What is Mathematics all about?

by Walter Noll

0.Introduction

This is the text of an Undergraduate Mathematics Colloquium lecture, given at Carnegie Mellon University on March 23, 2006. I started with taking a poll, asking the students to decide whether the following aspects of mathematics are unimportant, somewhat important, or very important.

1) The facility with numbers and calculations,
2) The art of avoiding unnecessary calculations,
3) Memorizing formulas and theorems,
4) Understanding and finding proofs,
5) Solving problems with well-described procedures,
6) Solving problems for which there are no well-described procedures,
7) Understanding abstract mathematical concepts,
8) Creating new mathematical concepts, and clarifying and refining old ones.

24 students participated in the poll. Here is the result:

aspect	**1)**	**2)**	**3)**	**4)**	**5)**	**6)**	**7)**	**8)**
unimportant	5	3	17	1	3	0	0	0
somewhat important	17	15	6	4	18	2	8	1
very important	2	6	1	18	3	22	16	23

Here is my own take on these aspects of mathematics:

Concerning 1): Most non-mathematicians think that this is the answer. Some mathematicians are good at this, others are terrible. (I cannot even remember my social security number, and it takes me a while to figure out what 7 times 8 is.)

Concerning 2): You probably know the story of Karl Friedrich Gauss (1777-1855) at age 7, more than 200 years ago. The teacher asked him to add all the numbers from 1 to 100 to keep him busy for a while. He immediately wrote down the answer (5050) and put it on the teacher's desk. He found a way to avoid this stupid calculation. When I graded homework problems in the past, I often used the comment UC (unnecessary calculation).

Concerning 3): I have poor memory and never memorized a single formula or theorem. Either I understood it or I didn't. When I did, I could remember it without memorizing, or at least was able to reconstruct it or knew where to look it up.

Concerning 4): This is a very important part of mathematics, but it is not all there is to it.

Concerning 5): This is a minor part of mathematics. If all you can do is to solve a problem according to some recipe, you will soon lose your job, because you can be replaced by a computer program. (An example is Mathematica.)

Concerning 6): When students are given a problem in a test, they sometimes complain: "You have never given us a problem like this before." The only problems that are worth doing by a human being rather than a computer program are the ones that require coming up with original ideas and tricks. Being able to do so is a very important part of mathematical ability and requires ingenuity. At its best, it is the ability to discover procedures (algorithms) which had never been thought of before.

Concerning 7): This is an extremely important part of mathematics. Without this ability, one cannot really be a mathematician.

Concerning 8): In my view, this is the highest form of mathematics. Consider the invention of the concepts of derivative and integral by Gottfried Leibniz (1646-1716) and Isaac Newton (1643-1727), the concept of a group originating with Évariste Galois (!811-1832), the concepts of sets and mappings introduced by Georg Cantor (1845-1918), and, more recently, the concepts of categories and functors introduced by Samuel Eilenberg (1913-1994) and Saunders MacLane (1909- 2005). (I had the privilege to meet both of them before they recently passed away.) I could give you a much longer list. In a small way, I have been able to introduce and clarify certain concepts. For example, in my doctoral thesis of 1954 I introduced the concept of a *constitutive law*. It has found very wide acceptance, and my thesis is now quoted in the Oxford English Dictionary.

I was somewhat dismayed about the discrepancy between the students' opinion and my own about the aspects 1) and 5), but I was gratified to learn that all but one think that the aspect 8) is very important. In the rest or the paper, I will elaborate on this aspect.

1.Abstraction

The discovery of features common to a variety of special situations can be used to formulate an *abstract concept*. I now give a concrete example of the creation of such an abstract concept:

In the operations listed below, we denote the set of all natural numbers (including zero) by $\mathbb{N}$, the set of all real numbers by $\mathbf{R}$, the set of all positive reals (including zero) by $\mathbb{P}$, the set of all integers by $\mathbf{Z}$, and the set of all strictly positive reals by $\mathbb{P}^{\times}$.

1) **addition:** formation of the sum $a+b$ when $a, b \in \mathbb{N}, \mathbf{R}, \mathbb{P}$, or $\mathbf{Z}$,

2) **multiplication:** formation of the product $a \times b$ when $a, b \in \mathbb{N}, \mathbf{R}, \mathbb{P}^{\times}$, or $\mathbf{Z}$,

3) **maximum:** formation of the maximum $\max(a, b)$ when $a, b \in \mathbb{N}, \mathbf{R}, \mathbb{P}$, or $\mathbf{Z}$,

4) **greatest common divisor:** formation of the greatest common divisor $\gcd(a, b)$ when $a, b \in \mathbb{N}$,

5) **union:** formation of the union $U \cup V$ when $U, V \in \operatorname{Sub} S$, the collection of all subsets of a given set S ,

6) **composition:** formation of the composite $f \circ g$ when $f, g \in \mathrm{Map}(S, S)$, the set of all mappings from a given set S to itself ,

7) **concatenation:** formation of a new word $a \asymp b$ by attaching a word b at the end of a given word a. Here, the term "word" is used to mean simply a string of symbols from a given set $\mathcal{S}$, for example $\mathcal{S} := \{0, 1\}$. We denote the set of all such words by $\mathfrak{W}$.

We observe that all of these operations satify an associative law:

$$(a + b) + c = a + (b + c) \quad \text{for all} \quad a, b, c \in \mathbb{N}, \mathbf{R}, \mathbb{P}, \text{or } \mathbb{Z} ,$$

$$(a \times b) \times c = a \times (b \times c) \quad \text{for all} \quad a, b, c \in \mathbb{N}, \mathbf{R}, \mathbb{P}^{\times}, \text{or } \mathbb{Z} ,$$

$$\max(\max(a, b), c) = \max(a, \max(b, c)) \quad \text{for all} \quad a, b, c, \in \mathbb{N}, \mathbf{R}, \mathbb{P}, \text{or } \mathbb{Z} ,$$

$$\gcd(\gcd(a, b), c) = \gcd(a, \gcd(b, c)) \quad \text{for all} \quad a, b, c \in \mathbb{N} ,$$

$$(U \cup V) \cup W = U \cup (V \cup W) \quad \text{for all} \quad U, V, W \in \mathrm{Sub}S ,$$

$$(f \circ g) \circ h = f \circ (g \circ h) \quad \text{for all} \quad f, g, h \in \mathrm{Map(S, S)} ,$$

$$(a \asymp b) \asymp c \ = \ a \asymp (b \asymp c) \quad \text{for all} \quad a, b, c \in \mathfrak{W} .$$

This is very boring. It is much better to introduce an abstract **mathematical structure** that captures the situation. Such a structure is called a **pre-monoid**. It is described by specifying two ingredients: A set M and a mapping $\mathrm{cmb} : M \times M \longrightarrow M$, called **combination**, and one **axiom**, namely the **associative law**

$$\mathrm{cmb}(a, \mathrm{cmb}(b, c)) = \mathrm{cmb}(\mathrm{cmb}(a, b), c) \quad \text{for all} \quad a, b, c \in M . \tag{1}$$

In the examples given above, the combination becomes addition, multiplication, maximum formation, formation of greatest common divisors, set-union, composition, and concatenation, respectively. It is easy to come up with many more examples of pre-monoids. Anything that can be proved about abstract pre-monoids applies to all of these examples. Thus, the use of the abstract concept can be viewed as a labor-saving device.

A somewhat more restrictive concept then that of a pre-monoid is that of a **monoid**, which may be obtained from a pre-monoid M by specifying an additional ingredient, namely an element $\mathrm{nt} \in M$, called the **neutral** and two additional axioms, namely the **neutrality laws**

$$\mathrm{cmb}(a, \mathrm{nt}) = a = \mathrm{cmb}(\mathrm{nt}, a) \quad \text{for all} \quad a \in M . \tag{2}$$

In the examples 1) and 4) given above, one designates the number 0 to be the neutral and thus obtains monoids rather than only pre-monoids. In the example 2) one designates the number 1, in 5) the empty set $\emptyset$, and in 6) the identity mapping 1_S of S to be the neutrals, respectively.

Given a pre-monoid M one can easily prove that there is *at most* one element $\mathrm{nt} \in M$ that satisfies the neutrality law (2). If such an element exists, we say that the pre-monoid M is **monoidable** and make it into a monoid by designating nt to be the neutral. If M contains no such element, one can take an object nt out of thin air, join it to M and extend the combination mapping to $M \cup \{\mathrm{nt}\}$ by *defining*

$$\mathrm{cmb}(a, \mathrm{nt}) := a =: \mathrm{cmb}(\mathrm{nt}, a) \quad \text{for all} \quad a \in M \cup \{\mathrm{nt}\} \ . \tag{3}$$

The pre-monoids described in the example 3) above fail to be monoidable except when $M := \mathbb{P}$ or $\mathbb{N}$, where 0 can serve as the neutral. In the other cases it is customary to join $-\infty$ as the neutral to make them monoids. In the example 7), one has to join an "empty word" ew as a neutral to $\mathfrak{W}$ to obtain a monoid. This empty word ew should not be confused with the empty set $\emptyset$.

A concept that is more restrictive than that of a monoid is that of a **group**, which may be obtained from a monoid M by specifying an additional ingredient, namely a mapping $\mathrm{rev} : M \longrightarrow M$, called the **reversion**, and two additional axioms, namely the **reversion laws**

$$\mathrm{cmb}(\mathrm{rev}(a), a) = \mathrm{nt} = \mathrm{cmb}(a, \mathrm{rev}(a)) \quad \text{for all} \quad a \in M \ . \tag{4}$$

In the example 1) above, in the case when $M := \mathbf{R}$ or $M := \mathbf{Z}$ one obtains groups when when the reversion is defined to be the process of taking the opposite. In the example 2) when $M := \mathbb{P}^{\times}$ one obtains a group when the reversion is defined to be the process of taking the reciprocal.

Given a monoid M one can easily prove that there is *at most* one mapping rev that satisfies the reversion law (4). If such an mapping exists, we say that the monoid M is **groupable** and make it into a group by designating rev to be the reversion. The procedure for converting a pre-monoid into a monoid has no counterpart for monoids and groups. In general, there is no procedure for converting a non-groupable monoid into a group. In all examples 1) to 7) listed above, except the ones described in the previous paragraph, the monoids or pre-monoids fail to be groupable.

A pre-monoid, monoid, or group M is said to be **commutative** if the additional axiom

$$\mathrm{cmb}(a, b) = \mathrm{cmb}(b, a) \quad \text{for all} \quad a, b \in M \tag{5}$$

is satisfied. The structures in the examples 1) to 5) are commutative, but those in 6) and 7) are generally not.

For all of the mathematical structures just described as well as many others, it is useful to consider the concept of a **substructure**, which is based on a subset of the underlying set, a subset that is **stable** with respect the ingredients specified. To make this explicit, we need a few basic notations:

1) Given a set S, we denote the set set of all subsets of S by $\mathrm{Sub}\, S$.

2) Given a mapping $f : A \longrightarrow B$ with domain A and codomain B, we define

the **image mapping** $f_> : \mathrm{Sub}\, A \longrightarrow \mathrm{Sub}\, B$ by

$$f_>(U) := \{f(x) \mid x \in U\} \quad \text{for all} \quad U \in \mathrm{Sub}\, A \ . \tag{6}$$

3) Let $U \in \mathrm{Sub}\, A$ and $V \in \mathrm{Sub}\, B$ be such that $f_>(U) \subset V$. Then the **adjustment** $f|_U^V : U \longrightarrow V$ of f is defined by

$$f|_U^V(x) := f(x) \quad \text{for all} \ \ x \in U \ . \tag{7}$$

Let M be a pre-monoid. To say that a subset H of M is stable under the combination cmb of M means that

$$\mathrm{cmb}_>(H \times H) \subset H \ . \tag{8}$$

If this is the case, H is endowed with the structure of a pre-monoid by designating the adjustment $\mathrm{cmb}|^H_{(H\times H)}$ to be the combination mapping of H. H is then called a sub-pre-monoid of M.

If M is a monoid, to say that a subset H is stable with respect to the ingredients cmb and nt of M means that, in addition to (8), we have

$$\mathrm{nt} \in H \tag{9}$$

If this is the case, H is endowed with the structure of a monoid by designating its neutral to be nt. H is then called a submonoid of M.

If M is a group, to say that a subset H is stable with respect to the ingredients cmb, nt, and rev of M means that, in addition to (8) and (9), we have

$$\mathrm{rev}_>(H) \subset H \tag{10}$$

If this is the case, H is endowed with the structure of a group by designating its reversion mapping to be the adjustment $\mathrm{rev}|^H_H$. H is then called a subgroup of M.

In the example 1) of addition when $M := \mathbf{R}$, the case when $H := \mathbb{Z}$ gives a subgroup, and the cases when $H := \mathbb{P}$ or $\mathbb{N}$ give submonoids that fail to be groupable. In the example 2) of multiplication when $M := \mathbf{R}$ the case when $H := \mathbb{P}^\times$ gives a groupable submonoid, and the cases when $H := \mathbb{Z}$ or $\mathbb{N}$ give submonoids that fail to be groupable. In the example 3) of maximum formation when $M := \mathbf{R}$, the cases when $H := \mathbb{P}$, $\mathbb{Z}$, or $\mathbb{N}$ give sub-pre-monoids. In the example of 5) of union formation, one obtains a submonoid by replacing the given set S by any one of its subsets. In the example 6) of composition, one obtains a groupable submonoid by considering only the *invertible* mappings in $\mathrm{Map}(S, S)$. The group so obtained is called the **permutation group** of S and is denoted by Perm$S$. In the example 7) of concatenation, one obtains a submonoid by considering words that are strings of symbols from a subset of $\mathcal{S}$.

We note that a monoidable sub-pre-monoid of a monoid need not be a submonoid. For example, the singleton $\{0\}$ is a sub-pre-monoid of the multiplicative

monoid $\mathbb{N}$ described in the example 2) of this section. It is monoidable, in fact trivially groupable, but is not a submonoid because it does not contain the multiplicative neutral 1. In the example 5), every singleton $\{U\}$ with $U \in \operatorname{Sub} S$ is a sub-pre-monoid that is trivially groupable, but itis not a submonoid unless $U = \emptyset$. However, one can prove that a groupable sub-pre-monoid of a group must be a subgroup.

2. Clarification

Here I wish to illustrate, by an example, how mathematics can serve to clarify and render precise a familiar concept, namely the concept of **volume**. My dictionary * lists 8 meanings for the word "volume", but we will consider only the one described by "the amount of space occupied in three dimensions; cubic contents or cubic magnitude".

Everybody, mathematician or not, is familiar with this concept. After all, we buy water or gasoline by the liter or gallon, and containers of milk, juice, or oil display their volume content on the label. We are intuitively familiar with the properties of this concept.

Volume is merely the three-dimensional version of a general concept, whose two-dimensional version is called *area*, described in the dictionary as "a measure of a bounded region on a plane". Actually, we now know that there is a version for any finite dimension. We also use the term **volume** for this general concept, so that *2-dimensional volume* becomes just another term for area.

The general concept of volume, in any dimension, is a mapping that assigns to every suitable region a positive number, called the volume of that region. The assignment depends on the choice of a unit, i. e. a fixed quantity used as a standard, such acre, hectare, etc in 2 dimensions, and liter, gallon, etc. in 3 dimensions. The following three principles are taken for granted:

(P1) If a region is divided into several pieces, then its volume is the sum of the volumes of the pieces.
(P2) Congruent regions have the same volume.
(P3) The volume of a rectangular box is the product of the lengths of all edges adjacent to some corner.

In the case of dimension 2, a rectangular box reduces to a rectangle.

The principle (P3) connects the concept of volume to that of length, which also depends on the choice of a unit, such as centimeter or inch. Therefore cubic centimeters or cubic inches can be used as units of 3-dimensional volume, and square centimeters or square inches can be used as units of area.

The three principles are sufficient to derive formulas for the volume of regions having a simple shape. Also, in dimension 2, one can give a proof of the Theorem of Pythagoras based only on these principles. I believe that this is the approach that should be presented in high-school geometry classes. I have elaborated on this idea in reference [2]. The principles are also sufficient for deriving

* Websters New World Dictionary, 1982

formulas for the areas and volumes of certain regions with curved boundaries, such as circular discs and spherical balls. However, such derivations involve sophisticated ideas, which were extensively used by Archimedes (287-212 B.C.), the greatest mathematician in antiquity. *

You probably heard the following story: Archimedes was in a public bath in Syracuse and suddenly had an insight. He jumped out into the street naked and shouted "eureka", which is Greek for "I found it". It is likely that his insight was the principle that the volume of a quantity of liquid, such as water, does not change if the liquid changes shape. This principle can be used to determine the volume of an odd-shaped solid object by immersing it in a container of water. By the priciple (P1) above, the volume of the object is the same as the volume of water displaced by the object, i.e. the difference between the volume after immersion and before immersion. This difference could be determined by measuring the change of water level in the container or the overflow if the container was originally full. Another way may be the application of what is now known as Archimedes' Principle: The weight of the displaced water is equal to the buoyancy, i.e., the difference between the weight of the object before and after immersion. Since the specific weight (weight per unit volume) of water is known, one can calculate the volume of the object by measuring the buoyancy. The king of Syracuse had ordered a crown, to be made of solid gold, and wanted to know whether it had been debased by a hidden use of another metal. Archimedes tested this possibility by determining the volume of the crown using one of the methods just described. By weighing the crown, he could then determine its specific weight. Since the specific weight of gold was know to be larger than the specific weight of other metals, he could test wether there was any other metal hidden in the crown.

A systematic mathematical method for determining the volume of odd-shaped regions was developed much later, with the invention of integral calculus by Newton and Leibniz. It was Leibniz, in 1675, who introduced the now standard integral symbol $\int$. Leibniz viewed integrals as limits of sums, as did Augustin-Louis Cauchy (1789-1857) in 1823, and Georg Friedrich Bernhard Riemann (1826-1866) in 1854, who gave more rigorous treatments of this concept. Integrals should be carefully distinguished from antiderivatives, i.e. functions whose derivatives are a given function. Unfortunately, such antiderivatives are often called "indefinite integrals", a very misleading term. A priori, antiderivatives have nothing to do with integrals; they are connected to integrals only by the Fundamental Theorem of Calculus, which has a non-trivial proof.

An important conceptual advance was made by Henri Lebesgue (1875-1941) in 1902, who introduced what is now called the *Lebesgue measure* of a subset of space. It is an extension of the concept of volume that applies not only to what we have called "regions" but to much more general (measurable) subsets of a point-space. This led to an entire new branch of mathematics, called *measure theory*, which has now become the title of a standard graduate course.

---

* In the historical remarks here, I rely on reference [1].

In my view, the treatment of volume and volume integrals is very unsatisfactory in all present textbooks. I have proposed a new approach in reference [3]. It is based on the following concepts:

1) A finite dimensional *flat space* (a.k.a *affine space*) $\mathcal{E}$ with translation space $\mathcal{V}$, which is a commutative subgroup of $\mathrm{Perm}\mathcal{E}$, described with additive notation.
2) The precise definition of a *negligible subset* of $\mathcal{E}$.
3) The collection $\mathrm{Bnb}\,\mathcal{E}$ of all subsets of $\mathcal{E}$ than are bounded and have a negligible boundary.
4) The concept of an *almost continuous function* $f : \mathcal{E} \longrightarrow \mathbf{R}$, which is a function whose set of discontinuities is negligible.
5) The collection $\mathrm{Bbac}\,\mathcal{E}$ of all functions that have bounded support and bounded range, and are almost continuous.

The characteristic function $\mathrm{ch}_{\mathcal{A}}$ of a subset $\mathcal{A}$ of $\mathcal{E}$ belongs to $\mathrm{Bbac}\,\mathcal{E}$ if and only if $\mathcal{A}$ belongs to $\mathrm{Bnb}\,\mathcal{E}$.

The collection $\mathrm{Bbac}\,\mathcal{E}$ is a function space, i.e., it is stable under value-wise addition and value-wise multiplication with real numbers. Given $f, g \in \mathrm{Bbac}\,\mathcal{E}$, we interpret $f \leq g$ as a value-wise inequality, i.e.

$$f \leq g :\Longleftrightarrow \ (f(x) \leq g(x) \quad \text{for all} \quad x \in \mathcal{E}) \ . \tag{11}$$

Here is my definition of integral and volume:

Definition 1. *A non-zero linear functional* Igl *on the function-space* $\mathrm{Bbac}\,\mathcal{E}$ *is called an* **integral** *on* $\mathcal{E}$ *if it is isotone in the sense that*

$$f \geq g \ \ \Longrightarrow \ \ \mathrm{Igl}(f) \geq \mathrm{Igl}(g) \qquad \text{for all} \ \ f, g \in \mathrm{Bbac}\,\mathcal{E} \tag{12}$$

and translation-invariant in the sense that

$$\mathrm{Igl}(f \circ \mathbf{v}) = \mathrm{Igl}(f) \qquad \text{for all} \quad f \in \mathrm{Bbac}\,\mathcal{E}, \ \mathbf{v} \in \mathcal{V}. \tag{13}$$

Definition 2. *Given an integral* Igl *on* $\mathcal{E}$*, the mapping* $\mathrm{vol} : \mathrm{Bnb}\,\mathcal{E} \longrightarrow \mathbb{P}$ *defined by*

$$\mathrm{vol}(\mathcal{A}) := \mathrm{Igl}(\mathrm{ch}_{\mathcal{A}}) \quad \text{for all} \ \ \mathcal{A} \in \mathrm{Bnb}\mathcal{E} \tag{14}$$

is called the **volume-function** *associated with* Igl.

The proof of the following theorem is very difficult and involves sophisticated approximations of integrals by sums:
Theorem on Existence and Uniqueness of Integrals. *There is an integral* Igl *on* $\mathcal{E}$*. A functional* $\mathrm{J} : \mathrm{Bbac}\,\mathcal{E} \longrightarrow \mathbf{R}$ *is also an integral if and only if* $\mathrm{J} = \mathrm{c}\,\mathrm{Igl}$ *for some* $c \in \mathbb{P}^{\times}$.

The indeterminacy of the factor c in this Theorem reflects the fact that one has to agree on a unit to give volumes definite numerical values.

Once the Theorem has been taken for granted, one can develop the entire theory of integrals and volumes by using no more than the two definitions above,

and one does not have to refer anymore to limits of sums. This has been done in reference [3].

One important insight can be gleaned from this approach. The definition of a negligible set I give in [3] does not require that the space have a Euclidean structure. In a Euclidean space one has the notions of distance between points and of lengths of line segments. In a flat space, one has no such notions. Hence the definition of volume above has, a priori, no connection with length. One can extend the volume function defined above to a Lebesgue measure. This Lebesgue measure does not depend on whatever Euclidean structure $\mathcal{E}$ may posess in special cases. One can prove that a set is negligible if and only if it is bounded and its closure has measure zero. (To *define* negligible sets to be bounded sets whose closure has measure zero is circular.)

This insight may be of importance when the space is the 4-dimensional event-world of special relativity. Here, distances between arbitrary events cannot be defined, yet one can define the 4-dimensional volume of an arbitrary region in the event-world.

References

[1] Kline, M.: *Mathematical Thought from Ancient to Modern Times*, 1238 pages, Oxford University Press 1972.

[2] Noll, W.: *Mathematics should not be boring*, 18 pages, lecture given before the mathematics teachers of the Catholic schools in the Diocese of Pittsburgh, 2003. Posted on the website math.cmu.edu /∼wn0g/noll.

[3] Noll, W.: *Volume Integrals*, 41 pages, to be included as Chapter 4 in a future book entitled *Finite-Dimensional Spaces, Vol.II*. (Vol.I was published in 1987.) Posted on the website math.cmu.edu/∼wn0g/noll.