Applied Categorical Structures **4**: 167–174, 1996.
© 1996 *Kluwer Academic Publishers. Printed in the Netherlands.*

Unity and Identity of Opposites in Calculus and Physics

F. WILLIAM LAWVERE
Department of Mathematics, SUNY, 106 Diefendorf Hall, Buffalo, NY 14214, U.S.A.

(Received: 5 December 1994; accepted: 6 May 1995)

Abstract. A significant fraction of dialectical philosophy can be modeled mathematically through the use of "cylinders" (diagrams of shape Δ_1) in a category, wherein the two identical subobjects (united by the third map in the diagram) are "opposite". In a bicategory, oppositeness can be very effectively characterized in terms of adjointness, but even in an ordinary category it may sometimes be given a useful definition. For example, an effective basis for teaching calculus is a ringed category satisfying the Hadamard–Marx property. The description in engineering mechanics of continuous bodies that can undergo cracking is clarified by an example involving lattices, raising a new questions about the foundations of topology.

Key words: cracking, dialectics, differentiation, Hadamard lemma, poll bodies, van der Waals fluids.

In early 1985, while I was studying the foundations of homotopy theory, it occurred to me that the explicit use of a certain simple categorical structure might serve as a link between mathematics and philosophy. The dialectical philosophy, developed 150 years ago by Hegel, Schleiermacher, Grassmann, Marx, and others, may provide significant insights to guide the learning and development of mathematics, while categorical precision may dispel some of the mystery in that philosophy. In any case, the structures described in the definition below occur frequently enough in mathematics to suggest their systematic study.

I

A *Unity and Identity* (UI) of two maps with a common codomain C is a third map with domain C which composes with both to give isomorphisms. The existence of such a third map obviously implies that the two maps are subobject-inclusions, that these two inclusions have isomorphic domains, and that C is retracted onto both of these subobjects, but moreover that there is a *common* retraction in the following sense: any UI in any category is canonically isomorphic to one in which both composite isomorphisms are actually the identity map. In this view, a UI is just a map equipped with two sections, or equivalently, is a common retraction for two subobjects whose underlying objects are identical. Such a diagram with two objects and three maps is of shape Δ_1, where Δ_1 is the category of order-preserving maps between totally-ordered sets of one and two elements, respectively; up to splitting idempotents, the diagram is determined by a left action of the three-element monoid generated by two elements e_0, e_1 satisfying the four equations $e_i e_j = e_i$. The existence of a common retraction implies a striking property not shared by most pairs of subobject inclusions with a common domain: the intersection of the subobjects is given by the equalizer of the inclusions! In such a UI diagram, i.e. $B \rightrightarrows C \to B$ with both composites the identity, we may unambiguously refer to B as the smaller and C as the larger of the two objects. Besides the names UI or Δ_1-diagram, a third name for such a structure is "cylinder". This name is suggested by the special class of examples constructed by starting with any bi-pointed object $1 \rightrightarrows I$ (where 1 is terminal), then taking the cartesian product with any "base" object B to obtain "left and right end" inclusions

$$B \rightrightarrows C = I \times B$$

which are united by the *projection map*. Thus any cylinder "unites" in one object C the left and right which are (although at "opposite" ends) "in themselves identical" (with B!). Even the generalized cylinders have an additional structure consisting of the uniting map's fibers, which may be pictured as threads running between the left and the right ends. But these threads are in general not necessarily isomorphic with each other, although in the special examples they are all isomorphic with the original bi-pointed object I.

In a 2-category, two parallel maps may be called *adjointly opposite* (AO) if there is a single third map which is right adjoint to one of the given pair and left adjoint to the other. (Here "opposite" does not necessarily imply non-isomorphic; for example, a Frobenius homomorphism of rings induces a forgetful functor between categories of modules whose left adjoint is isomorphic to its right adjoint.) In case the third map is *also* a UI for the adjointly opposites, then the AO are of course both full inclusions. Such a map, having both left and right adjoints which are moreover full inclusions, is a UIAO (unity and identity of adjointly opposites), also known as an "essential localization".

In any UIAO there is a canonical 2-map between the opposites. A composite $B \leftarrow C \leftarrow E$ of two UIAO's is again a UIAO; in such a composite we may consider that B is qualitatively smaller than C relative to E provided the left adjoint $B \to C$ inverts the canonical 2-map between the two larger opposites $C \rightrightarrows E$. For example, the category B of sets is qualitatively smaller than the category C of directed graphs relative to the category E of simplicial sets; in this example, discrete/codiscrete are AO as are 1-skeletal/1-coskeletal, the UI's strip away the higher-dimensional simplicies from a simplicial set or graph, and the "qualitatively smaller" judgement is justified because discrete simplicial sets are all (not only 1-skeletal, but also) 1-coskeletal.

Whenever we have a composite UIAO $B \leftarrow C \leftarrow E$ in which B is qualitatively smaller than C relative to E, if we consider all endomaps of E which are obtained by composing all opposites and uniters in all possible ways, we obtain a representation of a 6-element monoid which satisfies the identity $fgf = fg$.

Even in categories which are not 2-categories, it is sometimes nonetheless possible to give a useful meaning to the idea that the two subobjects united by a UI are "opposite". For example, a homotopy theory on a category might be specified by giving a class of idempotents to be considered as "deformation retracts"; a homotopy invariant is any functor which maps these distinguished idempotents to identities. A distinguished class of cylinders can be defined to consist of all UI's in which the two idempotents e_0, e_1 obtained by composing are distinguished idempotents; a given pair of maps $B \rightrightarrows Y$ can be called "homotopic" if there exists a distinguished cylinder

$$B \rightrightarrows C \longrightarrow B$$

with base B and a map h: $C \to Y$ which, when composed with $B \rightrightarrows C$ yields the given $B \rightrightarrows Y$. Every homotopy invariant is a functor which identifies homotopic maps. The converse will hold if the following axiom on the class of distinguished idempotents holds: for each distinguished idempotent $C \xrightarrow{e} C$ there exists a distinguished cylinder $C \rightrightarrows D \to C$ with base C and a map h: $D \to C$, showing that e is homotopic to 1_C. However, there are many examples in which only a weaker version of this axiom holds, as pointed out by Quillen. In the resulting theory the idea that two subobjects $B \rightrightarrows C$ be "opposite", becomes the condition that taken jointly they form a "cofibration" $B + B \to C$, which in many examples simply means that the two subobjects are *disjoint*. In a topos the classifying map h: $C \to \Omega$ of one of two such opposites will map the other to false; if the opposites are moreover part of a distinguished cylinder, then as pointed out by Grothendieck, true_B and false_B are homotopic maps $B \rightrightarrows \Omega$.

A favorite example in dialectical philosophy is the unity of the liquid and gaseous states which are both possible (for a given substance at a subcritical ambient pressure p) over a certain identical range of temperatures. That this exemplifies our mathematical definition may be seen as follows. According to van der Waals (1881),

$$T = \left(p + \frac{a}{V^2}\right)\left(\frac{V-b}{N}\right)$$

expresses temperature (in energy units) as a function of volume and pressure, where a, b, N are constants describing the attraction between molecules, the volume of the molecules themselves, and the number of molecules. There are three values of V for which $\mathrm{d}T/\mathrm{d}V = 0$, and with each such V there is a unique other V^* for which $T(V^*) = T(V)$. Let $V_1 < V_0$ be the two *positive* such V; then $V_0^* < V_1^*$ so we can consider the closed interval $C = [V_0^*, V_1^*]$ and the corresponding interval $B = [T_0, T_1]$ where $T_k = T(V_k) = T(V_k^*)$. Considering van der Waals' formula above as a map T: $C \to B$, we see that it is a UI for the liquid and gaseous maps $V_L < V_G$ which give volume as a function of temperature in each of the two phases:

$$T \circ V_L = 1_B = T \circ V_G.$$

In the above van der Waals example there is clearly also a third section of T in addition to the two opposites; this third section agrees with V_L at T_1, but agrees with V_G at T_0. Traversing this middle section of the S-shaped state space transforms some of the kinetic energy corresponding to the higher temperature $T_1 > T_0$ into potential energy of separation corresponding to the higher volume $V_0 > V_1$, thus doing $p(V_0 - V_1)$ of work. [For any thermodynamic substance in which the relation between pressure and temperature is affine-linear, the internal energy is a sum of two parts, a "kinetic" energy which depends on temperature only and a "potential" energy which depends on volume only, as can be proved by differentiating the free energy function. The amount of heat consumed by such a process is not determined by the van der Waals equation alone, because the kinetic energy function of a substance is an independent datum. By contrast, the potential energy function is $-a/V$ for all van der Waals substances.] It is not necessary to consider only isobaric processes: if we construe the endpoints as functions of p over a suitable range of pressures, the "volume interval" C becomes an S-shaped surface uniting liquid and gaseous copies of the "temperature interval" B as subsurfaces.

There seem to be many other situations in which a more or less canonical third section exists, mediating between two united identical "opposites". For example, in an essential localization (=UIAO) in Cat, if the larger of the two categories has image-factorisations, then taking the image of the canonical 2-map between the opposite inclusions defines often such a third inclusion functor of the smaller category into the larger.

II

Near the end of his life, Karl Marx wrote about the foundations of differential calculus. The essence of his line of thought, later rigorously established by Hadamard, yields an effective and simple basis for learning and developing the subject if made explicit. The problem may be stated as follows: Presupposing those laws of algebra which are equally valid for variable and constant quantities, what is additionally required in order to determine the derivatives of genuinely variable quantities and to establish the laws of the derivative? The answer is the unity and identity of opposites permitting a single variable to be split into two like variables and later collapsed again to one. How can we make this conclusion into precise mathematics?

The category of commutative rings may be taken to be a fair embodiment of the algebra of variable and constant quantities, so we consider a unity and identity of "opposites" in this category, that is, three homomorphisms, one of which is a retraction for both of the other two. An important kind of example is

$$\begin{tikzcd} C^\infty(M \times_N M) \arrow[d] \\ C^\infty(M) \arrow[u, bend left=60] \arrow[u, bend right=60] \end{tikzcd}$$

where $M \to N$ is a given morphism of smooth manifolds (for example $N = 1$). That is, we think of the smaller ring as the one to be studied and the bigger ring as consisting of the even-more-variable quantities resulting from allowing each quantity y to become a pair of quantities y_0, y_1. The single uniting homomorphism downward is the one which sets the variables again equal, i.e. restriction to the diagonal, whereas the two united homomorphisms upward indicate the two trivial ways of getting a function of two variables from a function of one variable, by composing with the respective projections.

Given y in the smaller ring, let y_0, y_1 denote the two resulting quantities in the bigger ring, while if the downward homomorphism takes q in the bigger ring to b in the smaller, we can write $q \to b$. For any y in the smaller ring $\Delta y =: y_1 - y_0$ is in the bigger ring, and $\Delta y \to 0$ because of the two basic retraction equations. The "oppositeness" we need to follow Marx and Hadamard follows from the existence of genuine variables in the sense of the

DEFINITION. Relative to a given diagram of shape Δ_1 in the category of rings, an element x of the smaller ring is called *variable* if for every q in the bigger ring

$$q\Delta x = 0 \Longrightarrow q \longrightarrow 0.$$

DEFINITION. $A(x)$ is the set of all y in the smaller ring for which there exists at least one g in the larger ring with $\Delta y = g\Delta x$ in the bigger ring. For any such y, $\mathrm{d}y/\mathrm{d}x$ denotes the quantity in the smaller ring such that for all g

$$\Delta y = g \cdot \Delta x \Longrightarrow g \longrightarrow \frac{\mathrm{d}y}{\mathrm{d}x}.$$

PROPOSITION. *If x is variable, then $A(x)$ is a ring, $\mathrm{d}/\mathrm{d}x$ is well defined on $A(x)$, and*

$$\frac{\mathrm{d}(uv)}{\mathrm{d}x} = u\frac{\mathrm{d}v}{\mathrm{d}x} + \frac{\mathrm{d}u}{\mathrm{d}x}v$$

for any u, v in $A(x)$.

Proof. Let $\Delta u = g\Delta x$, $\Delta v = h\Delta x$ in the big ring. Then

$$\Delta(uv) = u_0\Delta v + (\Delta u)v_1 = (u_0 h + g v_1)\Delta x.$$

But

$$u_0 h + g v_1 \longrightarrow u\,\frac{\mathrm{d}v}{\mathrm{d}x} + \frac{\mathrm{d}u}{\mathrm{d}x}\,v.$$

PROPOSITION. *If $y \varepsilon A(x)$ and if y is also a variable, then $A(y) \subseteq A(x)$ and*

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x}$$

for all $z \varepsilon A(y)$.

Proof. If $\Delta z = g\Delta y$, $\Delta y = h\Delta x$ then $\Delta z = (g \cdot h)\Delta x$, but $g \to \mathrm{d}z/\mathrm{d}y$ and $h \to \mathrm{d}y/\mathrm{d}x$, and the passage is a ring homomorphism.

Thus we have established the basic laws of one-variable calculus. A basic example is $y = x^3$, for then $\Delta y = g \cdot \Delta x$ with $g = x_0^2 + x_0 x_1 + x_1^2$ in any "cylinder in rings", since the two united maps are homomorphisms. But the concept is in no way limited to polynomial functions. There are many categories with products containing a ring object R such that the UI

$$(R^2, R) \substack{\longleftarrow \\ \longleftarrow \\ \longrightarrow} (R, R)$$

satisfies the two axioms: the identity map x is variable and all maps $R \xrightarrow{y} R$ are in $A(x)$. This is a reasonable basis for teaching, since it codifies the basic idea: the suitable categories are those in which the secant-slope function always exists and specializes unambiguously to a tangent-slope function.

III

Clearly, the above discussion does not require that the bigger ring consist of functions defined on the full cartesian square of the spectrum of the smaller ring; a suitable neighbourhood of the diagonal would do. Indeed we could even restrict to an infinitesimal neighbourhood of the diagonal by reducing the larger ring modulo the square of the kernel of the original UI. But there are other examples of a rather different, more particular, character.

When particular laws of "becoming" are taken to be described by vector fields, the zero vector field plays the role of a reference "non-becoming" relative to which "change" (Δ) can be defined. This remark is the basis of a class of examples of the calculus described in the previous section, but descriptive of particular dynamical systems. If a vector field on a manifold is regarded as a derivation ξ of the ring of smooth functions on the manifold, then by considering as unifier the ring of dual numbers $f+g\cdot\varepsilon$, $\varepsilon^2 = 0$ over that ring, we obtain a unity and identity of "opposites" in rings with $\varepsilon \mapsto 0$, $f_0 =: f+0\cdot\varepsilon$, $f_1 =: f+\xi(f)\cdot\varepsilon$. Thus $\Delta f = \xi(f)\cdot\varepsilon$ so a function x is variable in our sense relative to ξ iff $\xi(x)$ is a non-zero divisor. If for simplicity we assume $\xi(x)$ is invertible, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\xi(y)}{\xi(x)}$$

is the concept of "derivative along ξ with respect to x" which results from our general definition. Note that in this example $q\Delta x = 0$ for many $q = f + g\cdot\varepsilon$, but all satisfy $q \to 0$ as required if x is variable.

IV

Walter Noll recently proposed a generalization of the usual concept of continuous body, designed to describe motions during which the body can undergo cracking and self-contact. The usual concept is that of a single topological space which at each time is embedded in geometrical space; however, if during motion a body undergoes cracking, neither the point set nor the open-set frame of the body remains the same. Noll's proposal is that all these various topological spaces for various times nonetheless have something in common as a result of being the "same" body, namely the Boolean algebra $\mathcal{B}$ of "parts" of the body. At any given time t the body is placed in ordinary space by a locale map from $\mathcal{B}$ whose spatial image has a locale $\mathcal{U}_t$ of open sets; the requirement is that the algebra of regular open sets within $\mathcal{U}_t$ is isomorphic to $\mathcal{B}$ for all t. The condition that the body be "continous" is independent of the placement and even of the momentary locale $\mathcal{U}_t$, since it just means that $\mathcal{B}$ (as a locale in itself) has no points. Since metric spaces are regular, the locale $\mathcal{U}_t$ of opens at time t is a sublocale of $\mathbf{2}^{\mathcal{B}^{\mathrm{op}}}$, suggesting that the "states" of the body involve variable Grothendieck topologies on the lex poset $\mathcal{B}$. In any case, this new connection between lattice theory and engineering mechanics deserves to be further investigated. Here I will describe a connection with UIAO's.

The dialectical principle which says that each part is itself "and" its relationships, may be realized through the canonical isomorphism between the regular open sets and the regular closed sets (of the space occupied by the body at a given time). Indeed, the typical relationships of continuum mechanics are the contact forces which live on the intersections of *boundaries* of parts, whereas the mass of the body is carried on the opens. This double role of the parts can be made explicit by defining (at each time t)

$$\mathcal{S}_t = \{S \mid \operatorname{clo}(S) = \operatorname{clo}\operatorname{int}(S) \quad \text{and} \quad \operatorname{int}(S) = \operatorname{int}\operatorname{clo}(S)\}$$

where int and clo denote interior and closure, with respect to the topology $\mathcal{U}_t$, of the subsets S of the points

$$X_t = \text{Frame hom}(\mathcal{U}_t, \mathbf{2}) \subseteq \text{Lex}(\mathcal{B}, \mathbf{2}).$$

Then each S in $\mathcal{S}_t$ is associated with a unique subbody $B \varepsilon \mathcal{B}$; this map $\mathcal{S}_t \to \mathcal{B}$ is a UIAO with left adjoint as the regular open realization and with right adjoint as the regular closed realization. In more detail, if we start with the usual idea that there are two Boolean subposets $\mathcal{B}_O$, $\mathcal{B}_C$ of $\mathcal{S}_t$, then "closure of the interior of" is a retraction onto $\mathcal{B}_C$ which induces a canonical isomorphism from $\mathcal{B}_O$ to $\mathcal{B}_C$; thus, we should mathematically consider that there is but one $\mathcal{B}$ in itself, but of course two distinct inclusions into $\mathcal{S}_t$. As some other UIAO's deserve the traditional names "being vs. non-being", "becoming vs. non-becoming", etc., this example might deserve to be called "interacting vs. non-interacting"; in case the body is momentarily connected, the only B which are "both" (in the specific sense that the two inclusion maps actually give equal results) are $B = 0, 1$.

Now, with any UIAO we can associate, to each object B of the smaller category, the interval category of all objects S in the bigger category for which $S \mapsto B$; in our example, this is isomorphic to the Boolean algebra of all subsets of the momentary boundary of B; the topology of this boundary, i.e. which of these subsets are closed, is determined by the whole UIAO since the regular closed subsets of the boundary of B correspond to such S which are moreover of the form int $(B) \cup \text{clo}(C)$ for some C in $\mathcal{B}$. It may be of interest to investigate how a UIAO $\mathcal{S}$ over $\mathcal{B}$ induces a Grothendieck topology on $\mathcal{B}$ in such a way as to yield, as sheaves in our example, the regular locale $\mathcal{U}$.

References

1. Grassmann, H.: *Ausdehnungslehre von 1844*, Chelsea, 1969.
2. Grothendieck, A.: *Pursuing Stacks*, 400 p. Manuscript, 1983.
3. Kelly, G. M. and Lawvere F. W.: On the complete lattice of essential localizations, *Bull. Soc. Math. de Belgique* (*Serie A*) **41** (1989), 289–319.
4. Lawvere, F. W.: *Introduction to Categories in Continuum Physics*, Springer LNM 1174, 1986, 1–16.
5. Lawvere, F. W.: Intrinsic co-heyting boundaries and the Leibniz rule in certain toposes, *Category Theory Proceedings Como 1990*, Springer LNM 1488, 1991, 279–281.
6. Lawvere, F. W.: Categories of space and of quantity, in Echeverria *et al.* (eds), *The Space of Mathematics*, de Gruyter, 1992, pp. 14–30.
7. Lawvere, F. W.: Cohesive toposes and Cantor's "lauter Einsen", *Philosophia Mathematica* (*Series III*) **2** (1994), 5–15.
8. Marx, K.: *The Mathematical Manuscripts*, London New Park Publications Ltd., 1983.
9. Noll, Walter: The geometry of contact, separation, and reformation of continuous bodies, *Arch. Rational Mech. Anal.* **122** (1993), 197–212.
10. Quillen, D. G.: *Homotopical Algebra*, Springer LNM 43, 1967.