Philosophica 72 (2003) pp. 9-33

THOUGHT EXPERIMENTS IN MATHEMATICS: ANYTHING BUT PROOF[1]

Jean Paul Van Bendegem

1. Introduction

It is apparently not an easy task to understand what thought experiments (TEs) could be, what they are, how they function, and so on. There are many, quite different definitions around that seem to be in conflict with one another (as the contributions to this volume will no doubt illustrate). Usually all examples of TEs come from the natural and, more exceptionally, the social sciences: Galileo's falling bodies experiment, Newton's bucket, Einstein's light ray, Maxwell's Demon, are the prototypical cases. Occasionally, authors talk about mathematical thought experiments (MTEs). There the situation becomes even more complex: first, few authors actually believe that there are such things as MTEs and those that do believe so, put forward nearly contradictory definitions. Nevertheless, the aim of this paper is to suggest that, first, MTEs do exist, second that there is a wide class of such MTEs, and finally, that is necessary to have MTEs in order to understand a major part of mathematical practice.

The core thesis of this paper is this: if it is so that what mathematicians are searching for are proofs within the framework of a mathematical theory, then any consideration that (a) in the case where the

[1] This paper is a very close "cousin" of Van Bendegem (to appear) and it certainly belongs to the "family" of Van Bendegem (1993), (1998), (2000), and (2001). The core object of these papers is to formulate a theory of mathematical practice (as complete as possible). In contrast to the "cousin" however, I have searched for different examples. In that sense they are complimentary.

proof is not yet available, can lead to an insight as to what the proof could possibly look like, and, (b) in the case where the proof is available, can lead to a better understanding of that proof, can be considered to be a MTE.

This definition is very close in spirit to the approach of Imre Lakatos in his *Proofs and Refutations*[2]. I am not claiming it is the same thing for, as I will show in the sequel of this paper, Lakatos' MTEs are one of the possibilities I see for MTEs, so it is perhaps better to talk about an *extension* of Lakatos' ideas. This definition is also motivated by an analogy with TEs in the natural sciences. A scientific theory tries to deal with sets of facts by constructing frameworks for interpreting these facts, producing coherence, and so on. TEs, speaking quite generally, are ways of exploring "imaginary" facts, in order to better understand "real" facts and the theories that incorporate them. If we equal facts for the scientist to proofs for the mathematician, then a MTE is to be something like an "imaginary" proof, that should help us to understand the proofs we are looking for. Hence the idea that MTEs should provide insight into either what a proof could look like, or why it is convincing, explanatory, in short, why it functions as a proof.

The main part of this paper will be a presentation of a rather extensive list of MTEs (according to the proposal outlined here), as I believe that the proof of the pudding still remains in the eating. However, before doing that, I consider it necessary to say a few things about the very objects we are talking about: proofs. This will be important to be able to make the distinction between proofs and MTEs and to avoid a number of philosophical problems. In the final part of this paper I will return to these philosophical conundrums.

2. What is mathematics all about?

It is a rather safe bet that the question in the title of this chapter will be answered thus: "It is all about proofs." And what proofs are also seems to be rather clear. It is in the ideal case a connected series of statements,

---

[2] These ideas have been continued by Eduard Glas, see Glas (1999) as an excellent example. This paper actually carries thought-experimentation in the title.

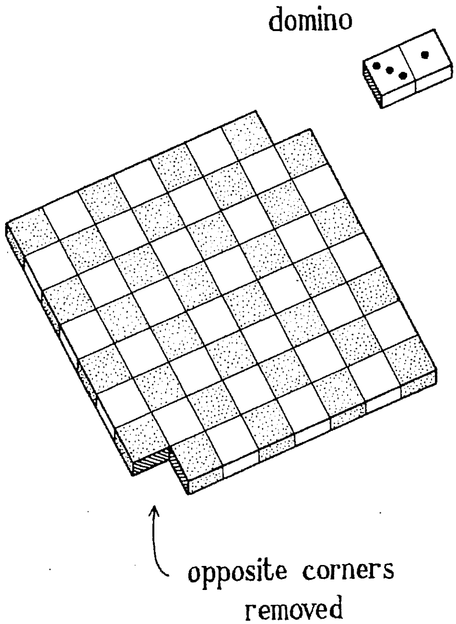
the last one being the statement to be proved and every step in the proof to be justified either because it is an axiom or the result of the application of one of the logical rules underlying the mathematical theory. It is usually acknowledged that proofs in handbooks, in journal papers, presented at conferences do not satisfy the standards of the ideal case, but that does not exclude that, if necessary, the proof could be rewritten in the required format. This rough picture is quite inaccurate on two points.

3.1 Rewriting proofs: when is it ever done?

The first point is that apparently mathematicians rarely feel the necessity to rewrite a proof. Usually reference is made to such books as Russell's and Whitehead's *Principia Mathematica* to emphasize the madness that would result if the ideal format were to be obligatory. Hence, in most if not nearly all cases, mathematicians deal with "real" proofs. In this sense I agree fully with James Robert Brown when he writes[3]: "If mathematical knowledge is based on proof, and proof is formal derivation, then in a vast number of cases, ..., mathematical knowledge is impossible. Obviously this is just absurd." (1991, p. 53)

At the same time, I believe that the example Brown presents to support his claim, is not the best one to use. As I will return to this example, let me devote some space to it. It is a very well known example: the covering of a particular chessboard by dominoes. The problem starts with the following initial configuration:

[3] For the reader who is familiar with Brown's position in the discussion about TEs, I am not sharing Brown's views on *why* mathematical knowledge is impossible. As I will defend at the end of this paper, I am not Platonist and that is precisely Brown's reason for this claim. I agree with his claim but for different reasons.



(from Brown (1991), p. 50)

One starts with a chessboard whereof two opposite corners have been removed. Given is also a set of dominoes, such that one domino covers exactly two squares. The question is asked whether it is possible to cover the whole board with dominoes. The answer to the question is no. A full formal proof would, as Brown argues, very likely have a kind of combinatorial structure. Along the following lines perhaps: label all squares of the board and add a label 0 or 1 as to whether or not a domino covers that square or not. Putting a domino on the board then translates into two zeros (belonging to the appropriate squares) changing into ones and a formal contradiction is derived if at some stage the same square has a 0 and a 1 assigned to it. It is straightforward to imagine that this kind of proof would, apart from extremely boring, also be extremely long. Whereas the proof preferred by all mathematicians says this: the two corners that have been removed have the same colour, a domino covers two squares of different colour, hence a full covering is impossible.

Brown remarks that this proof is not a formal proof. But I see no intrinsic difficulty to translate this proof into a formal proof in (the language of) *some* mathematical theory, i.e., not necessarily a theory that speaks about labels for the squares of the board and so on. Perhaps something along these lines (I present this proof in two parts, hence the labelling in terms of (a) and (b)):

(a1) A domino covers a black and a white square,
(a2) n dominoes cover n black and n white squares (generalization of (a)),
(a3) If the board is fully covered, a number of dominoes will have been used, say m,
(a4) Hence, the board consists of m black and m white squares,
(a5) Hence the number of black squares equals the number of white squares.
(b1) The board with the corners has the same number of white and black squares,
(b2) Removing two corners eliminates two squares of the same colour,
(b3) Hence the number of black squares does not equal the number of white squares.
As (a5) and (b3) contradict one another, it shows that the supposition of (a3) is wrong, hence, the board cannot be fully covered.

I am quite convinced that most if not all mathematicians will agree this is an acceptable proof (in the full mathematical sense of the term).

A much better example, to my mind, is Wiles' and Taylor's proof of Fermat's Last Theorem (see Wiles (1995) and Taylor and Wiles (1995)). Although the theorem itself is a statement belonging to the language of elementary arithmetic – it involves addition and exponentiation and natural numbers – the proof relies on methods and results that seem to lie light years away from number theory (although this statement is to a certain extent misleading, as will be clear in a moment): group theory, elliptic function theory, complex functions, modular forms, ... It is clearly a very legitimate question to ask whether a proof in elementary number theory is possible. The situation at the present moment is that the answer is (extremely likely) yes. The reason is this: Takeuti in Takeuti (1978) has shown that, if all definitions used are predicative, then a translation into elementary number theory is always possible[4]. At first sight, it seems that all definitions used in Wiles' and Taylor's proofs are predicative. But, at the same time, it is quite clear that no one seems to be interested to actually write down that proof, as it would probably have a length beyond all comprehension. I prefer this example, because in this case we do not have any idea whatsoever as to what the formal proof would look like, whereas in the chessboard problem we do. This, I believe, makes absolutely clear that the formal ideal is truly that: an ideal never to be reached, because one has more important things to do. Which brings me to the second observation.

3.2 There is more to proof than proof

The second point is that mathematics concerns much more than merely "looking" for proofs. Starting from the obvious observation that the mathematical activity can, to a certain extent, be seen as a problem-solving activity, we expect mathematics to have all the characteristics of such an activity. These involve, at least, strategies to find or construct proofs, strategies and heuristics to detect mistakes in proofs, techniques to develop new proof methods, criteria to judge the quality of proofs.

When all these aspects are taken into account, it becomes immediately clear that the mathematics building does count a huge number of rooms and, above all, that the building is immensely complicated as different rooms serve different purposes. In such a building there is plenty of room for thought experiments as will be shown in the next chapter. However, before doing that, let me just add two philosophical reflections on these general issues.

A first remark is that I will not defend a particular philosophical

---

[4] Takeuti's result is quite general: what he does is to show that, using only predicative definitions, it is possible to develop a sufficiently large portion of analysis within number theory (as a conservative extension) to be able to translate most mathematical results into number theory (on its own a quite impressive result). A definition is predicative if the definiens does not contain any reference to the definiendum. So, as Fermat's Last Theorem is translatable into analysis and if only predicative definitions are used, by Takeuti's result, it is translatable in number theory.

view of what mathematics is all about – i.e., in contrast to Brown, I believe that the issue of thought experiments in mathematics is *independent* of a Platonist view of mathematics – but that it is possible to take a neutral stand. After all, proofs can be seen as quite concrete objects: pieces of texts that have a very particular structure, serve very particular purposes and are treated and handled in very specific ways. Of course, it could very well be that the mathematician employing one or other strategy, increases the success of that strategy by actually believing that the proof she is looking for does actually exist, but in that case as well we only need to worry about a specific psychologically describable state of the mind of the mathematician. Another argument in support of this view is that the reader will at no time notice *my* particular philosophical view of mathematics, namely strict finitism. Hence what I present here, I consider to be independent from a strict finitistic point of view as well, bearing in mind of course that this does not exclude compatibility, but that is a different matter altogether.

A second remark concerns the *use* of mathematics. In this paper I do not discuss those cases of, say, a physical thought experiment that needs some mathematics to make its point. The most well known example would probably be Galileo's thought experiment about the connected heavy and light body, falling both slower and faster than the heavy body alone. If the mere fact that mathematics is used here as a tool is to be considered sufficient to count as a mathematical thought experiment, then the claim of this paper is trivial. The subject of this paper is whether or not some things mathematicians do when they are doing whatever it is mathematicians do when they do mathematics, can be considered to be thought experiments. Hence, on this point too I differ from Brown who writes: "My aim is to liken thought experiment to mathematical thinking" (1991, p. 49).

4. Candidates for thought experiments in mathematics

In this chapter I will present all the material promised. Some examples are well-known, actually they are very well-known, so much so that one is left with the impression that perhaps these are the one and only examples available. Therefore I have tried to find, besides the overrepresented classical examples, other cases to show that such examples are indeed in no way to be considered rare, but in fact are rather common in mathematical practice. There is one element I borrowed from Brown, namely the distinction between constructive and destructive TEs. This translates quite nicely into MTEs in favour of the existence of a proof or MTEs that show the way for a refutation.

4.1 Constructive MTEs of type 1: Informal "proofs".

Thought experiments of this type I consider to be reasonings that come very close to a formal proof (hence it is not conjectural as to what a proof might look like, *if there is one*), yet are, formally speaking, basically faulty. Often in the literature they are referred to as *informal* proofs (although that description is highly misleading as I have argued above). Obviously such MTEs tell you that a proof exists and give you quite some details of a correct proof.

Example 1. The most often quoted example is Euler's famous argument for the sum of the inverses of the squares, namely the argument that $\Sigma\ 1/n^2 = \pi^2/6$ (the summation taken over the natural numbers). I shall not present the full argument but its general structure. Euler first reasons about polynomials of finite even degree 2n, of the following form:

$b_o - b_1x^2 + b_2x^4 - \ldots + (-1)^n b_n x^{2n} = 0$,
with roots:
$r_1, -r_1, r_2, -r_2, \ldots, r_n, -r_n$.

He shows that the following holds:

$$b_1 = b_o(1/r_1^2 + 1/r_2^2 + \ldots + 1/r_n^2). \qquad (*)$$

All of this is quite regular mathematics. He then assumes that the same line of reasoning applies to polynomials of infinite degree. It is at this point that the reasoning goes astray, for there is no reason to suppose that the same result will hold for the infinite case. Thus the polynomial:

$1 - x^2/3! + x^4/5! - x^6/7! + \ldots = 0$,
with roots:
$\pi, -\pi, 2\pi, -2\pi, 3\pi, -3\pi, \ldots.$,

(as it is the series expansion of sin(x)/x), will satisfy (*), thus

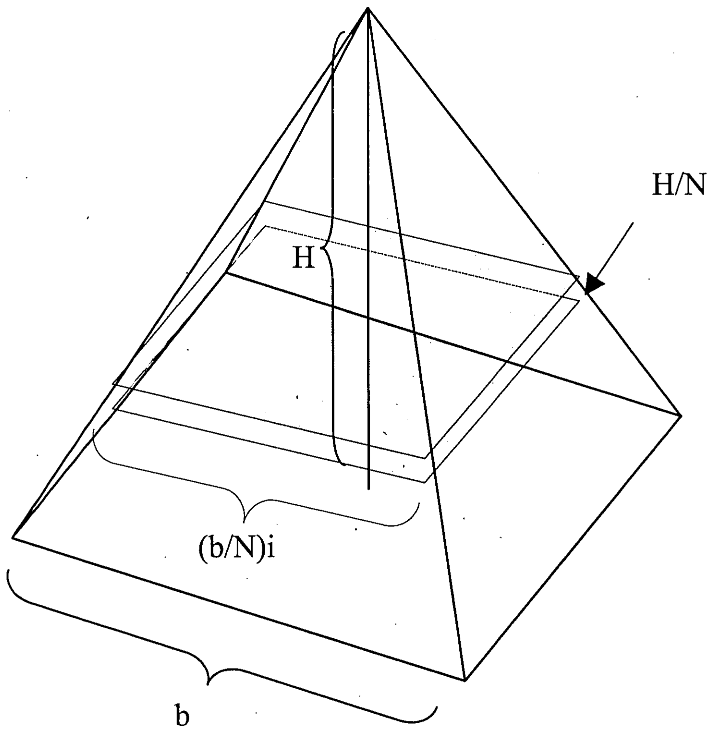
$$1/3! = 1/\pi^2 + 1/4\pi^2 + 1/9\pi^2 + \ldots, \text{ or:}$$
$$1 + 1/4 + 1/9 + \ldots = \pi^2/6. \qquad \text{QED (?)}$$

It is true that this specific example is quoted over and over again, nearly suggesting that we only have this one example, but such is not the case. Euler himself used similar arguments to show that (see Dunham (1990), pp. 207-222):

$\Sigma\ 1/(2n)^2 = \pi^2/24$, i.e., the sum of the reciprocals of all even squares,
$\Sigma\ 1/(2n+1)^2 = \pi^2/8$, i.e., the sum of the reciprocals of all odd squares, and
$\Sigma\ 1/n^4 = \pi^4/90$, i.e., the sum of the reciprocals of all fourth powers.

Example 2. Any argument that involves the use of infinitesimals (without the full apparatus of non-standard analysis[5]) can be considered to be an informal proof. Take a classic example: how to calculate the volume of a pyramid with a square basis. Assume the height of the pyramid is H and the side of the base is b. Cut up the pyramid in a series of slices, each of size (b/N)i, H/N.

[5] Non-standard analysis has been a quite intriguing approach to reintroduce infinitesimals into classical calculus. The basis idea is to consider an extension of the real numbers such that "around" every real number there is a structure of numbers distinct from the reals and that can play the part of infinitesimals. This is a full-blown mathematical theory, hence the proofs within this framework cannot be considered to be MTEs. What I am thinking of, is the way infinitesimals were used in the 16th, 17th century when calculus was in the making. For a nice treatment of non-standard analysis, see Keisler (1976).



The total volume is the sum of all the slices:

$$\sum_{i=1}^{N}(b/N)^2.i^2.(H/N), \text{ or}$$
$$= b^2.H.(1/N^3)\sum_{i=1}^{N} i^2, \text{ or}$$
$$= b^2.H.(1/N^3).N(N+1)(2N+1)/6, \text{ or}$$
$$= b^2.H.(1/N^3)(2N^3 + \ldots)/6$$

Ignoring the terms in N^2 and less, there remains:

$$b^2.H/3,$$

which is the correct answer[6].

Finally, one should not forget the most famous one of them all: Euler's "proof" the formula $V - E + F = 2$ for polyhedra. This "proof" (rather than a proof) was the starting point of Lakatos' famous study *Proofs and Refutations*. In this sense Lakatos' notion of a MTE falls into this category.

4.2 Constructive MTEs of type 2: "Career induction"

In this case the distance between the argument and the proof is larger, nevertheless as I will argue it does provide some insight into the proof itself (that is assumed to exist, although the certainty is less compared to the previous type). In the most frequent case the mathematician investigates a universal statement, usually over an infinite domain, by considering a finite number of cases and actually proving (in the mathematician's sense) these cases (including straightforward calculation). Famous examples are:

Example 1: Fermat's Last Theorem. It states that

$$(\forall n>2)(\forall x)(\forall y)(\forall z)(x^n + y^n \neq z^n).$$

Long before Andrew Wiles and Richard Taylor came up with their above-mentioned proof, it was known[7], e.g., in 1977, due to the work of Wagstaff, that the theorem held for n up to (at least) 125000. It is a rather intriguing question why mathematicians would do such a thing? There is, of course, a rather trivial answer. It is perfectly possible that there is a counterexample in the finite set of cases one investigates. But could there be other reasons? Actually there are. There are at least two reasons:

(a) It sets lower bounds. In the case of Fermat, there is a simple argument that shows that, for a given n, if there is a counterexample,

---

[6] One might wonder whether this kind of reasoning is not much more "fun" than differential and integral calculus reasoning. Alas, if one is not careful, all kinds of wrong results can be "deduced". For a very fine example, see Northrop (1961), pp. 201-204.

[7] See Ribenboim (1979), especially Lecture X, "Fresh Efforts" (pp. 199-223) for a splendid overview.

then x, y and z have to be larger than n:

> First, let us assume that $x \leq y < z$ (this is always possible). Suppose that
> $x^n + y^n = z^n$, then
> $x^n = z^n - y^n$, so
> $x^n = (z - y)(z^{n-1} + \ldots + y^{n-1})$.
> As $z - y \geq 1$, we have
> $x^n \geq z^{n-1} + \ldots + y^{n-1}$,
> and since $z, y \geq x$,
> $x^n \geq n.x^{n-1}$, or
> $x \geq n$.

In a sense this procedure shows the difficulty of finding counterexamples and does constitute important information as to the probability of finding such counterexamples.

(b) In a positive sense, the search for counterexamples very often necessitates special mathematical tools. Note, e.g., that in Fermat's case, even if n is fixed, the remaining equation still involves an infinite domain, viz., all triples (x, y, z) such that $x^n + y^n = z^n$. In other words, it is rarely a case of brute force calculations. Even in those cases, it is important to find ways to limit the time and space needed for the calculation. These tools can allow the mathematician to gain some insight into the kind of proof elements and proof concepts that will be needed if a proof of the (full) universal statement is ever to be found. In the case of Fermat, this is clear: the method of infinite descent was used in the special cases and it turned out to be a powerful method for dealing with the general case[8].

[8] Although I must add straight away that in the final proof by Andrew Wiles it is hard to see that this is a paper about number theory. Elliptic curves, group representations, Galois fields, ..., those are the ingredients needed to prove the statement, hence there is no direct use for infinite descent here, as it is a typical number-theoretic idea: if a solution in natural numbers exists, then a solution exists that is strictly smaller; this is impossible, because one would then have an infinite number of solutions, hence there is no solution. Infinite descent is very closely related (in some cases equivalent) to mathematical induction.

Example 2: Goldbach's Conjecture. The conjecture states that

$$(\forall n > 1)(2n = p_1 + p_2), \text{ where } p_1 \text{ and } p_2 \text{ are primes.}$$

A proof up to this point has still not been found[9], but it has been checked for n up to n (approximately) 10^{14}. As an illustration of the case (b) just mentioned above, one would expect that any calculation would consist simply of taking any even number 2n and to check all the cases to see whether there is a couple p_1, p_2 such that $2n = p_1 + p_2$. In Goldbach's conjecture, it turned out to be far more interesting to study the following function[10]:

$$G(2n) = \text{the number of different ways } 2n = p_1 + p_2.$$

Thus, $G(4) = 1$, $G(6) = 1$, $G(8) = 1$, $G(10) = 2$, ... Suppose that all the calculations show that $G(2(n+1)) \geq G(2n)$, i.e., G(2n) is an increasing function, then one will have a proof of Goldbach's conjecture if one can show that $G(4) \geq 1$ (but that is trivial of course). Thus we have a new statement to find a proof for: show that G is an increasing function.

Other famous examples include the Riemann Hypothesis (this problem concerns the non-trivial zeros of the complex function $\zeta(s) = \sum_{n=1}^{\infty} 1/s^n$, for s a complex number), the twin prime conjecture (the problem whether or not there are an infinite number of primes p_1 and p_2, such that

---

[9] The best results at the present moment are: every odd number is the sum of three primes on the one hand, and every even number is the sum of a prime and the product of two primes. Both results are implied by Goldbach's conjecture thus lending support to its truth. See Echeverria (1996) for a nice treatment, especially of the initial period, of this topic.

[10] Although not essential to the thesis of this paper, it is worth mentioning that Georg Cantor, the mathematician responsible for transfinite set theory, also spent some time on Goldbach's conjecture. The standard story is that a nervous breakdown made it impossible to work on serious matters, so Cantor "wasted" his time calculating decompositions in two primes for all even numbers up to 1000. However, as Echeverria (1996) shows, his contribution was very important. It was actually Cantor who proposed to study the function G(2n).

$p_2 - p_1 = 2$), and the Collatz problem (start with an arbitrary natural number and repeat the following instructions: if n is even, go to n/2, if n is odd, go to 3n+1; show that one must end with the cycle 4,2,1).

4.3. Constructive MTEs of type 3: Mathematical "experiments"

It is a very interesting and simultaneously very tangled question whether there can be such things as mathematical experiments. In my presentation here, I will try to avoid this nasty problem, but make ample use of quotation marks. Mathematicians themselves tend to speak about experiments and thus I will allow myself to use the term "experiments". There are at least two types to distinguish: visualisations and physical modelling.

Visualisations (see Hege & Polthier (1997) and (1998) for an overview) cannot be considered as formal proofs because, as we all know, the translation of a mathematical problem involving infinite domains (such as the real or complex numbers) to the computer screen consisting of a finite set of pixels must involve approximations. To be specific, suppose that a three-dimensional object, whereof an algebraic description is given, is visualized on the computer screen and the visual object has certain properties, then it would not be correct to conclude that the object actually does have that property[11]. In fact, as the literature shows, it is always necessary to establish estimations of the errors involved, but that needs to be proved mathematically, so, therefore, the image cannot add anything new. Or can it?

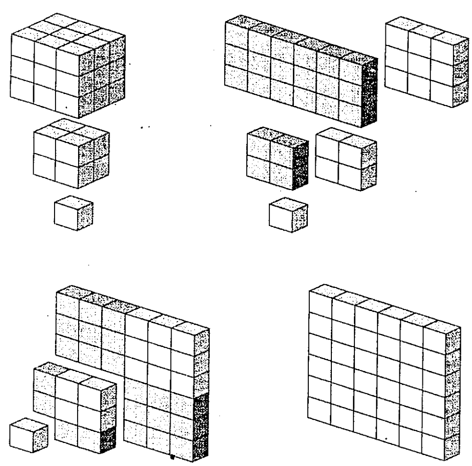
It is undoubtedly the case that an image can "reveal" certain aspects of a mathematical object. Seeing a mathematical object (or an approximation of one) does provide information in a different format. It is rather tempting to give a semiotic analysis at this point[12], but the fact

[11] The most striking example of recent times is of course the Mandelbrot set. The colourful visualisations show us without any doubt a wondrous world, but what it most certainly does not show us is the Mandelbrot set. To decide whether a point is inside or outside of the set, an infinite series has to be calculated to see whether it is convergent or divergent, but all computer simulations break off the series after a finite number of steps. Hence in the best of cases, what we see might be a nice approximation.

[12] I am thinking here of authors such as Michael Otte, see his (1997) or Brian Rotman, see his (2000).

that a formal text and a picture are not the same can hardly be a point of discussion. Even if it turns out that a property of the visualization is a computer artefact, this might still provide some insight.

On the other hand, visualisations can do something that the previous examples either fail to do or at best can only do indirectly: they continue being helpful even *after* the proof has been found. In terms of Brown's classification, they are constructive mediative TEs: "... it may act like a diagram in a geometrical proof in that it helps us to understand the formal derivation and may even have been essential in discovering the formal proof." (pp. 36-37). It is the role played by the drawing presented above when I derived the volume of a square pyramid. Here we touch upon the difficult problem of the explanatory value of a proof. If the proof has this quality, then usually it is possible to present a picture, a diagram, ..., that will make the explanation accessible, visible even. So-called "proofs by looking" have the property that the (formally correct) proof can be read off the drawing and thus explains what is happening. A classic example (though not the overrepresented one, viz. $\sum_{i=1}^{n} i$ = n(n+1)/2):



(from Nelsen (1993), p. 86)

This picture "shows" that $1^3 + 2^3 + \ldots + n^3 = (1 + 2 + \ldots + n)^2$. A "full" proof would proceed by mathematical induction and not explain anything at all.

The second type of "experiment" is a quite different process: physical modelling. Ivars Peterson (1988) discusses the Plateau problem: given a boundary curve B, what is the minimum surface S having B as its boundary? Mathematically this is a profound and difficult problem. Analytical methods are often insufficient. There is, however, a simple way to find solutions, though not necessarily the set of all solutions. Construct the boundary B in metal wire. Dip it in soapy water and a film will form having B as its boundary. Physics tells us that this film is a minimum surface. Hence, Peterson says: "They can explore shapes that are often too complicated to describe mathematically in a precise way. They can solve by experiment numerous mathematical problems associated with surfaces and contours." (p. 48). The relations between such experiments and mathematics is actually a quite profound philosophical issue[13]. In terms of TEs, it is, e.g., a difficult problem to decide whether we are doing a physical TE or indeed a MTE. Or are they both?

4.4. Constructive MTEs of type 4: Probabilistic arguments.

This type of MTE presents an intriguing problem. Probabilistic arguments often come in the form of a proof of a theorem and thus it seems strange to call them MTEs (as the title of this paper indicates: anything but proof). However the situation is slightly more complicated. One of the origins of probabilistic reasoning is to be found in number theory, more specifically in the problem to decide whether an arbitrary number n is a prime or not.

There are, of course, explicit algorithms, but they are exponentially complex. E.g., divide n by all prime numbers $p<\sqrt{n}$. However, an alternative is to make a random selection of k numbers out of this set of prime numbers and see what happens. Although such "primitive" tests are not used at all (as far as I know), the basic approach remains the same: on the basis of the random selection probabilities can be assigned

[13] I refer the reader to Van Bendegem (1998) for more details.

and thus it becomes possible to *prove* that the probability that the number is indeed prime is at least 0,999 (or something of that order). In this sense this proof tells us something about *another* proof we are looking for, namely, a proof that n is indeed prime or not. The probabilistic proof, although itself a proof, is a MTE for the other proof[14].

There is a weaker form where it is better to speak of an argument rather than of a proof. An example, accessible to the non-mathematician, may illustrate what I have in mind. The example is presented in Delahaye (2000):

"Consider the sequence of natural numbers defined as follows:
$x_1 = 2$
$x_2 =$ [the smallest prime factor of $x_1 + 1$] $= 3$
$x_3 =$ [the smallest prime factor of $x_1 x_2 + 1$] $= 7$,
...,
$x_{n+1} =$ [the smallest prime factor of $x_1 x_2 \ldots x_n + 1$] , ...

The sequence begins with 2, 3, 7, 43, 13, 53, 5, 6221671, 38709183810571, 139, 2801, 17, 5471, ... This Euclid-Mullin sequence lists only distinct primes: $y_n = x_1x_2\ldots x_n + 1$ is not divisible by x_1 (since if it were , then 1 would also be, as the difference of two numbers divisible by x_1.) The number y_n is not divisible by x_2 (otherwise 1 would be), etc. Since y_n is not divisible by any of the prime numbers $x_1, x_2, \ldots,$ x_n, x_{n+1} is a new prime number, and thus all of the x_i are distinct prime numbers. The question is knowing whether this sequence lists all prime numbers, omitting none (though to be sure they will be out of order.) It is thought that the answer is yes, and the following heuristic proof is proposed:

Suppose the sequence did not include all of the prime numbers. Let p be the smallest prime number not in the sequence. Beyond a certain N, all prime numbers less than p will be included in the numbers $x_1, \ldots, x_N$. If n is a randomly chosen whole number larger than N, then the number $y_n = x_1x_2\ldots x_n + 1$ can be considered a random number relative to p, and thus this number has one chance in p of being a multiple of p (since one

[14] A detailed and accessible treatment is presented in Ribenboim (1989), especially part XI of chapter 2.

out of every p whole numbers is a multiple of p.) The number y_n thus has probability (1 - 1/p) of not being a multiple of p, which is also the probability that x_{n+1} is different from p. The probability that neither x_{N+1}, nor $x_{N+2},\ldots,x_{N+k}$ is equal to p is thus $(1 - 1/p)^k$, which tends to 0 at infinity. Otherwise stated, the probability that p does not appear in the sequence x_n is zero. Thus p appears in the sequence, which contradicts its definition. Thus every prime number p appears in the sequence x_n, which is the same as the list of prime numbers, without repetition and written out of order.

Such a line of reasoning almost holds good, but it assumes that y_n is chosen at random, which is not the case, and thus without a complement (which no one has succeeded in discovering and which appears to be beyond the range of current mathematics) the heuristic proof is not an acceptable proof."

But one may add, it does retain its role as a MTE[15].

4.5 Destructive MTEs: Paradoxes yes, contradictions no?

Next to constructive MTEs we should distinguish destructive MTEs. A destructive MTE is an argument that refutes a given statement. It is important, I believe, to make a distinction between refutations backed up by a proof and refutations backed up by an argument. The former I do not consider to be MTEs: after all, what we show is the existence of a counterexample. Even if the counterexample itself is accompanied by a diagram, a drawing, a picture, whatever, these can be considered as MTEs but relative to the counterexample not relative to the statement that is being refuted. Hence, the argument-based refutation is a possible candidate for a MTE. The same reasoning applies to provable contradictions within a mathematical theory. Russell's set of all sets that do not belong to themselves is provably inconsistent within Frege's framework, hence there is no reason to call this a thought experiment. This raises the question, of course, whether anything at all can be considered a destructive MTE?

As the title of this chapter indicates, we should not look at

[15] In addition I do not address in this paper the further complication of proof verification systems that check the correctness of a formal proof by making a (random) selection of bits and pieces of the proof. We thereby introduce a metalevel.

contradictions, but at paradoxes. After all, a paradox does not claim that the theory is inconsistent; it claims that there is a tension between what is proved and our understanding of it (which brings us back to considerations similar to constructive MTEs of type 3).

Example: A good example of such a paradox is the (in)famous Banach-Tarski paradox. The Banach-Tarski paradox throws doubt (for some mathematicians at least) on the axiom of choice as it is used in set theory. The paradox states that it is always possible in three-dimensional real space, R^3, to decompose into five pieces a ball of volume V and reassemble the pieces into two balls of volume V, using only rigid motions (translations and rotations). There is no mathematical problem here, but the paradoxical character of the result is clear and serves as an argument against the axiom of choice[16]. (But note that there are at the same time arguments in favour of the axiom of choice as well, thus making the mathematician's activity so fascinating).

It is important to realize that we are entering now a grey area. Not all paradoxical results shed light on the truth or falsity of some mathematical statement. Very often, all there seems to be is a conflict between our expectations and the mathematical outcome. This does not plead against the mathematical result, it warns us to be very careful about our intuitions. In order to let the reader judge for herself, here is an example that I do not consider to be a MTE, as it does not refute anything, but merely challenges our intuitions. The example is pretty famous, namely Polya's urn.

Example: Given is an urn that contains a black and a white ball. One draws at random a ball, and returns the ball to the urn with one additional ball of the same colour. The question is what will be the limit distribution of white and black balls in the urn? The surprising answer is: anything at all. (One can get something of the flavour of the problem by considering the extreme cases: suppose that turn after turn you keep drawing black balls, then the limit distribution will be (black,white) = (1,0), whereas the other extreme scenario will produce the opposite result). In addition it can be shown that in a particular run of the experiment, quite soon the distribution tends to stabilize.

Since I have given an example of what I do not consider to be a

[16] See Moore (1982) for details.

MTE, let me develop this theme a bit further in the concluding chapter of this paper.

5. What are definitely not MTEs?

At the end of chapter 3 I already mentioned my disagreement with Brown who wants to liken TEs to mathematical thinking. Although of course in such TEs as Galileo's falling bodies, mathematics is used to derive a conclusion, it is not the mathematics itself that is under scrutiny, as it is merely applied. Therefore these are not MTEs. But there are other cases that, according to the core thesis presented here, are not to be considered as MTEs. I mention four cases (though not claiming completeness).

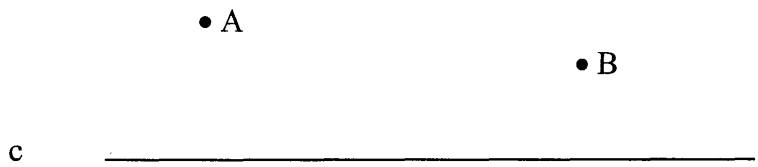
Case 1: A case of mathematical arguments that I do not consider to be MTEs are (partially or completely) computer-produced proofs. Without any doubt, the four-colour theorem (FCT) is the most famous example[17]. To many mathematicians the "proof" of FCT is not a proof, but an argument for a proof. To be a bit more specific: the theorem states that four colours are sufficient to colour any planar map in such a way that neighbouring areas are coloured differently. The first published proof consisted of two parts. The first part was a "classical" mathematical proof, in which it is shown that the set of all possible maps can be reduced to a finite set such that if all maps in that finite set can be coloured, so can the full set. It is actually a beautiful piece of mathematics. But the second part consists of a computer listing, presenting the details of a computer program that has actually coloured all the maps and said, "yes, I have coloured them all" at the end of the day. Although there are very good reasons to consider this "proof" as an argument, and it may help to convince us of the correctness of FCT, it does not provide an insight, let alone an explanation of why FCT holds. All we have is a list of cases, dealt with one by one.

Case 2: Related to the previous case, is the case of the "clever" proof. Here we are talking of proofs (no quotation marks) in the sense outlined in chapter 3, hence there is no reason to assume that they could be MTEs, but sometimes they involve a picture or diagram as part of the

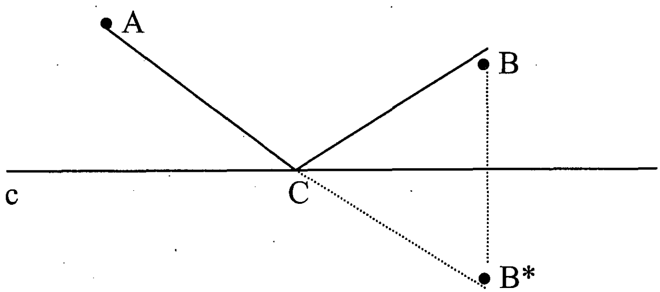
[17] See Tymoczko (1986) for details.

proof and in that sense these pictures and/or diagrams can be considered as MTEs, but then relative to the proof. Let me present one well-known example (see Kuipers (1991) for other examples):

Example: Given is the following situation:



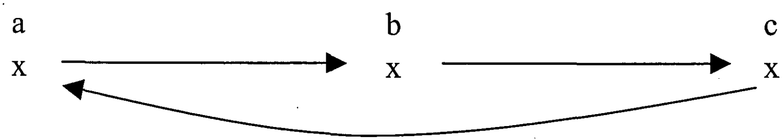
The question is to find the shortest distance from A to B that touches c. The "clever" answer is to introduce B*. The shortest distance from A to B* is simple: it is a straight line. This line intersects c in point C. One sees immediately that the distance ACB must be the shortest one.



Different proofs can now be derived from this drawing. Either one introduces a coordinate system and treats the problem algebraically or one observes that length(AB*) = length(ACB), for any C on c, hence as AB* is the shortest distance, so is ACB.

In short, the fact that it is a picture and as such irrelevant to the formal proof turns it into a MTE, not the cleverness (or beauty, as mathematicians are so fond of saying) of the proof.

Case 3: Leaving behind specific proofs and the arguments that surround them, it is worthwhile to mention that some authors consider entire mathematical theories as MTEs. A very good example would be Edwin Abbott's *Flatland*. The book presents us with a "real" two-dimensional world that we can visualize or picture (and, once again, of course, in that sense it does share properties with MTEs discussed above). It does perhaps contribute to a better understanding of two-dimensional geometry (in its relation, e.g., to higher-dimensional geometries), but it does not provide insights into proofs. In other words and somewhat more bluntly: it does not advance our mathematical knowledge. My worry, if we were to allow such stories as MTEs as well, is that one cannot avoid that any model will thereby become a MTE. Given a piece of syntactic mathematical work, then one can always construct a model (one selection out of a (usually, if interesting, infinite) set of possibilities). Suppose I am talking about functions from the set {a,b,c} into itself and it is given that f(a) = b, f(b) = c, and f(c) = a. A "nice" picture of this function would be:



And, of course, plenty of questions are now straightforward to answer: e.g., what is f(f(f(a)))? The answer is a, of course. But should one call this drawing a MTE? My answer would be no, in order to avoid triviality.

Case 4: Is there a case to be made for the idea that *all* of mathematics is one gigantic MTE? Sometimes it seems that James Robert Brown is defending such a position. Roughly speaking, if mathematics or, better, mathematical knowledge has an *a priori* status, then it has a relative independence of the real world, but, since mathematics is actually used in physics and physics does describe the real world, then, what we have been doing mathematically is one gigantic MTE. I think that we have here a similar difficulty as with the Plateau problem: are we doing a mathematical "experiment" or a physical experiment? Either the mathematics is "pure" or it is "applied" (to use an old distinction). In the latter case, since the mathematics is interpreted, we are dealing with a particular domain of human knowledge, but then that particular domain in a sense intervenes; in short, one can always consider the MTE as a TE in that domain. But in the former case, we remain within mathematics itself and it ceases to be a MTE. One might think that the Platonist has the final answer: if you accept the existence of a mathematical universe

(MU), then mathematics can be seen as a TE vis-à-vis this universe and there you have it. However I think that this position is self-defeating. If mathematics does describe the MU (much as our physical theories describe the physical universe), then it is no longer a TE, but a true description. That will not do. The last straw to hold on to could be that, just as Galileo's falling bodies, Newton's bucket, ..., one could imagine MTEs in relation to the MU. This however raises the question: how are we to decide what the "real" nature of the MU is? If a piece X of mathematics is accepted as a "decent" part of mathematics, then, for the Platonist, it is immediately part of the MU, hence no more MTEs. It is thus not really rewarding to be a Platonist in this case.

As a final thought, let me stress once again that the view about MTEs presented here derives from a view of mathematical *practice*, not from a philosophical view about the *nature* of mathematics. It shows that fundamental distinctions, e.g., what distinguishes mathematical from scientific knowledge, can also be drawn by looking at practices. At least for the question of MTEs it proved to be possible.

Vrije Universiteit Brussel
Universiteit Gent

REFERENCES

Abbot Edwin A. (2002), *The Annotated Flatland. A Romance of Many Dimensions*. Oxford: The Perseus Press (introduction and notes by Ian Stewart).

Brown James Robert (1991), *The Laboratory of the Mind. Thought Experiments in the Natural Sciences*. London: Routledge.

Delahaye Jean-Paul (2000), 'Shortcuts in Proof'. *La Lettre de la Preuve*, September/ October 2000, (electronic newsletter: http://www-didactique.imag.fr/preuve/).

Dunham William (1990), *Journey Through Genius. The Great Theorems of Mathematics*. New York: Wiley.

Echevarria Javier (1996), 'Empirical Methods in Mathematics. A Case-Study: Goldbach's Conjecture', in G. Munévar (ed.), *Spanish Studies in the Philosophy of Science*. Dordrecht: Kluwer, pp. 19-55.

Glas Eduard (1999), 'Thought-Experimentation and Mathematical Innovation', *Studies in the History and Philosophy of Science* **30**, pp. 1-19.

Hege H.C. & Polthier K. (eds.) (1997), *Visualization and Mathematics. Experiments, Simulations and Environments*. New York: Springer.

Hege H.C. & Polthier K. (eds.) (1998), *Mathematical Visualization. Algorithms, Applications and Numerics*. New York: Springer.

Keisler H. Jerome (1976), *Elementary Calculus*. Boston: Prindle, Weber & Schmidt.

Kuipers Theo A.F. (1991), 'Dat vind ik nou mooi', in Rien T. Segers (red.), *Visies op cultuur en literatuur. Opstellen naar aanleiding van het werk van J.J.A. Mooij*. Amsterdam: Rodopi, pp. 69-75.

Lakatos Imre (1976): *Proofs and Refutations*. Cambridge: Cambridge University Press.

Moore Gregory H. (1982), *Zermelo's Axiom of Choice. Its Origins, Development, and Influence*. New York: Springer.

Nelsen R.B. (1993), *Proofs without Words. Exercices in Visual Thinking*. Washington: MAA (Classroom Resource Materials, vol. 1).

Northrop Eugene P. (1961): *Riddles in Mathematics*. London: Pelican Books.

Otte Michael (1997), 'Mathematik und Verallgemeinerung. Peirce's semiotisch-pragmatische Sicht', *Philosophia Naturalis* **34** (Heft 2), pp. 175-222.

Peterson Ivars (1988), *The Mathematical Tourist. Snapshots of Modern Mathematics*. New York: Freeman.

Ribenboim Paulo (1989), *The Book of Prime Number Records*. New York: Springer.

Rotman Brian (2000), *Mathematics as Sign. Writing, Imagining, Counting*. Stanford: Stanford University Press.

Takeuti Gaisi (1978), *Two applications of logic to mathematics*. Princeton: Princeton University Press.

Taylor Richard & Wiles Andrew (1995), 'Ring-theoretic properties of certain Hecke algebras', in *Annals of Mathematics* (Second Series) **141**, pp. 553-572.

Tymoczko Thomas (1986), *New Directions in the Philosophy of Mathematics*. Stuttgart/Boston: Birkhauser.

Van Bendegem Jean Paul (1993), 'Real-Life Mathematics versus Ideal Mathematics: The Ugly Truth', in Erik C.W. Krabbe, Renée José Dalitz & Pier A. Smit (eds.), *Empirical Logic and Public Debate. Essays in Honour of Else M. Barth*. Amsterdam: Rodopi, pp. 263-272.

Van Bendegem Jean Paul (1998), 'What, if anything, is an experiment in mathematics?', In Dionysios Anapolitanos, Aristides Baltas & Stavroula Tsinorema (eds.) : *Philosophy and the Many Faces of Science*. London: Rowman & Littlefield, pp. 172-182.

Van Bendegem Jean Paul (2000), 'Analogy and Metaphor as Essentials Tools for the Working Mathematician', in: Fernand Hallyn (ed.), *Metaphor and*

Analogy in the Sciences (Origins: Studies in the Sources of Scientific Creativity). Dordrecht: Kluwer Academic, pp. 105-123.

Van Bendegem Jean Paul (2001): 'The Creative Growth of Mathematics', in *Philosophica* **63** (1999, date of publication: 2001), pp. 119-152.

Van Bendegem Jean Paul (to appear), 'Proofs and Arguments: The Special Case of Mathematics', in: *Logics of Scientific Cognition. Essays in Debate With Theo Kuipers*. Amsterdam: Rodopi (Poznan Studies).

Wiles Andrew (1995), 'Modular elliptic curves and Fermat's Last Theorem', in *Annals of Mathematics* (Second Series) **141**, pp. 443-551.