

A first exploration of effective reasoning

My youngest Concise Oxford Dictionary (7th edition, 1982) defines (pure) mathematics as "abstract science of space, number, and quantity". A long time ago, this definition may have been adequate, but for ~~more~~ ~~than~~ at least a century this definition is much too restrictive: there are simply too many people considered mathematicians — by themselves, by their colleagues, and by others — whose professional minds are hardly occupied with space, number, or quantity. The topics today's mathematicians think about are, in fact, so varied that it is no longer realistic to characterize mathematics by its subject matter.

Yet, despite this variation of topics, the mathematician very quickly recognizes his fellow mathematician as such: there is a unity that is not reflected in the diversity of the topics. The unity that constitutes the core of mathematics is not so much the "quod" as the "quo modo"; it is the precision of the concepts, the shape of the arguments and the nature of the use of symbols and language that are characteristic for mathematics. And that is why today a much better

definition of mathematics would be "the art and science of effective reasoning", independent of what that reasoning is supposed to be about.

But in order to reason effectively, the very least we should be able to do is establishing whether two statements mean the same thing or whether two definitions are equivalent, and here we immediately run into trouble with our so-called "natural languages", which (among other shortcomings) are not very well suited to precision. By human habit, our languages tend to become geared to the major purposes we want to use them for, and consequently they are excellent for being vague in, for telling jokes in, and for making love in; precision, however, is only rarely needed, and as soon as we try to be precise, as for instance in legal documents, the language used quickly becomes most unnatural.

Ambiguity is for instance caused by the unclarity as to how sentence components should be grouped together, the standard example being "pretty little girls school", which admits five different

parsings. (Does "pretty" qualify "little", "girls", or "school"? And is it about a little school or about little girls?)

A slightly more subtle example is provided by the two sentences:

"When Jack and Jill went up the hill, they picked 5 pounds of blueberries"

"When Jack and Jill went up the hill, they walked 5 miles"

I think that the first sentence permits the interpretation that Jack picked 2 pounds and Jill 3, but I am almost certain that the second sentence does not admit the analogous interpretation that Jack walked 2 miles and Jill 3.

Another source of difficulties, which I won't pursue here now, is that people sometimes don't distinguish between

"If the sun shines, I don't wear my coat." and

"If I don't wear my coat, the sun shines."

The above examples of linguistic problems should suffice at this stage; their only purpose was to convince the reader that

for truly effective reasoning, natural language is an inadequate vehicle, and I hope they have done so. The moral of the story is that the development of mathematics is partly - and perhaps for a major part! - a linguistic exercise, viz. the design of languages adequate for definitions of and arguments about the artefacts we would like to deal with.

A familiar example is provided by the names and notations for numbers as used in counting and arithmetic. Like the rest of natural language, the names of numbers are originally not the product of careful design: the French name "quatre-vingt-dix" for 90 - in which the first hyphen stands for a multiplication and the second one for an addition - is one of the worst offenders I know, and I can only applaud the later introduction of the simplifications "septante", "octante" and "nonante". The English names became more systematic when "four-and-twenty" - remember the blackbirds of the nursery song? - was replaced by "twenty-four".

Naming is one aspect, manipulation (in this case: calculation) is another one.

Though I can read the individual numbers, I have problems reading a sentence like "three hundred and twenty-eight plus one hundred twenty-three equals four hundred and fifty-one".

Roman numerals, whose general adoption prevented the rise of European science for 15 centuries, are all right for engraving on a building the year of its completion, but " $\text{CCCXXVIII} + \text{CXXIII} = \text{CDLI}$ " is at least as bad as the natural language version. In Roman numerals, the rules for calculation are just too complicated (and if you don't believe that, just try multiplication and long division in Roman numerals).

The great breakthrough came with the invention of the positional number systems. The sentence " $328 + 123 = 451$ " we can read and check because - particularly in the case of more digits - we can write the numbers on two successive lines:

$$\begin{array}{r} 328 \\ 123 \\ \hline 451 \end{array} +$$

Manipulation is immensely simplified by the

possibility of such a columnwise arrangement and the fact that the rules of digit-wise addition don't change while going from right to left. By an accident of history, I am also perfectly happy with the binary system

$$\begin{array}{r} 1010 \ 01000 \\ 11 \ 11011 \\ \hline 1110 \ 00011 \end{array} + ,$$

in which the rules for manipulation are even much simpler. (The binary addition table contains only $2^2 = 4$ entries:

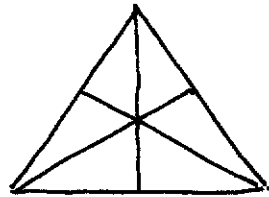
$$0+0 = 00 \quad 0+1 = 01 \quad 1+0 = 01 \quad 1+1 = 10 \quad ;$$

the decimal addition table contains $10^2 = 100$ entries, from $0+0$ through $9+9$, and as most designers of automatic computers quickly discovered, a factor of 25 is not something to be ignored.)

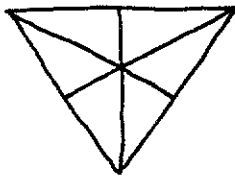
Another "language" most of us are familiar with is the one of the pictures we use in plane geometry, triangles, parallel lines and what have you. At first sight, that pictorial language seems completely adequate, for plane geometry is about things like triangles and parallel lines, isn't it? But

at closer inspection there are serious problems.

Take the following example. In connection with the theorem "In an equilateral triangle, the medians coincide with the altitudes", one could very well encounter the following picture



but we could equally well have had

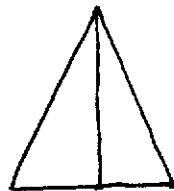


or



i.e. there is the almost always tacit convention that orientation and size of the picture don't matter. It is a little bit of a nuisance that we have to draw the picture in some size and in some orientation, and that in that sense our pictorial language forces us to be a little bit overspecific, but logical problems can be avoided by making the convention explicit that chosen size and orientation are "arbitrary", are not to be considered "essential features" of the pictures.

So far, so good, but it is not good enough! Because an equilateral triangle is in 3 different ways isosceles, the above theorem is a rather direct consequence of the theorem "In an isosceles triangle, the median from the top coincides with the altitude from the top." It is fully correct that this theorem says nothing about the top angle being 60° because the theorem holds for any top angle. Your geometry text will probably contain the picture



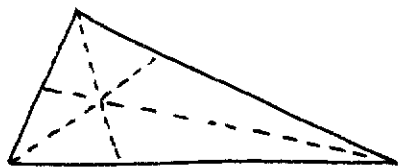
but could have contained the rather different



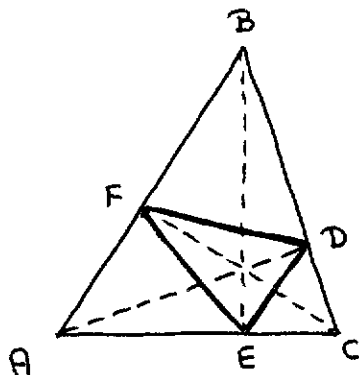
as well. Here is the problem: there is no picture of the general isosceles triangle in the sense that we cannot avoid choosing the top acute, right, or obtuse.

In the same sense there is no representative picture of the general triangle and the following example illustrates the

damage that may do. There exists the theorem that the 3 angle-bisectors of a triangle are copunctual (i.e. go through a single point), e.g



Using the above theorem, it has been shown that the altitudes are copunctual:



by showing that the altitudes of a triangle ABC - i.e. AD , BE , and CF - are the angle-bisectors of triangle DEF (and hence copunctual). The argument, however, breaks down when triangle ABC has an obtuse angle: in that case the altitudes - though still copunctual - are not the angle-bisectors of triangle DEF (at least not in the traditional sense), and a separate argument is required. (Fortunately, the case distinction forced upon us by the pictorial argu-

ment can be avoided by the use of vector calculus, which allows a one-line argument which is independent of the absence or presence of an obtuse angle in triangle ABC. It relies upon the algebraic identity

$$a \cdot (b-c) + b \cdot (c-a) + c \cdot (a-b) = 0 \quad .)$$

The main problem with the role of pictures in traditional plane geometry is that each theorem is applicable to an infinite set of figures of great variation, whereas we can only display one or two of them by way of example, leaving to the reader the conceptualization of the general case.

Above, we have given two first examples - one numerical, one geometrical - showing how ease and efficiency of reasoning depend on the "language" used. Later, when we have a clearer picture of the arguments we would like to render, we can and will be more explicit about language characteristics that ease or hamper reasoning. We now turn in this introductory chapter to another aspect of effective reasoning, viz. the subject matter to which the reasoning is applied.

Effective reasoning requires great precision and that explains why I won't deal with, say, arguments about the nature of the influence of the early singing of folk tunes on the patriotic feelings of mostly rural girls: the concepts involved would be several orders of magnitude too vague to do something rigorous with. Mathematicians prefer to deal with "simple" concepts such as numbers and triangles, rather than with messy "real-world" concepts such as patriotic feelings. This preference is not a matter of cowardice or laziness; on the contrary, and the layman should respect this self-constraint as the mathematician's analogue of the professional discipline that prevents the competent architect from erecting a high-rise building on quicksand.

The following two examples serve to illustrate that even in such "simple" contexts, our concepts should be chosen with care; again we shall give a numerical and a geometrical example.

We all know the numbers that we use in counting: $1, 2, 3, 4, \dots$. At the

right-hand side - the conceptually difficult one! - there is no difference of opinion: the sequence of what the mathematicians call "the natural numbers" goes on without end, amen. The problem is at the other end: where should it start? Should the sequence start at zero: $0, 1, 2, 3, 4, \dots$, or should it start at two: $2, 3, 4, 5, \dots$? For the classical Greeks, for instance, "two" was the smallest number; "one" existed, but was not considered a number.

Remark We are not that much different from the classical Greeks, for the sentence "Among my well-established colleagues, a number are of a different opinion." strongly suggests more than one dissenter! Please note that "number" (noun, singular) is followed by "are" (verb, plural). (End of Remark.)

As the term "multiplication" still faithfully reflects, 1 could not occur as a factor of a product - "it would not multiply" -, nor could it be a divisor - "it would not divide, would it?" -. But the Greek exclusion of 1 from the natural numbers created problems for them.

For us, any pair of positive natural numbers has a greatest common divisor, e.g. for the pair $(21, 33)$, the greatest common divisor is 3, for the pair $(21, 34)$ it is 1.

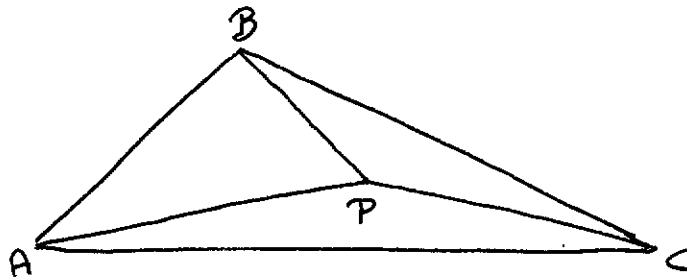
Remark "greatest common divisor" is both name and definition. The divisors of 21 are $\{1, 3, 7, 21\}$, the divisors of 33 are $\{1, 3, 11, 33\}$, their common divisors are $\{1, 3\}$, of which 3 is the greatest. (End of Remark.)

There is an efficient algorithm that determines for any pair of natural numbers their greatest common divisor (efficient in the sense that it does not require the, in general, much more laborious factorization of the two numbers). The Greeks knew it: the algorithm is, in fact, known as "Euclid's Algorithm". Euclid even proves the correctness of his algorithm, but has to give two different proofs, one for the case that — such as with $(21, 33)$ — the given numbers do have at least one common divisor, and a different proof that deals with the case — such as with $(21, 34)$ — that in Greek parlance the numbers have no common

divisor. Admission of 1 as divisor completely obviates this proof duplication.

Aside Definitions of concepts and formulations of theorems act in the modern parlance of the computing scientist as the interfaces between the modules of the argument. Computer scientists had to burn their fingers a number of times before they fully understood the critical importance of well-chosen interfaces, and Euclid's work clearly predates that experience. (End of Aside.)

Let us now turn to a geometrical example. Consider the following figure:

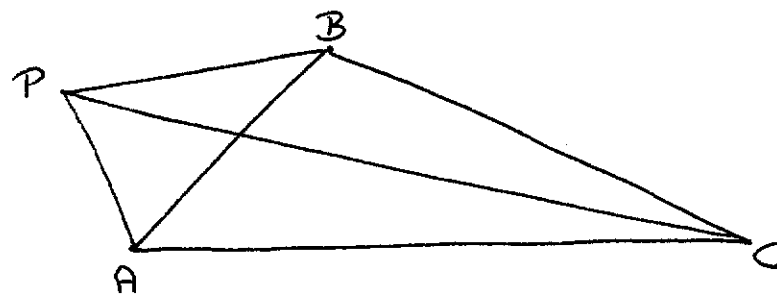


It contains 4 triangles, 3 smaller ones that together make up the big one, more precisely, the sum of the areas of the 3 small ones equals the area of the big one. Denoting with Δ the area, we have

$$\Delta PBC + \Delta APC + \Delta ABP = \Delta ABC$$

Remark Please note that this is a very nice formula: it has such a nice structure, that we can remember it. The three terms on the left are obtained by replacing in the right-hand term $\triangle ABC$ in turn A , B , and C by P . (End of Remark)

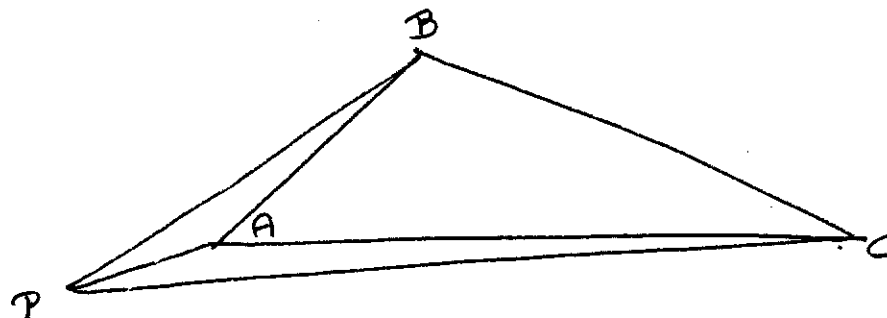
We run, however, in trouble when P does not lie inside triangle ABC . Consider for instance the figure



where we would have

$$\triangle PBC + \triangle APC - \triangle ABP = \triangle ABC,$$

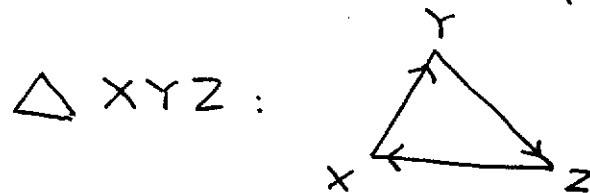
or the figure



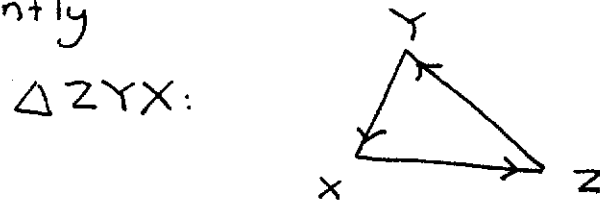
in which case we would have

$$\triangle PBC - \triangle APC - \triangle ABP = \triangle ABC.$$

The beautiful symmetry of our original formula has been destroyed in the latter ones by the occurrence of one or two minus signs at the left-hand side. Can we repair this? Yes, we can. To begin with, we let the triangle's name define a direction of traversal of the triangle's sides, to be precise, for $\triangle XYZ$ this will be from X to Y , from Y to Z , and from Z to X :



and similarly, naming the same triangle differently



Studying our three formulae and the corresponding figures, we see that in the triangles as named, the triangles with a minus sign in front of them are the ones whose perimeters are traversed counter-clockwise (or "widdershins", if you prefer). And this observation suggests the remedy: we introduce a new concept, called the "signed area", and denoted by Δ , which is the usual area in the case of clockwise

traversal, but is the negative area if the traversal is anti-clockwise.

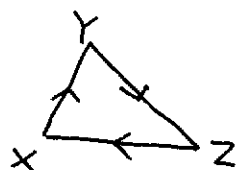
Remark Whether the traversal is clockwise, is undefined in the case of the degenerate triangle, i.e. a "triangle" with its vertices on a straight line, but in that case its area equals zero and its sign is irrelevant. Also, please note that the theorems

$$\triangle XYZ = -\triangle ZYX$$

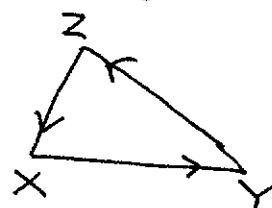
$$\triangle XYZ = \triangle YZX$$

hold independently of the sign of $\triangle XYZ$,
i.e.,

for both



and



A picture may be worth a thousand words,
a formula is worth a thousand pictures.
(End of Remark.)

In terms of the signed area, all three formulae are now expressed by

$$\triangle PBC + \triangle APC + \triangle ABP = \triangle ABC.$$

The user of the concept of the signed area has to pay the price of not listing after $\triangle$ the vertices in arbitrary order, but the concept is so much cleaner than the tradi-

tional area, that he should pay this price gladly. The traditional, unsigned area has the absolute value built in, and it is a clumsy concept because the absolute value is such an ugly function. The traditional unsigned area forces us to deal with different cases that should not have been distinguished in the first place.

The above two examples of how an unfortunate choice of concepts can complicate the reasoning should suffice at this stage, as we now pay some attention to a last aspect of effective reasoning, viz. the structure of the argument.

The purpose of the argument is always to draw a conclusion, but we wish to do so in the most trustworthy manner conceivable and we wish to present the reasoning that led us to the conclusion as convincing as possible. Mankind knows of only one way to reach those goals: constructing the argument as a finite sequence of small, explicit steps from a modest repertoire.

It has to be a finite sequence so that

the reader can check them the one after the other. The steps have to be small and explicit so that they can be checked without guesswork. Finally the repertoire of permissible steps has to be modest, firstly because it must be humanly possible to have it in our fingertips, and, secondly, because we have to accept all the steps in the repertoire as trustworthy.

Remark It is the finite repertoire of permissible steps that defines the class of valid arguments, that defines what a "proof" is. If the repertoire is empty, we can argue nothing, hence are automatically protected from the mortal sin of concluding nonsense. The larger the repertoire, the weaker the corresponding notion of "proof", and by the time the repertoire admits any step, we can conclude nonsense in a single step - we can prove anything in a single step! - and the notion of proof has become void. (End of Remark.)

Let me give here an informal illustration of a main reasoning step, known as "substituting equals for equals". Suppose that we are invited to conclude

(i) a fortnight is longer than a week .

Well, we certainly cannot do that if we don't have the foggiest notion of what fortnights and weeks are - coming from another country, we could have a very limited English vocabulary - , i.e. something about fortnights and weeks has to be given. Let these givens be

(ii) a fortnight is fourteen days

(iii) a week is seven days

Thanks to (ii) and (iii) we can eliminate the technical/mysterious/unknown terms/concepts "fortnight" and "week" from our demonstrandum (i), which, thanks to (ii) and (iii) is equivalent to

(iv) fourteen days is longer than seven days .

We won't pursue this example: if (iv) is an axiom, we are done. The point was to show that (ii) and (iii) yield the equivalence of (i) and (iv) .

Please notice that these manipulations have very little to do with the truth of the statements manipulated. I mean, (ii) and (iii) can be used as above, to yield the equivalence of (v) and (vi) with

- (v) a week is longer than a fortnight
- (vi) seven days is longer than fourteen days
or to yield the equivalence of (vii) and (viii)
with
- (vii) a fortnight is cheaper than a week
- (viii) fourteen days is cheaper than seven days.

It does not matter that (i) and (iv) seem right, that (v) and (vi) seem wrong, and (vii) and (viii) seem neither: in the context of the established equivalences, "is longer than" and "is cheaper than" are unanalysed relations, and what they seem is totally irrelevant.

As we continue our explorations, the distinction between "calculating" and "concluding" will get more and more blurred. This is to be expected; asked about the product of 17 and 323, we can "calculate" that that product is 5491, we can also "conclude" that that product is 5491: we cannot do the one without the other. We shall not hesitate to refer to the transformation of

$$a \cdot (b + c)$$

into

$$a \cdot b + a \cdot c$$

as "calculation", we shall not hesitate to regard decimal numbers such as 17, 323 and 5491 as "formulae", thus including calculation in the more general notion of "formula manipulation". This was largely a remark about terminology, our next remark, with which we close this chapter will be regarded by some as more fundamental.

The final remark is that the reasoning we are discussing deals with validity, not with truth. We may build up a theory about pyrodions, which we can do provided pyrodions have been defined in a useful manner. In order to be able to reason about pyrodions, earlier generations would say that we would need to know how to interpret statements about pyrodions, but we -as will become clear in subsequent chapters- just need to know the properties of pyrodions, i.e. no more and no less than how we may manipulate formulae containing elements that denote pyrodions.

For the purpose of our reasoning, the pyro-

dions are what our postulates define them to be, and it is about those that we can design a, let us hope beautiful, Theory of Pyrodions. And now someone familiar with our Theory encounters things very similar to what he always thought pyrodions looked like, and he would like to apply our Theory to them. If so, it is his responsibility to establish that his things enjoy all the characteristic properties that have been postulated for our pyrodions. Another way of saying that our reasoning does not deal with truth is that no theory can justify its own application, an observation that can create wonder that for any beautiful theory there seem to be Pyrodions all over the place.

Nuenen, 18 July 1996

prof. dr. Edsger W. Dijkstra
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712-1188
USA