# Chapter 4 Constructing Structures

## § 1. Structures

In chapter 3 a class of "funny numbers of the form $x \pm y$" was invented. I then said there were a great many questions about these numbers we might like to ask: What rules do we want for determining whether some particular $x \pm y$ is to be included in the class? Can we define an operation called "addition" for them? if so, how does it work? and is the result also a member of the class? Can we define "division" and "multiplication" operations that can be performed on this class? and, if so, how? Does the sum of three "funny numbers" obey an associative property of addition? Can we work out an algebra for them?

When a mathematician defines, works out and constructs rules for answering questions like these, he is said to be *structuring* them formally. The ideas and rules he comes up with constitute secondary quantities in what is called a "formal system." For the case of funny numbers of the form $x \pm y$, it would be called "the number system for $x \pm y$." Every useful species of numbers has defined for it a number system. Examples include the system of integers, the real number system, the imaginary number system, the complex number system, and the binary number system. Because there seems to be no limit to how many different classes of "numbers" human beings can invent, there is likewise no limit to how many different "number systems" there can be. The *set* of numbers in a number system is said to make up the "domain" of mathematical systems that are based on these numbers.

"Structure" is one of the most important ideas in mathematics. Indeed, mathematicians have definitions for many *species* of structures. These broadly include "mathematical structure," "mathematical logic structure," and "category theory structure." The association of mathematicians who wrote under the pseudonym "Nicolas Bourbaki" saw the notion of "structure" as being fundamental. Other examples of mathematical structures include "measures" (generalizations of the ideas of length, area, and volume), algebraic structures, topologies, metric structures (geometries), "orders," "events," equivalence relations, and differential structures. The Bourbaki tell us all of mathematics can be constructed out of just three fundamental "mother structures" and their combinations. These are: algebraic structures, ordering structures, and topological structures.

But what is a "structure" – the thing that stands as genus to these various species of structures? Mathematics has a formal definition of "structure" but how informative this definition is can be debated. As a genus, we are told,

> A **structure** can be defined as a triple $\mathcal{A} = (A, \sigma, I)$ consisting of a **domain** $A$, a **signature** $\sigma$, and an **interpretation function** $I$ that indicates how the signature is to be interpreted on the domain.

Provided one already knows what a domain, a signature, and an interpretation function are, this formal definition is constitutive and at least as clear as mud. For many people this definition would be equally informative if it were written in Mayan hieroglyphics.

The principal drawback of this formal definition is that it doesn't really provide a person who is first encountering it with a practical explanation of what it is. In Critical epistemology, a **structure** is

> a system of self-regulating transformations such that no new element engendered by their operation breaks the boundaries of the system and that the transformations of the system do not involve elements outside it; the system may have sub-systems differentiated within the whole of the system and have transformations from one sub-system to another.

This explanation of a "structure" was proposed by Piaget. He wrote,

> When we speak of 'structures,' in the most general sense of the term (mathematical, etc.), our

> definitions will nevertheless remain limiting in the sense that it will not include any static 'form' at all. We shall, indeed, give to this idea the following three characteristics: a structure implies first of all laws of totality distinct from those of its elements, which even permit complete disregard of those elements; secondly, its properties as a whole are laws of transformation as contrasted with any formal laws; thirdly, every structure implies an *autoréglage*[1] in the double sense that its combinations do not go outside its own frontiers and that they make no appeal to anything outside such frontiers. However, this does not prevent the structure from being able to divide itself into sub-structures which inherit its characteristics but at the same time show their own individual characteristics. [Piaget (1970), pg. 7]

A **transformation** is

> an action in which one representation is changed into another representation.

From these we can say that **structuring** is

> the act of putting into effect the operation of one or more of the self-regulating transformations in a structure.

This explanation of a "structure" is perhaps almost as murky as the formal mathematical definition cited above. Let us see if we can break it down a bit. In earlier chapters we already gave an explanation for what a "system" is (a set of interdependent relationships constituting an object with stable properties, independently of the possible variations of its elements). A "transformation" more or less corresponds to the "interpretation function" in the formal definition. "Laws of transformations" roughly correspond to "signature" in the formal definition. A "structure" is an example of a system characterized by the limited transformations carried out within it.

Even so, all of this might – and to most people probably does – remain somewhat meaningless because the "explanation" is non-practical, i.e., doesn't really tell you how to put it into effect. To really understand it a person needs to understand "structuring" first because a "structure" is an *outcome* of structuring *activities*. To put this another way, the explanations above are *abstract*. The idea of a structure didn't just jump into someone's imagination from out of nowhere. Instead, the idea of this object *grew out of* some small number of examples of things that shared some particular characteristics and exhibited some pattern preserved *in* the idea of a "structure." Critical epistemology tells us human beings learn *from* the particular *to* the general and not the other way around. Only *after* higher concepts have been synthesized can one find additional *lower* concepts to stand under it. Higher concepts are synthesized *from* lower ones.[2] Therefore, let us see if a more satisfactory explanation of what a "structure" is can be achieved by looking at some examples of things we will call "structures."

## §2. The Principle of Permanence

When I asked myself, "What examples should I choose in order to illustrate what structures are?" I was a bit surprised by how difficult it was for me to make these choices. There is no shortage of mathematical structures to choose *from*, and the ones we use most often are also the ones we most take for granted. It is not technically difficult to present the idea of, say, a "mathematical group structure" mathematically. But to explain *why* I would choose this example turned out to be a different, and more perplexing, kind of

[1] in English, "autotune": a device or facility for tuning something automatically.

[2] It is an unfortunate fact that most modern mathematics textbooks reverse this order. They tend to present "the idea in general" first and might (or might not) only afterwards present examples of it. Poincaré called this kind of pedagogy "contrary to all healthy psychology" [Poincaré (1914), pp. 144-145]. If you have found your own personal experience with mathematics textbooks and math classes more frustrating than illuminating, the real reason for this is that they are written and presented in exactly the opposite way that human beings learn new ideas.

question altogether. Mathematical structures are *constructed* but they are not constructed arbitrarily for no reason. Might it perhaps be better, I asked myself, to start with a principle that could be easily stated and understood by people who are not professional mathematicians? I convinced myself that it would.

The mathematical world is an invented world – a world invented by human beings to satisfy practical needs and purposes. "Numbers" are things mathematicians invent over and over again, but they are not their only inventions. Structures are invented to go along with them, and the ideas of structures tend to get progressively more abstract the more mathematicians develop them. Why is this? Rózsa Péter offered a quite neat and tidy explanation:

> Man created the natural number system for his own purposes, it is his own creation; it serves the purposes of counting and the purposes of the operations arising out of counting. But once created, he has no further power over it. The natural number series exists; it has acquired an independent existence. No more alterations can be made; it has its own laws and its own peculiar properties, properties such as man never even dreamed of when he created it. The sorcerer's apprentice stands in utter amazement before the spirits he has raised. The mathematician 'creates a new world out of nothing' and then this world gets hold of him with its mysterious, unexpected regularities. He is no longer a creator but a seeker; he seeks the secrets and relationships of the world which he has raised. [Péter (1957), pg. 22]

The more one adventures into this mathematical world, and the more questions it occurs to him to ask, the more complicated and – at the same time – the more simple the answers he comes up with become. He discovers new "laws" of mathematics and he invents new mathematical entities to go with them that bring greater generality to these "laws." But as the mathematical world grows and diversifies, we can identify an overriding principle that governs its creation. Péter wrote,

> Whenever we introduce new numbers or new operations, we should always see that they obey the old laws, since the reason we introduce them is to make the procedures more uniform. We do not want to have to split our operations into different types, depending on whether the new numbers or operations do or do not occur. This regard for the extension of old, established concepts is called the 'principle of permanence'.
>
> The natural-number series was a spontaneous creation. It was the breaking down of the structure, which had worked quite well in the past, that led to the conscious creation of new numbers. It is the shape of the structure which is helpful in the process. The framework for the new numbers is given exactly by the laws derived from the old numbers, and we are unwilling to abandon these laws if we can help it. This is the signpost in this conscious creation: we must shape the new numbers in such a way as to fit them well into this predetermined shape. [*ibid*., pg. 87]

Out of the idea of the natural numbers (1, 2, 3, etc.) there grew the ideas of the integers, fractions, rational numbers, irrational numbers, complex numbers, vectors, and even more exotic ideas such as the idea that polynomials could be seen as and used as numbers. Along with new numbers come new operations on them, and with these new operations come procedures ("algorithms") for carrying out these new operations. Watching over all of these we find the principle of permanence guiding our labors.

Let us look at an example most of us take for granted today: the idea of an operation called 'division'. This idea is of ancient origin, and it is beyond reasonable doubt that it had its beginnings in very concrete practical needs. Let us suppose that sometime in the remote past there was a herdsman who had three sons. We'll call them "eldest son" (ES), "middle son" (MS), and "youngest son" (YS). Let us also suppose he had eight goats he wanted to "divide up" among his three sons as, say, inheritances. Such situations probably came up countless times in antiquity. Our herdsman might have solved this "division problem" in the following way: one goat to ES; one goat to MS; one goat to YS; another goat to ES; another goat to MS; another goat to YS. At this point he has only two goats left. He could give a third goat to ES and a third goat to MS but this would mean YS would receive only two. What should our herdsman do?

This is a classic example of a "division problem" in antiquity. Let me use "÷" as an abbreviation for the phrase "divided by." The herdsman's problem could then be stated "in the language of mathematics" as

"8 ÷ 3 = 2 with 2 remaining" or "8 ÷ 3 = 2 remainder 2."

This is the "quotient and remainder" form of division we teach little schoolchildren to solve according to the time honored form "dividend divided by divisor equals quotient plus remainder." If we use "$a$" to represent the dividend, "$q$" to represent the quotient, "$b$" to represent the divisor, and "$r$" to represent the remainder then this division problem could be stated as "given $a$ and $b$, find $q$ and $r$ such that $r < b$ and

$$a = q \cdot b + r \, .$$

This is a problem simple enough that little children can understand it. But in the march of mathematical invention and discovery, *new* "numbers" were invented and the idea was generalized as a structure called a "Euclidean domain." Let me jump ahead to the modern definition of a Euclidean domain. It is:

> A Euclidean domain is a set (of numbers) $D$ with two binary operators "+" and "·" that satisfy the following:
>
> 1. $D$ forms an additive commutative ring with (additive) identity 0.
> 2. If $a \cdot b = b \cdot c$ with $b \neq 0$, then $a = c$.
> 3. Every element $a$ in the set $D$ has an associated metric $g(a)$ such that
>    a. $g(a) \leq g(a \cdot b)$ for all nonzero $b$ in the set $D$, and
>    b. for all nonzero $a$ and $b$ in the set $D$ with $g(a) > g(b)$, there exist $q$ and $r$ such that $a = q \cdot b + r$ with $r = 0$ **or** $g(r) < g(b)$. $q$ is called the quotient and $r$ is called the remainder.

Wow. There sure seems to be a huge difference between the herdsman's problem and the definition of the *structure* called a Euclidean domain, doesn't there? If you don't know what an "additive commutative ring" and an "associated metric" are, you probably can't even understand the definition. In between the herdsman and the "Euclidean domain" a lot of new things had to be invented: additive commutative ring; additive identity; metric (function). Even Euclid – in whose honor a Euclidean domain is named – didn't know this definition although he did know the idea of $a = q \cdot b + r$. Everything else in this definition was invented to handle new "funny numbers" as they were invented and to make them "behave the same way" as natural number division into quotients and remainders did. This is the principle of permanence in action.

An example of this is provided by polynomial long division – the division of one polynomial, $P_1(x)$, by another polynomial, $P_2(x)$. For specificity, suppose $P_1(x) = x^3+5x^2+7x^1+3x^0$ and $P_2(x) = x^1+2x^0$. $P_1(x)$ and $P_2(x)$ can be regarded as "funny numbers" for which arithmetic operations (addition, subtraction, multiplication, and division) can be defined. Algebra was invented by Islamic mathematicians beginning in the 9th century AD. The idea of applying arithmetic to polynomials was introduced in the 12th century by the Islamic mathematician Al-Samawal in his book *al-Bahir fi'l-jabr* ("The Brilliant in Algebra"). Al-Samawal's interest, as stated in his book, was "with operating on unknowns using all the arithmetical tools in the same way as the arithmetician operates on the known." His work began the study of what today we call "polynomial rings." Today the mathematics of polynomial rings, and polynomial division, is heavily used by engineers in developing error-detecting and error-correcting codes used by modern high-performance communication systems and by data storage devices such as the disk drive in your computer.

Suppose we wish to divide $P_1(x)$ by $P_2(x)$. This is done by performing long division in a Euclidean domain in a way not all that different from how you would perform long division on two integers. The result is

$$P_1(x) \div P_2(x) = (x^3+5x^2+7x^1+3x^0) \div (x^1+2x^0) = \text{quotient } q = (x^2+3x^1+x^0);$$
$$\text{remainder } r = x^0 \, .$$

The "metric" $g(P(x))$ of a polynomial $P(x)$ is merely the degree of the polynomial (i.e., the exponent of the highest nonzero power of $x$). Thus, $g(P_1(x)) = 3$ and $g(P_2(x)) = 1$ and so $g(P_1(x)) > g(P_2(x))$.

Al-Samawal's application of arithmetic to "unknowns" (polynomial expressions) is an excellent example of the principle of permanence in action. He probably didn't explicitly think of polynomials as "numbers" (funny or otherwise), but he clearly thought of them as being enough "like numbers" that arithmetic could be applied to them. But if they are "enough like numbers" for this, then as a practical matter they *are* "numbers." That they are *useful* "numbers" is a direct consequence of applying the principle of permanence to how they are defined and manipulated.

Most scientists and engineers tend to think of themselves as "users and consumers of mathematics" rather than as "developers of new mathematics." By doing so they greatly improve the efficiency with which they can carry out their scientific or engineering tasks-at-hand. Nonetheless, if a scientist or an engineer "closes his mind" to the notion that mathematics is invented and developed, and instead adopts an attitude that there is something so "sacred" about "numbers" that any notion of manipulating objects to *make* them "be like numbers" is heresy of the highest order, then he closes off any opportunity to make a scientific or engineering breakthrough that he might otherwise have been able to achieve. Permit me to indulge for a moment in a personal anecdote.

Many years ago, when I was a young engineer, I was working with a slightly younger colleague to see if we could find a way to invent a particular kind of new measuring instrument that could do something that had never been done before. Our boss liked to refer to what we were doing as "blue sky research" because he knew we'd need some "breakthrough idea" in order to succeed. One day I hit upon an idea by which it might be possible to look at the measurable signals we were working with in such a way that we might adapt some well-known measurement techniques to solve the problem *if* we looked at these signals in terms of them being a different kind of number ("funny" numbers). After about a week of playing around with the idea, I was pretty convinced it would work – barring some unexpected issue – and I tried to explain the idea to my colleague. To my surprise, he was horrified. "You can't do that !" he exclaimed. "You're messing with *the number system* !" His reaction was not unlike how an archbishop might react to an act of blasphemy committed in church. And yet, "numbers" are not sacred things.

I think a few remarks are in order concerning the notion of "manipulating" the members of a set of objects in accordance with the principle of permanence. This is because whether or not the members of an arbitrary set of objects can "be made to serve *as numbers*" depends on whether or not they can be manipulated in some way that is analogous to some other set of objects that *are* regarded as "numbers."

Mathematicians like to say they are "operating" on members of a set of objects instead of saying they are "manipulating" them, but this is more or less merely a question of semantics and reflects their prudent choice to conform with the conventions of a technical language mathematicians employ in their science. One good example of such manipulation – and one that I think clearly expresses the importance of what mathematicians call "limits" – is the manipulation used to make rational numbers "be equivalent to" fractions. When I was a little schoolboy first learning about "long division" and how to convert a fraction into a rational number, I ran into a perplexing question that, at the time, made me believe there was "something wrong with arithmetic" as it was being used for this conversion task. It bothered me quite a lot at the time, and it was only years later that I learned how the paradox is resolved.

There were two principles my teacher was teaching us. One was that a rational number is defined by carrying out long division on a fraction, and that the outcome of this "is the same as" the fraction. The second principle was that the "proof" this is true was the "fact" that multiplying a fraction by the divisor gave the dividend as the answer. It followed as a corollary that multiplying the rational number by this same divisor resulted in this same dividend. For example, $\frac{1}{4} = 0.25000\ldots$ and $0.25000\ldots \times 4 = 1$. Simple and obvious, neat and tidy, I thought.

The issue for me arose when I applied this idea to the fraction $\frac{1}{3}$. Long division gave me $0.3333\ldots$ with the "…" denoting that this sequence of threes never ended. Then I tried multiplying 0.3333… × 3. It seemed obvious to me that 0.3333… × 3 = 0.9999… and that 0.9999… was obviously not the same as 1. This alarmed me so much that I asked my teacher about it. She explained to me that "0.9999…" was "so close to 1 that there is no practical difference between them." I admitted this was true, and that gave me enough reassurance to trust that what she was teaching us was true; but the issue continued to nag at me for years to come until I finally learned how mathematics *makes* 0.9999… = 1 (even if this statement does sound like something very silly and even absurd). This is how it's done:

1. The whole number "1" used in the fraction is defined to be "the same" as the rational number 1.000… and we understand the symbol "…" means that this string of "0" numbers never ends. We further understand that a rational number like 1.000 is the same as what is indicated by 1.000… in every way.

2. Now we look at the subtraction 1.000 – 0.999 = 0.001; this is obviously true and there is nothing alarming about it.

3. Next we look at the subtraction 1.0000 – 0.9999 = 0.0001; again, this is clearly true and not at all mysterious. We compare this result with the 0.001 result from step 2 and we see that adding another "9" to the subtrahend resulted in adding another "0" to the resulting difference.

4. Now we do it again: 1.00000 – 0.99999 = 0.00001. We see this same numerical pattern once again. Every time we tack an additional "9" to the end of the subtrahend, the difference gets smaller by a factor of 10. Sooner or later, we realize that this characteristic of the operation repeats without limit, e.g.

$$1.000\ldots00 - 0.999\ldots99 = 0.000\ldots01$$

if there are the same number of "0"s in the minuend as there are "9"s in the subtrahend, and the number of "0"s in the difference is one less than the number of "0"s in the minuend.

5. Now we come to the key realization: There is nothing that, in theory, prevents us from continuing to do this again, and again, and again, and that the pattern of "the difference gets smaller by a factor of 10 each time" continues to be manifested. Every step in the procedure "pushes" the "1" at the end of the difference another step further to the right of the decimal point. We can say, as the mathematician does, "as the number of steps in this process continues without end, *in the limit* 1.000… – 0.999… = 0.000… and therefore the minuend 1.000… *is the same as* (is equal in value to) the subtrahend 0.999…".

Perhaps you have already recognized that this argument *is just like* the argument used earlier in this treatise to explain the idea of a "point" in geometry. We recognize *from the pattern* appearing in the difference that if the process never stops then the difference "becomes smaller than anything." In mathematics an argument like this is called "mathematical induction." It is a crucial way of manipulating the members of a set of mathematical objects. The Critical distinction in how we understand this notion of mathematical induction and the way some mathematicians since the days of Cantor have understood it is *the process can not be stopped after* ***any*** *number of steps* if the conclusion reached by induction is to be regarded as true. It is true only "in the limit of" a nonterminating process. In Critical epistemology the notion of mathematical infinity (symbolized as "∞") is not the idea of a "number"; it is the idea of a "becoming." By this I mean, in the case of this example, an actual difference "becomes more like 0" as more steps are undertaken.

Playing with the notion of "infinity" is "tricky" – all the more so because it is very easy to make mistakes while doing so. Here is one amusing example I have been known to spring on my engineering colleagues once in awhile on social occasions:

1. We all know that 0 = 0. We also all know that 0 + 0 = 0 and, in fact, 0 + 0 + 0 + … = 0 no matter how many 0s we add together.

2. But we also know that 1 – 1 = 0. Therefore 0 = 1 – 1 + 1 – 1 + … no matter how many times we

repeat adding the 1 – 1 operation.

3. Furthermore, we know addition and subtraction obey the associative property of arithmetic and therefore (1 – 1) + (1 – 1) + (1 – 1) + ... = 1 – (1 – 1) – (1 – 1) – … = 1 – 0 – 0 – 0 – …; therefore 0 = 1.

Even in cases where my colleagues have consumed a large quantity of malt beverages, every one of them instantly recognizes this conclusion is not merely false but is utterly absurd. But everyone I've ever pulled this on becomes perplexed about *how they know* it is false and absurd. I ask them to "point out the mistake in my reasoning" and they get a bit flummoxed when they can't find it.

Perhaps you have already spotted the mistake; mathematicians *instantly* spot it.[3] The mistake is in the misuse of the innocent looking little symbol "…". Every time we add another "1" to the sequence, the result *oscillates* back and forth between 0 and 1. For instance, 1 – 1 = 0; 1 – 1 + 1 = 1; 1 – 1 + 1 – 1 = 0 and so on. The "trick" in this little prank is in the 1 – 0 – 0 – 0 – … step just before saying 0 = 1. By putting that little "…" in, a great many people look at that step and assume "everything denoted by '…' is a 0." But that isn't true and, in fact, *misuses* the "…" symbol to *falsely* imply the pattern it denotes is always (1 – 1). An important part of pulling off the prank is to conduct the victim *step-by-step* through the argument so he fails to notice the forest for the trees. My colleagues who fall victim to this prank are every bit as flummoxed as I was that day in my arithmetic class when I "proved" 1 = 0.999999… to my great consternation.[4]

Whenever the symbol "∞" makes an appearance in mathematics – even if it is just the implied ∞ denoted by "…" – it is prudent for you to "go on high alert" for the possibility that you might be about to make some usually *very* subtle mistake. (If it was an obvious mistake, you wouldn't make it, would you?). In point of fact, there is an entire branch of mathematics, called "analysis," that is devoted to finding and fixing problems introduced by the use of "∞", infinite series, and other "limit" arguments. In carrying out their work, analysts are guided by the principle of permanence.

## §3. Algebraic Structures

Perhaps nothing better illustrates mathematicians' inventive structuring than those of its outcomes that are called "algebraic structures." Three things are required to define an algebraic structure. First, you have to define a set, *S*, of mathematical objects to work with. *S* is called the "domain" of the structure.

Second, you have to define a collection of one or more operations that can be carried out on these objects. Each operation will have one or more operands drawn from *S* on which the operation operates. You can think of an operand as an "input" to the operation. The number of operands an operation uses is called its "arity," and this number is always finite. Each different operation can have an arity that differs from those of other operations. For example, a unary operation has one operand, a binary operation has two, and an *n*-ary operation has *n* operands. The operations called "addition" and "multiplication" are binary operations.

Finally, you have to have a finite set of identities, called *axioms* of the structure, that each operation must satisfy. Axioms are probably best regarded as rules the operations are required to conform to by design. The word "axiom" derives from a Greek word (αξιωμα) meaning "that which is thought worthy or fit[5]."

[3] As a matter of fact, they are *trained* to spot it. It's part of a mathematician's education.
[4] To give credit where credit is due, I got this 0 = 1 prank from Péter's book, pg. 107.
[5] This word also has another connotation, i.e., "that which commends itself as evident." For the ancient Greeks, the axioms of Euclid were regarded as self-evident truths of nature, and for centuries that is how mathematicians regarded them. That was why it came as such a shock when non-Euclidean geometries (geometries that did not obey all of Euclid's axioms) were invented ("discovered").

If you are inventing a new algebraic structure to facilitate solving some scientific or engineering problem that does not seem to be well expressed using standard mathematics, *you* define or select axioms you think "best fit" the type of problem you want to solve, the operations you think will be useful, and the objects you decide to include in *S*. That is what Mandelbrot was doing when he invented "fractals" [Mandelbrot (1977)]. One important class of problems for which a new algebraic structure might prove to be useful or necessary is comprised of what are called "ill-posed problems" [Tikhonov & Arsenin (1977)]. These authors tell us,

> One important property of mathematical problems is the stability of their solutions to small changes in the initial data. Problems that fail to satisfy this stability are called, following Hadamard, *ill-posed*. For a long time mathematicians felt that ill-posed problems cannot describe real phenomena and objects. However, we shall show in the present book that the class of ill-posed problems include many classical mathematical problems and, most significantly, that such problems have important applications. [Tikhonov & Arsenin (1977), pg. xi]

Ill-posed problems have a nasty habit of popping up more often than you might think. Weather forecasting is one example, as are problems belonging to what has come to be called "chaos theory." Noise problems – which are problems electrical engineers frequently have to deal with – also, arguably, belong to the class of ill-posed problems. Broadly speaking, an ill-posed problem is one in which the secondary quantities of the theory violate Slepian's requirement that principal quantities not be sensitive to small changes in secondary quantities. If they are, the mathematical model is unreliable and might require the modeler to rethink his mathematical structuring of the problem.

Understanding the *construction* of algebraic structures is useful and enlightening because constructing them provides an example of how one might proceed to satisfy the maxim that the mathematics should be made to fit the problem rather than forcing the problem to fit the mathematics. If you make the problem fit the mathematics, then what you end up working on is not the problem you originally set out to work on but is, instead, some different problem – the solution for which might not solve the problem you really wanted to solve in the first place. Let us look at what goes in to structuring an algebraic structure.

## A. Mathematical Bookkeeping: Ordered Pairs, Operations, Functions, and Binary Relations

Suppose we have a set *S* of mathematical objects and a mathematical operation *f* that operates on two members of this set, *a* and *b*, to produce an outcome *c* that is also a member of *S*. Such an operation is called a "binary operation on *S*" and it is said to have the property of "closure" because the outcome *c* is also a member of *S*. In the symbolic language of mathematics, the statement "*c* is a member of set *S*" is written "$c \in S$". This can also be pronounced, "*c* belongs to *S*".

The statement "*f* operates on members *a* and *b* of set *S* to produce $c \in S$" can be written two different ways depending on whether we're talking about two specific members *a* and *b* or we're talking about the members of *S* generally. In the first case, we would write "$f\,(a, b) \to c$" to denote the specificity of the operation's "inputs" *a* and *b* with it being understood that $c \in S$ is implied. In the second case, the symbolic language is different and is written "$f$: $S \times S \to S$" to denote any two members of *S* can be used as "inputs" to the operation and that its "output" is also a member of *S*.

This sort of "hieroglyphic writing" is heavily utilized in mathematics textbooks and journal papers. It saves writing and, once a person is initiated into knowledge of math's symbolic language, it provides a more concise way to say what is said above in English. Thus you encounter statements like

> **Definition 1:** A binary operation *f* on a set *S* is a function $f$: $S \times S \to S$.

The notation $S \times S$ denotes *another* set and is called the *Cartesian product* of *S* with itself. The members of $S \times S$ are comprised of all *ordered pairs* of members of *S*, which are individually denoted by $(a, b)$. These pairs are *ordered* because the order in which they are written in $(a, b)$ might affect the outcome *c*.

For example, if $S$ is a set of 3 by 3 matrices and $f$ is matrix multiplication then the product $f(a, b)$ isn't equal to the product $f(b, a)$. Ordered pairs might be thought of as a kind of "bookkeeping method" in performing mathematical operations.

As the definition of a mathematical operation implies, an operation is a special case of a broader class of mathematical transformations called a *function*.

> **Definition 2:** A function $f: S_1 \rightarrow S_2$ from a set $S_1$ to a set $S_2$ is a binary relation from $S_1$ to $S_2$ such that for each $a \in S_1$ there is a unique $b \in S_2$ for which $(a, b)$ is a member of $S_1 \times S_2$.

A function $f: S_1 \rightarrow S_2$ is a rule which assigns to each member of $S_1$ some member of $S_2$. The two sets are called, respectively, the *domain* and *codomain* of $f$.

Functions likewise make up a special case of an even more general idea, namely, that of a binary relation.

> **Definition 3:** A binary relation $\alpha$ from a set $A$ to a set $B$ is a subset of the Cartesian product $A \times B$. The notation $a \,\alpha\, b$ indicates that $(a, b) \in A \times B$.

The "hieroglyphics" mathematicians use – for example, $(a, b) \in A \times B$ – is one of the reasons people who are not mathematicians often find mathematics textbooks and journal papers very indecipherable. Like every science, mathematics is expressed in its own technical language and these "hieroglyphics" are merely phrases and sentences expressed in that language. I assure you that the technical language used by, say, electrical engineers can be just as esoteric to a physicist as a physicist's technical language can be to an electrical engineer. For example, to an electrical engineer the word "phase" has a precise technical meaning, derived from the mathematics of Fourier transforms, and is used to express a particular property of signals; but to a physicist the word "phase" basically means "solid, liquid, and gas." Once years ago I had a lengthy and very confusing conversation with a physicist over an important technical issue. The reason the conversation was lengthy and confusing was because I was using the word "phase" in its electrical engineering context and he was using it in its physics context. As I said earlier, "context" is an extremely important idea in Critical epistemology. Success in projects that are interdisciplinary – that is, the project involves specialists from multiple disciplines – oftentimes requires specialists to learn each others' technical languages. Otherwise the project tends to become like a sort of Tower of Babel. On the shelves of my personal library there are more than a dozen technical dictionaries from ten different fields.

One thing I want to call to your attention is that the ideas of mathematical "operations," "functions," and "binary relations" are ideas progressing from a lower concept (operation) to higher concepts (function and binary relation). A binary relation is a more abstract idea than are the ideas of a function and an operation. When mathematicians write books they have a habit of starting out with their most abstract ideas and then presenting their lower concepts. But human beings learn new concepts in the opposite order of this. For this reason, trying to read a mathematics treatise can be a taxing and frustrating experience for people who are not trained mathematicians. If an idea is so abstract that the reader cannot think of examples of what that idea represents then that idea is useless to him because he knows no way to practically apply it. What you don't know how to apply, you can't use. All real meanings are, at root, practical.

The ideas of ordered pairs, Cartesian products, and definitions 1 through 3 provide a first look at how mathematical structures are constructed. Just as nails, lumber, and tile are components used in building a house, mathematical entities like these are components used in building a mathematical structure. Let us take a further look at some more mathematical entities that go into building algebraic structures.

## B. Groupoids

A *groupoid* is defined to be a non-empty set $S$ together with an operation that specifies a unique $s \in S$ for every ordered pair $(a, b) \in S \times S$. Now, this is a very abstract idea because it is very general. It doesn't say anything about properties or characteristics the operation must have *other than* the operation must have *closure*. Mathematicians regard a groupoid as a very simple thing – and *ontologically* it is. But I suggest it really is a very complicated idea *epistemologically* precisely because it is so very general. *Any* binary operation on $S$ will have the idea of a groupoid contained in it because the definition of a binary operation *specifies* that the operation has the property of closure.

C. Semigroups

A *semigroup* is a groupoid for which the binary operation has an additional property called the *associative property*. For example, you know 1 + 2 + 3 = (1 + 2) + 3 = 1 + (2 + 3) = 6, and this is none other than an example of the associative property of addition. It just means you can operate on the operands in any pairwise order you want to and the result will not change. A semigroup is still a pretty simple thing ontologically (though not quite as simple as a groupoid). *Epistemologically* it is a slightly *less* complicated idea than the more general (abstract) idea of a groupoid because *some* concepts that stand under the concept of a groupoid *do not* stand under the concept of a semigroup. This is because their operations do not obey the associative property. We saw an example of this in chapter 1, pg. 11, in regard to children's development of logical thinking ("ducks are birds; sparrows are birds" & etc.).

D. Identity Elements

For the next structures we will define, we need to introduce a special member of the set $S$, which will be denoted as $e$, called an *identity*. Let a semigroup $G$ be denoted as $G = [S, *]$ where * is its binary operation on set $S$. If $e \in S$ has the property that $e * s = s$ for every $s \in S$ then $e$ is called *the left-identity* of semigroup $G$. If $e \in S$ has the property that $s * e = s$ for every $s \in S$ then $e$ is called *the right-identity* of semigroup $G$. If $e \in S$ has the property that $s * e = e * s = s$ for every $s \in S$ then $e$ is called a *two-sided identity* of semigroup $G$. Another name for a two-sided identity is "*the* identity."

A semigroup does not necessarily have any identity member, but if it does that $e$ is often called an "identity element." It is possible to construct a semigroup that has more than one left identity member or more than one right identity member. However, if a semigroup has *both* a left identity member *and* a right identity member then it is easily proved: 1) they are the same member, i.e., $G$ has a two-sided identity $e \in S$; and 2) there is only one of them, i.e., a two-sided identity member $e$ is unique. In this case, $e$ is called "the identity" of the semigroup.

You are already familiar with the concept of "identities" in everyday arithmetic. In the operation of addition, the number 0 is often called "the additive identity." In multiplication, the number 1 is often called "the multiplicative identity." Note that whether or not a member of $S$ is an identity element depends on the binary operation. In other words, an identity $e$ is only an identity *with respect to* an operation *.

E. Monoids and Groups

A *monoid* is a semigroup $G = [S, *]$ with the identity, i.e. it has a unique two-sided identity $e \in S$ with respect to binary operation *. As a *thing*, a monoid is slightly more complicated (ontologically) than a semigroup because a monoid has an additional property – namely, it has the identity. But because this *is* an additional property not contained in the idea of a semigroup, the idea of a monoid is *epistemologically* a simpler idea (lower concept) than that of a semigroup. We can say it is *more structured* than a general semigroup is. Do you see the "theme" or "pattern" developing here? We "build more structure" into a mathematical construct as we add more properties to its definition. A semigroup has the associative property; a monoid has the associative property *and* the property of having a unique two-sided identity.

Adding a new property to a mathematical construct makes the resulting new construct an ontologically more complicated thing but, at the same time, an epistemologically more simple idea because it is less abstract than the higher concept from which it is specified. It inherits all the properties of the higher concept but is more specific because of its additional property or properties. As a non-mathematical example of this idea of concept structuring, consider the idea of a "cat." You know a lion is a cat, but a lion has properties that distinguish it from other kinds of cats, e.g., common house cats, tigers, leopards, or cheetahs. A North American mountain lion has different properties than an African lion. If you start with the idea of a "human being" and keep specifying additional properties you can eventually get down to an idea of a specifically unique human being, e.g., "myself," "my brother Bill," "my neighbor Joe," and so on. Once you get down to, say, yourself, you cannot get any further "down the chain" of concepts and still be thinking about "a human being" (although you can *widen* the concept by disjunction; for example, you can think of "me when I was five," "me when I was a volleyball player," and so on). When you get down to "me" you have arrived at an epistemologically most simple member of the set of "human beings" but an ontologically very complicated *thing* with many unique properties. So too it is with mathematical constructs.

Let us introduce a few more properties we might add to a monoid construct. If $M = [S, *, e]$ is a monoid with identity $e$ and it contains two members $a$ and $b$ in its set $S$ such that $a * b = e$ then $a$ is called a *left-inverse* of $b$ and $b$ is called a *right-inverse* of $a$. If $a$ is both a left-inverse and a right-inverse of $b$ then it is called a *two-sided inverse* of $b$. If $b$ has both a left-inverse and a right-inverse then they are the same member of $S$, i.e., a two-sided inverse. It is easy to prove the two-sided inverse of $b$ is unique. To see this, suppose $S$ contains members $a$ and $c$ such that $a$ is the left-inverse of $b$ and $c$ is the right-inverse of $b$. Then $a = a * (b * c) = (a * b) * c = e * c = c$, QED.

When we have a monoid $M$ in which each member of $S$ has a two-sided inverse, the monoid is said to have the *cancellation property*. In this case, it is conventional to label the inverse of $b$ as $b^{-1}$, i.e. $b^{-1} * b = b * b^{-1} = e$.

This brings us to the definition of another construct: A *group* is a monoid with the cancellation property.

Each time we "put more structure" into a mathematical construct by adding another property to it, there are more things we can say about it that are *always* mathematically true. These "mathematical truths" are generally called *theorems*. For example, if $G = [S, *, e]$ is a group then for any $a$ and $b \in S$ we can solve an equation with an unknown, $x$. Suppose $a * x = b$. Then $a^{-1} * a * x = e * x = x$. But because $a * x = b$, we then also have $a^{-1} * a * x = a^{-1} * b$. Therefore, $x = a^{-1} * b$. This is a quite simple theorem but it has an important consequence: *with it we can begin to solve equations to find unknowns*. That makes a group quite an important construct.

A *commutative group* is a group in which, for every pair of members $a, b$ in its set $S$, $a * b = b * a$. This is called the "commutative property" and it is an additional property not found in every group. We are all familiar with "the commutative property of addition" and "the commutative property of multiplication" in everyday arithmetic.

Most of us, however, are not familiar with non-commutative groups. A good example of one is shown in figure 1 below. This example is called a *permutation* (or "substitution") group. A permutation consists of reordering the arrangement of a set of objects. Symbols $a$ through $f$ denote these re-positionings and correspond to the *arrows* shown in the figure. It is *the arrows* that denote the members in the set $S$ of the permutation group and not the color of the balls. It is important to understand that the group itself is entirely independent of the color of the balls and has to do merely with re-ordering an initial arrangement, shown as the left-hand box in each of the six cases above, into a final arrangement, shown as the right-hand box. For example, permutation $c$ in the figure leaves the position of the top ball unchanged and swaps the positions of the middle and lower balls. So far as the permutation group is concerned, it does not at all matter what the color of the balls in the initial arrangement might be, nor what color they are in the final arrangement. All that matters is how the balls are repositioned. The permutations are "color blind."

To form a group we must also have an operation, denoted by "•", that operates on the members of $S$. In the permutation group this operation is *concatenation*, i.e. a succession of permutations. The operation $a \bullet b$ thus denotes performing permutation $a$ on the initial arrangement followed by performing permutation $b$

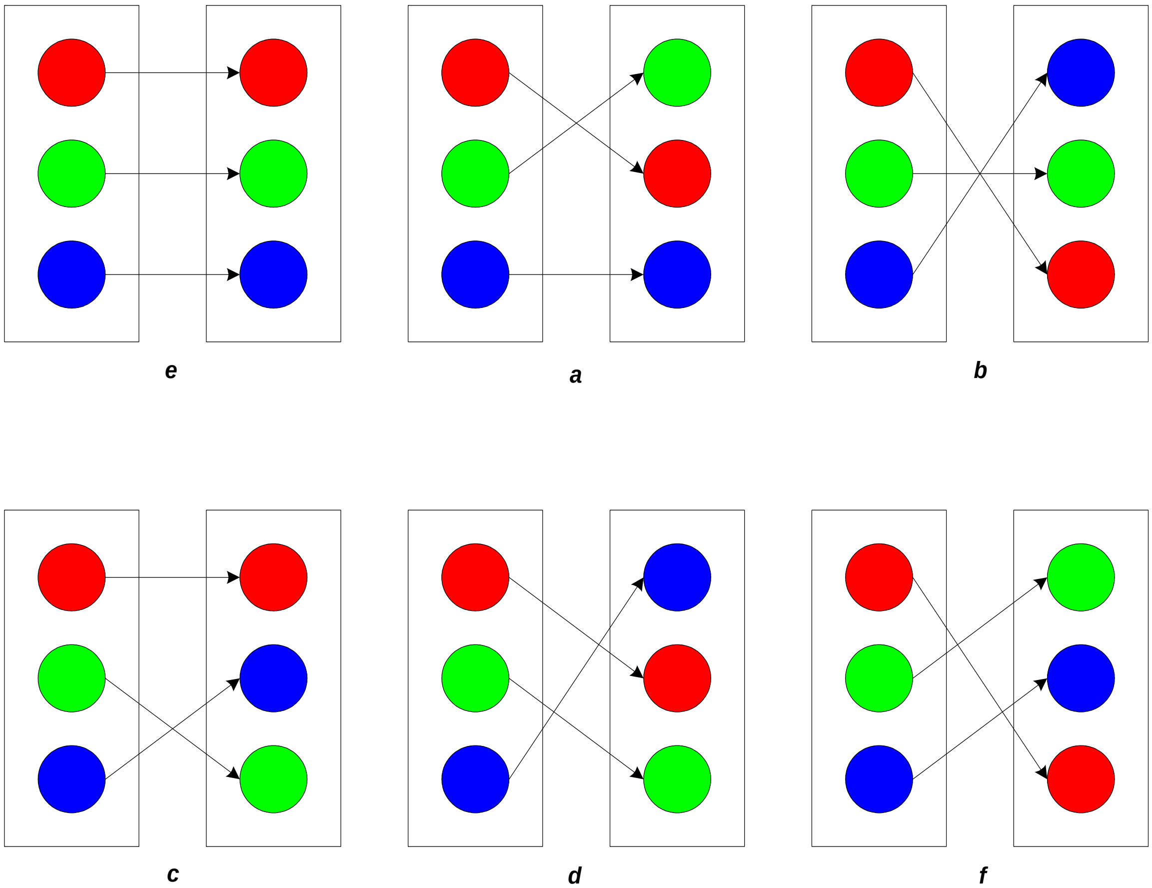



**Figure 1:** Example of a permutation group. The set *S* is composed of the six permutations shown in the figure, i.e., *S* = (*e*, *a*, *b*, *c*, *d*, *f*). The permutation operation • rearranges the objects {red ball, green ball, blue ball} into a new arrangement. The figures above show the initial arrangement on the left and the new arrangement on the right. The red, green, and blue balls are not the members of set *S*. The rearrangements (arrows) are the permutations in set *S*. Permutation *e* is the identity element in the structure.

on the results of the first permutation. Referring to Figure 1, *a* produces the arrangement (green, red, blue) from top to bottom; applying *b* to this new arrangement produces (blue, red, green). We now take note that the outcome of this concatenation of permutations is equivalent to what we would have gotten by applying permutation *d* to the initial arrangement; therefore *a* • *b* is equivalent to permutation *d*. Symbolically, *a* • *b* = *d*. We may next note that permutation *e* leaves the arrangement unaltered. Therefore for any permutation *x* belonging to *S* we will get *x* • *e* = *e* • *x* = *x*. Permutation *e* therefore is the identity element for the set *S*. Next, we note that some concatenations end up with the permutations canceling each other, i.e. *f* • *d* = *e* and *d* • *f* = *e*. Thus permutation *d* is called the inverse of *f*, and *f* is likewise the inverse of *d*. In some cases a permutation can cancel itself, i.e. *a* • *a* = *e*, thus *a* is its own inverse.

By taking the permutations in pairs we can construct an *operation table* for operation •. An operation table is similar to the addition table we all learned in elementary school. This operation table is shown in Figure 2 below. The table is arranged so that the order of the operations is from row to column, e.g. *a* • *b* = *d*, etc. Referring to the definition of a group, it is easily seen by inspecting the table that the closure property is satisfied. Inspection of the *e* row and the *e* column demonstrates that the identity property is likewise satisfied. Each row of the table has in it exactly one occurrence of *e*, and each column of the table has in it exactly one occurrence of *e*.

| $\bullet$ | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ |
|---|---|---|---|---|---|---|
| $a$ | $e$ | $d$ | $f$ | $b$ | $a$ | $c$ |
| $b$ | $f$ | $e$ | $d$ | $c$ | $b$ | $a$ |
| $c$ | $d$ | $f$ | $e$ | $a$ | $c$ | $b$ |
| $d$ | $c$ | $a$ | $b$ | $f$ | $d$ | $e$ |
| $e$ | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ |
| $f$ | $b$ | $c$ | $a$ | $e$ | $f$ | $d$ |

**Figure 2:** Operation table for the permutation group example.

This means that for each member in $S$ there is exactly one member that acts as its inverse, and so the "existence of inverses" property of the group is satisfied. That the associative property is also satisfied is not so self-evident from the operation table, but this can be demonstrated by taking all triplets of permutations and applying the operation term by term across the triplet. For example,

$$(a \bullet b) \bullet c = d \bullet c = b; \quad a \bullet (b \bullet c) = a \bullet d = b$$

$$\Rightarrow a \bullet b \bullet c = (a \bullet b) \bullet c = a \bullet (b \bullet c).$$

The same is found to be true for any triplet in $S$ and so the associative property is established. This completes the proof that the permutation group is indeed a group. However, it is not a commutative group because, e.g., $a \bullet b = d$ but $b \bullet a = f$.

F. Rings and Fields

Adding more properties is not the only way to "build more structure" into a mathematical construct. We can also do this by adding more operations to it. We might, for instance, take a construct $C' = [S, +]$ comprised of a set $S$ and an operation "+" and add another operation "$\bullet$" to it to build a new construct $C = [S, +, \bullet]$. We could add a third, fourth, fifth, or however many more operations we wish to it. There are two particular constructs that are extremely valuable in terms of their mathematical fruitfulness.

A *ring* is a construct $[S, +, \bullet]$ where the pair [S, +] is a commutative group and the pair $[S, \bullet]$ is a semigroup with the property that, for arbitrary members $a, b, c$ of $S$, the following two distributive laws hold:

$$a \bullet (b + c) = a \bullet b + a \bullet c \quad \text{and} \quad (b + c) \bullet a = b \bullet a + c \bullet a .$$

When we have such a structure, it is very common for mathematicians to call operation + by the name "addition" and operation $\bullet$ by the name "multiplication." When they do, $[S, +]$ and $[S, \bullet]$ are called the "additive" and the "multiplicative" structures, respectively, of the ring $R = [S, +, \bullet]$. The additive identity in $[S, +]$ is then called "zero" and symbolized by "0"; the multiplicative identity in $[S, \bullet]$ is called "one" and symbolized by "1". Perhaps these frequent naming conventions imply to you that rings are very important mathematical constructs. Perhaps they also might imply to you that there is more than one way to "add things" and more than one way to "multiply" things. If so, your anticipation is correct. There are.

A *field* is a ring $[S, +, \bullet]$ in which the nonzero members of $S$ form a commutative group under the multiplication operation $\bullet$.

Here we have added another property to ring structure. For a ring in general the only properties required of the construct $[S, \bullet]$ are that it be a semigroup and that $[S, +, \bullet]$ obeys the two distributive laws. For a field we require in addition that $[\{S - 0\}, \bullet]$ be a commutative group. The notation $\{S - 0\}$ denotes an operation called "set subtraction" and it means merely "remove 0 from $S$". The symbol 0 denotes the additive identity in the construct $[S, +]$. It would be hard to overstate the importance of a field construct. It underlies all of the arithmetic and algebra we routinely use every day.

## 4. Universal Algebras and Empirical Science

To summarize the foregoing: Mathematical structuring is carried out by starting with a set of objects (the set $S$) and defining properties and operations that can be carried out on the members of this set. Each additional property and each additional operation "build more structure" into the construct and specialize it for use in one or more specific applications. Unary operations (operations carried out using one member of $S$) are transformations of the form $f : S \rightarrow S$. Binary operations are transformations carried out using ordered pairs drawn from the Cartesian product $S \times S$. Ternary operations are operations carried out using ordered triples drawn from the Cartesian product $S \times S \times S$, and so on for 4-ary, 5-ary, etc. operations. As you can hopefully see, only the power of human imagination limits the number of structures a person is able to define and construct.

A *universal algebra* is a system $[S, f_1, f_2, \ldots, f_k]$ consisting of the set $S$ and one or more operations $f_1, f_2, \ldots f_k$ on $S$ such that each $f_i$ is an $n_i$-ary operation on $S$ for some $n_i > 0$.

Generally speaking, the more structure that is built into a system $[S, f_1, f_2, \ldots, f_k]$, the more statements that are *always* true of it (theorems) can be found. The emphasis mathematicians place on "proofs" is a consequence of searching for "things that are always true" of a particular mathematical construct.

However, these truths are truths only in regard to the mathematical system. They are not necessarily true in regard to nature because the mathematical construct is used to obtain a *model* of some set of natural phenomena. Richard Feynman said, "Mathematics is a language plus reasoning; it is like a language plus logic" [Feynman (1965), pg. 40]. Empirical science uses mathematics to make statements about empirical nature, and these statements are based upon scientists' observations and experiments about phenomena. It is important to clearly understand and keep distinct what mathematics is and what the phenomena of nature are. The mathematics is *not* the natural phenomenon. A person who forgets this and drifts into thinking his mathematics is somehow "more real" than the phenomena it is *designed* to describe shifts his "way of looking at the world" into an ontology-centered metaphysic and he is on his way to a land of mysticism and fantasy.

Mathematical descriptions of nature, when they are the most fecund, lead to *predictions* about phenomena that have not yet been empirically observed. To the extent that the mathematical description accurately describes "how phenomena behave," these predictions, when tested, are found to be true. To the extent that the mathematical description is an inaccurate description of "how phenomena behave," actual observation or experiment turns out to be different from what was predicted. If the difference is extreme enough, science is said to have made "a great discovery" necessitating a *new* mathematical description of empirical nature.

It often happens that scientists are tasked with finding relationships between one class of phenomenal objects (a set $A$) and another class of phenomenal objects (a set $B$). Mathematical descriptions of such relationships often are made by defining binary relations, $f : A \rightarrow B$. The description, in effect, represents phenomenal objects in $A$ as "funny numbers" of some sort, and represents phenomenal objects in $B$ as another sort of "funny numbers." To illustrate what I mean by this, let us look at an example from chemistry.

If you were a chemist, how would you describe the phenomenon of chemically compounding different kinds of molecules? For instance, how would you say "compounding hydrogen and oxygen produces water"? A chemist's training teaches him to say this in the following way:

$$2H_2 + O_2 \rightarrow 2H_2O$$

and he would verbalize this expression as "2 parts hydrogen molecule compounded with 1 part oxygen molecule yields 2 parts water." The mathematical expression takes for its set *S* "the set of all molecules" and takes for its operation "compounding." Molecules are the basis for the "funny numbers" (like $H_2$) in *S* in quantitative descriptions of molecular chemistry. In this illustration, the number of "funny numbers" in *S* (called the "cardinality of *S*") is as great as the number of different kinds of molecules that can be identified. The chemical "formula" above is representable as a groupoid structure since all of its "funny numbers" are "molecules" and the outcome of the compounding operation is a "molecule."

However, chemistry students soon learn that there is more (much more) to chemistry than this example illustrates. Consider, for example, the chemical reaction

$$CH_3OH + HC_2H_3O_2 \rightarrow CH_3C_2H_3O_2 + H_2O \ .$$

A chemist would verbalize this expression as, "methanol compounded with acetic acid produces methyl acetate mixed with water." His training teaches him that the "+" symbol on the left-hand side of the formula means something different from the "+" symbol on the right-hand side. Another difference in this example is that the outcome is not just one kind of molecule but, rather, involves two. Clearly some more powerful structure is called for in this example. Those of us who are not chemists might also object to using the "+" symbol to denote two different things in the same chemical "sentence." The chemist might answer this criticism by pointing out that the positioning of the "+" symbol relative to the "$\rightarrow$" symbol clearly enough distinguishes the meaning of the symbol. That is a fair enough answer; but the notational issue can also fairly be said to be one reason students in other fields sometimes complain that chemistry is "a fancy history course."

The phenomena studied in chemistry are even more varied and complex than this. To pick just one more example, the branch of chemistry called electrochemistry frequently deals with what are called oxidation-reduction reactions. Often these are understood, not by a single transformation formula like those above, but by multi-transformation formulae involving intermediate steps between the initial and final chemical states. One such example is

$$Ni + 2H^+ \rightarrow Ni^{2+} + 2e^- + 2H^+ \rightarrow Ni^{2+} + H_2 \ .$$

Here the reaction involves more than just molecules. It involves ionized atoms ($Ni^{2+}$) and electrons ($e^-$) as well. And, again, the "+" symbols have multiple interpretations. The universe of "funny number" definitions has to expand to, e.g., {molecules, ions, electrons} and the sort of mathematical structure(s) required are clearly much more complicated.

The desirability of such a development of new structures for "chemistry-math" begins to be manifested when chemistry theory must be united with theories from other disciplines to describe more complicated phenomenal objects. After all, chemists seem to have gotten along quite well with their present way of doing things for a considerable length of time. However, in studies, for instance, of computational neuroscience – where more than just chemistry is involved – there are advantages gained by more expanded mathematical expressions. A illustration of a modest beginning of such expansion is provided by Wells (2010).

Chemists, I have no doubt, would tell us that the system they use is much easier to understand than what we would get if we formally "broke it down" into more precisely structured mathematical terminology. I do in fact agree with them *provided* we make the slight modification "easier to understand *after* you have been trained in chemistry." But, on the downside, lacking a more general system of mathematical expression tends to make chemistry's mathematical hieroglyphics more esoterically isolated from people in fields outside of chemistry. After all, "Life is short, the art long, opportunity fleeting," as Hippocrates said. Chemistry is a much more difficult science than physics is. Chemists have not expressed a need for a more formal mathematical system to aid in their work, and so mathematicians have not endeavored to develop one. However, as the famous chemist Antoine Lavoisier wrote,

> The impossibility of separating the nomenclature of a science from the science itself is owing to this: that every branch of physical science must consist of three things: the series of facts which are the objects of the science; the ideas which represent these facts; and the words by which these ideas are expressed. Like three impressions of the same seal, the word ought to produce the idea, and the idea to be a picture of the fact. And, as ideas are preserved and communicated by means of words, it necessarily follows that we cannot improve the language of any science without at the same time improving the science itself; neither can we, on the other hand, improve a science without improving the language or nomenclature which belongs to it. However certain the facts may be, and however just the ideas we may have formed of these facts, we can only communicate false impressions to others while we want words by which these may be properly expressed. [Lavoisier (1789), pp. xiv-xv]

Mathematics is a language.

One can be pardoned for wondering what undiscovered properties of chemical nature there might be that await discoveries aided by additional development of "chemistry-math." As Timothy asked,

> Are there gallant dreams, noble challenges, and daring visions beckoning us in the daily pursuit of the best that is in us and around us? Or do we see just another pleasant view – something to be admired from distance and then forgotten as we sink back into a lukewarm pool of anesthetizing escape? [Timothy (2010), pg. 8]

## §5. Mathematical Infinity

As I noted earlier, mathematical infinity denoted by the symbol $\infty$ is not a number. It is not a member of any set *S* of objects and it denotes an unending process of "becoming."

Having said this, it must immediately be stated that there are symbolic notations often used in mathematics that easily mislead one into thinking $\infty$ is a number. Take, for example, the case of a sum of integers up to some final integer *n*,

$$\sigma = 1 + 2 + 3 + 4 + \ldots + n \, .$$

Here the "…" symbol just means "keep this pattern going until" you get to the integer *n*.

As you can easily surmise, the larger we allow *n* to become, the larger $\sigma$ will become. In particular, for any definite value of *n* this summation adds up to

$$\sigma = \sum_{k=1}^{n} k = \frac{n(n+1)}{2} \; .$$

When we want to say "let *n* grow without bound," the notation for this is written as $n \to \infty$. When we apply this imaginary operation (this "becoming") to the expression for $\sigma$, this is written

$$\frac{lim}{n\to\infty}\sigma = \frac{lim}{n\to\infty}\frac{n(n+1)}{2} \to \infty \; .$$

All this really says is "$\sigma$ increases without bound." But it is easy to look at this and think $\infty$ denotes a number. It does not. It is temptingly convenient, however, to incorporate $\infty$ into standard methods of mathematical notation *as if* it were a number like others. For example, when I was a schoolboy and was first introduced to the symbol "$\infty$" this introduction was accompanied by other statements that were written as $x + \infty = \infty$, $x \div 0 = \infty$ (so long as $x \neq 0$), etc. The problem with using the "=" symbol here instead of using the "$\rightarrow$" symbol was the following. Up until then, the "=" symbol always implied the answer was also a number. But now, in the notation, "=" was being used to denote something else, and this "something else" is an idea hard to put into words that a little boy or girl can understand. (For that matter, it is an idea hard to put into words that many adults can understand.) I remember being very puzzled by " $x + \infty = \infty$" because to me it was saying "$\infty$ is the only number you can't add anything to." In a manner of speaking, that's true; but it made "$\infty$" a "number that's not like any other number" and I didn't like that. It meant there were special rules that applied to it that did not apply to any other number, and all I could do was try to remember these special rules if I wanted to learn about mathematics. I hated having to memorize things because my memory isn't all that powerful[6]. If instead one writes "$x + \infty \rightarrow \infty$" this can be translated into English as something like "adding anything to something that is unbounded still results in something that is unbounded."

No person ever has a direct sensible experience with anything that is "unbounded." Take, for example, the idea of "the physical universe"; you have never had, and never will have, an encounter with "the physical universe"; you do encounter "things that are part of the physical universe" and, because there seems to be no limit to the number of things that exist, the *idea* of a physical universe "in which every thing exists" is an idea of something unbounded – and this idea belongs entirely to Slepian's Facet B.

Yet, "within the world" of Facet B, there is nothing to prevent a person from posing intriguing questions and riddles about the denizens of this mathematical world such as, "Are there more real numbers than there are integers?" If you draw a number line, look at the part of it between "1" and "2", and reason that there are an unlimited number of real numbers in between them, then it seems just a matter of common sense that "there must be more real numbers than there are integers." But since there is no limit to how many integers there are, what could it possibly *mean* to say there are "more" real numbers than there are integers?

I think it's probably fair to say most empirical scientists would just shrug at this question. But, being the sort of folks they are, mathematicians do get professionally interested in questions like this one. Their collective interest in them is called "the theory of transfinite numbers." I personally think it would be less confusing if they called it "the theory of transfinite sets." As it happened, though, when Dedekind began playing around with questions like this (in 1858) and Cantor extended the theory (in 1874), both of them used the word "number" in descriptions of their objects and that designation remains with us today.

Cantor was interested in what, at the time, was called "the theory of aggregates" and is today called "set theory" [Cantor (1897)]. Being a great mathematician, he knew that in order to ask "what could it mean to say there are 'more' real numbers than there are integers?" he would have to be more specific about what connotation the word "more" had to have for the question itself to make any kind of sense. It is instructive to take a look at how he attacked this issue:

> By an "aggregate" we are to understand any collection into a whole M of definite and separate objects *m* of our intuition or our thought. These objects are called the "elements" of M. In signs we express this thus:
>
> (1) $M = \{m\}$ .
>
> We denote the uniting of many aggregates M, N, P, …, which have no common elements, into a single aggregate by
>
> (2) (M, N, P, …) .
>
> The elements of this aggregate are, therefore, the elements of M, of N, of P,…, taken together. . . .

[6] Math paled in comparison to English in this regard. In spelling class I had to remember that things like the rule 'i before e except after c or when sounded like a as in neighbor and weigh' didn't apply to words such as "weird."

> Every aggregate M has a definite "power," which we will also call its "cardinal number."
>
> We will call by the name "power" or "cardinal number" of M the general concept which, by means of our active faculty of thought, arises from the aggregate M when we make abstraction of the nature of its various elements *m* and of the order in which they are given. . . .
>
> We say that two aggregates M and N are "equivalent" in signs
>
> (4) $M \sim N$ or $N \sim M$
>
> if it is possible to put them, by some law, in such a relation to one another that to every element of each of them corresponds one and only one element of the other. [Cantor (1897), pp. 85-87]

Cantor used the symbol convention $\bar{\bar{M}}$ to denote the "cardinal number" of an aggregate $M$. By the word "power" he meant, using today's terminology, "the set of all the subsets of $M$." This is an idea quite different from ideas non-mathematicians have when they use the word "power," but such is often the case in the technical language a particular area of scholarship uses. What he was leading up to is a way to define "greater" and "less" in comparisons between "aggregates." He tells us,

> If for two aggregates $M$ and $N$ with the cardinal numbers $a = \bar{\bar{M}}$ and $b = \bar{\bar{N}}$, both the conditions
>
> (a) There is no part of $M$ which is equivalent to $N$,
> (b) There is a part $N_1$ of $N$ such that $N_1 \sim M$,
>
> are fulfilled, it is obvious that these conditions still hold if in them $M$ and $N$ are replaced by two equivalent aggregates $M'$ and $N'$. Thus they express a definite relation of the cardinal numbers $a$ and $b$ to one another.
>
> Further, the equivalence of $M$ and $N$ and thus the equality of $a$ and $b$, is excluded; for if we had $M \sim N$, we would have, because $N_1 \sim M$, the equivalence $N_1 \sim N$, and then, because $M \sim N$, there would exist a part $M_1$ of $M$ such that $M_1 \sim M$, and therefore we should have $M_1 \sim N$, and this contradicts the condition (a). [*ibid*., pg. 89]

Certainly a grandiloquent way of expressing one's intuition about intervals on the number line, is it not? What, exactly, is Cantor doing here? Briefly put, *he is defining a new class of* "funny numbers" (the cardinal numbers $\bar{\bar{M}}$ and $\bar{\bar{N}}$) and defining a way to compare them to see if they are "equal." If they are not, then one of them must in some sense be "greater than" the other. Cantor went on to establish that the "cardinal number" of the set of integers was not equivalent to the "cardinal number" of the set of real numbers. From there he went on to, basically, define the former as "less than" the latter – which is a process of constructing an order structure for his new "funny numbers." The "cardinal number" of the set of integers is denoted $\aleph_0$ ("aleph-naught"); that of the real numbers is denoted $\aleph_1$ ("aleph-1")[7].

Cantor's funny "cardinal numbers" are "numbers" since he had a way to use them, and they are "trans-finite" in the sense that they are larger than all finite numbers. But this is not the same thing as saying they are "infinite" in the sense of what is meant by the symbol $\infty$. Funny numbers $\bar{\bar{M}}$ and the symbol $\infty$ simply do not mean the same *thing*. Mathematicians find Cantor's theory important, for instance, in developing axiom systems for transfinite sets; empirical scientists very rarely think they are important.

Davis & Hersch called infinity "the miraculous jar of mathematics." Of it they wrote,

> The contemporary stockpile of mathematical objects is full of infinities. The infinite is hard to avoid. . . . We have infinities and infinities upon infinities galore, infinities beyond the dreams of conceptual avarice. . . .

[7] Strictly speaking, mathematicians (including Cantor) use the symbol $c$ (for 'continuum') to denote the cardinality of the set of real numbers, and they tell us $\aleph_1 = c$ if and only if something called "the continuum hypothesis" is true. However, it has also been proved that the continuum hypothesis can neither be proved to be true nor proved to be false. Saying $\aleph_1 = c$ is basically equivalent to stating an axiom.

> This miraculous jar with all its magical properties, properties which appear to go against all experiences of our finite lives, is an absolutely basic object in mathematics, and thought to be well within the grasp of children in the elementary schools. Mathematics asks us to believe in this miraculous jar and we shan't get far if we don't. . . .
>
> The infinite is that which is without end. It is the eternal, the immortal, the self-renewable, the *apeiron* of the Greeks, the *ein-sof* of the Kabbalah, the cosmic eye of the mystics which observes us and energizes us from the godhead.
>
> Observe the equation
>
> $$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots=1\ ,$$
>
> or, in fancier notation, $\sum_{n=1}^{\infty} 2^{-n} = 1$. On the left-hand side we seem to have incompleteness, infinite striving. On the right-hand side we have finitude, completion. There is a tension between the two sides which is a source of power and paradox. There is an overwhelming mathematical desire to bridge the gap between the finite and the infinite. We want to complete the incomplete, to catch it, to cage it, to tame it. . . .
>
> Where there is power, there is danger. This is as true in mathematics as it is in kingship. All arguments involving the infinite must be scrutinized with especial care, for the infinite has turned out to be the hiding place of much that is strange and paradoxical. [Davis & Hersch (1981), pp. 152-155]

## §6. The Mathematically Infinitesimal and the Idea of Limits

Figuratively speaking, on the side opposite to infinity in the Facet B universe stands the infinitesimal. The infinitesimal is the idea of a number that is infinitely small but not zero. We got a taste of the idea of the infinitesimal when the Euclidean notion of a geometric "point" was discussed earlier – i.e., "a point is an object that is smaller than anything but such that if it got any smaller it wouldn't exist at all."

The idea of the infinitesimal has been an important one for empirical science since at least the days of Archimedes, but it has also been a subject of great controversy among mathematicians since at least the days of Isaac Newton and Gottfried Wilhelm Leibniz. Davis & Hersch remarked,

> In the nineteenth century infinitesimals were driven out of mathematics once and for all, or so it seemed. To meet the demands of logic the infinitesimal calculus of Newton and Gottfried Wilhelm von Leibniz was reformulated by Karl Weierstrass without infinitesimals. Yet today it is mathematical logic, in its contemporary sophistication and power, that has revived the infinitesimal and made it acceptable again. Robinson[8] has in a sense vindicated the reckless abandon of the eighteenth-century mathematicians against the strait-laced rigor of the nineteenth-century, adding a new chapter in the never ending war between the finite and the infinite, the continuous and the discontinuous. [Davis & Hersch (1981), pp. 237-238]

Controversies over "the infinitesimal" are not dead today even among empirical scientists. As noted previously, Richard Feynman remarked,

> On the other hand, I believe that the theory that space is continuous is wrong, because we get these infinities and other difficulties, and we are left with questions on what determines the size of all the particles. I rather suspect that the simple ideas of geometry, extended down to infinitely small space, are wrong. Here, of course, I am only making a hole and not telling you what to substitute. If I did, I should finish this lecture with a new law. [Feynman (1965), pp. 166-167]

As Kant would tell us, these controversies arise out of ontology-centered ways of looking at the world, including ideas of *objective* space and *objective* time regarded as things-in-themselves.

[8] Abraham Robinson, the founder of the branch of mathematics called "non-standard analysis."

Controversy over the idea of infinitesimals erupted in the eighteenth century shortly after Newton published his *Principia* [Newton (1687)]. Newton's development of calculus made use of "quantities which vanish"; he called these "evanescent quantities" [Newton (1687), pp. 31-39]. Using arguments drawn from geometry, he defined what we today call the *derivative* of a continuous function $f(x)$ in terms of what he called "the ratio of first and last evanescent quantities" and said their "ultimate ratio" is to be understood as "the ratio of the quantities not before they vanish, nor afterwards, but with which they vanish." Note that "vanish" means "to disappear," not "to become nothing."

Newton himself called derivatives by the name "fluxions" [Newton (1736)]. Newton used the symbol $\dot{y}$ to denote a fluxion; Leibniz used the notation $\frac{dy}{dx}$ to denote the derivative (fluxion) of y with respect to independent variable $x$. The two notations mean the same thing and are used interchangeably today. A present day student of calculus would be hard-pressed to recognize Newton's introduction of derivatives in his *Principia* because there Newton framed everything in terms of geometric arguments and diagrams. In today's notation, Newton's "fluxion" is written

$$\dot{f}(x) = \frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h} .$$

Both the numerator and the denominator on the right-hand side of this expression are "evanescent quantities." They "approach" but do not "equal" zero. Newton's term "evanescent" means "vanish*ing*," not "vanish*ed*."

The present-day mathematical concept of "limits" had not yet been formulated in the 18th century and so it isn't surprising that mathematicians and scientists alike both took Newton's argument to be equivalent to saying "set $h$ equal to zero." Note that if we set $h = 0$ in the expression above, we get $\frac{0}{0}$ . But, of course, this ratio is undefined in mathematics and that fact set off a strenuous objection to Newton's calculus by George Berkeley [Berkeley (1734)]. Berkeley's book was subtitled "A Discourse Addressed to an Infidel Mathematician" – who happened to be the astronomer Edmond Halley – so you can perhaps appreciate how heated the controversy was. On the other hand, Newton's physics was proving itself to be tremendously successful and scientists were not about to throw it away. Thus arose "the problem of infinitesimals" in the 18th century.

Integration, the anti-derivative operation in calculus, is likewise defined in terms of "evanescent quantities." The prototype problem for integrals is the problem of finding the area under a curve described by some function $f(x)$. Figure 3 illustrates the basic scheme for finding this area.

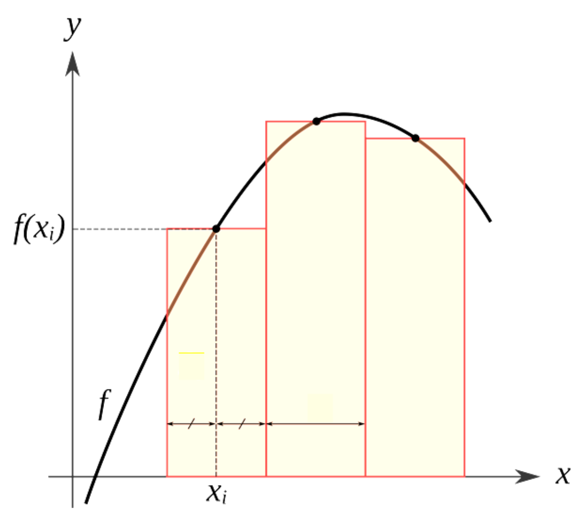



**Figure 3:** The scheme for calculating integrals. This scheme is known as the "midpoint rule."

To integrate $f(x)$ with respect to the independent variable $x$, the value $y$ of $f(x)$ is computed at a number of points $x_i$ and a rectangle of height $y = f(x_i)$ is superimposed at each point. The points are separated by intervals $\Delta x_i = x_{i+1} - x_i$ . Usually all these intervals are given the same length, $\Delta x$, but this is a matter of convenience rather than necessity. The width of the rectangle at each $x_i$ is made equal to $\Delta x$ when the intervals are all equal. The area under $f(x_i)$ is approximately equal to $f(x_i) \cdot \Delta x$ and the approximation becomes more accurate the smaller $\Delta x$ is made. The integral (area under $f(x)$) is then given by

$$F = \frac{lim}{\Delta x \to 0} \sum_i f(x_i)\Delta x = \int f(x)\, dx$$

and $dx$ denotes the infinitesimal $\Delta x \to 0$ while the symbol "∫" is called "the integral sign."

If one were to say $\Delta x = 0$ instead, an issue similar to the earlier one above would then appear because for any finite value of $f$ we would have $f \cdot 0 = 0$ and the sum of any number of zero terms is still zero. Hence, an "infinitesimals problem" also appears in the anti-derivative operation. The problem again is resolved by the idea of "limits" and the "vanishing" rather than "vanished" property of Newton's evanescent quantities. Because $dx \to 0$ means $f \cdot dx \to 0$ but the number of such terms being summed becomes unbounded, the summation indicated by $\int f \cdot dx \to \infty \cdot 0$, the product of which is undefined in *arithmetic* but not in the *calculation method* of Newton's calculus.

For example, suppose $f(x) = x$ and we wish to compute the area under $f(x)$ from $x = a$ to $x = b$. We divide the interval $[a, b]$ into $n$ intervals and then sum the areas of the rectangles as $n \to \infty$. With a bit of work, we find that for any integer $n > 1$ the interval is divided into segments

$$\Delta x = \frac{b-a}{n-1}$$

and the midpoints of the rectangles are located at points

$$x_i = a + (i-1)\frac{b-a}{n-1}$$

for $i = 1, 2, \ldots n$. Because $f(x_i) = x_i$ , evaluation of $F$ proceeds as

$$F = \frac{lim}{n \to \infty} \sum_{i=1}^{n} \left(\frac{b-a}{n-1}\right)\left[a + (i-1)\frac{b-a}{n-1}\right].$$

After grinding through some algebra, this summation reduces to

$$F = \frac{lim}{n \to \infty}\ \frac{na(b-a)}{n-1} - \frac{n(b-a)^2}{(n-1)^2} + \frac{n(n-1)(b-a)^2}{2(n-1)^2}$$

and as $n \to \infty$ this reduces to $F = \frac{1}{2}(b^2 - a^2)$, which is the correct solution for $\int_a^b x\, dx$ .

The reason this works is in the way $\Delta x \to 0$ as $n \to \infty$. It happens in such a way that some of the terms in $F$ are driven to zero (or, rather, made to approach ever closer to zero) as $n \to \infty$, while in other terms the numerator $n$ terms are cancelled out by denominator $n$ terms, leaving finite terms which eventually evaluate to the solution above. As Newton phrased it in his *Principia*,

> For these ultimate ratios with which quantities vanish are not truly the ratios of ultimate quantities, but limits toward which the ratios of quantities decreasing without limit do always converge; and to which they approach nearer than by any given difference, but never go beyond, nor in effect, attain to, till the quantities are diminished *in infinitum*. [Newton (1687), pg. 39]

Weierstrass, in fact, said this very same thing in the 19th century in his famous "epsilon-delta" definition of continuity, only Weierstrass said it using mathematical hieroglyphics,

> For any positive number ε a positive number δ depending on ε and c can be found such that
>
> $$|f(x) - f(c)| < \epsilon$$
>
> whenever $|x - c| < \delta$ .

Just as translating a statement in German into a statement in English is a matter of translating *ideas*, not *words*, so too it is when it comes to translating from mathematics' hieroglyphic language into a natural language, be that language English, French, German, or any other.

There are a fair number of people who find Newton's words and the hieroglyphs of Weierstrass equally opaque at first; but this is because those statements are generalized (that is, they express abstract ideas) and human beings *always* learn abstract concepts in the direction from the particular to the general (i.e., through prosyllogisms, not episyllogisms). That is the power of working through some number of "special cases" (like the example above); these provide examples from which the particular concepts are drawn prior to making the general concept by abstracting from these particulars.

Perhaps you have already surmised that working integration problems would be extremely tedious if we always had to start with the definition (figure 3) and proceed in the manner described in the example above. Fortunately, this is rarely necessary because mathematicians have, over time, compiled solutions for literally thousands of specific types of integrals. A first-year student of calculus spends a great deal of his or her time learning a small number of the basic ones plus algebraic and other techniques for getting solutions to others. Tables of integrals are published in mathematics handbooks where solutions can be looked up quite easily. These tables have even been placed on the Internet and "integral calculators" are also available, saving a person from even having to bother taking a book down from his shelf and opening it. With Internet-capable cell phones now available, scientists and engineers can even work integral problems at the dinner table in a restaurant – a technological capability that will probably result in many more lifetime bachelors and bachelorettes in science and engineering.

Newton is most often celebrated for his theory of gravitation – and that was indeed an accomplishment of great genius. However, Newton's invention of the calculus has had far more sweeping benefits for science and technology. I think celebrating him as the inventor of calculus would honor him much more. Yet this accomplishment was made possible by invention and construction of infinitesimals – another kind of funny number – and the invention and construction of limiting operations (or, more accurately, its reinvention and reconstruction; the Greeks were there first by almost two millennia).

## §7. Binary Relations and Transformations

There are many cases where we find that our ability to solve a mathematics problem is more easily done in one mathematical domain than another. One good example of this is provided by problems in electric circuit analysis. It very often is more difficult to solve for the voltages and currents in an electric circuit using "time" as an independent variable (the "time domain") than it is using "frequency" as the independent variable (the "frequency domain"). But in order for a change of domain to be useful in practice, it is obviously necessary to be able to restate results obtained in one domain (e.g. the "frequency domain") back into phenomenally equivalent results obtained for the other domain (e.g., the "time domain"). Mathematical transformations to or from one domain to another employ specific binary relations called by such names as homomorphisms and isomorphisms.

One venerable example of this is the function called "the logarithm." Logarithms were invented by John Napier in 1614 to transform multiplication of real numbers into addition of other real numbers, which was, in 1614, an easier way to perform multiplication. The term "logarithm" actually denotes a family of functions distinguished from each other by their inverse operation, "exponentiation."

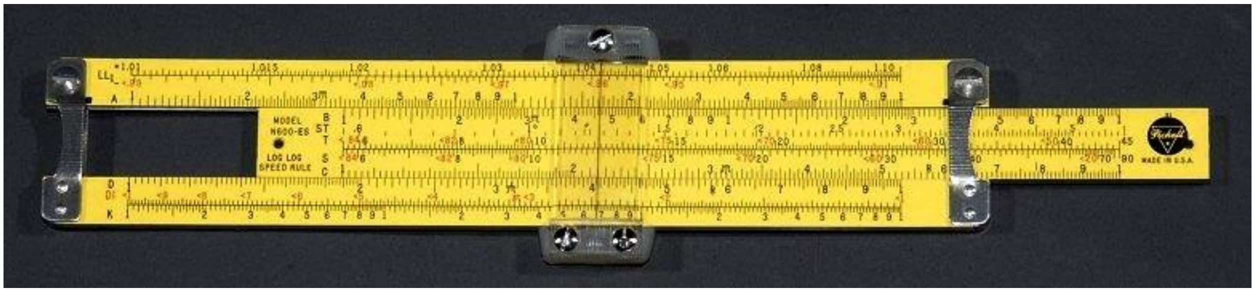


**Figure 4:** A slide rule.

The logarithm of a positive real number $x$ is denoted by $\log_b(x)$ where $b > 1$ is called the "base" of the logarithm and identifies its inverse operation, i.e.,

$$\text{if } y = \log_b(x) \quad \text{then} \quad x = b^{y}\ .$$

The most frequently used bases are b = 10 ($\log_{10}(x)$ or "common logarithm"), b = Euler's number, $e$ = 2.718281… ($\log_e(x)$ or "natural logarithm"), and b = 2 ($\log_2(x)$ or "binary logarithm"). In 1676, Leibniz[9] showed

$$\log_e(x) = \int_1^x \frac{du}{u}\ .$$

In 1620, Edmund Gunter of Oxford University invented a simple mechanical device for computing with logarithms without requiring extensive logarithm look up tables. His device is called a slide rule (figure 4). Additional scales added to Gunter's original idea added convenient capabilities for also computing trigonometric functions, square roots, and cube roots. Slide rules were heavily used by engineers until 1972 when the Hewlett Packard Company introduced the first "electronic slide rule," the HP 35 pocket calculator. The HP 35 and later pocket "scientific calculators" made the slide rule obsolete and today most freshmen entering college have never even heard of a slide rule[10].

The property that logarithms transform real number multiplication into real number addition is easily demonstrated. Let $u = \log_{10}(x_1)$ and $w = \log_{10}(x_2)$. Then observe that $10^{u+w} = 10^u \cdot 10^w$. Let $x$ be the anti-logarithm of $u + w$. Then $x = x_1 \cdot x_2$ since $x_1 = 10^u$ and $x_2 = 10^w$, QED.

Logarithms are one example of what mathematicians call "isomorphic transformations." Recall from before that a binary relation is a function $f$ that transforms (or "maps") a set $A$ to another set $B$,

$$f : A \to B$$

(page 71). If sets $A$ and $B$ are the same sets, then this is called a binary operation. An *isomorphism* is defined as follows:

> Given two semigroups $G_1 = [S, \circ]$ and $G_2 = [T, *]$, an invertible function $f : S \to T$ is said to be an *isomorphism* between $G_1$ and $G_2$ if, for every $a$ and $b$ in $S$,
>
> $f(a \circ b) = f(a) * f(b)$;
>
> $G_1$ and $G_2$ are said to be *isomorphic*.

9 There is a long-standing controversy over whether Leibniz independently invented calculus or whether he got the idea from seeing some of Newton's papers that had been written prior to publication of the *Principia* in 1687. However that may be, it is established that both men knew about calculus before 1687.

10 One time in my office, I happened to mention "slide rules" in passing to one of my freshmen. He asked, "What's that?" so I showed him my slide rule and how it worked. "Wow!" he exclaimed, "this is great!"

This definition specifies that the function $f$ be invertible. This means that there is some other function, denoted $f^{-1}$, such that for any $a \in S$ we have $f^{-1}[f(a)] = a$. If $f$ does not have an inverse function $f^{-1}$ but the definition above otherwise holds, then $f$ is called a *homomorphism* and semigroup $G_2$ is said to be the *homomorphic image* of $G_1$ under $f$.

There are many isomorphism structures of great importance in engineering and science. One of them that I mentioned earlier is the Fourier transform. It is defined by the pair of binary relations

$$F(i\omega) = \int_{-\infty}^{\infty} f(x) \cdot e^{-i\omega x}\, dx, \quad \text{and} \quad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(i\omega) \cdot e^{i\omega x} d\omega$$

where $i$ denotes the square root of -1. The Fourier transform is itself a special case of an even more general transformation called the "two-sided Laplace transform." As the symbol $\infty$ in the integral signs implies, these transforms are defined by means of limiting operations.

It would be difficult to overstate the importance of isomorphisms and homomorphisms in science and engineering. For example, Fourier and Laplace transforms are extensively used to solve differential equations – many of which would be exceedingly difficult to solve without them. Another transform, called the "$z$-transform," is used to solve difference equations. As it happens, the $z$-transform is another special case of the two-sided Laplace transform. Isomorphisms and homomorphisms comprise another case in which mathematical structures are invented and constructed to serve practical purposes.

## §8. Statistics and Probability

There are many natural phenomena for which their causes are unknown and their occurrences are therefore unpredictable. When they occur, they are said to be "random" or "stochastic." The term "random" just means "unpredictable." Nonetheless, these phenomena do appear to have describable properties and the branch of mathematics that deals with describing them is called *statistics*.

Statistics made its earliest known appearances in works of Arab scholars in the 9th to 13th centuries. Their primary application seems to have been cryptography. It underwent further development in the 16th and 17th centuries in attempts to analyze games of chance. Its development into a mathematical theory started in the 18th and 19th centuries, and its modern culmination came in 1931 in the work of Kolmogorov. With this came the introduction of the idea of "probability" – which I will soon be saying more about – and today it is usually called "probability theory" rather than "statistics theory." Indeed, it has become customary to regard the topic as being divided into "probability theory" and "statistics" (or "statistical theory") with an unspoken implication that "probability theory" is somehow more "basic" or "fundamental" and "statistics" is regarded as merely a theory of methodologies for applying "probability theory." The convention is what it is, but I am going to argue here that thinking about the topic in that way reverses the scientific priority proper to it and tends to endow a sort of modern number mysticism to the idea of "probability" that obscures its proper use in the empirical sciences.

At its root meaning, "a statistic" is a measurement of empirical events. This places objects of statistics in the position of being principal quantities of Facet B. In contrast, objects of probability are secondary quantities of Facet B. People encounter "statistics" nearly every day; but no has ever or will ever have a direct sensible encounter with the pure *noumenon* called "a probability." To more clearly understand this, let us take a look at the construction of statistical structures.

Statistics properly begins with some phenomenon that can be observed and quantitative measurements made on it that can be displayed in a graphical presentation called a *histogram*. Observed measured values, $x$, are displayed on one axis (the domain) in what are usually called "bins." Counts of how many times measurement $x$ is observed are plotted on the other axis (the range). This is illustrated in figure 5. The $x$-axis bins can be regarded as "funny numbers" of the type $a \pm b$ (as discussed earlier in this treatise). The counts are sometimes called the "frequencies" with which values falling into specific bins are seen.

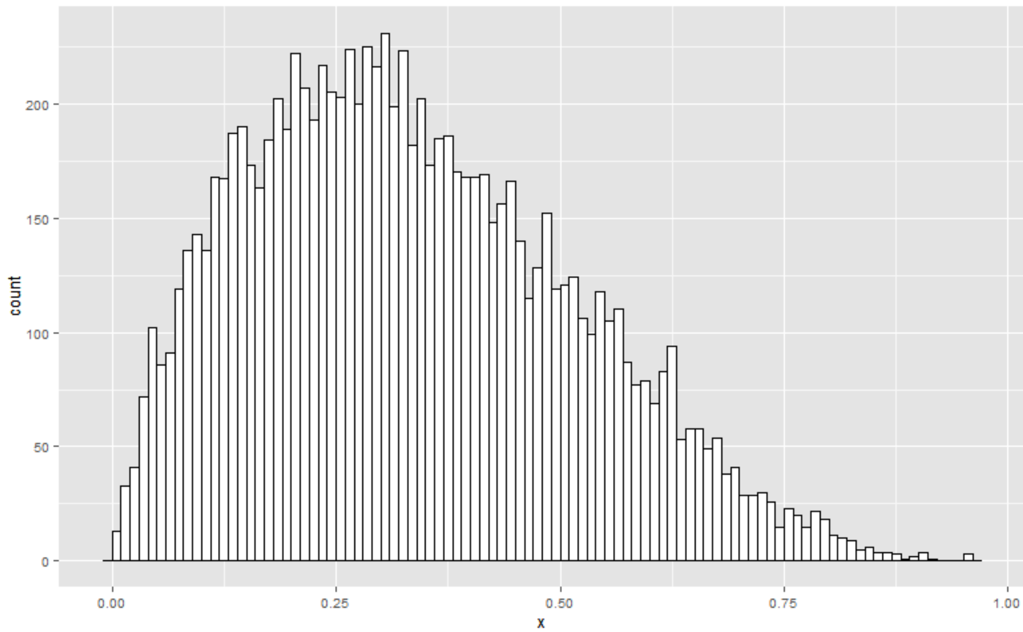



**Figure 5:** A histogram.

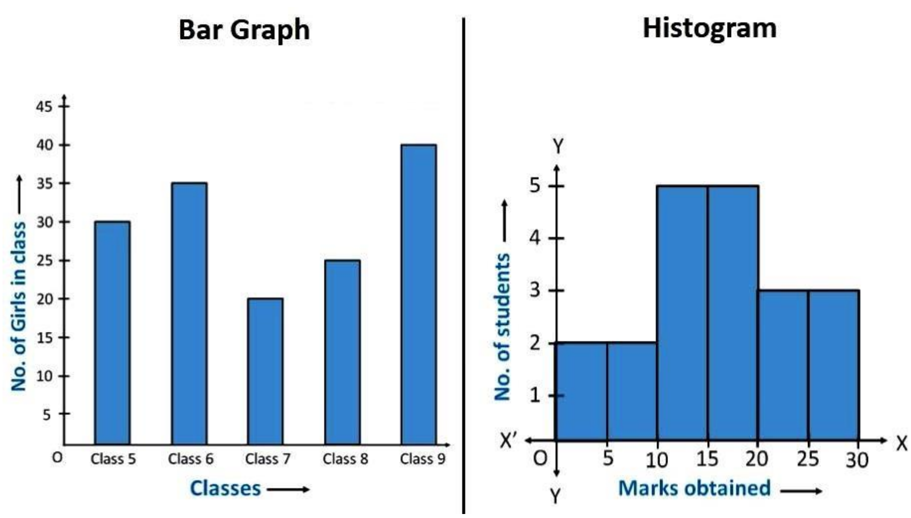



**Figure 6:** The difference between a bar graph and a histogram. Bar graphs depict frequencies of different categories of things. Histograms depict frequencies of the values obtained in measured results for observations of a single category of object.

It is important to distinguish between a histogram and another common graphical presentation called a bar graph. A bar graph is a visual presentation for comparing data among categories of things. A histogram is a display of the "shape and spread" of measured results occurring in a concrete particular phenomenon. Figure 6 illustrates this. The "x-axis" of a bar graph usually denotes a classification of some kind, e.g., different languages. The *x*-axis of a histogram denotes measured values. In both cases, the "y-axis" values are count data or measures derived from count data. It can sometimes be a little difficult to recognize the difference between a bar graph and a histogram, but it is important to be able to tell which is which. In this treatise we shall only be concerned with histograms.

Strictly speaking, histogram counts are not yet "statistics." They are the basis for *computing* statistics.

This is because a "statistic" represents some kind of average value exhibited by phenomenal data. The count data displayed in a histogram is often used to obtain different kinds of averages such as "mean values" and "variances." These values are technically referred to as "expected values" because they are measures of "what to expect" from a random phenomenon.

Actual measurements of "real world" Facet A phenomena rarely exhibit the sort of neat and tidy "shape and spread" often presented in textbooks. Real phenomena characterized as "random" or as having "random factors" in them are untidy, ragged, and, in a manner of speaking, "fuzzy." Indeed, there is a certain art involved in just *selecting* the $a \pm b$ intervals used in a histogram. Selection of different bin interval sizes can, and usually does, radically alter the "shape and spread" appearance of the data displayed in a histogram. For instance, the bin intervals used in figure 5 visually suggests that the data conforms to a certain kind of "distribution" of occurrences. In the case of figure 5, a scientist or engineer with adequate training in statistics would see the histogram "suggesting" a form of distribution called an "Erlang distribution," and this "suggestion" will tend to orient how he or she theorizes about the phenomenon. This "art part" of statistical analysis is very rarely brought up and discussed in statistics courses. Such a subjective topic is, of course, hard to teach. Nonetheless, failure to even bring it up is, I think, an important contributing factor to a generally high level of statistics-ignorance observable among engineers and scientists. Personally, I think learners would beneficially learn more about statistics if they undertook some experimental exercises that let them *see* how radically bin sizes affect histogram displays for the same set of experimental data. The first step in statistical analysis is selection of "funny numbers" $a \pm b$ for displaying bin data.

Another false impression many students take away from their statistics courses is the impression that random phenomena always exhibit a "shape and spread" statisticians call a "unimodal distribution." Figure 5 is an example of such a distribution. However, Facet A phenomena very often exhibit what is called a "multimodal distribution" such as that illustrated in figure 7. When measurement outcomes are "multimodal" this drastically affects the way in which a phenomenon is properly studied and analyzed. One exceptionally good example of an analysis of multimodal data is provided in a neural physiology study carried out and presented in Smetters (1995). Unfortunately, hers seems to be the exception rather than the rule in the practice of science for how to treat experimental data.

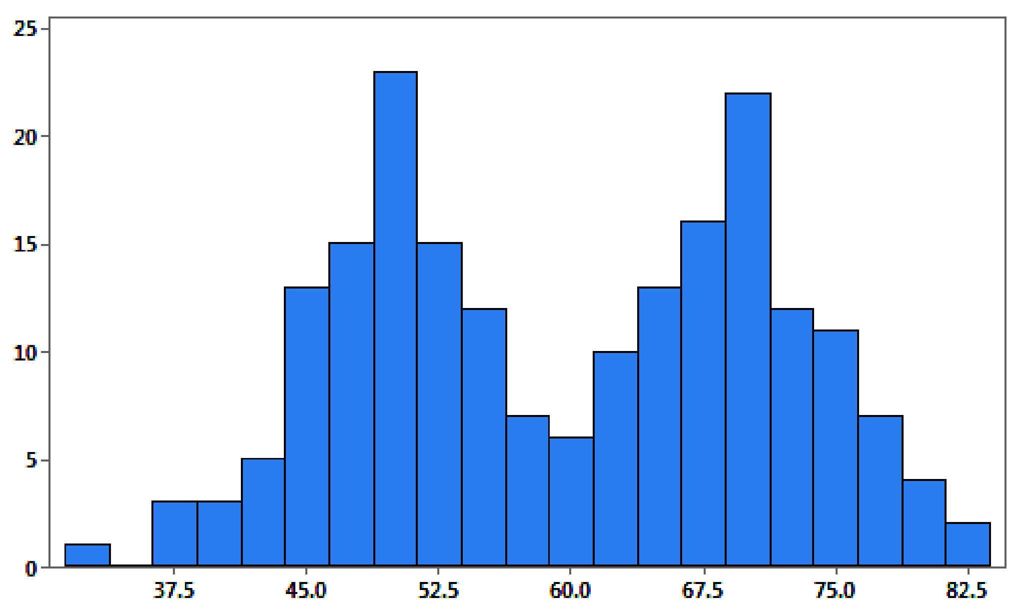



**Figure 7:** Illustration of a multimodal distribution of experimental data.

The ragged and untidy nature of Facet A measurement data severely limits what a scientist or engineer can do with this data unless mathematical approximations of it are constructed. This is where and why pure Facet B constructs called "probability distributions" are invented. I want to emphasize that these probability distributions *approximate the data*. The data does not "reflect" these distributions; *they* reflect the *data*. With very few exceptions – all of which are found in applications that have been extensively studied for years – it is a mistake to *first* assume a probability distribution and *then* look at the data. A probability distribution is a pure *noumenon* of a secondary quantity in Facet B and is nowhere found in Facet A. Over the years, many different distributions have been developed and studied.

If you assume a probability distribution without first looking at the data, it can result in some pretty silly outcomes. Allow me to share a true story with you. Many years ago, I was leading an engineering design team that was engaged in transferring a new product from the lab to production. Conjointly with a team of production engineers and operators, we were there to help wring out the last problems with the design of the product and of the production process that would manufacture it. My "opposite number" on the production side of it was a manager who I shall call "Joe." Part of the process involved a final test of the product that took 24 hours to carry out. 24 hours was regarded as a very long process step and Joe wanted to reduce the time it took if that was possible.

To that end, he gathered up data on how long it was taking products that failed this test to fail. He just assumed the failure pattern would conform to a particular probability distribution popularly known as "the bell-shaped curve." One day he came to me with a puzzled expression on his face and asked, "What does a negative time to fail mean?" I knew immediately what he had done. He had plugged his measured average time to failure and measured variance into the formula for the bell shaped curve and then observed that this curve exhibits "times to failure" that are less than zero – which, of course, made no sense because a product cannot exhibit failure before its test begins. I jokingly replied, "It means it was put together wrong in the assembly process. The negative time tells you what step in the process was where the error was made." I had expected him to chuckle and say something like, "Oh, c'mon, what does it really mean?" Instead, I was thunderstruck when he eagerly replied, "Really?!" Joe's analysis error was that he had chosen the wrong kind of probability distribution. He should have chosen a different one called an "Erlang distribution" – a distribution that cannot "report" a "negative time to failure."

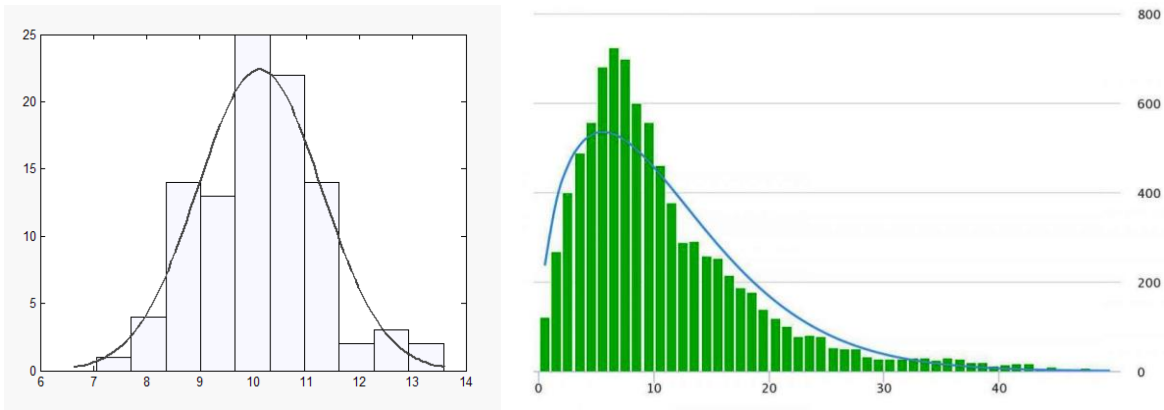
In this instance, the silliness brought about by selection of a probability distribution was obvious. There are other instances, though, where it is not. Probability analysis is not physics; it is modeling. Conclusions drawn from it belongs to a branch of knowledge called phenomenology, i.e., knowledge of how things appear to us or how they seem to be through judgments of sensory experience[11].

Figure 8 illustrates two histograms overlaid by two descriptive continuous probability distribution functions that have been scaled and fitted to the data. The distribution functions are said to be scaled because the area under their curves is made equal to the sum of the areas under the histogram rectangles. A probability distribution function always has an area under its curve equal to 1. Examination of figure 8 illustrates the fact that a probability distribution function is aimed at approximating the measured histogram since its smooth curve very obviously shows significant departures from the rectangular bins. The aim of the approximation is to allow the scientist or engineer to "calculate the probability" that an experiment or observation will yield a measured result $x \pm y$. If $Pr[(x-y), (x+y)]$ denotes this probability and $p(x)$ denotes the probability distribution function then

$$Pr[(x-y),(x+y)] = \int_{x-y}^{x+y} p(x)dx \ .$$

We can note that if $y \to 0$ then this probability becomes an infinitesimal.

[11] This is the Critical explanation of the term. Ontology-centered philosophers – e.g., Hegel, Husserl, or Heidegger – use the term somewhat differently.




**A** **B**

**Figure 8:** Two example histograms overlaid by probability distribution functions fitted to the data. A: Histogram modeled with a scaled normal distribution (also called a Gaussian distribution or a bell shaped curve). B: Histogram modeled with a scaled Erlang distribution.

The "expected value" (or "mean value" or "average value") of some function *f*(*x*) is computed as

$$E[f(x)] = \int_{-\infty}^{\infty} p(x)f(x)dx$$

given some particular probability distribution function *p*(*x*). Experimentally, if the measured empirical outcomes of some phenomenon disagree with computed expected values *this means the stochastic phenomenon is not described by the assumed probability distribution function*. When this happens, the scientist or engineer must look for a different probability distribution function in order to describe the phenomenon. Perhaps you find this statement obvious; but it underscores the fact that mathematics describes nature because *we make it* describe nature. Any mathematical statement is nothing else than a precise statement reflecting our *understanding* of natural phenomena. It is in this sense that mathematics is to be regarded as a "language" for saying things precisely and in such a way that consequences can be deduced from our statements. If experiment or observation disagree with the mathematical statement, this means the statement is wrong and we do not adequately understand the phenomenon.

Over the course of the history of statistics and probability theory, a fair number of different probability distribution functions have been derived by various mathematicians. New distribution functions are pretty rarely introduced today but breakthroughs in science and technology sometime motivate development of a new distribution. For example, the invention of the telephone and the development of telephone networks motivated the development of the Erlang distribution because telephone engineers needed to know how big a switchboard and how many telephone switchboard operators would be needed to handle the number of telephone calls that were likely to be made at any given time.

Probability densities and the integral expression for statistical expectation are the two primary new mathematical ideas that have been added to mathematical structure in this section. Let us take a look at how one might come to these ideas. Suppose we have the collection of 100 measurements, $x_i$, of some quantity as shown in Table 1 below (also see histogram figure 8A above). We might have obtained these measurements from, say, the readings from a 3-digit voltmeter. If we wanted to know the mean value of these readings, we would compute it as

$$\bar{x} = \frac{1}{100} \cdot \sum_{i=1}^{100} x_i = 10.14300 \ .$$

Note that this is equal to

**Table 1**

| $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ | $i$ | $x_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7.33 | 11 | 8.67 | 21 | 9.33 | 31 | 9.33 | 41 | 10.0 | 51 | 10.0 | 61 | 10.7 | 71 | 10.7 | 81 | 11.3 | 91 | 11.3 |
| 2 | 8.0 | 12 | 8.67 | 22 | 9.33 | 32 | 9.33 | 42 | 10.0 | 52 | 10.0 | 62 | 10.7 | 72 | 10.7 | 82 | 11.3 | 92 | 11.3 |
| 3 | 8.0 | 13 | 8.67 | 23 | 9.33 | 33 | 10.0 | 43 | 10.0 | 53 | 10.0 | 63 | 10.7 | 73 | 10.7 | 83 | 11.3 | 93 | 11.3 |
| 4 | 8.0 | 14 | 8.67 | 24 | 9.33 | 34 | 10.0 | 44 | 10.0 | 54 | 10.0 | 64 | 10.7 | 74 | 10.7 | 84 | 11.3 | 94 | 12.0 |
| 5 | 8.0 | 15 | 8.67 | 25 | 9.33 | 35 | 10.0 | 45 | 10.0 | 55 | 10.0 | 65 | 10.7 | 75 | 10.7 | 85 | 11.3 | 95 | 12.0 |
| 6 | 8.67 | 16 | 8.67 | 26 | 9.33 | 36 | 10.0 | 46 | 10.0 | 56 | 10.0 | 66 | 10.7 | 76 | 10.7 | 86 | 11.3 | 96 | 12.7 |
| 7 | 8.67 | 17 | 8.67 | 27 | 9.33 | 37 | 10.0 | 47 | 10.0 | 57 | 10.0 | 67 | 10.7 | 77 | 10.7 | 87 | 11.3 | 97 | 12.7 |
| 8 | 8.67 | 18 | 8.67 | 28 | 9.33 | 38 | 10.0 | 48 | 10.0 | 58 | 10.7 | 68 | 10.7 | 78 | 10.7 | 88 | 11.3 | 98 | 12.7 |
| 9 | 8.67 | 19 | 8.67 | 29 | 9.33 | 39 | 10.0 | 49 | 10.0 | 59 | 10.7 | 69 | 10.7 | 79 | 10.7 | 89 | 11.3 | 99 | 13.3 |
| 10 | 8.67 | 20 | 9.33 | 30 | 9.33 | 40 | 10.0 | 50 | 10.0 | 60 | 10.7 | 70 | 10.7 | 80 | 11.3 | 90 | 11.3 | 100 | 13.3 |

$\overline{x}$ = (0.01)(7.33)+(0.04)(8.0)+(0.14)(8.67)+(0.13)(9.33)+(0.25)(10.0)+(0.22)(10.7)+(0.14)(11.3)+ (0.02)(12.0)+(0.03)(12.7)+(0.02)(13.3) = 10.1430 .

The multiplicands {0.01, 0.04, 0.14, 0.13, 0.25, 0.22, 0.14, 0.02, 0.03, 0.02} are merely the bin counts $c_i$ from the histogram in figure 8A divided by the total number of measurements, $N$ = 100. Let us call these values $p_i$. The multipliers {7.33, 8.0, 8.67, 9.33, 10.0, 10.7, 11.3, 12.0, 12.7, 13.3} are merely the center values $b_i$ of the bins in figure 8A. Thus, the summation formula above can be equivalently written as

$$\bar{x} = \sum_{i=1}^{10} p_i \cdot b_i \ .$$

We can similarly compute the variance (square of the root mean squared value) $\sigma_x^2$ for this data set. When we do so we obtain $\sigma_x^2$ = 1.391659 .

Now, a 3-digit voltmeter has a display resolution of only 3 digits and one might well assume when it reads, say, 10.0 this only implies the voltage is within some range, e.g., 9.67 < $x$ < 10.33 (as suggested by the histogram of figure 8A). Put another way, a reading $x_i$ = 10.0 implies a "funny number" 10.0 ± $\frac{1}{3}$ . Let us suppose that our understanding of the system being measured leads us to assume the $x_i$ readings therefore actually represent "voltages" comprised of a "true value" $x$ plus some random "error" $\varepsilon$ that obeys a Gaussian distribution (as the probability curve fit in figure 8A suggests). The Gaussian distribution is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \cdot exp\left[-\frac{1}{2}\frac{(x-\bar{x})^2}{\sigma_x^2}\right] \quad .$$

Using the mean and variance calculated above, in $N$ = 100 measurements, the number of readings we would expect to see in the range of 10.0 ± $\frac{1}{3}$ would be

$$E(c) = N \cdot \int_{10.0-0.33}^{10.0+0.33} p(x)dx = 100 \cdot 0.2187 = 21.87 \ .$$[12]

Compare this outcome with the count scale of figure 8A, which gives a value at $x$ = 10 of about 22 counts for the Gaussian curve but 25 actual counts in the histogram. If the expected number of counts in the range between 7.67 and 8.33 is calculated, the result is 4.46 counts, which compares favorably with the actual count value of 4 seen in the histogram. Some of the other bin counts, especially out at the edges of the range of measurements, do not compare quite so favorably. For instance, bin value 12.0 ± $\frac{1}{3}$ gives an expected count value of 6.65 counts; the actual count in the histogram is 2 counts.

[12] The integral of the $p(x)$ function does not have a formula for its closed form solution but can be easily calculated numerically. The Internet provides several "on-line calculators" one can use for making this computation.

The point I would like you to note from this discussion is that the results obtained from the Gaussian probability fit are approximate in comparison with the actual histogram counts. If we took 1000 readings instead of 100 readings, would the two results draw closer together? There is simply no way to tell this *a priori*. All you can do is take more readings and see what happens. If your *model* of the system (which is what you represent with the $p(x)$ function) is accurate, the actual count data and the predictions from the probability function should draw closer to each other. If not, you will eventually see growing discrepancy between the two. The uncertainty being spoken of here is, of course, very undesirable for many reasons. That leads straightaway to the question: "Is there a reasonably economical way to reduce this uncertainty, recognizing that it can never be entirely eliminated?" The answer is yes, and is illustrated in the example following the next paragraph. There is also a second question implicit in the first one: "*How does one know* if the data and a statistical model 'disagree' with each other?" This, too, is discussed below.

The mathematical argument leading to the expressions for statistical expectation given above give another example of applying limit arguments to system models. When one is working "out in the tails" of a probability distribution, it is not at all uncommon to see larger discrepancies than one sees when working within one or two standard deviations from the mean of the distribution. Statisticians have come up with a notion called the confidence interval – a statistic that tries to inform the experimenter/observer of some measure of the uncertainty with which he should regard his results. They also have provided various methods of testing the agreement between a model and the observations of the Facet A system [Ott (1977)]. These are important pieces of information because *it is almost never the case* that the same theoretical accuracy of a statistical model holds across its entire range of operation. That in itself should be enough to give a person pause if he catches himself starting to think the mathematical model is somehow "more real" or "more true" than his actual observations of the Facet A system he is trying to understand.

Here is an example. Suppose the 100 data points in Table 1 are measurements taken from a voltage generator. Let us further assume this Facet A system is modeled as a 10 volt source plus an unwanted "noise" signal believed to be a zero-mean Gaussian noise source that generates a root mean squared voltage signal of 1.2 volts. An engineer would typically be interested in identifying the source of the noise and eliminating it as much as possible while maintaining the output of the 10 volt source at 10 volts.

The first question he would face is whether or not the data in Table 1 is consistent with these models. First he would compare the variance in his data (1.391659 volts-squared) with the variance of the modeled noise signal, $(1.2)^2 = 1.44$ volts-squared. To do this, he would compute what is called a "chi-squared statistic" [Ott (1977), chap. 12, pg. 342]. For this case, that statistic turns out to be $\chi^2 = 95.676$. He would then compare this result to a chi-squared range at a desired level of confidence; let us say he uses a typical confidence level of 95%. The range he would get in this case would be from 74.2219 to 129.561, and he would reject the hypothesis that the data and the model are consistent if his $\chi^2$ value fell outside this range. In this case, it falls within the range and so model and data are so far not-inconsistent.

This, however, doesn't finish the task. The mean value of the data must then be compared with the assumption that the mean value of the generator's output is 10 volts. To do so, he would compute another statistic; let us say he chooses a "*t*-statistic" [Ott (1977), chap. 5, pg. 102][13]. A *t*-statistic follows what is called "Student's T Distribution" and he would test to see if this test statistic falls too far out into the tail of this distribution at his desired 95% level of confidence. He will reject the model if his *t*-statistic exceeds a critical value of, in this case, 1.960 . For this set of data, his *t*-statistic calculation will give him $t = 1.2061$, which is less than the critical value. From this, he concludes that his model and the data are

[13] With 100 data points, he might also decide to base his test directly on the Gaussian distribution. However, in cases where the number of data points is less than 30, a Gaussian test isn't a reliable indicator and a *t*-test is called for. For this example, it turns out both tests return the same verdict: the data and the model are not-inconsistent. The great practical value of a *t*-test is that useful comparisons can be made on the basis of much less data than a Gaussian test requires. That means experiments are less costly and results are more quickly obtainable.

"not-inconsistent" at a 95% level of confidence[14]. With this level of confidence that he is indeed working on the right problem, he will then carry on with his task of finding out how to reduce the "noise" in his system.

Recognition that theoretical results can be quite sensitive to assumed probability distributions was one of the factors which drove the development of set membership theory – which does not assume *any* probability distribution – back in the late 1960s. It is generally unwise to ignore a significant discrepancy between a result from an experiment and a theoretical model by calling the measurement an "outlier" *unless* you have a strong reason to think something might have botched up the experimental trial that produced such an "outlier." I am not saying here that the theoretical methods are worthless; experience teaches us quite the opposite. I am saying that it is unwise to have more faith in your model than a saint has in Christ. Keep an open mind about things, occasionally do a few "reality checks," and remember the real object of interest lies in Facet A while *all* of the mathematical theory lies in Facet B.

Remember, too, that a probability is not a cause. The cause of a "random" phenomenon is unknown and there may even be multiple causes at work in one observable phenomenon. Margenau wrote,

> [A] cause becomes unique when it refers to a stage in a process involving the whole system under consideration. Or, to put it in terms of our previous analysis, it becomes unique when it refers to the entire *state* of a physical *system*.
>
> The reason why the causal assignment of the first examples in our list was somewhat indefinite is that the causes did not embrace a sufficiently large situation. They were what we shall henceforth call *partial causes*. [Margenau (1977), pg. 393]

Margenau's "list of causes" to which he refers did not contain probabilities. A probability does not cause *anything*. I cringe whenever I hear a cosmologist tell a student or layperson something like "the Big Bang was caused by a statistical fluctuation in the vacuum." I am far from convinced a "Big Bang" that "started the universe" ever really happened at all – "Big Bang theory" is a hypothesis, not a fact – but if it did, I am certain it wasn't caused by any sort of *noumenal* "probability object."

## §9. Other Structures

The brief survey presented above touches upon only some of the many different kinds of structures that collectively make up the constituents of mathematics. The world of Facet B is unbounded in imagination and mathematical structuring is constantly adding to this world. I would feel remiss if I did not at least mention a few more examples of mathematical structures: order structures; topological structures; geometry structures; graph structures; measure structures; and more still besides these. Entire books have been written about each of the ones just named. Topics in applied mathematics generally contain ideas gathered from many of the structures studied in pure mathematics and combined in specific ways suitable to the purposes one has for studying particular applications. The Bourbaki mathematicians of the mid-20th century advanced a proof that *all* of mathematics was constructed out of just three basic "mother structures" – namely, topological structures, order structures, and algebraic structures – and their combinations. Stated like this, the Bourbaki view seems comfortingly simple but – as is the case for all highly abstract ideas – this formal summation of mathematics masks a great amount of specific detail.

For example, measure theory deals with generalizations of the ideas of length, area, and volume. Our

[14] Using 95% as the confidence level is a heuristic for scientific practice. What I mean by this is that over the course of many years, scientists and engineers have found that they obtain more satisfactory results using this confidence level than they do using, say, 90% or 99%. "The 95% rule" has become a sort of *de facto* standard for "doing good science" and a scientist or engineer is expected to be able to defend using some different level of confidence. Such a defense is necessary because by "steering" the level of confidence a researcher can make the outcome be anything he *wants* it to be. This way of making the answer come out the way you want it to is called "lying with statistics."

everyday notions of these things, so familiar in geometry, turn out to be starting points for mathematical descriptions of many other things such as "the distance between" two code words in error correcting codes, how similar or dissimilar strings of letters are in written languages, and many other things as well. Any function that assigns a non-negative real number to the subsets of a set is called a "metric function" if that function satisfies the following three conditions:

> Given a set of points, a metric function is a function which gives any pair of points $x$ and $y$ a non-negative number $d(x, y)$, called the distance between $x$ and $y$, such that
>
> Condition 1: $d(x, y) = 0$ if and only if $x = y$;
> Condition 2: $d(x, y) = d(y, x)$; and
> Condition 3: $d(x, y) + d(y, z) \geq d(x, z)$ for any points $x, y, z$ of the set.

Mathematically, there are a great many ways to define "distances." The set of points plus the metric function defines what is called a "metric space." A couple examples of this are given below. Note, too, that the "points in the space" need not necessarily be the geometric points of Euclidean geometry. The mathematical concept of a "distance" construct is very general and very abstract.

Order theory is a formal mathematical framework for expressing concepts such as "$x$ is less than $y$" or "$x$ precedes $y$," or "house cats are smaller than lions; lions are smaller than elephants," or "an obtuse triangle contains exactly one angle that is greater than 90 degrees and less than 180 degrees." It is probably fair to say that applied mathematics doesn't get very far before you need to introduce the idea of ordering the objects or properties of objects to which mathematics is being applied. To construct an order structure one begins with a binary relation on a set, called a partial order, that has three general properties. If we let the symbol "≤" denote this binary relation, then a partial order is

> a relation "≤" between the members of a set $S$ that satisfies the following three conditions:
>
> condition 1: *the reflexive condition*: $a \leq a$ for each $a$ in $S$;
> condition 2: *the antisymmetric condition*: for $a$ and $b$ in $S$, $a \leq b$ and $b \leq a$ can both hold only if $a = b$;
> condition 3: *the transitive condition*: if $a$, $b$, and $c$ are in $S$, then $a \leq b$ and $b \leq c$ together imply $a \leq c$.

Although we are accustomed to "automatically" interpreting the symbol ≤ to mean "less than or equal to," in the definition above ≤ could equally well mean "greater than or equal to," "less than," or "greater than" and the definition would still hold. And this is not all. You could use $a \leq b$ to mean "$a$ is inside $b$" or "straight flush ($a$) beats four-of-a-kind ($b$)" or "Alex is braver than Ben" or "Alice is prettier than Beth." The idea of an order relation is very general. We can use it to make all kinds of comparisons. If a pair $a$ and $b$ in the set is given an ordering relation (either $a \leq b$ or $b \leq a$), then they are said to be "comparable." If every $a$ and $b$ in the set are comparable then the set is said to be a "totally ordered set." Otherwise, it is said to be a "partially ordered set" or "poset."

Topology is the study of properties of objects (often but not always geometric objects) that are unchanged by continuous deformation – i.e., deformations that do not tear, cut, or "paste together" the object. In a sense, topology is concerned with studying properties that elude description by geometry. Examples include such notions as: unquantified perceptions of proximity or separation; enclosure or surrounding of one object by another; or continuity between objects. In topology theory one is interested in recognizing "topologically equivalent" objects. These are objects for which transformations $f(a) \rightarrow b$ and $f^{-1}(b) \rightarrow a$, called "homeomorphisms," exist that "map" the description of the one object into that of the other. In this case, $a$ and $b$ are said to be "in the same neighborhood." Homeomorphisms play the role of "continuous deformations" of the objects.

To construct a topology, you must first define some property or set of properties exhibited by your "topological objects." Two objects $a$ and $b$ are then said to be "topologically equivalent" is there is an invertible homeomorphism that preserves these properties. For each object $x$ in your overall set of objects

$X$, you define a set $u_x$ of neighborhoods which each contain $x$ and are constructed such that,

(1) every member $x$ of $X$ is in some neighborhood; and

(2) the intersection of any two neighborhoods of $x$ contains a neighborhood of $x$.

This is called a "neighborhood system at $x$." The assignment of neighborhood systems for each object $x$ in your set of objects $X$ is called a "topology" of $X$.

Perhaps the example most favored by topologists is deforming a donut shape into the shape of a coffee cup. To a topologist, a circle and a square are topologically equivalent, as are the letters S and U when printed in Arial font but not when printed in Edwardian Script font, *S* and *U*. If you wanted to, you could construct a "pronunciation topology" in which the words 'weigh', 'play', 'may', 'say', and 'neighbor' are topologically equivalent (because they all exhibit the "long A" sound in American dialects of English). In such a topology, 'father' would not be topologically equivalent to them. If you wanted to, you could construct a topology for poetry around basic elements of a poem such as: a) metrical patterns of lines of poem ("meter"); b) rhythm; (c) rhyme and harmony; and d) tone and intonation. Indeed, Aristotle's *Poetics* (*c*. 335 BC) discourses on some basic elements of "poetry topology" without mathematical hieroglyphics.

Graph theory is the study of mathematical structures used to model pairwise relations between objects. Mathematically, a graph is a set of "vertices" ($V$) and a set of "edges" ($E$) connecting the vertices to one another. Because a graph defines how its vertices are in some way connected, a graph is also describable as a topological space. We find it used in a great many applications in biology, chemistry, physics, computer science, electrical engineering, the social sciences, and even in linguistics (e.g., when one "diagrams sentences"). Computer languages (such as C, Fortran, Algol, etc.) use graphs to define the syntax structure of their language constructions. In the study of neural networks, one finds ideas of graph theory and topology theory co-mingled rather seamlessly in modeling both the nervous system as well as the "artificial neural networks" studied by engineers and computer scientists.

Were you to cast about looking for a common theme to unite the variety of ways by which we construct mathematical structures, you might do no better than the following: Definition and construction of mathematical structures aims to find ways to take already-familiar ideas and models and adapt them to the study of new things or the study of old things in new ways.

For instance, you are already familiar with the idea of the distance between one side of a street and the other. The everyday *measure* of this is called "the Euclidean distance" by mathematicians. Now ask yourself: does it make any sense to talk about the "distance" between the strings of letters "Mary had a little lamb" and "please pay at the desk"? It turns out that *if you want it to make sense*, *you can **make** it make sense*, and you can make it make sense in more ways than one.

One way is to define a metric function called a "Hamming distance" based on the number of letters in the strings that are not the same in the same positions in the two strings (using the "blank" symbol as a "letter"). If you do, then the Hamming distance between these two strings is $d_H = 19$ because there are 19 places where a letter in one of them is not the same as the letter in that same position in the other. The idea of Hamming distance is heavily utilized in the design of computer logic circuits and in the design and analysis of error correcting codes in communication systems theory.

But this is not the only "funny way" to do it. You could, instead, define a "string metric" based on the first position where the letters in the two strings differ. For instance, these two strings happen to differ in their first letters and so a distance metric of, say, $(0.5)^1$ defines a "distance" between the two strings. So, in fact, would another "metric" defined as $(0.2501)^1$. More generally, if $r$ is any real number between 0 and 1 (not including either of them), and the first position where two strings differ is a positive integer $j$, then the quantity $(r)^j$ is a valid "distance metric for strings." String metrics are sometimes called "inverse similarity" metrics because the bigger the number is, the less alike the strings are. Applications for them are found in information theory, computer science, and linguistics theory.

In light of this, let us reconsider Feynman's statement that "mathematics is language plus reasoning." Before you can reason about something you have to "be able to say things" about that something. To reason about it logically, you have to be precise in "the things you say" about your object. In a nutshell, that is what mathematics regarded as a language is: a way to say things precisely about an object and in such a way that inferences and deductions may be drawn from what you have said.

There is an interesting analogy that can be made between mathematics and its uses in science and engineering, on the one hand, and the learned nonfiction literature of the 18th century. It is an easily observable fact that, in "living languages" (languages actually spoken by living people), the meanings of words shift and evolve over time. For example, the word "dungeon" originally meant "a massive tower in the interior of a castle." Later it *also* came to mean "a dark and usually underground prison or vault." Quite obviously, the objects described by these two connotations of "dungeon" are very different things.

European scholars in the 18th century knew this about living languages. They also knew the evolving character of living languages could – and probably would – eventually render their writings ambiguous, obscure, or even nonsensical to readers as the decades and centuries passed.[15] To guard against having what they said misinterpreted by later readers, they adopted the precaution of *using a dead language* when they wanted to say something in a way that preserved the ideas they wanted to convey. In Europe, the dead language most often chosen was Latin because Latin was a language that: (1) was no longer an everyday spoken language used by people in any country; and (2) all educated people in Europe studied and learned Latin in school. Because it wasn't a language ordinary people used, the meanings of Latin words would not change over time. So it was that John Locke wrote,

> *Salus populi suprema lex* is certainly so just and fundamental a rule that he who follows it cannot dangerously err. [Locke, *Concerning Civil Government*, paragraph 158]

When a writer in this era dropped a Latin phrase into one of his sentences, he did not provide a footnote to translate his Latin phrase. Indeed, doing that would defeat his purpose for using Latin in the first place. The flaw in this tactic appears when "educated people" no longer understand Latin.

Regarded in this same spirit, mathematics is "the new Latin" for science and engineering. Mathematical phrases always can be translated into a living language, much as the hieroglyphics of the ancient Egyptians or the Mayans can be translated into English. A person who wants to become proficient in the use of mathematics must perforce learn its very peculiar "dead language" (even though it isn't really "dead"; scientists, engineers, and mathematicians are constantly adding to it by introducing new objects, ideas and structures). One important benefit the principle of permanence brings to mathematics is that it helps preserve the ideas expressed in mathematics originally constructed long ago.

If the aim is to find ways to take already-familiar ideas and models and adapt them to the study of new things or the study of old things in new ways, how does one "take aim" at this? This question takes us into considerations of the last topic of this treatise: mathematics and human aesthetical judgments.

## References

**Author's website for citations:** http://www.mrc.uidaho.edu/~rwells/techdocs/

[15] Translators – especially those who translate ancient texts – have to face this issue every working day. That is why translations of these texts by different translators can look very different from one another. The language doesn't even have to be all that ancient. You will find radically different translations of the "Shakespearean German" that Kant spoke in the 18th century. The task of a translator is to translate *ideas*, not words. One very prominent example is found in translations of the New Testament from ancient Greek; it is why different Bibles read differently.

Berkeley, George (1734), *The Analyst*, in *The Works of George Berkeley*, vol. 3, pp. 1-60, Oxford, UK: the Clarendon Press.

Cantor, Georg (1897), *Contributions to the Founding of the Theory of Transfinite Numbers*, NY: Dover Publications, 1955.

Davis, Philip J. & Reuben Hersch (1981), *The Mathematical Experience*, Boston, MA: Houghton Mifflin Co.

Feynman, Richard P. (1965), *The Character of Physical Law*, Cambridge, MA: The MIT Press, 21st printing, 1994.

Lavoisier, Antoine (1789), *Elements of Chemistry*, NY: Dover Publications, 1965.

Mandelbrot, Benoit (1977), *The Fractal Geometry of Nature*, NY: W.H. Freeman and Co.

Margenau, Henry (1977), *The Nature of Physical Reality*, Woodbridge, CN: Ox Bow Press.

Newton, Isaac (1687), *Mathematical Principles of Natural Philosophy*, Seattle, WA: printed by CreateSpace, a division of Amazon (2010).

Newton, Isaac (1736), *A Treatise on the Method of Fluxions and Infinite Series*, Gale ECCO Print Editions, 2018.

Ott, Lyman (1977), *An Introduction to Statistical Methods and Data Analysis*, North Scituate, MA: Duxbury Press.

Péter, Rózsa (1957), *Playing With Infinity*, NY: Dover Publications, 1976.

Piaget, Jean (1970), *Main Trends in Interdisciplinary Research*, NY: Harper & Roe, 1973.

Poincaré, Henri (1914), *Science and Method*, Bristol, England: Thoemmes Press, 1996.

Smetters, Diana Kathryn (1995), *Electrotonic Structure and Synaptic Integration in Cortical Neurons*, Ph.D. dissertation, Massachusetts Institute of Technology, June, 1995.

Tikhonov, Andrey N. & Vasiliy Y. Arsenin (1977), *Solutions of Ill-posed Problems*, NY: John Wiley & Sons.

Timothy, Eva Koleva (2010), *Lost in Learning: The Art of Discovery*, Athenaeum Publishing.

Wells, Richard B. (2006), *The Critical Philosophy and the Phenomenon of Mind*, available from the author's web site.

Wells, Richard B. (2009), *The Principles of Mental Physics*, available free of charge from the author's web site.

Wells, Richard B. (2010), "A lumped element modeling schema for metabotropic signaling in neurons," available free of charge from the author's web site.