Draft of a paper published in *Studies in History and Philosophy of Science*

The Nature of Mathematical Explanation

Carlo Cellucci

University of Rome La Sapienza, Department of Philosophical and Epistemological Studies, Rome, Italy

Abstract

Although in the past three decades interest in mathematical explanation revived, recent literature on the subject seems to neglect the strict connection between explanation and discovery. In this paper I sketch an alternative approach that takes such connection into account. My approach is a revised version of one originally considered by Descartes. The main difference is that my approach is in terms of the analytic method, which is a method of discovery prior to axiomatized mathematics, whereas Descartes's approach is in terms of the analytic-synthetic method, which is a heuristic pattern in already axiomatized mathematics.

1. The Aristotle-Pólya tradition

In two recent papers (Cellucci, 2005, 2006b) I challenged a claim of a long tradition, from Aristotle to Pólya, according to which there is a sharp distinction between two kinds of reasoning, demonstrative reasoning, that is, the deductive derivation of conclusions from premisses which are primitive and true in some sense of 'true', and non-demonstrative reasoning, that is, the non-deductive (inductive, analogical, etc.) derivation of conclusions from premisses which are not known to be true but are only 'accepted opinions', in the sense of Aristotle's *éndoxa*. The former is essentially superior to the latter since it is cogent, whereas the latter is not cogent.

This claim is untenable because, by Gödel's incompleteness results, knowing that the premisses of demonstrative reasoning are true – either in Gödel's strong sense that they express properties of objects independent of us or in Hilbert's weak sense that they are consistent – is generally impossible. Thus premisses are only 'accepted opinions' in the sense explained above, and so have the same status as the premisses of non-demonstrative arguments (Cellucci, 2005, pp. 158-159). Moreover, deductive inferences can be justified only in the same 'external' non-absolute sense as non-deductive inferences (Cellucci, 2006b, pp. 231-232).

Here I will consider another claim of the Aristotle-Pólya tradition, according to which, within demonstrative reasoning, there is a sharp distinction between two kinds of reasoning, the reasoning which shows why something is the case and the reasoning which only shows that something is the case. The former is essentially superior to the latter since it shows the cause, or reason, of the thing, thus providing an explanation of it, whereas the latter does not.[1]

According to the Aristotle-Pólya tradition, 'there are proofs and proofs, there are various ways of proving' (Pólya, 1962-65, II, p. 126). Specifically, there is a sharp

[1] The standard English translation for Aristotle's 'aitia' is 'cause', which however has strong connotations. In what follows I will use 'reason' in place of 'cause' when possible.

distinction between the reasoning which 'shows why something is the case' and the reasoning which 'does not show why something is the case but only shows that something is the case' (Aristotle, *An. Post.*, A 13, 78 a 37). The reasoning which shows why something is the case gives us the cause of the thing, for 'to know why something is the case is to know it through its cause' (ibid., A 6, 75 a 35). Giving us the cause of the thing, this kind of reasoning gives us an explanation of it, and so a full understanding of it, for 'we think we understand something absolutely' only 'when we think we know the cause on which the thing depends' (ibid., A 2, 71 b 9-11). On the contrary, the reasoning which 'shows that something is the case but does not state why it is the case', generally 'does not tell us its cause' (ibid., A 13, 78 b 14-15). Thus it gives no explanation of it and hence no full understanding of it. The reasoning which shows why something is the case is essentially superior to the reasoning which only shows that something is the case, for 'we only have scientific knowledge about something when we know its cause' (ibid., A 2, 71 b 30-31).

Also this claim of the Aristotle-Pólya tradition is untenable. However, it is not untenable in itself but only in conjunction with the basic assumption of such tradition that the method of mathematics is the axiomatic method and, specifically, that both the reasoning which shows why something is the case and the reasoning which only shows that something is the case are given by that method.

I will call 'view of the Aristotle-Pólya tradition on explanation' the claim of the Aristotle-Pólya tradition on explanation plus the basic assumption of such tradition.

The view of the Aristotle-Pólya tradition on explanation is untenable because it is self-contradictory. The basic assumption of such tradition implies that both the reasoning which shows why something is the case and the reasoning which only shows that something is the case depend on the very same ultimate premisses. For in the axiomatic method all demonstrations of a given mathematical theory ultimately start from the principles of that theory, so from the very same principles.[2] Thus all demonstrable propositions of a given mathematical theory have the same ultimate reason, that is, the principles, hence the same explanation. Therefore the reasoning which shows why something is the case is the same as the reasoning which only shows that something is the case. This contradicts the claim of the Aristotle-Pólya tradition on explanation, that there exists a sharp distinction between the reasoning which shows why something is the case and the reasoning which only shows that something is the case.

Against the conclusion that the basic assumption of the Aristotle-Pólya tradition implies that the reasoning which shows why something is the case is the same as the reasoning which only shows that something is the case, it might be objected that one may distinguish between two kinds of demonstrations, direct demonstrations and reductio ad absurdum demonstrations. Direct demonstrations are explanatory whereas reductio ad absurdum demonstrations are non-explanatory. Such objection, however, does not hold because, on the one hand, there are direct demonstrations, such as the long demonstrations-as-computations of finitary mathematics, that are non-explanatory since, being mere computations, they don't show the reason of the result. On the other hand, there are reductio ad absurdum demonstrations, such as the demonstration of the fact that the square root of 2 is not rational considered in Section 10 below, that are explanatory since they show the reason of the result. So the distinction between explanatory and non-explanatory demonstrations cannot amount to that between direct demonstrations and reductio ad absurdum demonstrations. Moreover, usually a reductio ad absurdum demonstration can be converted into a direct demonstration based on essentially the same idea.

2. The Popper-Balacheff tradition

In addition to the Aristotle-Pólya tradition on explanation there is another, more radical and more recent tradition, from Popper to Balacheff, which claims that there are no two different kinds of demonstrative reasoning but only one kind, the reasoning which only

[2] In this paper I use the expression 'demonstration' instead of 'proof' to include both arguments based on the axiomatic method and arguments based on the analytic method.

shows that something is the case. To give an explanation of something is to deduce it from given principles, since the principles can be viewed as the causes, or reasons, of the thing. Therefore the reasoning which shows why something is the case is the same as the reasoning which only shows that something is the case.

This view is often credited to Hempel-Oppenheim (1948) but actually goes back to Popper (1934). Balacheff (1987) extended it to mathematics, but similar statements can be found in other authors.

According to the Popper-Balacheff tradition on explanation, 'we call proof an explanation accepted by a given community at a given time', although 'only explanations of a special form can be accepted as proofs', that is, 'sequences of sentences organized by well defined rules: a sentence is either known to be true or is derived from previous ones by a deduction rule belonging to a well defined set of rules' (Balacheff, 1987, pp. 147-148). A 'fully explicit explanation always consists in pointing out the logical derivation (or the derivability) of the explicandum from the theory strengthened by some initial conditions' (Popper, 1994, pp. 76-77). Thus 'every explanation consists of a logical deductive inference whose premisses consist of a theory and some initial conditions, and whose conclusion is the explicandum' (ibid., p. 77). This is 'the concept of causal explanation' (ibid., p. 76).

The Popper-Balacheff tradition on explanation takes as its own viewpoint what is a perhaps unintended consequence of the view of the Aristotle-Pólya tradition on explanation: the reasoning which shows why something is the case is the same as the reasoning which only shows that something is the case. Thus the Popper-Balacheff tradition makes a virtue of what is actually a defect of the Aristotle-Pólya tradition.

The view of the Popper-Balacheff tradition on explanation has been very influential and indeed, for some time, has been the 'received view'. Nevertheless it is untenable. For to deduce something from given principles is not a necessary nor a sufficient condition for giving an explanation of it.

1) *To deduce something from given principles is not a necessary condition for giving an explanation of it*. For instance, consider the following fact:

$$\text{(A)} \quad \frac{1}{4}+\frac{1}{4^2}+\frac{1}{4^3}\ldots=\frac{1}{3}\,.$$

A demonstration of (A) is given by the diagram shown in Figure 2.1.

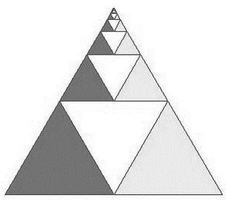

Figure 2.1

The biggest white triangle is $\frac{1}{4}$ the whole triangle, the white triangle immediately smaller is $\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{4^2}$ the whole triangle, the white triangle immediately smaller is $\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{4^3}$ the whole triangle, and so on *ad infinitum*. Thus the series of white triangles represents the series $\frac{1}{4}+\frac{1}{4^2}+\frac{1}{4^3}\ldots$. Moreover, by construction, the series of white triangles is $\frac{1}{3}$ the whole triangle. That establishes (A).

The diagram shows a possible reason for (A) and thus gives an explanation of it. But such explanation is not an explanation in the sense of the Popper-Balacheff tradition, since it does not deduce (A) from given principles. It involves an induction rather than a deduction, since it infers (A) from the finite, and indeed very small, number of triangles actually shown in the diagram. Moreover it involves recovering visual data from a diagram, an operation which could hardly be viewed as deductive. Therefore, to deduce (A) from given principles is not a necessary condition for giving an explanation of it.

2) *To deduce something from given principles is not a sufficient condition for giving an explanation of it*. For instance, consider:

(B) the Pythagorean theorem,

and a demonstration of (B) from the axioms of set theory.

Such demonstration gives an explanation of (B) in the sense of the Popper-Balacheff tradition, since it deduces (B) from given principles. But it does not give an explanation of (B), since it deduces (B) from very general principles which have no special connection with (B). Therefore, to deduce (B) from given principles is not a sufficient condition for giving an explanation of it.

The fact that to deduce something from given principles is not a necessary nor sufficient condition for giving an explanation of it, shows that the view of the Popper-Balacheff tradition on explanation is untenable.

On the other hand, as we have seen, under the basic assumption of the Aristotle-Pólya tradition that the method of mathematics is the axiomatic method, the reasoning which shows why something is the case is the same as the reasoning which only shows that something is the case. Therefore the view of the Aristotle-Pólya tradition on explanation collapses into that of the Popper-Balacheff tradition. Since the latter is untenable, this provides further evidence for the fact that the view of the Aristotle-Pólya tradition on explanation is untenable.

Such view is thus untenable not only because it has a consequence that contradicts it – the reasoning which shows why something is the case is the same as the reasoning which only shows that something is the case – but also because such consequence is incompatible with the fact that to deduce something from given principles is not a necessary nor sufficient condition for giving an explanation of it.

3. The Aristotle-Pólya tradition and set theory

As it is implicit in what we have already said, the view of the Aristotle-Pólya tradition on explanation has an even stronger consequence: to give an explanation of a mathematical fact is to deduce it from the principles of set theory. This depends on the fact that any present branch of mathematics is reducible to set theory, in the sense that its primitive notions are definable in terms of the primitive notions of set theory, and its theorems are demonstrable from the axioms of set theory. Therefore, from the viewpoint of the Aristotle-Pólya tradition, the principles of set theory provide an explanation of any mathematical truth.

Such consequence is actually drawn by Gödel who, in conversations with Mehlberg, opposes the view that 'an axiomatization of classical mathematics on a logical basis', that is, in terms of systems like *Principia Mathematica*, 'or in terms of set theory', is 'a foundation of the relevant mathematics' (Mehlberg, 1960, p. 397). According to Gödel, 'the role of these alleged foundations is rather comparable to the function discharged, in physical theory, by explanatory hypotheses'. The 'actual function of postulates or axioms occurring in a physical theory is to explain the phenomena described by the theorems of this system rather than to provide a genuine 'foundation' for such theorems'. Similarly, the 'so-called logical or set-theoretical foundation for number-theory, or any other well established mathematical theory, is explanatory, rather than really foundational, exactly as in physics'.

Against the claim that the view of the Aristotle-Pólya tradition on explanation has the consequence that to give an explanation of a mathematical fact is to deduce it from the principles of set theory, it might be objected that it is incompatible with Aristotle's theory of genera, according to which each science 'is concerned with a single genus' (Aristotle, *An. Post.*, A 28, 87 a 38). Thus 'arithmetical demonstration always is of the genus with which the demonstration is concerned, and so do all other demonstrations' (ibid., A 7, 75 b

7-8). Therefore 'it is impossible to demonstrate a proposition of a certain science by another science' (ibid., A 7, 75 b 14). For instance, 'it is impossible to demonstrate something geometrical by arithmetic' (ibid., A 7, 75 a 39). Conversely, it is impossible to demonstrate something arithmetical by geometry, for instance, that 'two cubes make a cube' (ibid., A 7, 75 b 13-14). Thus no principles can possibly exist 'from which everything will be demonstrated' (ibid., A 32, 88 a 37). Therefore one cannot demonstrate all mathematical facts from the principles of set theory.

To such objection one may answer that, while the assumption of Aristotle's theory of genera, that demonstrations of different mathematical theories must start from different principles, was reasonable at Aristotle's time, it is no longer reasonable today when axiomatic set theory has been successful at least to the extent that it permits the derivation of modern mathematics. Admittedly, the theory of categories seems to trascend the concept of set, as becomes apparent, for instance, by the self-applicability of categories. But, as Gödel himself points out, it does not seem 'that anything is lost from the mathematical content of the theory if categories of different levels are distinguished' (Gödel, 1986-2002, II, p. 258, footnote 12). Thus the theory of categories reduces to the concept of set. Therefore the assumption of Aristotle's theory of genera is no longer justified.

But such answer is unnecessary since the above objection is not justified even from the viewpoint of Aristotle's theory of genera. For while maintaining that 'it is impossible to demonstrate a proposition of a certain science by another science' Aristotle adds: 'except when' such sciences 'are so related to one another that one falls under the other, as for instance optics is related to geometry and harmonics to arithmetic' (ibid., A 7, 75 b 14-17). Now, since axiomatic set theory permits the derivation of modern mathematics, all present mathematical theories can be viewed as applied set theory, and hence as falling under set theory. Therefore, even from the viewpoint of Aristotle's theory of genera, one is justified in saying that the view of the Aristotle-Pólya tradition on explanation has the consequence that to give an explanation of a mathematical fact is to deduce it from the principles of set theory.

4. Explanation and Gödel's incompleteness theorems

That the view of the Aristotle-Pólya tradition on explanation is untenable depends on the fact that the basic assumption of such tradition, that the method of mathematics is the axiomatic method, is unfounded. A method can be said to be the method of mathematics if only if all mathematical truths can be obtained by means of it. Now, by Gödel's first incompleteness theorem, for any given mathematical theory satisfying certain minimal conditions, there are truths in the language of that theory which are unprovable from the axioms of the theory. In particular, for any formulation of set theory satisfying certain minimal conditions, there are elementary arithmetical truths, indeed finitary ones, which are unprovable from the axioms of the theory. Therefore the axiomatic method cannot be said to be the method of mathematics, in the sense stated above.

This was already stressed by Post who (in 1941) expressed his 'continuing amazement that ten years after Gödel's remarkable achievement current views on the nature of mathematics are thereby affected only to the point of seeing the need of many formal systems, instead of a universal one'. Post considered 'inevitable that these developments will result in a reversal of the entire axiomatic trend of the late 19th and early 20th centuries', and that 'postulational thinking will then remain as but one phase of mathematical thinking' (Post, 1994, p. 378).

The basic assumption of the Aristotle-Pólya tradition that the method of mathematics is the axiomatic method is an expression of Aristotle's theory of genera. For, if demonstrations of different mathematical theories must start from different principles, then all demonstrations of results of a given mathematical theory must start from the principles of that theory, and so must be based on the axiomatic method. But Aristotle's theory of genera is untenable because, by Gödel's first incompleteness theorem, not all demonstrations of results of a given mathematical theory satisfying certain minimal conditions can use only principles of that theory. Therefore, that the view of the Aristotle-

Pólya tradition on explanation is untenable is a corollary of the fact that Aristotle's theory of genera is untenable.[3]

5. An alternative approach

Since the views on explanation of both the Aristotle-Pólya tradition and the Popper-Balacheff tradition are untenable, the question naturally arises: Is there any more adequate view on mathematical explanation?

As we have already pointed out, the claim of the Aristotle-Pólya tradition on explanation – that there is a sharp distinction between the reasoning which shows why something is the case and the reasoning which only shows that something is the case – is not untenable in itself but only in conjunction with the basic assumption of such tradition that the method of mathematics is the axiomatic method. It is untenable since the basic assumption of that tradition is untenable.

There is nothing, however, against the claim of the Aristotle-Pólya tradition on explanation in itself. Indeed, it seems desirable to retain it, since the distinction between the reasoning which shows why something is the case and the reasoning which only shows that something is the case is an important feature of mathematical activity. Mathematicians not only look for demonstrations of new results, but often look for new demonstrations of results for which demonstrations are already available. They do so mainly because they are not satisfied with the existing demonstrations as they does not seem to show why the result holds. (For further reasons, see Dawson, 2006).

Of course, if we retain the claim of the Aristotle-Pólya tradition on explanation, then we must drop the basic assumption of such tradition that the method of mathematics is the axiomatic method. In particular, we must drop the assumption that both the reasoning which shows why something is the case and the reasoning which only shows that something is the case are given by the axiomatic method.

6. The analytic method

Alternatively we may assume that, while the reasoning which only shows that something is the case is given by the axiomatic method, the reasoning which shows why something is the case, or explanatory reasoning, is given by the analytic method.

The analytic method is the method by which, to solve a mathematical problem, we formulate a hypothesis that is a sufficient condition for its solution. The hypothesis is obtained from the problem, and possibly other data, by some non-deductive inference: inductive, analogical, etc.. (Several kinds of non-deductive inferences by which hypotheses can be obtained are discussed in Cellucci, 2002, pp. 235-295). The hypothesis must not only be a sufficient condition for the solution of the problem but must also be plausible, that is, compatible with the existing data, in the sense that, comparing the reasons for and the reasons against the hypothesis on the basis of the existing data, the reasons for prevail over those against. Plausibility is distinct from probability, as it appears from the fact that induction from a single instance often leads to hypotheses that are plausible but whose probability is zero when the number of possible instances is infinite. Since hypotheses are merely plausible, their status is similar to that of Aristotle's *éndoxa*.

[3] Incidentally, the same applies to the failure of Hilbert's foundational program, which depends on two assumptions. The first assumption is that any mathematical theory must satisfy the 'requirement of the purity of methods of demonstration', according to which any truth in the language of that theory must be demonstrated using only principles of the theory, thus it is fixed once for all 'what axioms, assumptions or means are necessary for demonstrating any truth' (Hilbert, 1962, p. 125). The second assumption is that any sentence of finitary mathematics demonstrable in infinitary mathematics must be already demonstrable in finitary mathematics, which ensures that the axioms of infinitary mathematics 'can never lead to a provably false result' (Hilbert-Bernays, 1968-70, I, p. 44). These assumptions are expressions of Aristotle's theory of genera. Therefore the failure of Hilbert's program is a corollary of the fact that Aristotle's theory of genera is untenable.

Being merely plausible, a hypothesis is in turn a problem that must be solved, and will be solved in the same way, that is, formulating another hypothesis that is a sufficient condition for its solution, it is obtained from the previous hypothesis, and possibly other data, by some non-deductive inference, and must be plausible. And so on, *ad infinitum*. Therefore, the solution of a mathematical problem is a potentially infinite process. (Further details on the analytic method can be found in Cellucci, 1998, pp. 270-304; 2002, pp. 174-182; 2005).

This depends on the fact that hypotheses must be plausible. In the process of investigating the plausibility of a hypothesis, new facts may emerge which show that certain previous hypotheses, until then considered as plausible, are no longer plausible and so must be revised or even dropped. Therefore, while in the axiomatic method principles are given once for all, in the analytic method hypotheses are provisional and will be ultimately replaced by other ones. The axiomatic method is what results from the analytic method when the hypotheses formulated at a certain stage are considered as absolute starting points for which no justification is provided. Thus the axiomatic method is an unjustified truncation of the analytic method. Since axioms are hypotheses for which no justification is provided, they are mere conventions and, as Plato – the harshest critic of the axiomatic method – pointed out, how can one 'imagine that such a fabric of convention can ever become science?' (Plato, *Rep*., VII 533 c 4-5).

There is also another difference between the analytic and the axiomatic method. While in the axiomatic method axioms serve to demonstrate all demonstrable propositions of a given mathematical theory and hence do not depend on the specific proposition being considered, in the analytic method hypotheses are used to solve a specific problem and depend on it. Thus distinct problems will generally require distinct hypotheses.

Against this conclusion one might argue, for instance, that Peano's axioms are plausible, so they may be explanatory; but all theorems of number theory can be explained using only Peano's axioms; hence distinct problems within the same branch of mathematics will not generally require distinct hypotheses. But this argument depends on the assumption that all theorems of number theory can be explained using only Peano's axioms, an assumption which is refuted by Gödel's first incompleteness theorem.

Not only distinct problems will generally require distinct hypotheses, but hypotheses are not uniquely determined by problems. One may use distinct hypotheses to solve the very same problem, since the same thing may have distinct reasons, thus distinct explanations.

As Hanson puts it, 'suppose that Galileo's carriage strikes a pedestrian in the darkened streets of Padua' (Hanson, 1958, p. 52). Then 'the cause of death' of the pedestrian 'might have been set out by a physician as multiple haemorrhage, by the barrister as negligence on the part of the driver, by a carriage-builder as a defect in the brakeblock construction, by a civic planner as the presence of tall shrubbery at that turning (ibid., p. 54).

That the very same thing may have distinct reasons, thus distinct explanations, depends on the fact that problems have many sides, so they can be seen from distinct perspectives, each of which suggests a distinct hypothesis, thus a distinct explanation. Any hypothesis establishes a connection between the problem and a distinct body of knowledge, thus revealing a new essential aspect of the problem.

While there are basic differences between the analytic and the axiomatic method, there is no opposition between reductio ad absurdum demonstrations and demonstrations based on the analytic method. Indeed reductio ad absurdum is a form of the analytic method, although of a very special kind (see Cellucci, 2002, pp. 198-202).

From antiquity to the Eighteenth century the analytic method has been viewed as the method for solving mathematical problems. Afterwards mathematicians continued using it but, owing to the influence of the axiomatic ideology, this fact is not generally acknowledged.

For instance, it is usually claimed that Wiles and Taylor solved Fermat's Problem. Such view depends on the axiomatic ideology. In fact it was Ribet who solved Fermat's Problem showing that the Taniyama-Shimura's hypothesis – any elliptic curve on rational numbers is a modular form – is a sufficient condition for its solution. Such hypothesis is not only a sufficient condition for the solution of the problem but, when Ribet used it, was also plausible, that is, compatible with the existing data. However, the Taniyama-Shimura's hypothesis was in turn a problem to be solved, and was solved by Wiles and Taylor in the same way, that is, using hypotheses that were a sufficient condition for its solution and were plausible. Therefore Wiles and Taylor solved the Taniyama-Shimura's problem rather than Fermat's Problem. The solution of Fermat's Problem is a typical example of use of the analytic method.

There is no point in objecting that, at the time when Ribet showed that the Taniyama-Shimura's hypothesis is a sufficient condition for the solution of Fermat's Problem, the Taniyama-Shimura's hypothesis was merely a hypothesis, it had not been proved yet, so Ribet's proof cannot be considered a solution of Fermat's Problem. For the Wiles and Taylor proof of the Taniyama-Shimura's hypothesis ultimately depends on the axioms of set theory, which are merely a hypothesis, they have not been proved yet. Thus, from the viewpoint of this objection, one would have to conclude that the Wyles and Taylor proof cannot be considered a solution of Fermat's Problem.

7. Explanation and the analytic method

In terms of the analytic method it is straightforward to state what an explanation is. A hypothesis provides an explanation of a problem if it plays an essential role in solving it, in the sense that it reveals an aspect of the problem that is essential to a solution of the problem. In that sense an explanatory hypothesis is strictly connected with the problem.

Thus, although the Pythagorean theorem can be proved from the axioms of set theory, the latter do not provide an explanation of the problem posed by the Pythagorean theorem since they play no essential role in solving the problem, in the sense explained above.

Identifying explanatory reasoning with the one given by the analytic method, we escape the difficulty of the axiomatic method that all demonstrable propositions of a given mathematical theory have the same ultimate reason – the principles – thus the same explanation. For, as we have seen, in the analytic method distinct problems will generally require distinct hypotheses, so they will have distinct reasons – the hypotheses – thus distinct explanations. This prevents the notion of explanation from collapsing into that of the Popper-Balacheff tradition.

Moreover, the analytic method need not satisfy Aristotle's theory of genera. For hypotheses need not be of the same genus as the problem, they can be of any genus. Any mathematical theory is an open system, that is, a set of problems whose solution will generally require hypotheses which are not given once for all but may involve concepts and methods from other mathematical theories (cf. Cellucci, 1998, pp. 309-346; 2000; 2002, pp. 203-211). Thus Gödel's incompleteness theorems not only do not refute but even support this notion of explanation.

Since the analytic method is a method for solving mathematical problems, the fact that explanatory reasoning is given by it implies that such reasoning is the same as the one by which we solve problems. Only this kind of reasoning shows the reason of the problem and hence is explanatory. For, starting from the problem, and possibly other data, it introduces a hypothesis that is strictly connected with the problem, in the sense explained above.

Unlike principles in the axiomatic method, which are used to establish any proposition of a given mathematical theory, in the analytic method hypotheses serve only to solve a specific problem, so they are aimed at the problem and are strictly connected with it. This depends on the fact that they are obtained from the problem, although possibly in addition to other data, by non-deductive inferences.

8. Heuristics vs. essence

The assertion that an explanatory hypothesis is strictly connected with the problem may be open to misunderstanding, so the following remarks may be useful.

1) That an explanatory hypothesis is strictly connected with the problem is not to be meant in Peirce's sense that an explanatory hypothesis adds 'the least to what has been observed' (Peirce, 1931-58, 6.477). For in the analytic method hypotheses add much to what has been observed, since they are obtained from it by non-deductive inferences, which go well beyond the data.

2) That an explanatory hypothesis is strictly connected with the problem is not to be meant in Aristotle's sense that an explanatory hypothesis reveals the essence of the thing being explained.

According to Aristotle, to know why something is the case is to know it through its cause, and 'the cause is the substance and the essence' (Aristotle, *Metaph.,* A 3, 983 a 27-28). An explanatory hypothesis reveals properties of the thing being explained that 'belong to their subject as elements in its essential nature' (Aristotle, *An. Post.*, A 4, 73 a 34-35).

Aristotle's view is supported by Steiner who claims that 'to explain the behavior of an entity, one deduces the behavior from the essence or nature of the entity', except that, instead of 'essence', Steiner prefers to speak of 'characterizing property', saying that 'an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem' (Steiner, 1978, p. 143).

The trouble with the Aristotle-Steiner view is however that characterizing properties are scarce. For instance, in terms of this view, a demonstration of a theorem of number theory makes reference to a characterizing property of the structure of natural numbers, but there is no such characterizing property. To begin with, Peano's second-order axioms have (non-full) non-standard models. Moreover, not all their full models are isomorphic but only those belonging to the same model of set theory, so they are categorical only relative to a given model of set theory (cf. Cellucci, 2007, pp. 106-107). Thus they fail to characterize the structure of natural numbers.

3) That an explanatory hypothesis is strictly connected with the problem is to be meant in the sense that it reveals an aspect of the problem that is essential to its solution, where such aspect does not necessarily consist in a characterizing property of an entity or structure mentioned in the theorem. It may consist in a connection between such entity or structure and other entities or structures not mentioned in the problem. An elementary but significant example is provided by the above solution of (A), which uses a hypothesis about triangles although (A) is a problem concerning a certain series of rational numbers.

What is crucial in a mathematical explanation is not a characterizing property of an entity or structure mentioned in the theorem, but rather the heuristic value of the hypothesis, its effectiveness as a means of discovery. While characterizing properties are properties that entities or structures mentioned in the problem are supposed to possess, the heuristic value of a hypothesis may depend on entities or structures not mentioned in the problem.

9. Descartes on explanation

The claim that explanatory reasoning is the same as the analytic method, that is, the method by which we solve mathematical problems, is implicit in Descartes.

According to Descartes, 'analysis shows the true way by means of which a thing has been discovered methodically, and permits seeing how the effects depend on the causes' (Descartes, 1996, IX-1, p. 121). Here 'the causes from which I deduce' the effects 'do not serve so much to prove these effects as to explain them' (ibid., VI, p. 76). On the contrary synthesis – that is, the axiomatic method – 'clearly proves what is contained in its conclusions', but 'does not fully satisfy the minds of those who are eager to learn' (ibid., IX-1, p. 122). For it does not 'show adequately to the mind why these things should be so and how they were discovered' (ibid., X, p. 375). Therefore 'there is a great difference between proving and explaining' (ibid., II, p. 198).

Descartes's assertion that there is a great difference between proving and explaining corresponds to the above assertion that there is a sharp distinction between the reasoning which shows why something is the case and the reasoning which only shows that something is the case.

Moreover, Descartes's assertion that, while the synthetic method does not show adequately to the mind why things should be so and how they were discovered, in the analytic method the causes from which the effects are deduced do not serve so much to prove these effects as to explain them, corresponds to the above assertion that, while the reasoning which only shows that something is the case is given by the axiomatic method, explanatory reasoning is given by the analytic method.

There is, however, a basic difference between Descartes's approach and the present one. By 'analytic method' Descartes does not mean the method described above, which is a method of discovery prior to axiomatized mathematics, but rather the 'analytic-synthetic method' (going back to Aristotle, *Eth. Nic*., Γ 3, 1112 b 15-27), which is merely a heuristic pattern in already axiomatized mathematics.

In fact Descartes says that by analysis 'we gradually reduce involute and obscure propositions to simpler ones' (Descartes, 1996, X, p. 379). We thus ultimately arrive at the simplest ones of all, the principles, 'which can be intuited first and per se, independently of any others' (ibid., X, p. 383). Then, 'from the intuition of the simplest ones of all, we will try to ascend through the same steps to a knowledge of all the others' (ibid., X, p. 379). On the other hand, 'all the others can be perceived only by deducing them from those' (ibid., X, p. 383). Thus 'the things which are put forward first', that is, the principles, 'must be known without the aid of what comes later, and the things which follow must be arranged in such a way that they are demonstrated solely by the things which come before them' (ibid., IX-1, p. 121).

This is a very neat statement of the analytic-synthetic method. While the analytic method is a method for discovering hypotheses to solve problems, the analytic-synthetic method is a method for discovering proofs of theorems from given principles. Principles are so important for Descartes that he sharply criticizes Galileo since he, 'without first considering the first causes in nature, merely looked for the reasons of some special effects, and thus constructed without foundation' (ibid., II, p. 380).

10. Examples

It may be useful to illustrate the claim that explanatory reasoning is given by the analytic method by some elementary although historically significant examples.

Problem I. Show the Pythagorean theorem.

To solve this problem, given a right triangle, construct the square on its shorter leg. Make a copy of the triangle and construct the square on its longer leg. Put the two resulting figures together as shown in the diagram on the left in Figure 10.1. Then slide up both the triangle on the left and the triangle on the right, forming a quadrilateral as shown in the diagram on the right in Figure 10.1.

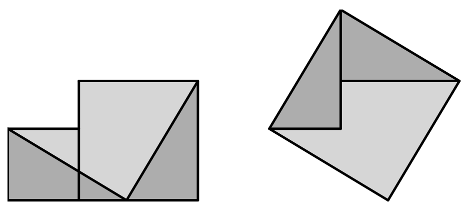

Fig. 10.1

Such quadrilateral equals the sum of the squares on the two legs of the triangle. Therefore, to show that the square on the hypotenuse equals the sum of the squares on the two legs, we need only show that the quadrilateral is the square on the hypotenuse. To show that we state the following hypothesis:

(A) The three interior angles of a triangle are equal to two right angles.

Hypothesis (A) is capable of solving the problem since it implies that the two non-right angles of the right triangles add up to a right angle. Thus each angle of the quadrilateral is a right angle, so the quadrilateral is a square – the square on the hypotenuse of the right triangle. Since hypothesis (A) plays an essential role in solving Problem I, it explains why the square on the hypotenuse equals the sum of the squares on the two legs.

Problem II. Show that if, starting from a right triangle, we draw a semicircle on the hypotenuse and on each of its legs, as shown in Fig. 10.2,

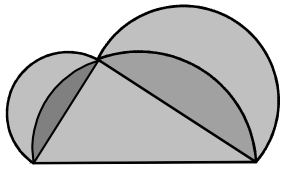

Fig. 10.2

then the right triangle equals the sum of the two lunules which arch over the legs – a generalization of a problem originally solved by Hippocrates of Chios.

To solve this problem, notice that the semicircle on the hypotenuse consists of the right triangle plus two segments of circle, and the semicircles on the legs consist of the lunules plus those two segments of circle. Thus, to show that the right triangle equals the sum of the two lunules, we need only show that the semicircle on the hypotenuse equals the sum of the two semicircles on the legs. Since, by the solution of Problem I, the square on the hypotenuse equals the sum of the squares on the two legs, to show this fact we state the following hypothesis:

(B) Circles are as the squares on their diameters.

Hypothesis (B) is capable of solving the problem since it, together with the solution of Problem I, implies that the semicircle on the hypotenuse equals the sum of the two semicircles on the legs. Since hypothesis (B) plays an essential role in solving Problem II, it explains why the right triangle equals the sum of the two lunules which arch over its legs.

Problem III. Show that the square root of 2 is not rational, that is, there are no integers *a* and *b* such that $a^2 = 2b^2$.

To solve this problem notice that, if every positive integer ≥ 2 has a unique prime factorization, then, in the prime factorization of a^2, the prime 2 will have an exponent double the exponent it has in the prime factorization of *a*, thus an even exponent. Similarly, in the prime factorization of b^2, the prime 2 will have an exponent double the exponent it has in the prime factorization of *b*, thus again an even exponent. Hence in the prime factorization of $2b^2$ the prime 2 will have an odd exponent. But then $a^2 = 2b^2$ is impossible. Therefore we state the following hypothesis:

(C) Every positive integer ≥ 2 has a unique prime factorization.

Hypothesis (C) is capable of solving the problem since it implies that *a*, *b*, a^2, b^2 have unique prime factorizations. Since hypothesis (C) plays an essential role in solving Problem III, it explains why the square root of 2 is not rational.

11. Explanation and generality

It is widely held that there is a strict connection between explanation and generality: the most explanatory proof is the most general.

This view goes back to Aristotle who claims that 'if demonstration is syllogism that gives the cause and the reason why, and if the universal is more causative', then 'universal demonstration is better. For it is rather it which gives the cause and the reason why' (Aristotle, *An. Post.*, A 24, 85 b 23-24, 26-27). The 'universal is precious because it makes clear the cause' (ibid., A 31, 88 a 5-6).

However, the view that the most explanatory proof is the most general is refuted by the following simple counterexample.

Problem IV. Show that, if $a^2 = nb^2$, then *n* is a perfect square.

To solve this problem, assume $a^2 = nb^2$. With no loss of generality we may also assume that *a* and *b* have no common factors (if they have common factors, we cancel them both in *a* and *b*). We show that no prime divides *b*. Suppose that some prime *p* divides *b*. Then *p* will divide b^2, hence $b^2 = kp$ for some *k*. Thus $a^2 = nb^2 = nkp$, hence *p* divides a^2. But, if a prime *p* divides the product of two integers, then *p* will divide at least one of them. Then, from the fact that *p* divides a^2, it follows that *p* divides *a*. Thus *p* divides both *a* and *b*. On the other hand, since *a* and *b* have no common factors, *p* cannot divide *a* and *b*. Contradiction. Therefore we conclude that no prime divides *b*. Then $b = 1$, so $a^2 = nb^2 = n$, hence *n* is a perfect square.

This solution of Problem IV also provides a new solution of Problem III since, in the case $n = 2$, it yields that, if $a^2 = 2b^2$, then 2 is a perfect square, so, since 2 is not a perfect square, $a^2 \neq 2b^2$.

Such solution of Problem III is more general than the one given in Section 10 since it accounts for infinitely many cases rather than for the single case $n = 2$, but it would be hard to claim that it is more explanatory. Hypothesis (C) provides a deeper reason for the fact that the square root of 2 is not rational than the hypothesis that if a prime *p* divides the product of two integers, then *p* will divide at least one of them.

One may even claim that hypothesis (C) provides a deeper reason also for the fact that, if $a^2 = nb^2$, then *n* is a perfect square. For an alternative solution of Problem IV can be given along the lines of the solution of Problem III given in Section 10. Assuming (C), if $a^2 = nb^2$, then in the prime factorization of a^2 each prime will have an exponent double the exponent it has in the prime factorization of *a*, so an even exponent, and similarly, in the prime factorization of b^2, each prime will have an exponent double the exponent it has in the prime factorizations of *b*, so again an even exponent. Thus, since $a^2 = nb^2$, in the prime factorization of *n* each prime will have an even exponent, hence *n* must be a perfect square. Therefore hypothesis (C) is capable of solving Problem IV.

12. Inference to the best explanation

Explanation in the sense considered here is somehow related to inference to the best explanation, by which 'we infer that what would best explain our evidence is likely to be true, that is, that the best potential explanation is likely to be an actual explanation' (Lipton, 2001, p. 97). For 'a potential explanation is something that satisfies all the conditions on actual explanation, with the possible exception of truth' and 'the actual explanations must be true' (ibid., pp. 96-97). Inference to the best explanation 'can be seen as an extension of the idea of self-evidencing explanations, where the phenomenon that is explained in turn provides an essential part of the reason for believing that the explanation is correct' (ibid., p. 96). Admittedly, 'self-evidencing explanations exhibit a curious circularity, but this circularity is benign' since, although 'hypotheses are supported by the very observations they are supposed to explain', the 'observations support the hypothesis precisely because it would explain them'.

There are, however, some basic differences between explanation in the sense considered here and inference to the best explanation.

1) In explanation in the sense considered here, hypotheses are not obtained by abduction but by other kinds of non-deductive inferences (inductive, analogical, etc.).

2) One cannot really say that what would best explain our evidence is likely to be true. For hypotheses can never be known to be true since, by Gödel's incompleteness results, knowing that they are true is generally impossible. Hypotheses can only be plausible. That, of course, does not mean that there is no mathematical knowledge, but only that there is no mathematical knowledge which is absolutely certain. Like all other knowledge, mathematical knowledge cannot be absolutely certain since it depends on hypotheses, such as the axioms of set theory, for which no absolute justification can be given (cf. Cellucci, 2005, pp. 161-163; 2006a, pp. 30-31).

3) Explanation in the sense considered here is not such that the phenomenon that is explained in turn provides an essential part of the reason for believing that the explanation is correct. For hypotheses are not supported by the observations they are supposed to explain, since even implausible hypotheses could explain such observations.

4) Explanation in the sense considered here exhibits no circularity, not even a benign one. For hypotheses are not supported by the observations they are supposed to explain but by their plausibility, that is, compatibility with the existing data, which include things distinct from the observations they are supposed to explain.

Admittedly, Descartes attributes circularity to the analytic method, since he claims that in such method 'the reasons follow one another in such a way that, as the last ones are demonstrated by the first ones, which are their causes, these first ones are conversely demonstrated by the last ones which are their effects' (Descartes, 1996, VI, p. 76). But Descartes is wrong here since, in the analytic method, hypotheses are not demonstrated by their effects but their acceptance depends on their compatibility with the existing data. Thus there is no circularity, not even a benign one.

13. Conclusion

The approach to mathematical explanation sketched here is related to a general view on mathematics presented in Cellucci (2002, 2006a). It differs from other approaches in the recent literature on the subject (Kitcher, 1989; Resnik-Kushner, 1987; Sandborg, 1998; Steiner, 1978) since they seem to neglect the strict connection between explanation and discovery. Excellent surveys and discussions of recent and older literature on mathematical explanation can be found in Mancosu (2000, 2001).

Hafner and Mancosu (2005, pp. 221-222) maintain that there are two approaches to mathematical explanation. In the first approach one presents a view on mathematical explanation and then tries to see how well it accounts for the practice. In the second approach one begins by avoiding any commitment to a particular view on mathematical explanation and rather gives a taxonomy of recurrent patterns of mathematical explanation through case studies. Then one tries to see whether these patterns are heterogeneous or can be subsumed under a general account.

Mancosu and Hafner favour the second, 'inductive' approach since, in their opinion, proceeding by the first approach would mean forcing the evidence from mathematical practice into a predefined mould, thereby narrowing the perspective from the outset and disregarding proofs, theories, methods etc. which do not satisfy that view on explanation.

The trouble with the second approach is however that there are no raw data, to recognize a pattern of mathematical explanation one must know what one is looking for, so one must already have a view of what a mathematical explanation is. Thus the second approach ultimately reduces to the first one.

As to the first approach, if finding an explanation is one of the main aims of solving a problem, then the question of finding an explanatory hypothesis for a problem is strictly related to that of finding a hypothesis which yields a solution to the problem. Therefore there is a strict connection between explanation and discovery.

Acknowledgments

An earlier version of this paper was presented at the conference 'Philosophy of Mathematics Today', held at the Scuola Normale Superiore of Pisa in January 2006. Thanks to Andrea Cantini, Charles Chihara, Per Martin-Löf, Dag Prawitz, William Tait for helpful remarks in the discussion following my presentation. I would also like to thank Donald Gillies, Robert Thomas and two anonymous referees for very helpful comments.

References

Balacheff, N. (1987). Processus de preuves et situations de validation. *Educational Studies in Mathematics*, *18*, 147-176.

Cellucci, C. (1998). *Le ragioni della logica*. Rome: Laterza.

Cellucci, C. (2000). The growth of mathematical knowledge: an open world view. In E. Grosholz & H. Breger (Eds.), *The growth of mathematical knowledge* (pp. 153-176). Dordrecht: Kluwer.

Cellucci, C. (2002). *Filosofia e matematica*. Rome: Laterza.

Cellucci, C. (2005). Mathematical discourse vs. mathematical intuition. In C. Cellucci & D. Gillies (Eds.), *Mathematical reasoning and heuristics* (pp. 137-165). London: College Publications.

Cellucci, C. (2006a). "Introduction" to *Filosofia e matematica*. In R. Hersh (Ed.), *18 unconventional essays on the nature of mathematics* (pp. 17-36). New York: Springer.

Cellucci, C. (2006b). The question Hume didn't ask: why should we accept deductive inferences? In C. Cellucci & P. Pecere (Eds.), *Demonstrative and non-demonstrative reasoning in mathematics and natural science* (pp. 207-235). Cassino: Edizioni dell'Università.

Cellucci, C. (2007). *La filosofia della matematica del Novecento*. Rome: Laterza.

Dawson, J. W. (2006). Why do mathematicians re-prove theorems? *Philosophia Mathematica*, *14*, 269-286.

Descartes, R. (1996). *Oeuvres* (C. Adam and P. Tannery, Eds.). Paris: Vrin.

Gödel, Kurt (1986-2002). *Collected works* (S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay & J. van Heijenoort, Eds.). Oxford: Oxford University Press.

Hafner, J. & Mancosu, P. (2005). The varieties of mathematical explanation. In P. Mancosu, K. F. Jørgensen & S. A. Perdersen (Eds.), *Visualization, explanation and reasoning styles in mathematics*. Dordrecht: Springer (pp. 215-250).

Hanson, N. R. (1958). *Patterns of discovery. An inquiry into the conceptual foundations of science*. Cambridge: Cambridge University Press.

Hempel, C. G. & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*, 135-175.

Hilbert, D. (1962). *Grundlagen der Geometrie*. Stuttgart: Teubner.

Hilbert, D. & Bernays, P. (1968-70). *Grundlagen der Mathematik I-II*. Berlin: Springer.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation* (pp. 410-505). Minneapolis: Minnesota Studies in the Philosophy of Science.

Lipton, P. (2001). Is explanation a guide to inference? A reply to Wesley C. Salmon. In G. Hon & S. S. Rakover (Eds.), *Explanation. Theoretical approaches and applications* (pp. 93-120). Dordrecht: Kluwer.

Mancosu, P. (2000). On mathematical explanation. In E. Grosholz & H. Breger (Eds.), *The growth of mathematical knowledge* (pp. 103-119). Dordrecht: Kluwer.

Mancosu, P. (2001). Mathematical explanation: problems and prospects. *Topoi, 20*, 97-117.

Mehlberg, H. (1960). The present situation in the philosophy of mathematics. *Synthese*, *12*, 380-414.

Peirce, C. S. (1931-58). *Collected papers* (C. Hartshorne & P. Weiss, Eds.). Cambridge MA: Harvard University Press.

Pólya, G. (1962-65). *Mathematical discovery. On understanding, learning and teaching problem solving*. New York: John Wiley & Sons.

Popper, K. R. (1934). *Logik der Forschung*. Wien: Julius Springer.

Popper, K. R. (1994). *In search of a better world. Lectures and essays from thirty years*. London: Routledge.

Post, E. (1994). Absolutely unsolvable problems and relatively undecidable propositions. Account of an anticipation. In Martin Davis (Ed.), *Solvability, provability, definability: the collected works of Emil L. Post* (pp. 375-441). Boston : Birkhäuser.

Resnik, M. & Kushner, D. (1987). Explanation, independence and realism in mathematics. *British Journal for the Philosophy of Science, 38*, 141-158.

Sandborg, D. (1998). Mathematical explanation and the theory of why-questions. *British Journal for the Philosophy of Science, 49*, 603-624.

Steiner, M. (1978). Mathematical explanation. *Philosophical Studies, 34*, 135-151.