Toposes of Laws of Motion

F. William Lawvere

Transcript from Video, Montreal September 27, 1997

Individuals do not set the course of events; it is the social force. Thirty-five or forty years ago it caused us to congregate in centers like Columbia University or Berkeley, or Chicago, or Montreal, or Sydney, or Zurich because we heard that the pursuit of knowledge was going on there. It was a time when people in many places had come to realize that category theory had a role to play in the pursuit of mathematical knowledge. That is basically why we know each other and why many of us are more or less the same age. But it's also important to point out that we are still here and still finding striking new results in spite of all the pessimistic things we heard, even 35 or 40 years ago, that there was no future in abstract generalities. We continue to be surprised to find striking new and powerful general results as well as to find very interesting particular examples.

We have had to fight against the myth of the mainstream which says, for example, that there are cycles during which at one time everybody is working on general concepts, and at another time anybody of consequence is doing only particular examples, whereas in fact serious mathematicians have always been doing both.

1. Infinitesimally Generated Toposes

In fact, it is the relation between the General and the Particular about which I wish to speak. I read somewhere recently that the basic program of infinitesimal calculus, continuum mechanics, and differential geometry is that all the world can be reconstructed from the infinitely small.
One may think this is not possible, but nonetheless it's certainly a program that has been very fruitful over the last 300 years. I think we are now finally in a position to actually make more explicit what that program amounts to. As you know 30 years ago I made certain proposals in Chicago and then again

15 years ago in Buffalo. There has since been a lot of work on what came to be called synthetic differential geometry. At least 20 people in the world have made important advances in synthetic differential geometry; indeed several of these people are here. And there are also very encouraging developments about the simplification of functional analysis. So I think that on the basis of these developments we can focus on this question of making very explicit how continuum physics etc. can be built up mathematically from very simple ingredients.

To say that a topos can be built up from an object T will mean here that every object is a direct limit of finite inverse limits of exponentials of T. By exponentials of T we mean T^T, T^{T^2} etc. and of course the inverse limits involve equalizers of maps between finite products of these. Such equalizers may be considered as varieties, and in particular the equalizers of maps between finite products of T itself are intended to be infinitesimal varieties. There are actually many interesting useful toposes which are built up in that way from an object T which in some of several senses is infinitely small. Of course T is not just a single point; but it may *have* only a single point, or more generally the set of components functor may agree with the functor represented by 1 on T and its products and sums. One of these senses is that it is a space whose algebra of functions is linearly finite-dimensional; of course that presupposes that we have some linear algebra in the topos, in particular a base rig. But actually it turns out that the base rig itself can be constructed from T.

I'm going to assume T to be a pointed object

$$1 \xrightarrow{0} T$$

This arrow itself is a kind of contradiction expressing that an instant of time involves a point and yet is more than a point. A crucial role is played by the internal endomorphism monoid T^T of T. Also very important is the submonoid of that consisting (in the internal sense) of those endomorphisms which preserve the point.

I am actually going to define R to be that part. If this works we can consider

$$R^X$$

as the space of the simplest kind of intensively variable quantities. We can also consider the R-homogeneous part of the space of functionals

$$\mathrm{Hom}_R(R^X, R)$$

as representing the simplest kind of extensively variable quantities over the domain X; typically this means something like the space of distributions of compact support in X. The basic spaces which are needed for functional analysis and theories of physical fields are thus in some sense available in any topos with a suitable object T. It would be nice if we could prove that R is commutative, but I don't know how to do that from more basic assumptions. You might ask "couldn't T^T itself be commutative?" But there is a very general fact about cartesian closed categories: If an object T has a commutative internal endomorphism monoid, then T itself is a subobject of 1. Intuitively, T^T always includes constants, $T \longrightarrow T^T$ and if constants commute, they are equal. Thus although T itself may be very small, we must have that T^T is a little bit bigger than R. The idea is that a real quantity λ is just a temporal speed-up or retardation

$$1 \xrightarrow{\;\rho\;} T^{c\lambda}$$

As we will see, although R is in a sense the more familiar, the bigger monoid T^T and its actions also play very important roles. Again, there is a general fact about cartesian closed categories. For any monoid M in such a category we can consider also the category of all internal right actions of M. There are, of course, the co-free actions $(\quad)^M$, but we can also consider the action on $(\quad)^T$ where T is the space of constants of M; this functor will be right adjoint to the fixpoint functor if T has a point. In case M = T^T, this right adjoint is actually a full inclusion. We want to think of X^T as the tangent bundle of X with the evaluation at the point of T as the bundle projection. (This idea is already described by Gabriel in SGA3, for example.) The fullness is in contrast with the situation obtained, if we consider that the tangent bundles are equipped only with the R-action, in which case maps between them are essentially contact transformations, not necessarily induced by differentiating (i.e. exponentiating) maps between the configuration spaces. The space can be recovered from its tangent space as the zero section, but even the maps between the spaces can be recovered if we take into account the action of this slightly larger monoid T^T.

As we know, there are many examples of such categories: algebraic geometry, smooth geometry, analytic geometry (real or complex), and many variations on those; actually, in my Chicago lectures I pointed out that there are many potentially interesting intermediate examples of such toposes, for example obtained by adjoining the single operation

$$\exp(-(1/(\ )))$$

to the ordinary theory of real polynomials, so that we can obtain the typical partitions of unity of smooth geometry and yet work in a concrete "algebraic extension" context. (But these are algebraic extensions of systems of quantities of various arieties, so appropriately modeled as a category, rather than just as a differential ring.) All these examples have something in common, and part of the program was to figure out what that "something" is, while at the same time providing a language powerful enough to make all significant distinctions between them. Part of what they have in common is that they are all defined over a simple base topos, the classifying topos for a pointed instant T acted on by a pointed monoid M and satisfying some rather remarkable special axioms. In this base topos M probably is T^T since there is nothing else between, although without additional hypotheses that isomorphism will not persist into arbitrary toposes defined over this base. In the standard examples, exponentiation by T (the tangent bundle concept) is actually preserved up to isomorphism by the inverse of the classifying morphism although this is not a general topos-theoretic fact. Those examples differ mainly in the higher types, that is, in the precise determination of the maps whose domain is the finite-dimensional space T^T. Anyway, since R is a basic definable sub-object of M, we see why the usual examples are infinitesimally generated in the sense of this lecture; smoothness of morphisms between infinite-dimensional spaces has been successfully tested via smooth maps from finite-dimensional varieties for 300 years. In the standard examples the functor to this infinitesimal base topos has additional adjoints so that the latter is actually even an essential sub-topos; this implies that there is a comonad on the big topos which deserves the name of infinitesimal skeleton.

In all the standard examples the object T is isomorphic to the spectrum of the algebra of dual numbers. This implies some rather remarkable axioms that the pair M,T may be required to satisfy. For example, T has a fixed point operator

$$T^T \longrightarrow T$$

assigning to each endomap a fixed point of it; in this case this operator is nothing but the bundle projection (evaluation at zero) and can be interpreted as a map with domain M. This is a consequence of a more general remarkable axiom, namely if β is in M and if λ is in the zero-preserving part R of M, then

$$\beta\lambda\beta = \beta^2\lambda$$

Of course, this striking commutation relation is interpreted as a commutative square whose domain is $M \times R$ in our infinitesimal base topos.

The speedups/retardations λ of the temporal instant $1 \xrightarrow{0} T$ should actually form a rig R. The multiplication is just composition of speedups, but the addition is also intrinsic, somewhat as in Mac Lane's 1950 analysis of linear categories. More precisely, the requirement is that the extensive-quantity functor

$$\mathrm{Hom}_R(R^X,R)$$

is additive in the sense that it takes finite coproducts of spaces X to cartesian products, the needed projections coming ultimately from the point 0 of T. Of course, it suffices to assume this for the case $X = 2$. It is the expectation, that in the smooth world R-homogeneous maps are automatically linear, that underlies this axiom.

Many people have thought about related questions. For example, Peter Freyd had many unpublished ideas, and David Yetter's thesis develops some of our suggestions quite far. The part of R consisting of elements of square zero may be called D as has become customary in synthetic differential geometry. An isomorphism between T and D should amount to the same thing as a choice of a unit of time.

There is another striking property which seems to be frequently correlated with being very small. In order to settle once and for all the various terminological differences, perhaps we can use

a. t. o.m.

as an abbreviation for "amazingly tiny object model". Whatever we call it, the property is that of the exceptional existence of an additional right adjoint. Since I first wrote about this in 1980, it occurred to me that a suggestive name for this adjoint is fractional exponentiation. Briefly, certain very special objects T may be not only exponentiable, but also fractionally so in the sense that there is an adjointness where the new functor is denoted as the fractional exponent $1/T$.

$$\frac{X \longrightarrow Y^{1/T}}{X^T \longrightarrow Y}$$

Of course, $(\ )^T$ itself is defined by another adjointness (lambda conversion)

$$\frac{W \longrightarrow X^T}{T \times W \longrightarrow X}$$

More generally, if A is any object (not necessarily an a.t.o.m.) which is exponentiable, then we even have fractional exponents A/T. These fraction symbols compose as right adjoint right operators. When the denominators are trivial, this composition is just represented by cartesian product of the numerators. When the denominators are general a.t.o.m.'s the composition is not commutative, but nonetheless can be reduced to the simple fractional form. That is, it follows from the assumed adjointnesses that (right actions)

$$(1/T)B = B^T/T$$

These fractional exponents will play a crucial role in what follows.

2. Galilean 'monoids' for 2nd order ODE's in toposes

Nowadays many mathematicians study abstract objects that are called dynamical systems. Dynamical systems conceptually are intended as what one might call an analysis of Becoming. Already with Aristotle it became customary to analyze Becoming into two aspects, Time and States, with the Time somehow acting on the States. There are many variants on this model, but it is the one we still have. The action of Time on the States is the particular law of motion. More precisely, a given model of Time (a discrete monoid, a continuous group, etc. etc.) serves as an Abstract General which is accompanied by the Concrete General which is the category of all dynamical systems, i.e. systems of states, acted on by that model of Time.

But in a sense much of the current work on dynamical systems is within a framework that still hasn't caught up with Galileo. Galileo made a big advance on the basic idea. In his book "Dialogues concerning two new sciences" written toward the end of his life, Galileo put forward the dynamical refinement of the Time/States analysis which involves the following features.

(1) States are states of Becoming. This again admits many variants: the States may involve velocities, or memories, or destinies, but in any case they themselves should be more structured than just points which abstract static Being particularized as configurations. [In Birkhoff's 1919 Palermo paper which gave rise to the theory of fiber bundles, states are fibered over configurations.]

(2) The particular law of actual motion is accompanied by another law which is not the actual law, but which "would be if there were no forces", as Newton put it. This accompanying law is called inertial or geodesic or spray. The latter merely means that the law is homogenous with respect to the monoid R of time–speedups. Thus the Abstract General itself is more detailed and refined than just a group. (It is of course not excluded that the actual law may be itself homogenous in some particular cases, but the accompanying inertial law always is, it seems.) Although the notion of affine connection is given a much more complicated explanation in most text books, in fact it just expresses this homogeneity idea: if we speed up by a factor λ, then move ahead inertially in time for duration t, we arrive at the

same state as if we had proceeded without speedup for a duration λt and then sped up.

(3) Typically, all the laws constitute an infinite-dimensional affine space which is not a vector space, but the specification of the inertial law provides an origin in this space. Thus we can define the specific force to be the difference between an actual law and the inertial law, and the forces can be added vectorially.

There could be no science or technology without something like feature (3). The actual motion of a piece of chalk thrown into the air is influenced by Jupiter's third moon and any number of other things. But the most important thing is gravity, or the most important thing is wind resistance, or wind resistance and gravity, etc., i.e. we can make an understandable theory of a law of motion which depends only on a few forces, and to the extent that other forces really are negligible that theory will lead to workable technological design. It is the specification of the inertial law as a zero in the affine space of all laws which permits the vectorial addition of individual simplified laws (or rather of their corresponding specific forces). For example, a viable law s of Becoming might be (without mentioning forces explicitly) an alternating sum

$$s = s_1 - s_0 + s_2$$

(so that the coefficients add up to 1 as required for an affine combination), wherein the inertial law s_0 is homogeneous with respect to all of R , whereas the others (without themselves being linear in the usual sense), might enjoy some restricted homogeneities. For example, s_1 might be a purely reactive law homogeneous with respect to those lambdas in R which are involutions, while s_2 might be a purely dissipative law homogeneous for the lambdas in R which are idempotent; the first is often expressed by saying that a purely reactive force (no friction present) enjoys reversibility in time, while the second expresses roughly that pure friction or viscosity is inoperative when the velocity happens to be zero. Each of these three laws s_1, s_0 , or s_2 could be considered as a simple model in its own right, but the alternating sum will often be more accurate. Expressed in terms of the specific force laws this combination is

$$s_i - s_0 = f_i \qquad i = 1, 2$$
$$f_1 + f_2 = f$$
$$s - s_0 = f$$

What, more precisely, do we mean by a 'law'? and how could the laws possibly form a topos as promised in my title? First, note that the usual 'dynamical systems' involving for example the smooth actions of a monoid, if properly construed, will surely form a topos with all the virtues that that entails such as internal logic, good exactness, function space of 'dynamical systems', etc. Likewise, the infinitesimal version of such systems i.e. vector fields or first-order ODE's, will also form a topos as I pointed out in my Chicago lecture. But what about actual dynamical systems in the spirit of Galileo, for example, second-order ODE's? [Of course, the symplectic or Hamiltonian systems that are also much studied do address this question of states of Becoming versus locations of Being, but in a special way which it may not be possible to construe as a topos; in any case, most systems arising in engineering are not conservative.]

To specify an Abstract General whose corresponding Concrete General will consist of state spaces, each equipped with its own genuinely dynamical law, I propose the following. Consider any given map

$$T \longrightarrow A$$

in a topos of spaces, subject only to the restriction that its domain should be an a.t.o.m. (the codomain A need not be an a.t.o.m., although it often will be). If I have any space X I can consider the restriction

$$X^T \longleftarrow X^A$$

along the map induced by my given map on the map spaces. The kind of structure that I want to consider is that of a given section s to this restriction map

$$X^T \xrightarrow[s]{} X^A$$

This section will serve as a law of Becoming in the sense that given a map from T to X, considered as a state of instantaneous Becoming, the law will provide a definite extension to A considered as a distinctly longer instant.

$$\begin{tikzcd} T \arrow[r] \arrow[dr, "x"'] & A \arrow[d, "s(x)"] \\ & X \end{tikzcd}$$

The standard example has T as a first-order infinitesimal instant and A as a second-order instant. In that case the choice of a unit of time would identify A with the part D_2 of R consisting of elements of cube zero in T^T. An actual motion following a law s would be a map (in the dynamical topos) whose domain is a relatively small object idealizing the state space of a clock, i.e. an interval of time equipped with its own (often homogeneous) law which the map must preserve.

Theorem: For any given map $T \longrightarrow A$ in a topos with natural number object, where T is an a.t.o.m., the category of all pairs X, s as above, with the obvious motion of morphism, is a topos lex-comonadic over the given topos. In fact, the resulting 'surjective' geometric morphism is essential.

Proof: By the basic adjointness such a section is equivalent to a map

$$X \longrightarrow X^{A/T}$$

which, when 'evaluated at' the given $T \longrightarrow A$ reduces to the identity on X. The co-pointed endofunctor (pullback of the fractional exponentiation) has a left adjoint. Therefore, iterating it and passing to the sequential limit yields a lex-comonad which even has a left adjoint monad. Since the comonad is left exact, the coalgebra/algebras for this pair constitute a new topos, but they are equivalent to the laws of motion s under discussion. The essentiality is thus a special case of the Eilenberg-Moore theorem about adjoint monads.

Thus we see that there is a Galilean generalization of the notion of monoid. Recall that any monoid in a 'cartesian' closed category is equivalent to a pair consisting of an adjoint monad with its right adjoint comonad. But the converse of that statement is true only if the adjointnesses are internal. In our case the right adjointness of the comonad is defined only over some lower topos (as was discussed in my 1981 Cambridge lecture and investigated in Yetter's thesis). This is, however, still a very special kind of lex comonad since it is generated by this fractional exponent.

To sum up, the actions of such a Galilean 'monoid' thus constitute a topos of laws of motion in the Galilean dynamical sense. For example, if A consists of second-order infinitesimals, all the usual smooth dynamical systems, including the infinite-dimensional ones, (elasticity, fluid mechanics, and Maxwellian electro-dynamics) are included as special objects.

3. Infinitesimals bodies too?

Galileo's second new science, as interpreted by Noll, concerns the particularity of the ways in which constitutive relations of actual materials give rise to laws of motion s on the configuration space X *of a body*, when the body is subjected to arbitrary external conditions. While there was no time at the AMS lecture to elaborate on that, I did discuss it in my 1992 lecture in the engineering faculty at Pisa and in my talk at the 1993 Nollfest. In addition to providing a flexible general conceptual setting for considerations of materials science, the methods involving infinitesimal objects in toposes seem to also offer a definite particular model for the kind of surfaces studied by the Cosserat brothers and for the pseudorigid or zero-dimensional bodies studied by Muncaster and Cohen.