Stud. Hist. Phil. Sci. 39 (2008) 202–210

Contents lists available at ScienceDirect

Stud. Hist. Phil. Sci.

journal homepage: www.elsevier.com/locate/shpsa

The nature of mathematical explanation

Carlo Cellucci

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

A B S T R A C T

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.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

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, and so on) 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, 1962–1965, Vol. 2, p. 126). Specifically, there is a sharp 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).

E-mail address: carlo.cellucci@uniroma1.it

[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.

0039-3681/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2008.03.006

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 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.

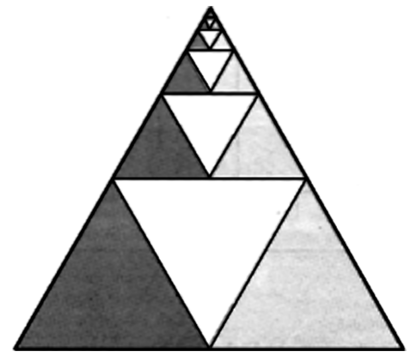

Fig. 1.

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' (ibid., 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:

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

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

[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.