Remarks on simple proofs

Rosalie Iemhoff[*]

March 15, 2015

Abstract

This note consists of a collection of observations on the notion of simplicity in the setting of proofs. It discusses its properties under formalization and its relation to the length of proofs, showing that in certain settings simplicity and brevity exclude each other. It is argued that when simplicity is interpreted as purity of method, different foundational standpoints may affect which proofs are considered to be simple and which are not.

Keywords: proof, formalization, purity of method
MSC: 03A99, 03F03

1 Introduction

Mathematics is the science par excellence that can be simple and complex at the same time. Complex in its intricate arguments, yet simple in the structure that proofs are required to have, or in the theories that underlie these proofs. In contrast with the use of the word in daily life, in mathematics, a simple argument does not necessarily mean that it is easy to find. There exist simple yet ingenious proofs that took some years to be discovered. Paradoxes are an example, where Frege was unaware of a paradox in his formal system until Russell proved there to be one. Also, it often is the case that the first proof found for a theorem is not the most simple one and that only later simpler proofs than the original one are discovered.

What then are simple proofs? That question is not easy to answer, and I will not attempt to do so here. Rather, this note addresses certain aspects of that question. In particular, its possible meaning in the setting of formal systems, and its relation to the notion of purity of method. It is furthermore claimed that in certain settings some forms of simplicity exclude brevity or shortness of argument in that there are theorems for which no proof can be both short and simple.

[*] Department of Philosophy, Utrecht University, The Netherlands, R.Iemhoff@uu.nl, http://www.phil.uu.nl/~iemhoff. Support by the Netherlands Organisation for Scientific Research under grant 639.032.918 is gratefully acknowledged

This note is loosely based on a talk that I presented at the conference *Simplicity. Ideals of Practice in Mathematics & the Arts* that took place at City University of New York, April 3-5, 2013. It is not meant to be a philosophical account of simplicity in mathematics, but rather a collection of observations from a working mathematician on the matter.

2 Formalization

Most of us will agree that Carl Friedrich Gauss' famous argument that the sum of the first n natural numbers is equal to ${}^{n(n+1)}\!/_{2}$ is simple:
Proof The following sum shows that $2\sum_{i=1}^{n} i = n(n+1)$.

$$\begin{array}{ccccccccc}
1 & + & 2 & + & \dots & + & n & & \\
n & + & n-1 & + & \dots & + & 1 & & + \\
\hline
n+1 & + & n+1 & + & \dots & + & n+1 & = & n(n+1)
\end{array}$$

Therefore $\sum_{i=1}^{n} i = {}^{n(n+1)}\!/_{2}$. *qed*

And most of us will also agree that the proof by Andrew Wiles of *Fermat's Last Theorem* is complex (even without having seen it).

The two proofs illustrate many aspects of simplicity: the first is short and the reasoning is elementary, the second one is long and complicated, too complicated for most mathematicians to understand, actually. Gauss' proof also illustrates something else, namely that the simplicity of a proof may depend on the background theory that in the practice of mathematics is mostly kept implicit. It uses, for example, several facts about the operations of addition and multiplication on the natural numbers that are not explicitly mentioned.

The proof given above we call *informal* to contrast it with a proof in a formal theory in which every step is made explicit. Now although a formal proof in general does not look like an informal proof, still, given a theory, one can speak of an informal proof being *expressed formally* in the theory. This means that what one considers the *proof idea* in the informal proof, its essence, is faithfully translated into the formal setting. For example, in Gauss' proof above, one could require of a faithful formalization that the idea of summing up the first n numbers twice is part of the formalization.

One would expect that the simplicity of an informal proof is somehow reflected in faithful formalizations. On the other hand, the form of the foundational theory very much influences the form of its proofs. Clearly, in a formal theory that is minimal, proofs of even the simplest facts may be long and cumbersome. And the richer and stronger the theory, the simpler the proofs will be. In the extreme case a proven statement could be added to a theory as an axiom and the statement thereby receives a trivial and certainly simple proof in the new system thus obtained.

This, however, does not seem to be a strong argument against the independence of the notion of simplicity from formalization, as the theories in which we wish to carry out the formalization should be foundational theories, meaning that on the one hand they consist of axioms and rules which are evidently true and on the other hand are strong enough to formalize all or almost all of mathematics. Therefore we will only consider foundational theories as formal theories in which to formalize mathematical proofs. In general, adding a theorem to a formal system will result in a theory not satisfying the first requirement of a foundational theory, which is why the extreme case described above does not have to be considered.

One could require of the foundational theory that the *idea* of Gauss' proof as given above is expressible in a natural way and then claim that that proof is simple and will be so in every sufficiently strong foundational theory. But although this sounds perfectly natural at first, it may not always be easy or possible to determined which foundational theories satisfy these constraints. To explain my point, let us consider two foundational theories: type theory and set theory.

Over the last years type theory, and in particular homotopy type theory, has gained increasing attention as a foundational theory for mathematics, while set theory has been considered by many to be the main foundational theory for already a long time. Interestingly, fundamental concepts such as the natural numbers are treated very differently in type theory and set theory. Thus it is conceivable that certain intuitive proof ideas can in one theory be captured by simple and natural formulations and in the other theory only by complicated ones or cannot be captured in a faithful way at all. For such statements the notion of simplicity still makes sense for the informal proof, but it is not quite clear how to transfer it to their formalized versions, as it seems to depend very much on whether one works in the type–theoretic of the set–theoretic framework. Thus the above argues that while simplicity, even though hard to define, seems to be a genuine property of informal proofs that some satisfy and others do not, it may be hard to determine in how far such a property is preserved under formalization and to establish which foundational theory captures the informal arguments best.

3 Foundational theories

What about foundational theories themselves? Is there a way to distinguish the simple from the complex as it comes to foundational theories? Do there exist simple foundations for mathematics? Albert Einstein, in a famous quote has said: *I have deep faith that the principle of the universe will be beautiful and simple*[1]. By which, I think, he meant that the foundations

[1] Another quote on simplicity by Einstein that I love but that is somewhat beside the point here: Everything should be made as simple as possible, but not simpler.

of physics could be captured in simple laws. Mathematicians and philosophers have shown similar believe in the simplicity of the fundamentals of mathematics. By trying to reduce mathematics to logic, for example. Here simplicity should, I think, be read as self-evident.

The existence of a self-evident foundational theory would, of course, not exclude the possibility that some theorems have complicated proofs, but it would show that ultimately, truths can be reduced to a set of simple principles. Under the strict interpretation, meaning that the theory should be complete and really elementary, Kurt Gödel has proved this to be impossible. But under the weaker interpretation, meaning that the theory, although maybe not elementary or complete, is evident and large parts of mathematics can be carried out in it, such theories do indeed exist.

Given such foundational theories, the question naturally arises, which is the most fundamental, or self-evident, or simplest one. Three questions that although not strictly equal are intimately linked. The discussion above about set theory and type theory indicates that it may be hard to conclusively state which theory is more fundamental or self-evident than the other. And it may well be that it depends on the subject one wishes to capture in the foundational theory which is the better choice in terms of simplicity and self-evidence.

4 Purity of method

Closely related to simplicity is the notion of *purity of method* which refers to the property of proofs of being pure, where, following Detlefsen (2008), proofs are called *pure* if they concern themselves only with the concepts contained in the theorems proved. Such pure proofs should, for example, not contain reasoning about geometrics objects when the conclusion of the proof is a statement about the natural numbers. The *Prime Number Theorem* roughly stating that the asymptotic behavior of the number of primes not exceeding a given number n is $n/\log n$, illustrates this phenomenon nicely. The first proofs of this theorem were given by Jacques Hadamard and Charles Jean de la Vallée Poussin independently (Hadamard, 1896; de la Vallée Poussin, 1896). These proofs were not elementary in that they referred to objects far more complex than numbers, using techniques from complex analysis, while the elementary proofs later found independently by Atle Selberg (1949) and Paul Erdös (1949) did not.[2]

But is it correct to consider the later proofs more elementary than the first ones? In some cases it seems more or less clear that a proof method is not

[2] Interestingly, David Goldfeld (2004) cites Godfrey Harold Hardy from a lecture to the Mathematical Society of Copenhagen in which he says about the theorem (Bohr, 1952): *A proof of such a theorem, not fundamentally dependent on the theory of functions, seems to me extraordinarily unlikely.*

pure, as is the case for the use of mechanics in analysis. As pointed out in (Arana and Detlefsen, 2011), the mathematician Joseph–Louis Lagrange (1736–1813), who tried to liberate analysis from such impure notions thus pursued what might be called a purification program. However, in other cases it is not so clear what purity means, as it seems to depend on the way in which a theorem is presented. If a theorem about natural numbers has a proof that uses geometry of the plane, then by restating the theorem in terms of line segments its proof may, after all, be considered to be pure. A counter argument to this kind of reasoning could be that the theorem should be stated in its purest form. But whether there exists such a form is not easy to establish. Consider, for example, category theory or algebra versus proof theory. Logical notions such as theories, unifiers, and interpolants have very different definitions in these two settings, and it is therefore hard to conceive that there is a unique purest form for these notions. The fact that proof–theorists, like myself, sometimes reprove theorems for which the original proof is categorical and category–theorists do the converse, seems to support the idea that there is not one purest version of a theorem or proof.

In category theory one aims to put a notion into its proper categorical context in order to start reasoning about it. In proof theory one does the same, but then for a proof–theoretic rather than a categorical context. These contexts are very different in nature. Broadly, one could say that in category theory one provides a lot of structure and then considers a notion as part of that large framework. Proof theory, on the other hand, is in general more concerned with the generation of structure from below: one supplies some minimal principles that should be satisfied and then reasons about the notion on the basis of these principles. Any of the two approaches is superior over the other with respect to some theorems in that these theorems have shorter proofs in that foundational system than in the other. Therefore it seems at present hard to decide which foundational view is likely to produce more or purer proofs than the other, and it may well be that the outcome depends on the theorem or subject at hand.

Arana and Detlefsen (2011) discuss the epistemological significance of a conception of purity that they call *topical*. They argue convincingly that this significance lies in providing stable means of reducing certain ignorance in investigations. I would like to add that the distinction between proof–theorists and category–theorists elaborated on above may be exploited here as well. It namely shows that in practice it may be hard to settle whether there has been sufficient reduction of ignorance, since an argument may reduce the ignorance of a logician of the first kind more than of a logician of the second kind, or vice versa. Still, given a specific view on mathematics, be it proof–theoretical or categorical or otherwise, I think the notion of purity is meaningful against such a background and the theory developed in (Arana and Detlefsen, 2011) insightful and plausible.

5 Brevity

Jean Dieudonné (1969)[3] wrote: *…and that it is good discipline for the mind to seek not only economy of means in working procedures but also to adapt hypotheses as closely to conclusions as possible.* When interpreting closeness as not containing notions that are not directly related to those in the theorem, then the question arises whether what Dieudonné aims for can always be achieved, that is, whether proofs can be both short and simple. In this section, a proof is considered to be simple if it is close to the theorem it proves, and thus does not contain notions with no or only a distant relation to the ones in the theorem.

There are examples in the literature that suggest that at least in certain settings proofs cannot be both short and simple in the sense just defined. For example, in the setting of predicate logic a somewhat restrictive but reasonable interpretation of closeness could be that of being cut–free, where proofs are presented in a sequent calculus. The sequent calculus is a proof system (or rather a family of proof systems) that manipulates sequents, expressions consisting of and corresponding to formulas, in an elegant, concise manner, which renders it convenient for reasoning about meta–mathematics. Without defining what cut–free means, what is important for this exposition is that in cut–free proofs all formulas are, in some sense of the word, subformulas of those in the conclusion of the proof, which is why being cut–free may be considered a reasonable interpretation of closeness. Since it has been shown that there exist tautologies that have short proofs but no short proofs that are cut–free, under this interpretation there are theorems for which there do not exist proofs that are both short and simple.

This phenomenon, that proofs cannot be both short and close to the theorem they prove, also occurs elsewhere. In propositional logic it is possible to express for every n the n-th Pigeonhole Principle stating that when $n+1$ pigeons are placed in n holes at least one hole must contain more than one pigeon. These principles have simple proofs in the sense of being close to the principle they prove, but these proofs are long, of size exponential in n. Sam Buss (1987) developed an ingenious method to express and use counting in propositional logic and obtained short proofs of the Pigeonhole Principles that are of size polynomial in n. These proofs, however, are complicated and could be considered less close to the Pigeonhole Principle and thus less simple. Therefore the example of the Pigeonhole Principle suggests that also under this reading of simplicity shortness and simplicity may in certain settings exclude each other.

This tension between length and simplicity is well-known to everybody who has ever tried to write down an interesting proof in a formal system or in a programming language. The more the individual steps in the proof are

[3] page 11

reduced to evident logical inferences, the longer the proof gets. On the other hand, the more high–level concepts are allowed in the reasoning, the shorter the proof may become. But that such notions as shortness and simplicity have formal counterparts that actually display behavior that the informal notions seem to suggest, illustrates in a striking manner that, at least in this case, formalization can capture certain aspects of shortness and simplicity faithfully.

6 Proofs, short and simple

In conclusion, drawing from my own experience I have argued in the above that it may be hard to establish whether simplicity is preserved under formalization, and whether a foundational theory is simpler or more fundamental than another. I have discussed aspects of the concept of purity in the setting of proofs and provided examples illustrating that for certain interpretations of simplicity, shortness and simplicity exclude each other in that there are true statements that cannot have proofs that satisfy both properties, as least in certain settings. These observations are meant as fruit for thought rather than as a full account of the notion of simplicity of proofs. The notion is a natural albeit complicated one, which is why the title of this section is meant slightly ironically, as in the study of simplicity in the context of proofs hardly anything is ever short and simple.

References

Bohr, H. Address of Professor Harold Bohr. In *Proc. Internat. Congr. Math. (Cambridge, 1950)* vol 1, Amer. Math. Soc., Providence, R.I.: 127–134 (1952)

Buss, S. Polynomial size proofs of the propositional pigeonhole principle. *Journal of Symbolic Logic* 52: 916–927 (1987)

Dieudonné, J. *Linear Algebra and Geometry.* Boston, Mass.: Houghton Mifflin Co. (1969)

Detlefsen, M. Purity as an Ideal of Proof. In P. Mancosu (ed.), *The Philosophy of Mathematical Practice*: 179–197 (2008)

Detlefsen, M. and Arana, A. Purity of Methods. *Philosophers' Imprint* 11 (2): 1–20 (2011)

Erdös, P. Démonstration élémentaire du théorème sur la distribution des nombres premiers. *Scriptum 1, Centre Mathématique*, Amsterdam (1949)

Hadamard, J. Sur la distribution des zéros de la fonction zeta(s) et ses conséquences arithmétiques. *Bull. Soc. math. France* 24: 199–220 (1896)

Goldfeld, D. The Elementary Proof of the Prime Number Theorem: An Historical Perspective. In D. Chudnovsky, G. Chudnovsky, and M. Nathanson (eds.), *Number Theory*: 179–192 (2004)

Selberg, A. An Elementary Proof of the Prime Number Theorem. *Ann. Math.* 50: 305–313 (1949)

de la Vallée Poussin, C.J. Recherches analytiques la théorie des nombres premiers. *Ann. Soc. scient. Bruxelles* 20: 183–256 (1896)