Cauchy [1813].

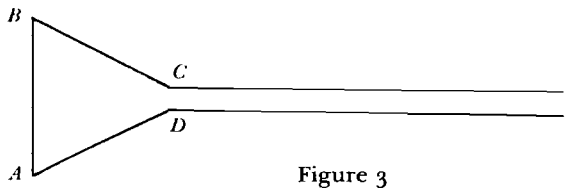


Figure 3

Suppose P and Q lie anywhere inside the figure or on the boundary, even including the possibility that they may be at any of the four vertices A, B, C, D. [Unless PQ is exactly the side AB, a nearby point P' can be found within the figure such that the distance $P'Q$ is greater than the distance PQ.] Just as in the earlier cases, for each pair of points P, Q we can find a nearby pair that are further apart in every case except when the pair is A, B. No pair other than A, B can give a maximum. If we now follow the previous argument strictly, we must conclude that AB is the maximum.

The falsification of our argument ran along the same lines as in the case of Euler's theorem for *all* polyhedra. We thought we showed more than we actually did. In our second case, we showed only that the maximum must be such and such *if the maximum exists at all.* In the case of Euler's theorem we only showed the truth of the theorem for the case where our rubber sheet could really be stretched out to the plane without any holes in it.

I should like to emphasize that the correction of such mistakes can be accomplished on the level of the pre-formal theory, by a new pre-formal theory.

The thought experiments I have just presented constitute only one type of pre-formal proof. There are others, basically different; ones for instance with the rather exciting property that in a certain sense we may say that contrary to the thought experiments we have just considered, they may be verified but not falsified. They give quite an insight into the nature of rules in a pre-formal theory and in pre-formal rigour.*

But now let us turn to axiomatized theories. Up to now no informal mathematical theory could escape being axiomatized. We mentioned that when a theory has been axiomatized, then any competent logician can formalize it. But this means that proofs in axiomatized theories can be submitted to a peremptory verification procedure, and this can be done in a foolproof, mechanical way. Does this mean that for instance if we prove Euler's theorem in Steenrod's and Eilenberg's fully formalized postulate system[1] it is impossible to have any counter-example? Well, it is certain that we won't have any counterexample formalizable in the system [assuming the system is consistent]; but we

* We have been unable to find out what Lakatos had in mind here (*eds.*).

[1] Eilenberg and Steenrod [1952].

have no guarantee at all that our formal system contains the full empirical or quasi-empirical stuff in which we are really interested and with which we dealt in the informal theory. There is no formal criterion as to the correctness of formalization.

Well-known examples of 'falsified' formalizations are (1) the formalization of the theory of manifolds by Riemann, where there is no account of Möbius-strips; (2) the Kolmogorov-axiomatization of probability theory, in which you cannot formalize such intuitive statements as 'every number turns up in the set of natural numbers with the same probability'.* As a final but most interesting example I should mention (3) Gödel's opinion that the Zermelo–Fraenkel and kindred systems of formalized set theory are not correct formalizations of pre-formal set theory as one cannot disprove in them Cantor's continuum-hypothesis.†

I will show with a trivial example how little formalization may add to the demonstrative or convincing force of informal thought experiments. You remember the proof of Euler's theorem? A formalist will certainly reject it. But it won't be easy for him to reject the following 'proof': set up a formal system, with one axiom: A; no rules [except that all axioms are theorems!]. The interpretation of A is Euler's theorem. This system I think complies with the strictest demands of formalism.

Does all this mean that proof in a formalized theory does not add anything to the certainty of the theorem involved? Not at all. [In the informal proof it may turn out that we failed to make some assumption explicit which results in there being a counterexample to the theorem. But, on the other hand, if we manage to *formalize* a proof of our theorem within a formal system, we know that there will never be a counterexample to it which could itself be formalized within the system, as long as that system is consistent.] For instance, if we had a formal proof of Fermat's last theorem, then if our formalized number theory is consistent it would be impossible for there to be a counterexample to the theorem formalizable within the system.

Now we see that if formalization (we shall use this term from now on as essentially having the same meaning as axiomatization) conforms with some informal requirements, such as enough intuitive counter-examples being formalized in it and so on, we gain quite a lot in the value of proofs. But if we try to formalize a pre-formal theory too early, there can be unfortunate results. I wonder what would have happened if probability theory had been axiomatized just in order to supply 'foundations' for probability theory, before the discovery of Lebesgue-measure. Or, to take another example, it is clear that it would have been wasted time and effort to formalize meta-mathematics at the time

* See Renyi [1955] (*eds.*).

† For more detail on this point and references to Gödel's opinions, see this volume, chapter 2 (*eds*).

of finitary illusionism, because later it turned out that the only useful methods must reach not only just beyond finitary tools but even beyond the object-theory in question. In an immaturely axiomatized algebra – axiomatized so as not to allow for complex numbers, say – we could never prove for instance that an equation of nth degree cannot have more than n real roots. Sometimes a well-formed formula of a theory T may be undecidable in the theory, but it may well be decided if suitably interpreted in a different theory, which may not even be an extension of the original theory. It is very difficult to decide in which theory a mathematical statement is really provable: for instance just take some theorems formalizable in the theory of real functions but provable only in the theory of complex functions, or theorems formalizable in measure theory, but provable only in the theory of distributions and so on. Even after a theory has been fruitfully axiomatized, there may arise issues which can bring about a change in axiomatization. This is now going on in probability theory. Axiomatization is a big turning point in the life of a theory, and its importance surpasses its impact on proofs; but its impact on proofs is immense in itself. While in an informal theory there really are unlimited possibilities for introducing more and more terms, more and more hitherto hidden axioms, more and more hitherto hidden rules in the form of new so-called 'obvious' insights, in a formalized theory imagination is tied down to a poor recursive set of axioms and some scanty rules.

Let me finally turn to the third part of my classification: to *post*-formal proofs. Here I shall just make a few programmatic remarks.

Two types of post-formal proofs are well-known. The first type is represented by the Duality Principle in Projective Geometry which says that any properly-worded valid statement concerning incidences of points and lines on a projective plane gives rise to a second valid statement when the words 'point' and 'line' are interchanged. For instance if the statement 'Any two distinct lines in the same plane determine a unique point' is valid, then so is the statement 'Any two distinct points in the same plane determine a unique line'. But then in proving the second statement we use a theorem of the system and another theorem, a meta-theorem, which we cannot specify, and still less prove, without specifying the concepts of provability in the system, theorem in the system and so on. This meta-theorem which we use like a lemma in our proof of an informal mathematical theory is not just about lines or points but about lines, points, provability, theorem-hood and so on. Although projective geometry is a fully axiomatized system, we cannot specify the axioms and rules used to prove the Principle of Duality, as the meta-theory involved is informal.

The second class of post-formal proofs I should mention is the class of proofs of undecidability. As students of mathematical logic know,

in the last few years it has turned out that formal proofs really prove much more than we want them to prove. Namely, to put it very roughly indeed, axioms in the most important mathematical theories implicitly define not just one, but quite a family of structures. For instance, Peano's axioms may be satisfied not only by our familiar natural numbers, but by some quite queer structures, Skølem's functions, which are far from being isomorphic with the set of natural numbers. Thus it turns out that when we fight hard to prove an arithmetical theorem, we prove at the same time some theorem in this other absolutely unintended structure. Now there are always statements, which are true in one structure but false in the other. Such statements are undecidable in the common formal structure. Are we helpless in such a situation? To see the point better, let us take a concrete, though hypothetical example. If we could prove that Fermat's theory is undecidable, then are we forever helpless to say anything about the truth of Fermat's theorem? Not at all. We may again call informal reasoning to our help, and try to operate informally *only* in the intended model. A concrete example of this is Gödel's proof [that his undecidable sentences are *true* (i.e. true in the standard model)]. But such post-formal proofs are certainly informal and so they are subject to falsification by the later discovery of some not-thought-of possibility.

Now at the present stage of our mathematical knowledge undecidable sentences occur only in rather artificial examples and do not affect the bulk of mathematics. But this situation may turn out similar to the case of transcendental numbers, which occurred first rather as exceptions and later turned out to be the more general case. So post-formal methods may gain in importance as undecidability encroaches more and more on mathematics.

And now a brief summary. We saw that mathematical proofs are essentially of three different types: pre-formal; formal; post-formal. Roughly the first and third prove something about that sometimes clear and empirical, sometimes vague and 'quasi-empirical' stuff, which is the real though rather evasive subject of mathematics. This sort of proof is always liable to some uncertainty on account of hitherto unthought-of possibilities. The second sort of mathematical proof is absolutely reliable; it is a pity that it is not quite certain – although it is approximately certain – what it is reliable about.