Logique & Analyse **179–180** (2002), 299–307

EXPLANATORY PROOFS IN MATHEMATICS

ERIK WEBER AND LIZA VERHOEVEN*

1. *Introduction*

1.1. *Steiner*

Mark Steiner (1978, p. 143) uses the concept of *characterizing property* to draw a distinction between explanatory and non-explanatory proofs. A characterizing property is a property unique to a given entity or structure within a *family* or domain of such entities or structures. The concept of a family is left undefined. According to Steiner, an explanatory proof always makes reference to a characterizing property of an entity or structure mentioned in the theorem. Furthermore, it must be evident that the result depends on the property (if we substitute the entity for another entity in the family which does not have the property, the proof fails to go through) and that by suitably 'deforming' the proof while holding the 'proof-idea' constant, we can get a proof of a related theorem. Though many of Steiner's concepts (family, deformation, proof-idea) are vague, we can construct examples which beyond any doubt would classify as explanatory proofs by his criterion. Take the following proof of the Pythagorean Theorem:

PROOF T_1

(1) For every triangle ABC: $c^2 = a^2 + b^2 - 2ab.\cos(a, b)$. PREM

(2) For every angle θ: $\cos(\theta) = 0$ if $\theta = 90^\circ$. PREM

(3) For every right-angled triangle ABC with hypotenuse c: $(a, b) = 90^\circ$. PREM

(4) For every right-angled triangle ABC with hypotenuse c: $c^2 = a^2 + b^2$. 1,2,3

*Liza Verhoeven is research assistant of the Fund for Scientific Research – Flanders (Belgium). The research for this paper was made possible by the Research Council of Ghent University (grant BOF/2001/GOA/008). We thank the audience of CENSS2002 (Ghent, May 2002), ECAP 4 (Lund, June 2002) and PMP (Brussels, October 2002) for their comments of earlier versions of this text.

This proof makes reference to characterizing properties of right-angled triangles with hypotenuse c, namely that $(a, b) = 90°$. It is evident that the proof fails to go through if another kind of triangle is considered, since (3) is false for all other types of triangles. Furthermore, we can easily imagine similar proofs of related theorems. E.g., obtuse triangles contain exactly one angle $\theta > 90°$. Because $\cos(\theta) < 0$ if $90° < \theta < 180°$, we can derive that for all obtuse triangles, $c^2 > a^2 + b^2$.

1.2. *Resnik & Kushner*

According to Michael Resnik and David Kushner (1987, p. 153–154), there is no objective distinction between explanatory and non-explanatory proofs. Every proof allows us to answer at least one why-question (the question why the proven theorem is true). However, some proofs contain more information: they allow us to answer more why-questions; the more why-questions a proof can help to answer, the more explanatory it is. A why-question that is often asked is: why is this theorem true for this class of objects, but not for other classes of objects in the same larger family of objects? They agree that Steiner's proofs that use a characterizing property are ideally suited to presenting the information for answering this kind of why-questions. However, Resnik and Kushner say, there may be other why-questions, for which explanatory proofs in Steiner's sense are not well suited, because they provide the wrong kind of additional information.

1.3. *Aim of this paper*

We agree with Resnik & Kushner that Steiner's account is too narrow, because it is confined to questions of the form:

(C) Why do mathematical objects of class X have property Q, while those of class Y have property Q′?

However, we think that even within this restricted domain, Steiner's theory must be corrected and completed. This brings us to the first two aims of this article:

(1) Steiner assumes that proofs that satisfy his criterion are explanatory in their own virtue: by giving the proof, we automatically answer a question of type (C). In Section 2 we show that this assumption is wrong: proofs can become parts of answers, but they don't answer the question in their own right.

(2) Why is it important to answer questions of type (C)? What's the benefit? Steiner does not consider these questions. In Section 3 we try to complete his theory by answering these questions.

The "refined Steiner theory" that results from Sections 2 and 3 still faces the problem pointed out by Resnik & Kushner: not all explanatory questions have form (C). In Section 4 we will take a first step in developing a more complete theory by discussing questions of the form:

(P) Why do mathematical objects of class X have property Q, but not property Q′?

In this question it is assumed that Q′ implies Q. Note that in questions of type (C) we only assume that there is some connection between Q and Q′: no specific relation is required.

In Section 5 we will sketch some general characteristics of the theory proposed in 2–4, while in Section 6 we give an overview of the research that remains to be done.

2. *How are contrasts between classes to be explained?*

2.1. *A criterion*

Questions of the form (C) can be answered by means of a couple of two proofs:

> The couple (P_1, P_2) explains why mathematical objects of class X have property Q, while those of class Y have property Q′ if and only if
> (1) P_1 is a proof for the theorem that mathematical objects of class X have property Q;
> (2) P_2 is a proof for the theorem that mathematical objects of class Y have property Q′;
> (3) P_1 and P_2 use the same axioms and theorems as premises;
> (4) P_1 and P_2 have the same logical rules;
> (5) P_1 uses a defining property of X, but not of Y, as premise; and
> (6) P_2 uses a defining property of Y, but not of Y, as premise.

According to this criterion proofs have explanatory power only if we put them together in the right kind of couples. Furthermore, there is no objective distinction between explanatory and non-explanatory proofs, only between explanatory and non-explanatory couples of proofs. The reason is that conditions (3) and (4) refer to relations between proofs, not to properties of individual proofs.

2.2. *Examples*

The proof T_1 can be 'deformed' in Steiner's way to proofs of related theorems. An example of such a 'deformed' proof is the following.

PROOF T_2

(1) For every triangle ABC: $c^2 = a^2 + b^2 - 2ab.\cos(a, b)$.	PREM
(2) For every angle θ: $-1 < \cos(\theta) < 0$ if $90° < \theta < 180°$.	PREM
(3) For every obtuse-angled triangle ABC with obtuse angle in C: $90° < (a, b) < 180°$.	PREM
(4) For every obtuse-angled triangle ABC with obtuse angle in C: $c^2 > a^2 + b^2$.	1,2,3

This proof makes reference to the characterizing property of obtuse-angled triangles with obtuse angle in C, namely that $90° < (a, b) < 180°$. It can also be 'deformed' back to proof T_1. The couple (T_1,T_2) obviously is an explanatory couple of proofs according to our criterion. It answers a question of the form (C): why does it hold for right-angled triangles ABC with hypotenuse c that $c^2 = a^2 + b^2$, while for obtuse-angled triangles ABC with obtuse angle in C it holds that $c^2 > a^2 + b^2$? It is also clear that both proofs are individually explanatory according to Steiner's criterion, though their explanatory power depends on the possibility of 'deforming' the proofs, or in practice on the existence of 'deformed' proofs.

The same theorems can also be proven by the following proofs.

PROOF T_3

(1) For every n-tuple similar figures with corresponding sides $a_1, \ldots, a_n$, there is a factor $k \neq 0$ for which the areas of the figures can be represented as $k.a_1^2, \ldots, k.a_n^2$.	PREM
(2) For every triangle, the sum of its angles amounts to 180°.	PREM
(3) For every right-angled triangle ABC with hypotenuse c: $(a, b) = 90°$.	PREM
(4) For every right-angled triangle ABC with hypotenuse c: one can construct two similar triangles on the sides a and b that cover together an area that equals the original triangle.	2,3
(5) For every right-angled triangle ABC with hypotenuse c, there is a factor $k \neq 0$ such that $k.a^2 + k.b^2 = k.c^2$.	1,4
(6) For every right-angled triangle ABC with hypotenuse c: $a^2 + b^2 = c^2$.	5

PROOF T_4

(1) For every n-tuple similar figures with corresponding sides $a_1, \ldots, a_n$, there is a factor $k \neq 0$ for which the areas of the figures can be represented as $k.a_1^2, \ldots, k.a_n^2$.	PREM
(2) For every triangle, the sum of its angles amounts to 180°.	PREM
(3) For every obtuse-angled triangle ABC with obtuse angle in C: $90° < (a, b) < 180°$.	PREM
(4) For every obtuse-angled triangle ABC with obtuse angle in C: one can construct two similar triangles on the sides a and b that cover together an area that is smaller than the original triangle.	2,3
(5) For every obtuse-angled triangle ABC with obtuse angle in C, there is a factor $k \neq 0$ such that $k.a^2 + k.b^2 < k.c^2$.	1,4
(6) For every obtuse-angled triangle ABC with obtuse angle in C: $a^2 + b^2 < c^2$.	5

The defining properties used here are the same as in T_1 and T_2. For the couple (T_3,T_4) the same observations can be made. What we claimed is that a proof cannot be explanatory in its own virtue. The reader can only accept a proof as explanatory in Steiner's sense after he has 'deformed' it into a proof of a related theorem, i.e., from the moment he has in mind one other proof that together with the original proof forms a couple that meets our criterion. It is obvious that the couples (T_1,T_4) and (T_2,T_3) do not satisfy our criterion, neither are they 'deformable' into each other; the theorems are related, but their proofs are not.

3. *Unification*

One of the generally accepted aims of explanation in the natural and social sciences is *unification*. Unifying events consists in showing that two or more different events are instances of the same (set of) law(s) of nature. As an example, assume that we have observed the following:

> Pendulum a has a period in the interval [1.99s, 2.02s].
> Pendulum b has a period in the interval [2.44s, 2.47s].

These events can be unified by deriving them from the pendulum law. The first derivation is:

> (1) For all pendulums $P = 2\pi\sqrt{L/g}$.
> (2) a has a length in the interval [0.99m, 1.01m].

(3) All pendulums that have a length in the interval $[0.99m, 1.01m]$, have a period in the interval $[1.99s, 2.02s]$.
(4) Pendulum a has a period in the interval $[1.99s, 2.02s]$.

(1) and (2) are premises. (3) is derived from (1), the explanandum (4) is derived from (2) and (3). The second derivation is:

(1) For all pendulums: $P = 2\pi\sqrt{L/g}$.
(2′) b has a length in the interval $[1.49m, 1.51m]$.
(3′) All pendulums that have a length in the interval $[1.49m, 1.51m]$, have a period in the interval $[2.44s, 2.47s]$.
(4′) Pendulum b has a period in the interval $[2.44s, 2.47s]$.

Note that the two derivations have the same structure, use the same law and differ only in the "characterizing property" of each pendulum: its length.

A detailed analysis of unification in the empirical sciences, including a criticism of Philip Kitcher's influential views on unification (cf. his 1981 and 1989), can be found in Weber (1999). However, even without further analysis it is clear that answers to questions of type (C) that satisfy our criterion of adequacy, unify mathematical facts (in casu: theorems) in a similar way as the couple of physical explanations unifies the two physical facts. So answering questions of type (C) has at least one benefit: unification. Unification can be considered an intrinsic value, but also has practical and pedagogical advantages (applying and teaching a unified system is easier than applying and teaching a non-unified one).

4. *Contrasts between properties*

4.1. *Proofs and 'disproofs'*

Now that we have refined Steiner's theory we will take a step towards completing it. We discuss questions of the following form:

(P) Why do mathematical objects of class X have property Q, but not Q′?

As mentioned before, we assume that Q′ implies Q. Questions of type (P) can be answered by a couple of proofs (P_1,P_2):

The couple (P_1, P_2) explains why mathematical objects of class X have property Q, but not property Q′ if and only if:
(1) P_1 is a proof for the theorem that mathematical objects of class X have property Q, and

(2) P_2 is a proof for the theorem that not all mathematical objects of class X have property Q′.

The proof P_2 in this construction can have at least two different formats: (i) a counterexample, (ii) a proof that a subclass of X does not have Q′. These possible formats are clarified in the example we give in section 4.2.

4.2. *The Fundamental Theorem of Algebra*

The Fundamental Theorem of Algebra (Girard-d'Alembert-Gauss) states that a non-constant complex polynomial $P(z) = c_0 + c_1 z + \ldots + c_n z^n$ (all c_i complex numbers, $n \geq 1$ and $c_n \neq 0$) has at least one complex zero point. A consequence is that a non-constant complex polynomial $P(z) = c_0 + c_1 z + \ldots + c_n z^n$ (all c_i complex numbers, $n \geq 1$ and $c_n \neq 0$) can be factorized as $P(z) = c_n(z - z_1)\ldots(z - z_n)$ with all z_i complex numbers. Hence, if we consider the class X of the non-constant real polynomials $R(z) = r_0 + r_1 z + r_2 z^2$ (all r_i real numbers and $r_2 \neq 0$) of degree two, we know from the factorization $R(z) = r_2(z - z_1)(z - z_2)$, with all z_i complex numbers, that all such polynomials have two *complex* zero points (property Q). Call the proof of the latter fact F_1. We can prove that not all such polynomials have two *real* zero points (property Q′) by constructing counterexamples as follows. If a real polynomial $R(z) = r_0 + r_1 z + r_2 z^2$ has a complex zero point $a + ib$ (a and b real numbers), then $a - ib$ will also be a zero point, because real polynomials map conjugate complex numbers onto conjugate complex numbers. The factorized polynomial $r_2(z - (a + ib))(z - (a - ib))$ can be expanded to $r_2((z - a)^2 + b^2)$, which is a real polynomial for all values of a and b. So we have found a range of examples of non-constant real polynomials of degree two that have two complex zero points that are not real (those for which $b \neq 0$). These are the desired counterexamples that give us the proof F_2. The couple (F_1,F_2) explains the contrast that non-constant real polynomials of degree two have two complex zero points, but do not all have two real zero points.

One can investigate further which polynomials do have two real zero points and which do not. After some calculations, one finds that for the polynomial $R(z) = r_0 + r_1 z + r_2 z^2$ a simple characterisation can be made in terms of $\Delta = r_1^2 - 4r_0 r_2$. For $\Delta \geq 0$, $R(x)$ has two real zero points, whereas for $\Delta < 0$, $R(x)$ has two complex zero points that are not real. By elaborating on the contrast, we have found a subclass that corresponds with the property Q′.

4.3. *Why explain contrasts between properties?*

Asking question (P) is part of Lakatos' method of proofs and refutations. After exceptions to a primitive conjecture are found, the question why the conjecture is not valid in those cases arises. Trying to answer that question involves searching for hidden assumptions that have been made on the base of wrong intuitions. Once these implicitly made assumptions are identified, they can be incorporated into the formulation of the conjecture as an extra condition. This means that answering questions of type (P) leads to the discovery of new theorems. The process of improving the conjecture and its proof may also give rise to the discovery of new concepts (proof-generated concepts as Lakatos calls them) which means that so new fields of inquiry are opened up.

5. *Conclusions*

In this paper we have refined Steiner's theory with respect to questions of type (C), and developed a criterion for identifying satisfactory answers to questions of type (P). Two points of clarification are useful for preventing misunderstanding:
(1) In our view, the way in which the theorems which are used as premises in a proof are formulated, is important: in order to produce proofs that satisfy our criteria, some formulation of a theorem may be better than a different one which is logically equivalent.
(2) We have looked at explanation in the context of axiomatic mathematical theories: we have tried to find out how traditional formal proofs can explain. Our criteria do not apply to non-traditional proofs (e.g. picture proofs). However, this does not mean that we think that such proofs cannot be explanatory. We are convinced that they can, but that the criteria will be completely different.

6. *Questions for future research*

The theory of mathematical explanation that we have presented is far from complete. The following list of types of explanation-seeking questions should suffice to convince the reader that there is a lot of work to be done (even if we confine ourselves to traditional proofs):
(1) There is an interesting variant of type (C): questions of the form "Why do mathematical objects of class X have property Q, while those of class Y do not have property Q?". Our guess is that such questions can be answered by giving the proof for class X and showing that an adapted proof for Y goes

wrong somewhere.
(2) There are also interesting variants of (P): questions of the form "Why do mathematical objects of class X have property Q, rather than property Q′ (where Q and Q′ are mutually exclusive)" and similar questions of with other relations among Q and Q′.
(3) Explanation-seeking questions with respect to existence proofs.
(4) Explanation-seeking questions with respect to uniqueness proofs.
(5) Explanation-seeking questions with respect to identity proofs.
On top of that, there is also the question of how non-traditional proofs explain.

Centre for Logic and Philosophy of Science
Ghent University
E-mail: Erik.Weber@UGent.be
Liza.Verhoeven@UGent.be

REFERENCES

Kitcher P. (1981), 'Explanatory Unification', *Philosophy of Science 48*, pp. 507–531.

Kitcher P. (1989), 'Explanatory Unification and the Causal Structure of the World', in P. Kitcher & W. Salmon (eds.), *Scientific Explanation*, University of Minnesota Press, Minneapolis, pp. 410–505.

Lakatos I. (1979), *Proofs and Refutations. The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.

Resnik M. & Kushner D. (1987), 'Explanation, Independence and Realism in Mathematics', *British Journal for the Philosophy of Science 38*, pp. 141–158.

Steiner M. (1978), 'Mathematical Explanation', *Philosophical Studies 34*, pp. 135–151.

Weber E. (1999), Unification: What Is it, How Do We Reach it and Why Do We Want it?, in *Synthese 118*, pp. 479–499.