From Dirac

to

Stokes

Dirac Coordinates

Light Cone Coordinates

Light Front Coordinates

God used beautiful mathematics in creating the world.

This result is too beautiful to be false; it is more important to have beauty in one's equations than to have them fit experiment.

This balancing on the dizzying path between genius and madness is awful. --Einstein

It is fun and productive to study Dirac's papers. They are like poems. --Y. S. Kim

Dirac on poetry

Oppenheimer was working at Göttingen and the great mathematical physicist, Dirac, came to him one day and said: "Oppenheimer, they tell me you are writing poetry. I do not see how a man can work on the frontiers of physics and write a poetry at the same time. They are in opposition. In science you want to say something that nobody knew before, in words which everyone can understand. In poetry you are bound to say...something that everybody knows already in words that nobody can understand.

For each value of ρ, $Z^{\rho\sigma\tau}$ is of the form of a multiple of the metric tensor plus a self-dual tensor, that is of the form

$$Wg^{\sigma\tau}+W^{\sigma\tau}.$$

However, if such a tensor is multiplied by

$$Wg_{\tau\lambda}-W_{\tau\lambda},$$

and summed on τ, we obtain

$$(W^2+\tfrac{1}{4}W^{\tau\mu}W_{\tau\mu})\delta_\lambda{}^\sigma.$$

This means that we may solve Eqs. (5.4) for $T_{\kappa,\lambda}$ as functions of T_ν and $T_{\rho,\lambda}$ $(\rho\neq\kappa)$. The discussion of the existence of solutions of the resulting system of equations reduces to a discussion of the integrability conditions.

To solve Eqs. (5.4) for $T_{\kappa,\lambda}$ for one value of κ, we must multiply Eqs. (5.4) by

$$W^\epsilon(g_{\tau\lambda}\delta_\epsilon{}^\kappa-g_{\alpha\tau}g_{\beta\lambda}\eta_{\epsilon\delta}{}^{\alpha\beta}g^{\delta\kappa}),$$

and sum on τ. We obtain for the right hand side,

$$\begin{aligned}W^\epsilon(g_{\tau\lambda}\delta_\epsilon{}^\kappa-g_{\alpha\tau}g_{\beta\lambda}\eta_{\epsilon\delta}{}^{\alpha\beta}g^{\delta\kappa})\bar{F}_\nu{}^\tau T^\nu\\ =H_{\kappa\lambda}=W^\kappa\bar{F}_{\lambda\nu}T^\nu+T^\kappa\bar{F}_{\lambda\nu}W^\nu+W^\nu T_\nu\bar{F}_{\kappa\lambda}\\ +W_\lambda\bar{F}_\nu{}^\kappa T^\nu-T_\lambda\bar{F}_\nu{}^\kappa W^\nu-\delta_\lambda{}^\kappa\bar{F}_{\sigma\nu}W^\sigma T^\nu.\end{aligned}\quad(5.6)$$

The left-hand side involves

$$\begin{aligned}W^\mu(g^{\sigma\tau}\delta_\mu{}^\rho+\eta_{\mu\nu}{}^{\sigma\tau}g^{\rho\nu})(g_{\tau\lambda}\delta_\epsilon{}^\kappa-g_{\alpha\tau}g_{\beta\lambda}\eta_{\epsilon\delta}{}^{\alpha\beta}g^{\delta\kappa})W^\epsilon\\ =a^2\delta_\lambda{}^\sigma g^{\kappa\rho}+\delta_\lambda{}^\rho(2W^\sigma W^\kappa-g^{\sigma\kappa}a^2)-\delta_\lambda{}^\kappa(2W^\sigma W^\rho-g^{\sigma\rho}a^2)\\ +2(g^{\sigma\kappa}W^\rho W_\lambda-g^{\sigma\rho}W^\kappa W_\lambda).\end{aligned}\quad(5.7)$$

The differential Eq. (5.4) may then be written as

$$\begin{aligned}T_{\rho,\sigma}[a^2\delta_\lambda{}^\sigma g^{\kappa\rho}+\delta_\lambda{}^\rho(2W^\sigma W^\kappa-g^{\sigma\kappa}a^2)-\delta_\lambda{}^\kappa(2W^\sigma W^\rho\\ -g^{\sigma\rho}a^2)+2(g^{\sigma\kappa}W^\rho W_\lambda-g^{\sigma\rho}W^\kappa W_\lambda)]=H_{\kappa\lambda}.\end{aligned}\quad(5.8)$$

Multiplying this by $g^{\kappa\lambda}$ and summing, we obtain

$$T_{\rho,\sigma}(g^{\rho\sigma}a^2-2W^\rho W^\sigma)=-2\bar{F}_{\rho\sigma}W^\rho T^\sigma.\quad(5.9)$$

Thus (5.8) becomes

$$\begin{aligned}a^2(T_{\kappa,\lambda}-T_{\lambda,\kappa})+2(T_{\lambda,\sigma}W^\sigma W_\kappa+T_{\sigma,\kappa}W^\sigma W_\lambda\\ -T_{,\rho}{}^\rho W_\kappa W_\lambda)=G_{\kappa\lambda},\end{aligned}\quad(5.10)$$

where

$$\begin{aligned}G_{\kappa\lambda}=W_\kappa\bar{F}_{\lambda\nu}T^\nu+T_\kappa\bar{F}_{\lambda\nu}W^\nu+W^\nu T_\nu\bar{F}_{\kappa\lambda}+W_\lambda\bar{F}_{\kappa\nu}T_\nu\\ -T_\lambda\bar{F}_{\kappa\nu}W^\nu+g_{\kappa\lambda}\bar{F}_{\sigma\nu}W^\sigma T^\nu.\end{aligned}\quad(5.11)$$

Setting $\kappa=4$ in Eqs. (5.10), and assuming that the coordinate system is a galilean one in which $V^\sigma=\delta_4{}^\sigma$, we obtain the four equations

$$\left.\begin{aligned}T_{4,i}&=-T_{i,4}-2(T_{i,j}W^j-T_{,j}{}^jW_i)+\frac{G_{4i}}{W_4},\\ T_{4,4}&=2\frac{T_{i,j}W^iW^j}{W_4}-2(1-W_4)T_{,j}{}^j+\frac{G_{44}-2G_{4i}W^i}{2(W_4)^2},\end{aligned}\right\}\quad(5.12)$$

where G_{4i} and G_{44} are obtained from (5.11). The existence of solutions of these equations depends on the nature of the functions χ_σ and their derivatives.

REVIEWS OF MODERN PHYSICS VOLUME 21, NUMBER 3 JULY, 1949

Forms of Relativistic Dynamics

P. A. M. Dirac
St. John's College, Cambridge, England

For the purposes of atomic theory it is necessary to combine the restricted principle of relativity with the Hamiltonian formulation of dynamics. This combination leads to the appearance of ten fundamental quantities for each dynamical system, namely the total energy, the total momentum and the 6-vector which has three components equal to the total angular momentum. The usual form of dynamics expresses everything in terms of dynamical variables at one instant of time, which results in specially simple expressions for six or these ten, namely the components of momentum and of angular momentum. There are other forms for relativistic dynamics in which others of the ten are specially simple, corresponding to various sub-groups of the inhomogeneous Lorentz group. These forms are investigated and applied to a system of particles in interaction and to the electromagnetic field.

1. INTRODUCTION

EINSTEIN'S great achievement, the principle of relativity, imposes conditions which all physical laws have to satisfy. It profoundly influences the whole of physical science, from cosmology, which deals with the very large, to the study of the atom, which deals with the very small. General relativity requires that physical laws, expressed in terms of a system of curvilinear coordinates in space-time, shall be invariant under all transformations of the coordinates. It brings gravitational fields automatically into physical theory and describes correctly the influence of these fields on physical phenomena.

Gravitational fields are specially important when one is dealing with large-scale phenomena, as in cosmology, but are quite negligible at the other extreme, the study of the atom. In the atomic world the departure of space-time from flatness is so excessively small that there would be no point in taking it into account at the present time, when many large effects are still unexplained. Thus one naturally works with the simplest kind of coordinate system, for which the tensor $g^{\mu\nu}$ that defines the metric has the components

$$g^{00}=-g^{11}=-g^{22}=-g^{33}=1$$
$$g^{\mu\nu}=0 \quad \text{for} \quad \mu\neq\nu.$$

Einstein's restricted principle of relativity is now of paramount importance, requiring that physical laws shall be invariant under transformations from one such coordinate system to another. A transformation of this kind is called an inhomogeneous Lorentz transformation. The coordinates u_μ transform linearly according to the equations

$$\left.\begin{aligned} &u_\mu{}^*=\alpha_\mu+\beta_\mu{}^\nu u_\nu \\ \text{with} \quad &\beta_\mu{}^\nu\beta^{\mu\rho}=g^{\nu\rho}, \end{aligned}\right\}(1)$$

the α's and β's being constants.

A transformation of the type (1) may involve a reflection of the coordinate system in the three spacial dimensions and it may involve a time reflection, the direction du_0 in space-time changing from the future to the past. I do not believe there is any need for physical laws to be invariant under these reflections, although all the exact laws of nature so far known do have this invariance. The restricted principle of relativity arose from the requirement that the laws of nature should be independent of the position and velocity of the observer, and any change the observer may make in his position and velocity, taking his coordinate system with him, will lead to a transformation (1) of a kind that can be built up from infinitesimal transformations and cannot involve a reflection. Thus it appears that restricted relativity will be satisfied by the requirement that physical laws shall be invariant under infinitesimal transformations of the coordinate system of the type (1). Such an infinitesimal transformation is given by

$$\left.\begin{aligned} &u_\mu{}^*=u_\mu+a_\mu+b_\mu{}^\nu u_\nu \\ \text{with} \quad &b_{\mu\nu}=-b_{\nu\mu}, \end{aligned}\right\}(2)$$

the a's and b's being infinitesimal constants.

A second general requirement for dynamical theory has been brought to light through the discovery of quantum mechanics by Heisenberg and Schrödinger, namely the requirement that the equations of motion shall be expressible in the Hamiltonian form. This is necessary for a transition to the quantum theory to be possible. In atomic theory one thus has two over-riding requirements. The problem of fitting them together forms the subject of the present paper.

The existing theories of the interaction of elementary particles and fields are all unsatisfactory in one way or another. The imperfections may well arise from the use of wrong dynamical systems to represent atomic phenomena, i.e., wrong Hamiltonians and wrong interaction energies. *It thus becomes a matter of great importance to set up new dynamical systems and see if they will better describe the atomic world.* In setting up such a new dynamical system one is faced at the outset by the two requirements of special relativity and of Hamiltonian equations of motion. The present paper is intended to make a beginning on this work by providing the simplest methods for satisfying the two requirements simultaneously.

2. THE TEN FUNDAMENTAL QUANTITIES

The theory of a dynamical system is built up in terms of a number of algebraic quantities, called dynamical variables, each of which is defined with respect to a system of coordinates in space-time. The usual dynamical variables are the coordinates and momenta of particles at particular times and field quantities at particular points in space-time, but other kinds of quantities are permissible, as will appear later.

In order that the dynamical theory may be expressible in the Hamiltonian form, it is necessary that any two dynamical variables, ξ and η, shall have a P.b. (Poisson bracket) $[\xi, \eta]$, subject to the following laws,

$$\left.\begin{aligned} [\xi,\eta]&=-[\eta,\xi] \\ [\xi,\eta+\zeta]&=[\xi,\eta]+[\xi,\zeta] \\ [\xi,\eta\zeta]&=[\xi,\eta]\zeta+\eta[\xi,\zeta] \\ [[\xi,\eta],\zeta]+[[\eta,\zeta],\xi]&+[[\zeta,\xi],\eta]=0. \end{aligned}\right\}(3)$$

A number or physical constant may be counted as a special case of a dynamical variable, and has the property that its P.b. with anything vanishes.

Dynamical variables change when the system of coordinates with respect to which they are defined changes, and must do so in such a way that P.b. relations between them remain invariant. This requires that with an infinitesimal change in the coordinate system (2) each dynamical variable ξ shall change according to the law

$$\xi^*=\xi+[\xi, F], \qquad (4)$$

where F is some infinitesimal dynamical variable independent of ξ, depending only on the dynamical system involved and the change in the coordinate system. We are thus led to associate one F with each infinitesimal transformation of coordinates.

Let us apply two infinitesimal transformations of coordinates in succession. Suppose the first one changes the dynamical variable ξ to ξ^* according to

$$\xi^*=\xi+[\xi, F_1],$$

and the second one changes ξ^* to $\xi^\dagger$ according to

$$\xi^\dagger=\xi^*+[\xi^*, F_2{}^*]=\xi^*+[\xi, F_2]^*.$$

The two transformations together change ξ to $\xi^\dagger$

according to

$$\xi^{\dagger}=\xi+[\xi, F_1]+[\xi, F_2]+[[\xi, F_2], F_1],$$

to the accuracy of the order F_1F_2 (with neglect of terms of order F_1^2 or F_2^2). If these two transformations are applied in the reverse order, they change ξ to $\xi^{\dagger\dagger}$ according to

$$\xi^{\dagger\dagger}=\xi+[\xi, F_2]+[\xi, F_1]+[[\xi, F_1], F_2].$$

Thus

$$\begin{aligned}\xi^{\dagger\dagger}&=\xi^{\dagger}+[[\xi, F_1], F_2]-[[\xi, F_2], F_1]\\ &=\xi^{\dagger}+[\xi, [F_1, F_2]],\end{aligned}$$

with the help of the first and last of Eqs. (3). This gives the change in a dynamical variable associated with that change of the coordinate system which is the commutator of the two previous changes. It is of the standard form

$$\xi^{\dagger\dagger}=\xi^{\dagger}+[\xi^{\dagger}, F],$$

with an F that is the P. b. of the F's associated with the two previous changes of coordinates. Thus the commutation relations between the various infinitesimal changes of coordinates correspond to the P.b. relations between the associated F's.

The F associated with the transformation (2) must depend linearly on the infinitesimal numbers a_μ, $b_{\mu\nu}$ that fix this transformation. Thus we can put

$$\left.\begin{aligned}F&=-P^{\mu}a_{\mu}+\tfrac{1}{2}M^{\mu\nu}b_{\mu\nu}\\ M^{\mu\nu}&=-M^{\nu\mu},\end{aligned}\right\}\quad(5)$$

where P^{μ}, $M^{\mu\nu}$ are finite dynamical variables, independent of the transformation of coordinates.

The ten quantities P_μ, $M_{\mu\nu}$ are characteristic for the dynamical system. They will be called the ten *fundamental quantities*. They determine how all dynamical variables are affected by a change in the coordinate system of the kind that occurs in special relativity. Each of them is associated with a type of infinitesimal transformation of the inhomogeneous Lorentz group. Seven of them have simple physical interpretations, namely, P_0 is the total energy of the system, P_r ($r=1$, 2, 3) is the total momentum, and M_{rs} is the total angular momentum about the origin. The remaining three M_{r0} do not correspond to any such well-known physical quantities, but are equally important in the general dynamical scheme.

From the commutation relations between particular infinitesimal transformations of the coordinate system we get at once the P.b. relations between the ten fundamental quantities,

$$\left.\begin{aligned}&[P_\mu, P_\nu]=0\\ &[M_{\mu\nu}, P_\rho]=-g_{\mu\rho}P_\nu+g_{\nu\rho}P_\mu\\ &[M_{\mu\nu}, M_{\rho\sigma}]=-g_{\mu\rho}M_{\nu\sigma}+g_{\nu\rho}M_{\mu\sigma}-g_{\mu\sigma}M_{\rho\nu}+g_{\nu\sigma}M_{\rho\mu}.\end{aligned}\right\}\quad(6)$$

To construct a theory of a dynamical system one must obtain expressions for the ten fundamental quantities that satisfy these P.b. relations. *The problem of finding a new dynamical system reduces to the problem of finding a new solution of these equations.*

An elementary solution is provided by the following scheme. Take the four coordinates q_μ of a point in space-time as dynamical coordinates and let their conjugate momenta be p_μ, so that

$$[q_\mu, q_\nu]=0,\quad [p_\mu, p_\nu]=0$$
$$[p_\mu, q_\nu]=g_{\mu\nu}.$$

The q's will transform under an infinitesimal transformation of the coordinate system in the same way as the u's in (2). This leads to

$$P_\mu=p_\mu,\quad M_{\mu\nu}=q_\mu p_\nu-q_\nu p_\mu, \quad (7)$$

and provides a solution of the P.b. relations (6). The solution (7) does not seem to be of any practical importance, but it may be used as a basis for obtaining other solutions that are of practical importance, as the next three sections will show.

The foregoing discussion of the requirements for a relativistic dynamical theory may be generalized somewhat. We may work with dynamical variables that are connected by one or more relations for all states of motion that occur physically. Such relations are called *subsidiary equations*. They will be written

$$A\approx 0 \quad (8)$$

to distinguish them from dynamical equations. They are less strong than dynamical equations, because with a dynamical equation one can take the P.b. of both sides with any dynamical variable and get another equation, while with a subsidiary equation one cannot do this in general. The lesser assumption is made, however, that from two subsidiary equations $A\approx 0, B\approx 0$ one can infer a third

$$[A, B]\approx 0. \quad (9)$$

A subsidiary equation must remain a subsidiary equation under any change of coordinate system. This enables one to infer from (8)

$$[P_\mu, A]\approx 0,\quad [M_{\mu\nu}, A]\approx 0. \quad (10)$$

A dynamical variable is of physical importance only if its P.b. with any subsidiary equation gives another subsidiary equation, i.e., its P.b. with A in (8) must vanish in the subsidiary sense. Such a dynamical variable will be called a *physical variable*. The P.b. of two physical variables is a physical variable. Equations (10) show that the ten fundamental quantities are physical variables.

The physical variables are the only ones that are really important. One could eliminate the non-physical variables from the theory altogether and one could then make the subsidiary equations into dynamical equations. However, the elimination may be awkward and may spoil some symmetry feature in the scheme of equations, so it is desirable to retain the possibility of subsidiary equations in the general theory.

3. THE INSTANT FORM

The ten fundamental quantities for dynamical systems that occur in practice are usually such that some of them are specially simple and the others are complicated. The complicated ones will be called the Hamiltonians. They play jointly the rôle of the single Hamiltonian in non-relativistic dynamics. Since the P.b. of two simple quantities is a simple quantity, the simple ones of the ten fundamental quantities must be those associated with some sub-group of the inhomogeneous Lorentz group.

In the usual form of dynamics one works with dynamical variables referring to physical conditions at some instant of time, e.g., the coordinates and momenta of particles at that instant. An instant in the four-dimensional relativistic picture is a flat three-dimensional surface containing only directions which lie outside the light-cone. The simplest instant referred to the u coordinate system is given by the equation

$$u_0=0. \tag{11}$$

The effect of working with dynamical variables referring to physical conditions at this instant will be to make specially simple those of the fundamental quantities associated with transformations of coordinates that leave the instant invariant, namely P_1, P_2, P_3, M_{23}, M_{31}, M_{12}. The remaining ones, P_0, M_{10}, M_{20}, M_{30}, will be complicated in general and will be the Hamiltonians. We get in this way a form of dynamics which is associated with the sub-group of the inhomogeneous Lorentz group that leaves the instant invariant, and which may appropriately be called the *instant form*.

Let us take as an example a single particle by itself. The ten fundamental quantities in this case are well known, but they will be worked out again here to illustrate a method that can be used also with the other forms of dynamics.

We take as dynamical coordinates the three coordinates of the particle at the instant (11). Calling these coordinates q_r, we can base our work on the scheme (7), with the additional equation

$$q_0=0. \tag{12}$$

With this equation p_0 no longer has a meaning. We must therefore modify the expressions for the ten fundamental quantities given by (7) so as to eliminate p_0 from them, without invalidating the P.b. relations (6).

Let us change the expressions for the ten fundamental quantities by multiples of $p^\sigma p_\sigma - m^2$, where m is a constant, i.e., let us put

$$\left.\begin{aligned} P_\mu &= p_\mu + \lambda_\mu(p^\sigma p_\sigma - m^2) \\ M_{\mu\nu} &= q_\mu p_\nu - q_\nu p_\mu + \lambda_{\mu\nu}(p^\sigma p_\sigma - m^2), \end{aligned}\right\} \tag{13}$$

where

$$\lambda_{\mu\nu} = -\lambda_{\nu\mu}$$

and the coefficients λ are functions of the q's and p's that do not become infinitely great when one puts $p^\sigma p_\sigma = m^2$ with $p_0>0$. Since $p^\sigma p_\sigma - m^2$ has zero P.b. with all the expressions (7), the modified expressions (13) must still satisfy the P.b. relations (6), apart from multiples of $p^\sigma p_\sigma - m^2$, with any choice of the λ's. If we now choose the λ's so as to make the P_μ, $M_{\mu\nu}$ given by (13) independent of p_0, the P.b. relations (6) must be satisfied apart from terms that are independent of p_0 as well as being multiples of $p^\sigma p_\sigma - m^2$. Such terms must vanish, so we get in this way a solution of our problem.

The λ's have the values

$$\left.\begin{aligned} &\lambda_r = 0, \quad \lambda_0 = -\{p_0 + (p_s p_s + m^2)^{\frac{1}{2}}\}^{-1}, \\ &\lambda_{sr} = 0, \quad \lambda_{r0} = -q_r\{p_0 + (p_s p_s + m^2)^{\frac{1}{2}}\}^{-1}, \end{aligned}\right\} \tag{14}$$

and Eqs. (13) become

$$P_r = p_r, \qquad M_{rs} = q_r p_s - q_s p_r, \tag{15}$$

$$P_0 = (p_s p_s + m^2)^{\frac{1}{2}}, \quad M_{r0} = q_r(p_s p_s + m^2)^{\frac{1}{2}}, \tag{16}$$

with the help of (12). Equations (15) and (16) give all the ten fundamental quantities for a particle with rest-mass m. Those given by (15) are the simple ones: those given by (16) are the Hamiltonians.

For a dynamical system composed of several particles, P_r and M_{rs} will be just the sum of their values for the particles separately,

$$P_r = \sum p_r, \quad M_{rs} = \sum(q_r p_s - q_s p_r). \tag{17}$$

The Hamiltonians P_r, M_{r0} will be the sum of their values for the particles separately plus interaction terms,

$$\left.\begin{aligned} P_0 &= \sum(p_s p_s + m^2)^{\frac{1}{2}} + V, \\ M_{r0} &= \sum q_r(p_s p_s + m^2)^{\frac{1}{2}} + V_r. \end{aligned}\right\} \tag{18}$$

The V's here must be chosen to make P_0, M_{r0} satisfy all the P.b. relations (6) in which they appear.

Some of these relations are linear in the V's and are easily fulfilled. The P.b. relations for $[M_{rs}, P_0]$ and $[M_{rs}, M_{t0}]$ are fulfilled provided V is a three-dimensional scalar (in the space u_1, u_2, u_3) and V_r a three-dimensional vector. The P.b. relation for $[P_r, P_0]$ will be fulfilled provided V is independent of the position of the origin in the three-dimensional space u_1, u_2, u_3. The P.b. relation for $[M_{r0}, P_s]$ will be fulfilled provided

$$V_r = q_r V + V_r', \tag{19}$$

where the q_r are the coordinates of any one of the particles and the V_r' are independent of the position of the origin in three-dimensional space.

The remaining conditions for the V's are quadratic, involving $[V, V_r]$ or $[V_r, V_s]$. These conditions are not easily fulfilled and provide the real difficulty in the problem of constructing a theory of a relativistic dynamical system in the instant form.

4. THE POINT FORM

One can build up a dynamical theory in terms of dynamical variables that refer to physical conditions on some three-dimensional surface other than an instant. The surface must satisfy the condition that the world-line of every particle must meet it, otherwise the particle could not be described by variables on the surface, and preferably the world-line should meet it only once, for the sake of uniqueness.

To get a simple form of theory one should take the surface to be such that it is left invariant by some sub-group of the group of inhomogeneous Lorentz transformations. A possible sub-group is the group of rotations about some point, say the origin $u_\mu=0$. The surface may then be taken to be a branch of a hyperboloid

$$u^\rho u_\rho=\kappa^2, \quad u_0>0, \tag{20}$$

with κ a constant. The fundamental quantities associated with the infinitesimal transformations of the sub-group, namely the $M_{\mu\nu}$, will then be specially simple, while the others, namely the P_μ, will be complicated in general and will be the Hamiltonians. A new form of dynamics is thus obtained, which may be called the *point form*, as it is characterized by being associated with the sub-group that leaves a point invariant.

To illustrate the new form, let us take again the example of a single particle. The dynamical coordinates must determine the point where the world-line of the particle meets the hyperboloid (20). Let the four coordinates of this point in the u system of coordinates be q_μ. Only three of these are independent, but instead of eliminating one of them, it is more convenient to work with all four and introduce the subsidiary equation

$$q^\rho q_\rho\approx\kappa^2. \tag{21}$$

It is then necessary that the ten fundamental quantities, and indeed all physical variables, shall have zero P.b. with $q^\rho q_\rho$. The condition for this is that they should involve the p's only through the combinations $q_\mu p_\nu - q_\nu p_\mu$.

The ten fundamental quantities may be obtained by a method parallel to that of the preceding section, with the subsidiary equation (21) taking the place of Eq. (12). We again assume Eqs. (13), and now choose the λ's so as to make their right-hand sides have zero P.b. with $q^\rho q_\rho$. The resulting expressions for the ten fundamental quantities will again satisfy the P.b. relations (6), as may be inferred by a similar argument to the one given in the preceding section.

We find at once

$$\lambda_{\mu\nu}=0.$$

To obtain λ_μ, instead of arranging directly for the P_μ to have zero P.b. with $q^\rho q_\rho$, it is easier to make $q_\mu P_\nu - q_\nu P_\mu$ and $P^\mu P_\mu$ have zero P.b. with $q^\rho q_\rho$. Now

$$q_\mu P_\nu - q_\nu P_\mu = q_\mu p_\nu - q_\nu p_\mu + (q_\mu\lambda_\nu - q_\nu\lambda_\mu)(p^\sigma p_\sigma - m^2),$$

so we must have

$$q_\mu\lambda_\nu - q_\nu\lambda_\mu = 0,$$

and hence

$$\lambda_\mu = q_\mu B,$$

where B is some dynamical variable independent of μ. Further

$$\begin{aligned} P^\mu P_\mu &= \{p^\mu + q^\mu B(p^\sigma p_\sigma - m^2)\}\{p_\mu + q_\mu B(p^\rho p_\rho - m^2)\} \\ &= m^2 + \{1 + 2p^\mu q_\mu B + q^\mu q_\mu B^2(p^\sigma p_\sigma - m^2)\}(p^\rho p_\rho - m^2). \end{aligned}$$

In order that $P^\mu P_\mu$ shall have zero P.b. with $q^\rho q_\rho$, we must take

$$1 + 2p^\mu q_\mu B + q^\mu q_\mu B^2(p^\sigma p_\sigma - m^2) = 0,$$

so that

$$B(p^\sigma p_\sigma - m^2) = (q^\rho q_\rho)^{-1}\{[(p^\nu q_\nu)^2 - q^\lambda q_\lambda(p^\sigma p_\sigma - m^2)]^{\frac{1}{2}} - p^\nu q_\nu\}$$

The right-hand side here tends to zero as $p^\sigma p_\sigma - m^2 \to 0$, so it is a multiple of $p^\sigma p_\sigma - m^2$, as it ought to be. We now get finally

$$\left.\begin{aligned} P_\mu &= p_\mu + q_\mu\kappa^{-2}\{[(p^\nu q_\nu)^2 - \kappa^2(p^\sigma p_\sigma - m^2)]^{\frac{1}{2}} - p^\nu q_\nu\} \\ M_{\mu\nu} &= q_\mu p_\nu - q_\nu p_\mu, \end{aligned}\right\} \tag{22}$$

in which the expression for P_μ has been simplified with the help of (21).

It is permissible to take $\kappa=0$ and so to have a light-cone instead of a hyperboloid. The expression for B then becomes much simpler and gives

$$P_\mu = p_\mu - \tfrac{1}{2}q_\mu(p^\nu q_\nu)^{-1}(p^\sigma p_\sigma - m^2). \tag{23}$$

instead of the first of Eqs. (22).

For a dynamical system composed of several particles, the $M_{\mu\nu}$ will be just the sum of their values for the particles separately,

$$M_{\mu\nu} = \sum (q_\mu p_\nu - q_\nu p_\mu). \tag{24}$$

The Hamiltonians P_μ will be the sum of their values for the particles separately plus interaction terms,

$$P_\mu = \sum \{p_\mu + q_\mu B(p^\sigma p_\sigma - m^2)\} + V_\mu. \tag{25}$$

The V_μ must be chosen so as to make the P_μ satisfy the correct P.b. relations. The relations for $[M_{\mu\nu}, P_\rho]$ are satisfied provided the V_μ are the components of a 4-vector. The remaining relations, which require the P_μ to have zero P.b. with one another, lead to quadratic conditions for the V_μ. These cause the real difficulty in the problem of constructing a theory of a relativistic dynamical system in the point form.

5. THE FRONT FORM

Consider the three-dimensional surface in space-time formed by a plane wave front advancing with the velocity of light. Such a surface will be called a *front* for brevity. An example of a front is given by the equation

$$u_0 - u_3 = 0. \tag{26}$$

We may set up a dynamical theory in which the dynamical variables refer to physical conditions on a front. This will make specially simple those of the fundamental quantities associated with infinitesimal transformations of coordinates that leave the front invariant, and will give a third form of dynamics, which may be called the *front form*.

If A_μ is any 4-vector, put

$$A_0+A_3=A_+,\quad A_0-A_3=A_-.$$

We get a convenient notation by using the $+$ and $-$ suffixes freely as tensor suffixes, together with 1 and 2. They may be raised with the help of

$$g^{++}=g^{--}=0,\quad g^{+-}=\tfrac{1}{2},$$
$$g^{i+}=g^{i-}=0,\quad \text{for}\quad i=1,2,$$

as one can verify by noting that these g values lead to the correct value for $g^{\mu\nu}A_\mu A_\nu$ when μ and ν are summed over $1, 2, +, -$.

The equation of the front (26) becomes in this notation

$$u_-=0.$$

The fundamental quantities P_1, P_2, P_-, M_{12}, M_{+-}, M_{1-}, M_{2-} are associated with transformations of coordinates that leave this front invariant and will be specially simple. The remaining ones P_+, M_{1+}, M_{2+} will be complicated in general and will be the Hamiltonians.

Let us again work out the example of a single particle. The dynamical coordinates are now q_1, q_2, q_+. We again assume Eqs. (13), and add to them the further equation $q_-=0$. We must now choose the λ's so as to make the right-hand sides of (13) independent of p_+. The resulting expressions for the ten fundamental quantities will then again satisfy the required P.b. relations.

We find

$$\lambda_+=-1/p_-,\quad \lambda_{i+}=-q_i/p_-,$$

the other λ's vanishing. Thus

$$\left.\begin{aligned}&P_i=p_i,\quad P_-=p_-,\\&M_{12}=q_1p_2-q_2p_1,\quad M_{i-}=q_ip_-,\quad M_{+-}=q_+p_-,\end{aligned}\right\}\quad(27)$$

$$\left.\begin{aligned}&P_+=(p_1^2+p_2^2+m^2)/p_-,\\&M_{i+}=q_i(p_1^2+p_2^2+m^2)/p_--q_+p_i.\end{aligned}\right\}\quad(28)$$

Equations (27) give the simple fundamental quantities. Equations (28) give the Hamiltonians.

For a dynamical system composed of several particles, P_i, P_-, M_{12}, M_{+-}, M_{i-} will be just the sum of their values for the particles separately. The Hamiltonians P_+, M_{i+} will be the sum of their values for the particles separately plus interaction terms,

$$\left.\begin{aligned}&P_+=\sum(p_1^2+p_2^2+m^2)/p_-+V\\&M_{i+}=\sum\{q_i(p_1^2+p_2^2+m^2)/p_--q_+p_i\}+V_i.\end{aligned}\right\}\quad(29)$$

The V's must satisfy certain conditions to make the Hamiltonians satisfy the correct P.b. relations.

As before, some of these conditions are linear and some are quadratic. The linear conditions for V require that it shall be invariant under all transformations of the three coordinates u_1, u_2, u_+ of the front except those for which du_+ gets multiplied by a factor, and for the latter transformations V must get multiplied by the same factor. The linear conditions for V_i require it to be of the form

$$V_i=q_iV+V_i',\quad(30)$$

where q_i are the coordinates 1, 2 of any one of the particles, and V_i' has the same properties as V with regard to all transformations of the three coordinates of the front except rotations associated with M_{12}, and under these rotations it behaves like a two-dimensional vector. The quadratic conditions for the V's are not easily fulfilled and give rise to the real difficulty in the construction of a theory of a relativistic dynamical system in the front form.

6. THE ELECTROMAGNETIC FIELD

To set up the dynamical theory of fields on the lines discussed in the three preceding sections, one may take as dynamical variables the three-fold infinity of field quantities at all points on the instant, hyperboloid, or front, and use them in place of the discrete set of variables of particle theory. The ten fundamental quantities P_μ, $M_{\mu\nu}$ are to be constructed out of them, satisfying the same P.b. relations as before.

For a field which allows waves moving with the velocity of light, a difficulty arises with the point form of theory, because one may have a wave packet that does not meet the hyperboloid (20) at all. Thus physical conditions on the hyperboloid cannot completely describe the state of the field. One must introduce some extra dynamical variables besides the field quantities on the hyperboloid. A similar difficulty arises, in a less serious way, with the front form of theory. Waves moving with the velocity of light in exactly the direction of the front cannot be described by physical conditions on the front, and some extra variables must be introduced for dealing with them.

An alternative method of setting up the dynamical theory of fields is obtained by working with dynamical variables that describe the Fourier components of the field. This method has various advantages. It disposes of the above difficulty of extra variables, and it usually lends itself more directly to physical interpretation. It leads to expressions for the ten fundamental quantities that can be used with all three forms. For a field by itself, there is then no difference between the three forms. A difference occurs, of course, if the field is in interaction with something. The dynamical variables of the field are then to be understood as the Fourier components that the field would have, if the interaction were suddenly cut off at the instant, hyperboloid or front, after the cutting off.

Let us take as an example the electromagnetic field, first without any interaction. We may work with the four potentials $A_\lambda(u)$ satisfying the subsidiary equation

$$\partial A_\lambda/\partial u_\lambda \approx 0. \tag{31}$$

Their Fourier resolution is

$$A_\lambda(u)=\int \{A_{k\lambda}\exp(ik^\mu u_\mu) + A^\dagger_{k\lambda}\exp(-ik^\mu u_\mu)\}k_0^{-1}d^3k, \tag{32}$$

with

$$k_0=(k_1^2+k_2^2+k_3^2)^{\frac{1}{2}},\quad d^3k=dk_1dk_2dk_3.$$

The factor k_0^{-1} inserted in (32) leads to simpler transformation laws for the Fourier coefficients $A_{k\lambda}$, since the differential element $k_0^{-1}d^3k$ is Lorentz invariant. We now take the $A_{k\lambda}$, $A^\dagger_{k\lambda}$ as dynamical variables.

Under the transformation of coordinates (2) the potential $A_\lambda(u)$ at a particular point u changes to a potential at the point with the same u-values in the new coordinate system, i.e., the point with the coordinates $u_\mu - a_\mu - b_\mu{}^\nu u_\nu$ in the original coordinate system. This causes a change in $A_\lambda(u)$ of amount

$$-(a_\mu+b_\mu{}^\nu u_\nu)\partial A_\lambda/\partial u_\mu.$$

There is a further change, of amount $b_\lambda{}^\nu A_\nu$, owing to the change in the direction of the axes. Thus, from (4) and (5)

$$\begin{aligned}[A_\lambda(u), -P^\mu a_\mu+\tfrac{1}{2}M^{\mu\nu}b_{\mu\nu}] &= A_\lambda(u)^*-A_\lambda(u)\\ &= -(a_\mu+b_\mu{}^\nu u_\nu)\partial A_\lambda/\partial u_\mu + b_\lambda{}^\nu A_\nu,\end{aligned}$$

and hence

$$\left.\begin{aligned}&[A_\lambda(u), P^\mu]=\partial A_\lambda/\partial u_\mu,\\ &[A_\lambda(u), M_{\mu\nu}]=u_\mu\partial A_\lambda/\partial u^\nu - u_\nu\partial A_\lambda/\partial u^\mu + g_{\lambda\mu}A_\nu - g_{\lambda\nu}A_\mu.\end{aligned}\right\}\tag{33}$$

Taking Fourier components according to (32), we get

$$\left.\begin{aligned}&[A_{k\lambda}, P_\mu]=ik_\mu A_{k\lambda},\\ &[A_{k\lambda}, M_{\mu\nu}]=(k_\mu\partial/\partial k^\nu - k_\nu\partial/\partial k^\mu)A_{k\lambda} + g_{\lambda\mu}A_{k\nu} - g_{\lambda\nu}A_{k\mu},\end{aligned}\right\}\tag{34}$$

in which $A_{k\lambda}$ may be considered as a function of four independent k's for the purpose of applying the differential operator $k_\mu\partial/\partial k^\nu - k_\nu\partial/\partial k^\mu$ to it.

The Maxwell theory gives for the energy and momentum of the electromagnetic field

$$P_\mu=-4\pi^2\int k_\mu A_k{}^\lambda A^\dagger_{k\lambda}k_0^{-1}d^3k, \tag{35}$$

the $-$ sign being needed to make the transverse components contribute a positive energy. In order that this may agree with the first of Eqs. (34), we must have the P.b. relations

$$\left.\begin{aligned}&[A_{k\lambda}, A_{k'\mu}]=0\\ &[A_{k\lambda}, A^\dagger_{k'\mu}]=-ig_{\lambda\mu}/4\pi^2\cdot k_0\delta(k_1-k_1')\delta(k_2-k_2')\delta(k_3-k_3').\end{aligned}\right\}\tag{36}$$

The second of Eqs. (34) then leads to

$$M_{\mu\nu}=4\pi^2 i\int\{A^\dagger_{k\lambda}(k_\mu\partial/\partial k^\nu - k_\nu\partial/\partial k^\mu)A_k{}^\lambda + A^\dagger_{k\mu}A_{k\nu} - A^\dagger_{k\nu}A_{k\mu}\}k_0^{-1}d^3k. \tag{37}$$

Equations (35) and (37) give the ten fundamental quantities.

For the electromagnetic field in interaction with charged particles, the ten fundamental quantities will be the sum of their values for the field alone, given by (35) and (37), and their values for the particles, given in one of the three preceding sections, with interaction terms involving the field variables $A_{k\lambda}$, $A^\dagger_{k\lambda}$ as well as the particle variables. One usually assumes that there is no direct interaction between the particles, only interaction between each particle and the field. The ten fundamental quantities then take the form

$$\left.\begin{aligned}P_\mu&=P_\mu{}^F+\textstyle\sum_a P_\mu{}^a\\ M_{\mu\nu}&=M_{\mu\nu}{}^F+\textstyle\sum_a M_{\mu\nu}{}^a,\end{aligned}\right\}\tag{38}$$

where $P_\mu{}^F$, $M_{\mu\nu}{}^F$ are the contributions of the field alone, given by (35) and (37), and $P_\mu{}^a$, $M_{\mu\nu}{}^a$ are the contributions of the a-th particle, consisting of terms for the particle by itself and interaction terms. For point charges, the interaction terms will involve the field variables only through the $A_\lambda(q)$ and their derivatives at the point q where the world-line of the particle meets the instant, hyperboloid or front. The expressions for $P_\mu{}^a$, $M_{\mu\nu}{}^a$ may easily be worked out for this case by a generalization of the method of the three preceding sections, as follows.

Suppose there is only one particle, for simplicity. We must replace Eqs. (13) by

$$\left.\begin{aligned}P_\mu&=P_\mu{}^F+p_\mu+\lambda_\mu(\pi^\sigma\pi_\sigma-m^2)\\ M_{\mu\nu}&=M_{\mu\nu}{}^F+q_\mu p_\nu - q_\nu p_\mu+\lambda_{\mu\nu}(\pi^\sigma\pi_\sigma-m^2),\end{aligned}\right\}\tag{39}$$

where

$$\pi_\sigma=p_\sigma-eA_\sigma(q),$$

and $P_\mu{}^F$, $M_{\mu\nu}{}^F$ are the right-hand sides of (35) and (37). From (33),

$$\begin{aligned}&[A_\lambda(q), P_\mu{}^F+p_\mu]=0\\ &[A_\lambda(q), M_{\mu\nu}{}^F+q_\mu p_\nu - q_\nu p_\mu]=g_{\lambda\mu}A_\nu(q)-g_{\lambda\nu}A_\mu(q),\end{aligned}$$

and hence

$$\begin{aligned}&[\pi_\lambda, P_\mu{}^F+p_\mu]=0\\ &[\pi_\lambda, M_{\mu\nu}{}^F+q_\mu p_\nu-q_\nu p_\mu]=g_{\lambda\mu}\pi_\nu-g_{\lambda\nu}\pi_\mu.\end{aligned}$$

It follows that $\pi^\sigma\pi_\sigma$ has zero P.b. with each of the quantities $P_\mu{}^F+p_\mu$, $M_{\mu\nu}{}^F+q_\mu p_\nu-q_\nu p_\mu$. One can now infer, by the same argument as in the case of no field, that if the λ's in (39) are chosen to make P_μ, $M_{\mu\nu}$ have zero P.b. with q_0, $q^\rho q_\rho$ or q_-, the P.b. relations (6) will all be satisfied. Such a choice of λ's, in conjunction with one of the equations $q_0=0$, $q^\rho q_\rho\approx\kappa^2$, $q_-=0$, will provide the ten fundamental quantities for a charged particle in interaction with the field in the instant, point and front forms, respectively. The subsidiary Eq. (31) must be modified when a charge is present.

The point form will be worked out as an illustration. In this case we have at once $\lambda_{\mu\nu}=0$. We can get λ_μ conveniently by arranging that $q_\mu(P_\nu-P_\nu^F)-q_\nu(P_\mu-P_\mu^F)$ and $\{P^\mu-P^{\mu F}-eA^\mu(q)\}\{P_\mu-P_\mu^F-eA_\mu(q)\}$ shall have zero P.b. with $q^\rho q_\rho$. The first condition gives $\lambda_\mu=q_\mu B$. The second then gives

$$1+2\pi^\mu q_\mu B+q^\mu q_\mu B^2(\pi^\sigma\pi_\sigma-m^2)=0.$$

Thus we get finally

$$\left.\begin{aligned}P_\mu&=P_\mu^F+p_\mu+q_\mu\kappa^{-2}\{[(\pi^\nu q_\nu)^2\\&\qquad-\kappa^2(\pi^\sigma\pi_\sigma-m^2)]^{\frac{1}{2}}-\pi^\nu q_\nu\}\\M_{\mu\nu}&=M_{\mu\nu}^F+q_\mu p_\nu-q_\nu p_\mu.\end{aligned}\right\}\quad(40)$$

The above theory of point charges is subject to the usual difficulty that infinities will arise in the solution of the equations of motion, on account of the infinite electromagnetic energy of a point charge. The present treatment has the advantage over the usual treatment of the electromagnetic equations that it offers simpler opportunities for departure from the point-charge model for elementary particles.

8. DISCUSSION

Three forms have been given in which relativistic dynamical theory may be put. For particles with no interaction, any one of the three is possible. For particles with interaction, it may be that all three are still possible, or it may be that only one is possible, depending on the kind of interaction. If one wants to set up a new kind of interaction between particles in order to improve atomic theory, the way to proceed would be to take one of the three forms and try to find the interaction terms V, or to find directly the Hamiltonians, satisfying the required P.b. relations. The question arises, which is the best form to take for this purpose.

The instant form has the advantage of being the one people are most familiar with, but I do not believe it is intrinsically any better for this reason. The four Hamiltonians P_0, M_{r0} form a rather clumsy combination.

The point form has the advantage that it makes a clean separation between those of the fundamental quantities that are simple and those that are the Hamiltonians. The former are the components of a 6-vector, the latter are the components of a 4-vector. Thus the four Hamiltonians can easily be treated as a single entity. All the equations with this form can be expressed neatly and concisely in four-dimensional tensor notation.

The front form has the advantage that it requires only three Hamiltonians, instead of the four of the other forms. This makes it mathematically the most interesting form, and makes any problem of finding Hamiltonians substantially easier. The front form has the further advantage that there is no square root in the Hamiltonians (28), which means that one can avoid negative energies for particles by suitably choosing the values of the dynamical variables in the front, without having to make a special convention about the sign of a square root. It may then be easier to eliminate negative energies from the quantum theory. This advantage also occurs with the point form with $\kappa=0$, there being no square root in (23).

There is no conclusive argument in favor of one or other of the forms. Even if it could be decided that one of them is the most convenient, this would not necessarily be the one chosen by nature, in the event that only one of them is possible for atomic systems. Thus all three forms should be studied further.

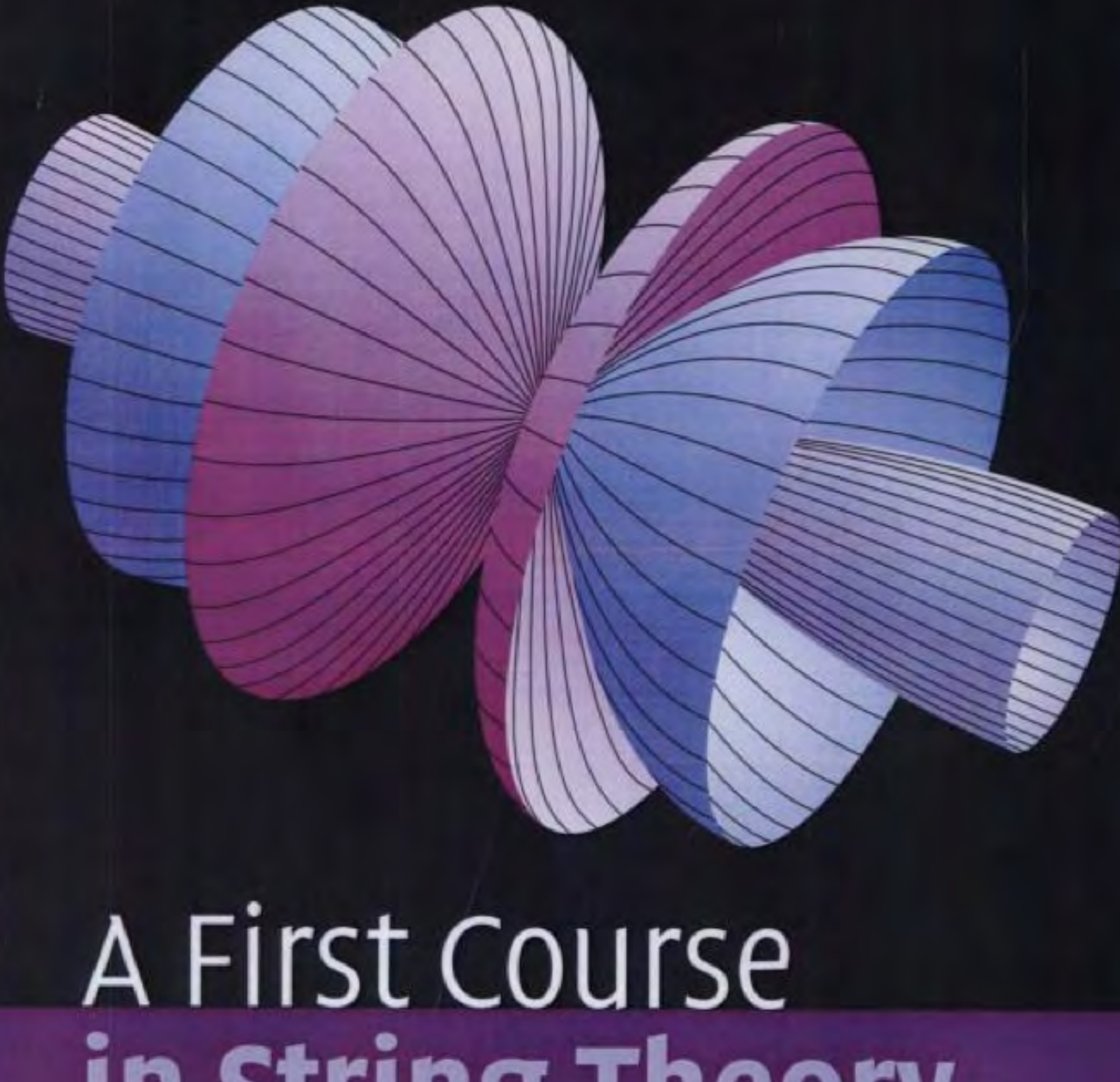
The conditions discussed in this paper for a relativistic dynamical system are necessary but not sufficient. Some further condition is needed to ensure that the interaction between two physical objects becomes small when the objects become far apart. It is not clear how this condition can be formulated mathematically. Present-day atomic theories involve the assumption of localizability, which is sufficient but is very likely too stringent. The assumption requires that the theory shall be built up in terms of dynamical variables that are each localized at some point in space-time, two variables localized at two points lying outside each other's light-cones being assumed to have zero P.b. A less drastic assumption may be adequate, e.g., that there is a fundamental length λ such that the P.b. of two dynamical variables must vanish if they are localized at two points whose separation is space-like and greater than λ, but need not vanish if it is less than λ.

I hope to come back elsewhere to the transition to the quantum theory.

Mathematics, rightly viewed, possesses not only truth, but supreme beauty---a beauty cold and austere, like that of sculpture.

~Bertrand Russell

Many people would sooner die than think; in fact, they do so.



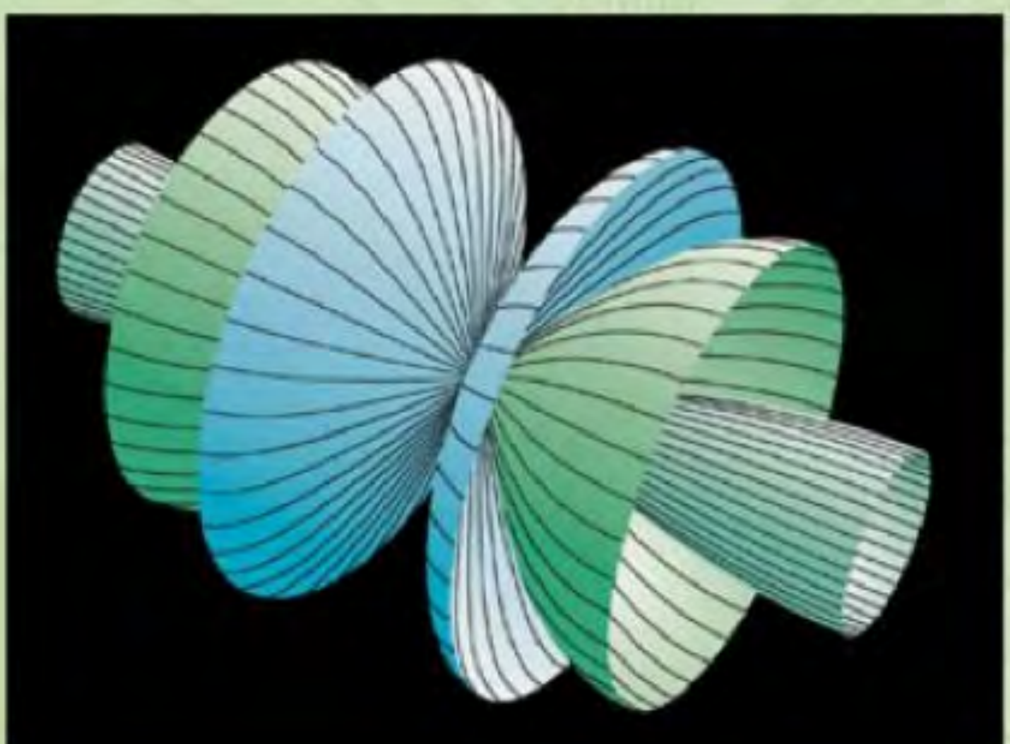
A First Course in String Theory

Barton Zwiebach

CAMBRIDGE

Second Edition

A First Course in STRING THEORY



Barton Zwiebach

between timelike-separated events define a timelike vector. Similarly, the coordinate differences between spacelike-separated events define a spacelike vector, and the coordinate differences between lightlike-separated events define a null vector.

Quick calculation 2.1 Verify that the invariant ds^2 is indeed preserved under the Lorentz transformations (2.36).

Quick calculation 2.2 Consider two Lorentz vectors a^μ and b^μ. Write the Lorentz transformations $a^\mu \to a'^\mu$ and $b^\mu \to b'^\mu$ analogous to (2.36). Verify that $a^\mu b_\mu$ is invariant under these transformations.

2.3 Light-cone coordinates

We now discuss a coordinate system that will be extremely useful in our study of string theory, the light-cone coordinate system. The quantization of the relativistic string can be worked out most directly using light-cone coordinates. There is a different approach to the quantization of the relativistic string, in which no special coordinates are used. This approach, called Lorentz covariant quantization, is discussed briefly in Chapter 24. Lorentz quantization is very elegant, but a full discussion requires a great deal of background material. We will use light-cone coordinates to quantize strings in this book.

We define the two light-cone coordinates x^+ and x^- as two independent linear combinations of the time coordinate and a chosen spatial coordinate, conventionally taken to be x^1. This is done by writing

$$
\begin{aligned}
x^+ &\equiv \frac{1}{\sqrt{2}}(x^0 + x^1), \\
x^- &\equiv \frac{1}{\sqrt{2}}(x^0 - x^1).
\end{aligned}
\tag{2.50}
$$

The coordinates x^2 and x^3 play no role in this definition. In the light-cone coordinate system, (x^0, x^1) is traded for (x^+, x^-), but the other two coordinates x^2, x^3 are kept. Thus, the complete set of light-cone coordinates is (x^+, x^-, x^2, x^3).

The new coordinates x^+ and x^- are called light-cone coordinates because the associated coordinate axes are the world-lines of beams of light emitted from the origin along the x^1 axis. For a beam of light moving in the positive x^1 direction, we have $x^1 = ct = x^0$, and thus $x^- = 0$. The line $x^- = 0$ is, by definition, the x^+ axis (Figure 2.2). For a beam of light moving in the negative x^1 direction, we have $x^1 = -ct = -x^0$, and thus $x^+ = 0$. This corresponds to the x^- axis. The $x^\pm$ axes are lines at 45° with respect to the x^0, x^1 axes.

Can we think of x^+, or perhaps x^-, as a new time coordinate? Yes. In fact, both have equal right to be called a time coordinate, although neither one is a time coordinate in the standard sense of the word. Light-cone time is not quite the same as ordinary time.

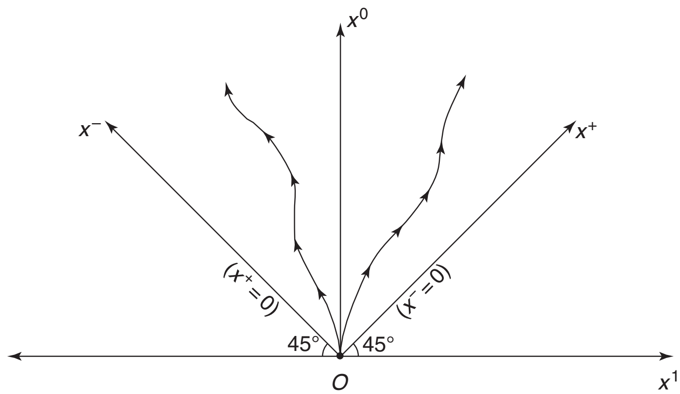


Fig. 2.2 **A spacetime diagram with x^1 and x^0 represented as orthogonal axes. Shown are the light-cone axes $x^{\pm} = 0$. The curves with arrows are possible world-lines of physical particles.**

Perhaps the most familiar property of time is that it goes forward for any physical motion of a particle. Physical motion starting at the origin is represented in Figure 2.2 as curves that remain within the light-cone and whose slopes never go below 45°. For all these curves, both x^+ and x^- increase as we follow the arrows. The only subtlety is that, for special light rays, light-cone time will freeze! As we saw above, x^+ remains constant for a light ray in the negative x^1 direction, while x^- remains constant for a light ray in the positive x^1 direction.

For definiteness, we will take x^+ to be the *light-cone time* coordinate. Accordingly, we will think of x^- as a spatial coordinate. Of course, these light-cone time and space coordinates will be somewhat strange.

Taking differentials of (2.50), we readily find that

$$2\,dx^+dx^- = (dx^0 + dx^1)\,(dx^0 - dx^1) = (dx^0)^2 - (dx^1)^2\,. \tag{2.51}$$

It follows that the invariant interval (2.13), expressed in terms of the light-cone coordinates (2.50), takes the form

$$-ds^2 = -2\,dx^+dx^- + (dx^2)^2 + (dx^3)^2\,. \tag{2.52}$$

The symmetry in the definitions of x^+ and x^- is evident here. Notice that, if we are given ds^2, solving for dx^- or for dx^+ does not require us to take a square root. This is a very important feature of light-cone coordinates, as we will see in Chapter 9.

How do we represent (2.52) with index notation? We still need indices that run over four values, but this time the values will be called

$$+\,,\,-\,,\,2\,,\,3. \tag{2.53}$$

Just as we did in (2.21), we write

$$-\,ds^2 = \hat{\eta}_{\mu\nu} dx^\mu dx^\nu . \tag{2.54}$$

Here we have introduced a light-cone metric $\hat{\eta}$ which, like the Minkowski metric, is also defined to be symmetric under the exchange of its indices. Expanding this equation, and comparing with (2.52), we find

$$\hat{\eta}_{+-} = \hat{\eta}_{-+} = -1\,, \quad \hat{\eta}_{++} = \hat{\eta}_{--} = 0\,. \tag{2.55}$$

In the $(+, -)$ subspace, the diagonal elements of the light-cone metric vanish, but the off-diagonal elements do not. We also find that $\hat{\eta}$ does not couple the $(+, -)$ subspace to the $(2, 3)$ subspace:

$$\hat{\eta}_{+I} = \hat{\eta}_{-I} = 0\,, \quad I = 2, 3\,. \tag{2.56}$$

The matrix representation of the light-cone metric is

$$\hat{\eta}_{\mu\nu} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{2.57}$$

The light-cone components of any Lorentz vector a^μ are defined in analogy with (2.50):

$$\begin{aligned} a^+ &\equiv \frac{1}{\sqrt{2}}\,(a^0 + a^1)\,, \\ a^- &\equiv \frac{1}{\sqrt{2}}\,(a^0 - a^1)\,. \end{aligned} \tag{2.58}$$

The scalar product between vectors, shown in (2.29), can be written using light-cone components. This time we have

$$a \cdot b = -a^- b^+ - a^+ b^- + a^2 b^2 + a^3 b^3 = \hat{\eta}_{\mu\nu}\, a^\mu b^\nu . \tag{2.59}$$

The last equality follows immediately from summing over the repeated indices and using (2.57). The first equality needs a small computation. In fact, it suffices to check that

$$-\,a^- b^+ - a^+ b^- = -a^0 b^0 + a^1 b^1 . \tag{2.60}$$

This is quickly done using (2.3) and the analogous equations for $b^\pm$. We can also introduce lower light-cone indices. Consider the expression $a \cdot b = a_\mu b^\mu$, and expand the sum over the index μ using the light-cone labels:

$$a \cdot b = a_+ b^+ + a_- b^- + a_2 b^2 + a_3 b^3 . \tag{2.61}$$

Comparing with (2.59), we find that

$$a_+ = -a^-\,, \quad a_- = -a^+ . \tag{2.62}$$

When we lower or raise the zeroth index in a Lorentz frame, we get an extra sign. In light-cone coordinates, the indices of the first two coordinates switch and we get an extra sign.

Since physics described using light-cone coordinates looks unusual, we must develop an intuition for it. To do this, we will look at an example where the calculations are simple but the results are surprising.

Consider a particle moving along the x^1 axis with speed parameter $\beta = v/c$. At time $t = 0$, the positions x^1, x^2, and x^3 are all zero. Motion is nicely represented when the positions are expressed in terms of time:

$$x^1(t) = vt = \beta x^0, \quad x^2(t) = x^3(t) = 0. \tag{2.63}$$

How does this look in light-cone coordinates? Since x^+ is time and $x^2 = x^3 = 0$, we must simply express x^- in terms of x^+. Using (2.63), we find

$$x^+ = \frac{x^0 + x^1}{\sqrt{2}} = \frac{1+\beta}{\sqrt{2}}\, x^0. \tag{2.64}$$

As a result,

$$x^- = \frac{x^0 - x^1}{\sqrt{2}} = \frac{(1-\beta)}{\sqrt{2}}\, x^0 = \frac{1-\beta}{1+\beta}\, x^+ . \tag{2.65}$$

Since it relates light-cone position to light-cone time, we identify the ratio

$$\frac{dx^-}{dx^+} = \frac{1-\beta}{1+\beta} \tag{2.66}$$

as the light-cone velocity. How strange is this light-cone velocity? For light moving to the right ($\beta = 1$) it equals zero. Indeed, light moving to the right has zero light-cone velocity because x^- does not change at all. This is shown as line 1 in Figure 2.3. Suppose you have a particle moving to the right with high conventional velocity, so that $\beta \simeq 1$ (line 2 in the figure). Its light-cone velocity is then very small. A long light-cone time must pass for this particle to move a little in the x^- direction. Perhaps more interestingly, a static particle in standard coordinates (line 3) is moving quite fast in light-cone coordinates. When $\beta = 0$ the particle has unit light-cone speed. This light-cone speed increases as β grows negative: the numerator in (2.66) is larger than one and increasing, while the denominator is smaller than one and decreasing. For $\beta = -1$ (line 5), the light-cone velocity is infinite! While this seems odd, there is no clash with relativity. Light-cone velocities are just unusual. The light-cone is a frame in which kinematics has a nonrelativistic flavor and infinite velocities are possible. Note that light-cone coordinates were introduced as a change of coordinates, not as a Lorentz transformation. There is no Lorentz transformation that takes the coordinates (x^0, x^1, x^2, x^3) into coordinates $(x'^0, x'^1, x'^2, x'^3) = (x^+, x^-, x^2, x^3)$.

Quick calculation 2.3 Convince yourself that the last statement above is correct.

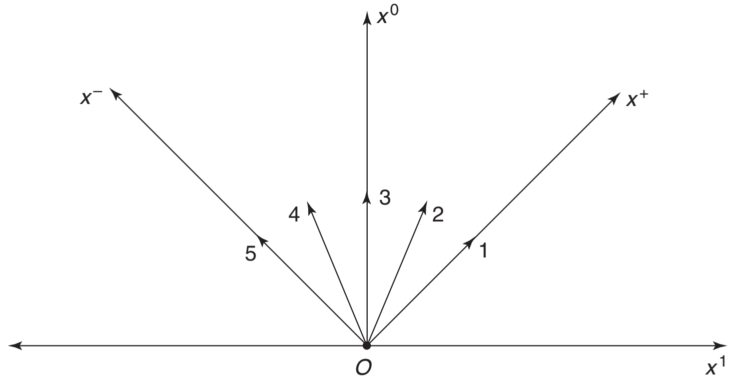


Fig. 2.3 World-lines of particles with various light-cone velocities. Particle 1 has zero light-cone velocity. The velocities increase through that of particle 5, which is infinite.

2.4 Relativistic energy and momentum

In special relativity there is a basic relationship between the rest mass m of a point particle, its relativistic energy E, and its relativistic momentum $\vec{p}$. This relationship is given by

$$\frac{E^2}{c^2} - \vec{p} \cdot \vec{p} = m^2 c^2. \tag{2.67}$$

The relativistic energy and momentum are given in terms of the rest mass and velocity by the following familiar relations:

$$E = \gamma m c^2 , \quad \vec{p} = \gamma m \vec{v}. \tag{2.68}$$

Quick calculation 2.4 Verify that the above E and $\vec{p}$ satisfy (2.67).

Energy and momentum can be used to define a momentum four-vector, as we will prove shortly. This four-vector is

$$p^\mu = (p^0, p^1, p^2, p^3) \equiv \left(\frac{E}{c}, p_x, p_y, p_z\right). \tag{2.69}$$

Using the last two equations, we have

$$p^\mu = \left(\frac{E}{c}, \vec{p}\right) = m\gamma (c, \vec{v}). \tag{2.70}$$

We use (2.28) to lower the index in p^μ:

$$p_\mu = (p_0, p_1, p_2, p_3) = \eta_{\mu\nu} p^\nu = \left(-\frac{E}{c}, p_x, p_y, p_z\right). \tag{2.71}$$

The above expressions for p^μ and p_μ give

$$p^\mu p_\mu = -(p^0)^2 + \vec{p} \cdot \vec{p} = -\frac{E^2}{c^2} + \vec{p} \cdot \vec{p} , \tag{2.72}$$

and, making use of (2.67), we have

$$p^{\mu} p_{\mu} = -m^2 c^2. \tag{2.73}$$

Since $p^{\mu} p_{\mu}$ has no free index it must be a Lorentz scalar. Indeed, all Lorentz observers agree on the value of the rest mass of a particle. Using the relativistic scalar product notation, condition (2.73) reads

$$p^2 \equiv p \cdot p = -m^2 c^2. \tag{2.74}$$

A central concept in special relativity is that of *proper time*. Proper time is a Lorentz invariant measure of time. Consider a moving particle and two events along its trajectory. Different Lorentz observers record different values for the time interval between the two events. But now imagine that the moving particle is carrying a clock. The proper time elapsed is the time elapsed between the two events *on that clock*. By definition, it is an invariant: all observers of a particular clock must agree on the time elapsed on that clock!

Proper time enters naturally into the calculation of invariant intervals. Consider an invariant interval for the motion of a particle along the x axis:

$$-\,ds^2 = -c^2 dt^2 + dx^2 = -c^2 dt^2\,(1-\beta^2). \tag{2.75}$$

Now evaluate the interval using a Lorentz frame attached to the particle. This is a frame in which the particle does not move and time is recorded by the clock that is moving with the particle. In this frame $dx = 0$ and $dt = dt_p$ is the proper time elapsed. As a result,

$$-\,ds^2 = -c^2\, dt_p{}^2. \tag{2.76}$$

We cancel the minus signs and take the square root (using (2.20)) to find

$$ds = c\, dt_p. \tag{2.77}$$

This shows that, for timelike intervals, ds/c is the proper time interval. Similarly, cancelling minus signs and taking the square root of (2.75) gives

$$ds = c dt\sqrt{1-\beta^2} \;\longrightarrow\; \frac{dt}{ds} = \frac{\gamma}{c}. \tag{2.78}$$

Being a Lorentz invariant, ds can be used to construct new Lorentz vectors from old Lorentz vectors. For example, a velocity four-vector u^{μ} is obtained by taking the ratio of dx^{μ} and ds. Since dx^{μ} is a Lorentz vector and ds is a Lorentz scalar, the ratio is also a Lorentz vector:

$$u^{\mu} = c\frac{dx^{\mu}}{ds} = c\left(\frac{d(ct)}{ds}, \frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds}\right). \tag{2.79}$$

The factor of c is included to give u^{μ} the units of velocity. The components of u^{μ} can be simplified using the chain rule and (2.78). For example,

$$\frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds} = \frac{v_x \gamma}{c}. \tag{2.80}$$

Back in (2.79), we find

$$u^\mu = \gamma(c, v_x, v_y, v_z) = \gamma(c, \vec{v}). \tag{2.81}$$

Comparing with (2.70), we see that the momentum four-vector is just mass times the velocity four-vector:

$$p^\mu = m u^\mu. \tag{2.82}$$

This confirms our earlier assertion that the components of p^μ form a four-vector. Since any four-vector transforms under Lorentz transformations as the x^μ do, we can use (2.36) to find that under a boost in the x-direction the p^μ transform as

$$\begin{aligned} \frac{E'}{c} &= \gamma\left(\frac{E}{c} - \beta\, p_x\right), \\ p'_x &= \gamma\left(-\beta \frac{E}{c} + p_x\right). \end{aligned} \tag{2.83}$$

2.5 Light-cone energy and momentum

The light-cone components p^+ and p^- of the momentum Lorentz vector are obtained using the rule (2.3):

$$\begin{aligned} p^+ &= \frac{1}{\sqrt{2}}(p^0 + p^1) = -p_-, \\ p^- &= \frac{1}{\sqrt{2}}(p^0 - p^1) = -p_+. \end{aligned} \tag{2.84}$$

Which component should be identified with light-cone energy? The naive answer would be p^+. In any Lorentz frame, both the time and energy are the zeroth components of their respective four-vectors. Since light-cone time was chosen to be x^+, we might conclude that light-cone energy should be taken to be p^+. This is not appropriate, however. Light-cone coordinates do not transform as Lorentz ones do, so we should be careful and examine this question in detail. Both $p^\pm$ are energy-like, since both are positive for physical particles. Indeed, from (2.67), and with $m \neq 0$, we have

$$p^0 = \frac{E}{c} = \sqrt{\vec{p}\cdot\vec{p} + m^2c^2} > |\,\vec{p}\,| \geq |p^1|\,. \tag{2.85}$$

As a result, $p^0 \pm p^1 > 0$, and thus $p^\pm > 0$. While both are plausible candidates for energy, the physically motivated choice turns out to be $-p_+$, which happens to coincide with p^-.

Before we explain this choice, let us first evaluate $p_\mu x^\mu$. In standard coordinates,

$$p \cdot x = p_0 x^0 + p_1 x^1 + p_2 x^2 + p_3 x^3. \tag{2.86}$$

In light-cone coordinates, using (2.61),

$$p \cdot x = p_+ x^+ + p_- x^- + p_2 x^2 + p_3 x^3. \tag{2.87}$$

In standard coordinates, $p_0 = -E/c$ appears together with the time x^0. In light-cone coordinates, p_+ appears together with the light-cone time x^+. We would therefore expect p_+ to be minus the light-cone energy.

Why is this pairing significant? Energy and time are conjugate variables. As you learned in quantum mechanics, the Hamiltonian operator measures energy and generates time evolution. The wavefunction of a point particle with energy E and momentum $\vec{p}$ is given by

$$\psi(t, \vec{x}) = \exp\Big(-\frac{i}{\hbar}(Et - \vec{p}\cdot\vec{x})\Big). \tag{2.88}$$

Indeed, this wavefunction satisfies the Schrödinger equation

$$i\hbar\,\frac{\partial\psi}{\partial x^0} = \frac{E}{c}\,\psi. \tag{2.89}$$

Similarly, light-cone time evolution and light-cone energy $E_{\rm lc}$ should be related by

$$i\hbar\,\frac{\partial\psi}{\partial x^+} = \frac{E_{\rm lc}}{c}\,\psi. \tag{2.90}$$

To find the x^+ dependence of the wavefunction, we recognize that

$$\psi(t, \vec{x}) = \exp\Big(\frac{i}{\hbar}\,(p_0 x^0 + \vec{p}\cdot\vec{x})\Big) = \exp\Big(\frac{i}{\hbar}\;p\cdot x\Big), \tag{2.91}$$

and, using (2.87), we have

$$\psi(x) = \exp\Big(\frac{i}{\hbar}\,(p_+ x^+ + p_- x^- + p_2 x^2 + p_3 x^3)\Big). \tag{2.92}$$

We can now return to (2.90) and evaluate:

$$i\hbar\frac{\partial\psi}{\partial x^+} = -p_+\psi \;\longrightarrow\; -p_+ = \frac{E_{\rm lc}}{c}. \tag{2.93}$$

This confirms our identification of $(-p_+)$ with light-cone energy. Since, presently, $-p_+ = p^-$, it is convenient to use p^- as the light-cone energy in order to eliminate the sign in the above equation:

$$p^- = \frac{E_{\rm lc}}{c}. \tag{2.94}$$

Some physicists like to raise and lower $+$ and $-$ indices to simplify expressions involving light-cone quantities. While this is sometimes convenient, it can easily lead to errors. If you talk with a friend over the phone, and she says "... p-plus times ...," you will have to ask, "plus up, or plus down?" In the rest of this book we will not lower the $+$ or $-$ indices. They will always be up, and the energy will always be p^-.

We can check that the identification of p^- as light-cone energy fits together nicely with the intuition that we have developed for light-cone velocity. To this end, we confirm that a particle with small light-cone velocity also has small light-cone energy. Suppose we have a particle moving very fast in the $+x^1$ direction. As discussed below (2.66), its light-cone velocity is very small. Since p^1 is very large, equation (2.67) gives

$$p^0 = \sqrt{(p^1)^2 + m^2c^2} = p^1\sqrt{1 + \frac{m^2c^2}{(p^1)^2}} \simeq p^1 + \frac{m^2c^2}{2p^1}. \tag{2.95}$$

The light-cone energy of the particle is therefore

$$p^- = \frac{1}{\sqrt{2}}(p^0 - p^1) \simeq \frac{m^2c^2}{2\sqrt{2}\,p^1}. \tag{2.96}$$

As anticipated, both the light-cone velocity and the light-cone energy decrease as p^1 increases.

2.6 Lorentz invariance with extra dimensions

If string theory is correct, we must entertain the possibility that spacetime has more than four dimensions. The number of time dimensions must be kept equal to one – it seems very difficult, if not altogether impossible, to construct a consistent theory with more than one time dimension. The extra dimensions must therefore be spatial. Can we have Lorentz invariance in worlds with more than three spatial dimensions? Yes. Lorentz invariance is a concept that admits a very natural generalization to spacetimes with additional dimensions.

We first extend the definition of the invariant interval ds^2 to incorporate the additional space dimensions. In a world with five spatial dimensions, for example, we would write

$$-ds^2 = -c^2dt^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2 + (dx^5)^2. \tag{2.97}$$

Lorentz transformations are then defined as the linear changes of coordinates that leave ds^2 invariant. This ensures that every inertial observer in the six-dimensional spacetime will agree on the value of the speed of light. With more dimensions, come more Lorentz transformations. While in four-dimensional spacetime we have boosts in the x^1, x^2, and x^3 directions, in this new world we have boosts along each of the five spatial dimensions. With three spatial coordinates, there are three basic spatial rotations: rotations that mix x^1 and x^2, those that mix x^1 and x^3, and finally those that mix x^2 and x^3. The equality of the number of boosts and the number of rotations is a special feature of four-dimensional spacetime. With five spatial coordinates, we have ten rotations, which is twice the number of boosts.

The higher-dimensional Lorentz invariance includes the lower-dimensional one: if nothing happens along the extra dimensions, then the restrictions of lower-dimensional Lorentz invariance apply. This is clear from (2.97). For motion that does not involve the extra dimensions, $dx^4 = dx^5 = 0$, and the expression for ds^2 reduces to that used in four dimensions.

Dirac's light-cone coordinate system

Y. S. Kim
Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
Marilyn E. Noz
Department of Radiology, New York University, New York, New York 10016

(Received 30 June 1981; accepted for publication 15 October 1981)

It is shown that Dirac's light-cone coordinate system provides an effective method for treating the geometry of Lorentz transformation in a rectangular coordinate system. Transformation properties of the coordinate variables and those of the derivatives are discussed in detail. The Lorentz boost along a given direction is shown to be a coordinate transformation in which "cross products" are preserved. It is pointed out that the Lorentz boost is a symplectic transformation.

I. INTRODUCTION

In learning or teaching special relativity, the most uncomfortable step is to deal with the skew coordinate system associated with Lorentz transformations. Many attempts have been made in the past to rectify this situation. Among them, the most common practice has been to use a four-dimensional Euclidian geometry with three real spatial coordinates and one imaginary time variable. However, the recent trend has been to phase out this imaginary-time metric from the physics curriculum.

In an attempt to formulate a "relativistic dynamics of atoms," Dirac[1] introduced a coordinate system which allows us to describe the geometry of Lorentz transformations using only a rectangular coordinate system commonly known as the light-cone coordinate system. The purpose of the present paper is to provide a pedagogical elaboration of Dirac's light-cone coordinate system with the following points in mind:

(a) There are many students who wish to learn special relativity without relying solely on the Minkowskian geometry.

(b) The recent trend in teaching relativity has been to use space-time diagrams more extensively.[2–9] We are therefore constantly looking for "pictures," i.e., space-time diagrams which may be more appealing to students.

(c) When we learn/teach orthogonal transformations, we recite the fact that the scalar or "dot" product of two vectors is preserved. Since the students who know the dot product are likely to know also the "cross" product of two vectors, it is not uncommon for them to ask whether there are transformations which preserve the cross product.[8,10]

We are not the first ones to introduce the light-cone coordinate system to this Journal. In 1970, Parker and Schmieg published two papers containing the basic concepts and applications of the light-cone variables.[6] However, it is important to note that the light-cone coordinate system gained its prominence in physics research in the 1970s, after the papers of Ref. 6 were published.

The light-cone coordinate system is now a standard language for high-energy research.[11] It serves also as a concrete physical example for the symplectic geometry which is becoming increasingly important in many branches of physics, including classical mechanics,[12] geometrical optics,[13] general relativity,[14] quantum mechanics,[15] and relativistic quantum mechanics.[16] Therefore, it is not unreasonable to discuss a few more basic pedagogical aspects of the light-cone coordinate system which can serve as a bridge between the current physics research and the early works of Dirac[1] and of Parker and Schmieg.[6]

In Sec. II, we discuss the light-cone coordinate system as it is defined by Dirac, and quantities which remain invariant under Lorentz transformations. Section III deals with derivatives with respect to the light-cone variables. It is shown in Sec. IV that Lorentz invariant quantities can be defined in terms of cross products, and that the Lorentz boost is a cross-product-preserving transformation. In Sec. V, the mathematics of Secs. II and IV is translated into a matrix language. It is pointed out in Sec. IV that the Lorentz boost along a given direction is a symplectic transformation.

II. LIGHT-CONE COORDINATE SYSTEM

Let us consider a coordinate system (z,t), and another system (z',t') which is moving with respect to the (z,t) system with velocity β. Then the transformation law leads to

$$z' = z \cosh \eta - t \sinh \eta, \qquad t' = t \cosh \eta - z \sinh \eta, \tag{1}$$

where

$$\sinh \eta = \beta/(1-\beta^2)^{1/2}.$$

We use here the unit system in which $c = 1$, and ignore transverse coordinates which are not affected by the Lorentz boost.

Dirac observed in 1949 that it may be more convenient to use the variables

$$u = (t+z)/\sqrt{2}, \qquad v = (t-z)/\sqrt{2}, \tag{2}$$

which are commonly called the "light-cone variables." In terms of these variables, the Lorentz transformation of Eq. (1) becomes

$$u' = (t'+z')/\sqrt{2} = e^{-\eta}(t+z)/\sqrt{2} = e^{-\eta} u, \qquad v' = (t'-z')/\sqrt{2} = e^{+\eta}(t+z)/\sqrt{2} = e^{+\eta} v. \tag{3}$$

The Lorentz transformation indeed takes a simple form in this coordinate system. The most remarkable feature is that

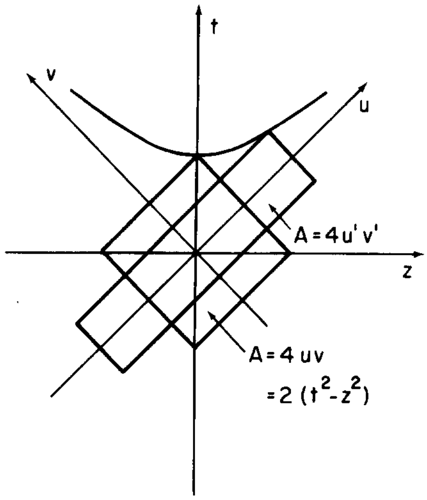


Fig. 1. Graphical illustration of the Lorentz-invariant product (t^2-z^2). Under Lorentz boosts, the space-time point (z,t) traces a hyperbola. The Lorentz invariance of (t^2-z^2) is equivalent to the invariance of the area of the rectangle "inscribed" by the hyperbola.

the u and v variables do not become linearly mixed under this transformation.[1,6,8] They simply undergo scale transformations in such a way that the product uv remains constant:

$$uv = u'v'. \tag{4}$$

The transformation property of the u and v variables appears mathematically different from the conventional form given in Eq. (1). However, the light-cone variables cannot be strange to us. The transformation from z and t to u and v is simply a rotation by 45° in the Euclidean zt plane. In terms of the t and z variables, the Lorentz invariance of the above quantity can be written as

$$uv = (t+z)(t-z)/2 = (t^2-z^2)/2. \tag{5}$$

Figures 1 and 2 illustrate the transformation property in the light-cone coordinate system. In both of these figures, the above formula describes the well-known hyperbola. However, a closer investigation reveals that the area of the rectangle "inscribed" by this hyperbola is a Lorentz-invariant constant. The Lorentz invariance of the quantity

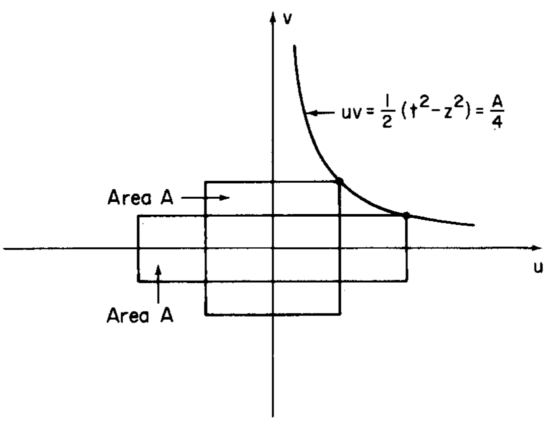


Fig. 2. Space-time diagram of a Lorentz boost in the light-cone coordinate system. This figure is only a 45° rotation of Fig. 1, but gives a much clearer picture of the coordinate transformation associated with the Lorentz boost.

(t^2-z^2) is well known. It is therefore interesting to note that we can associate this invariance with the invariance of the area of the rectangle. It is also interesting to note that this rectangular deformation is an effective language in high-energy hadronic physics.[11]

III. DERIVATIVES

One of the most difficult steps in teaching the Minkowskian geometry is to explain that the derivative with respect to a contravariant four-vector transforms like a covariant vector. We can bypass this difficulty by using the light-cone coordinate system. Because the u and v variables do not mix with each other, their differential operators transform like

$$\partial/\partial u' = e^{+\eta}\,\partial/\partial u, \qquad \partial/\partial v' = e^{-\eta}\,\partial/\partial v. \tag{6}$$

From the above formulas, it is clear that the operator $(\partial^2/\partial u\,\partial v)$ is Lorentz invariant. In terms of the u and v variables, the conventional differential operators can be written as

$$\partial/\partial t = (1/\sqrt{2})\,(\partial/\partial u + \partial/\partial v), \qquad \partial/\partial z = (1/\sqrt{2})\,(\partial/\partial u - \partial/\partial v). \tag{7}$$

It is then easy to prove that

$$(\partial^2/\partial u\,\partial v) = \partial^2/(\partial t)^2 - \partial^2/(\partial z)^2. \tag{8}$$

The Lorentz invariance of the right-hand side of the above form is well known.

The immediate application of the above formula is the wave equation, which, in terms of the light-cone variables, can be written as

$$\frac{\partial^2}{\partial u\,\partial v}\,\psi(u,v) = 0. \tag{9}$$

This equation tells us that solutions of the above equation depend either on u or v, not on both. Thus the most general form of solution takes the form

$$\psi(u,v) = f(u) + g(v), \tag{10}$$

where $f(u)$ is a function of $(t+z)$, and $g(v)$ depends on $(t-z)$. The above form is indeed a solution of the wave equation.

It would be interesting to see how we can write and solve Maxwell's equations. This important task has been carried out in Ref. 6.

IV. LORENTZ INVARIANCE OF CROSS PRODUCTS

In Sec. II, we considered only the transformation properties of the coordinate variables. In general, we have to consider transformation of four-vectors, such as the energy–momentum four-vector, which also satisfy the transformation law given in Eq. (1). Let us consider two four-vectors

$$A^\mu = (A_t, A_z), \qquad B^\mu = (B_t, B_z). \tag{11}$$

Here again, we assume that the Lorentz transformation is made along the z direction, and ignore the x and y components which are not affected by the transformation. We can define the u and v components of the above four-vectors:

$$A_u = (A_t + A_z)/\sqrt{2}, \quad A_v = (A_t - A_z)/\sqrt{2};$$
$$B_u = (B_t + B_z)/\sqrt{2}, \quad B_v = (B_t - B_z)/\sqrt{2}. \tag{12}$$

They then satisfy the transformation equations:

$$A'_u = e^{-\eta} A_u, \quad A'_v = e^{+\eta} A_v;$$
$$B'_u = e^{-\eta} B_u, \quad B'_v = e^{+\eta} B_v. \tag{13}$$

Therefore $A_u B_v$ and $A_v B_u$ remain invariant under the Lorentz transformation. Consequently, the following linear combinations are also Lorentz invariant.

$$A_u B_v + A_v B_u = A_t B_t - A_z B_z, \tag{14}$$
$$A_u B_v - A_v B_u = A_t B_z - A_z B_t. \tag{15}$$

We can explain the above invariances in terms of the cross product. If we introduce the unit vectors $\mathbf{e}_u$ and $\mathbf{e}_v$, we can write the "vector" $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{e}_u A_u + \mathbf{e}_v A_v, \tag{16}$$

with a similar expression for $\mathbf{B}$.

Then the invariance of the quantity given in Eq. (15) is simply the invariance of the cross product:

$$\mathbf{A} \times \mathbf{B} = (A_u B_v - A_v B_u)(\mathbf{e}_u \times \mathbf{e}_v). \tag{17}$$

Since the u and v components do not become mixed during the process of Lorentz transformation, the vectors

$$\mathbf{A}_u = \mathbf{e}_u A_u, \quad \mathbf{A}_v = \mathbf{e}_v A_v;$$
$$\mathbf{B}_u = \mathbf{e}_u B_u, \quad \mathbf{B}_v = \mathbf{e}_v B_v. \tag{18}$$

can be regarded as independent vectors for mathematical purposes, and the cross product of any pair is a Lorentz-invariant quantity. For instance,

$$\mathbf{A}_u \times \mathbf{B}_u = \mathbf{A}_v \times \mathbf{B}_v = 0,$$
$$\mathbf{A}_u \times \mathbf{B}_v = (A_u B_v)(\mathbf{e}_u \times \mathbf{e}_v), \tag{19}$$
$$\mathbf{A}_v \times \mathbf{B}_u = -(A_v B_u)(\mathbf{e}_u \times \mathbf{e}_v).$$

The following linear combination is therefore boost invariant:

$$\mathbf{A}_u \times \mathbf{B}_v - \mathbf{A}_v \times \mathbf{B}_u, \tag{20}$$

which leads to the invariance of the quantity in Eq. (14).

From Eqs. (17) and (20), we can conclude that the Lorentz boost is indeed a cross-product-preserving transformation.[8]

V. MATRIX FORMS

In terms of matrices, the Lorentz transformation of Eq. (13) can be written as

$$\begin{pmatrix} A'_u \\ A'_v \end{pmatrix} = \begin{pmatrix} e^{-\eta} & 0 \\ 0 & e^{+\eta} \end{pmatrix} \begin{pmatrix} A_u \\ A_v \end{pmatrix}. \tag{21}$$

Let us call the above transformation matrix $L(\eta)$:

$$L(\eta) = \begin{pmatrix} e^{-\eta} & 0 \\ 0 & e^{+\eta} \end{pmatrix}; \tag{22}$$

and define the metric matrix J:

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{23}$$

The L matrix then has the following interesting property:

$$\tilde{L}(\eta) J L(\eta) = J, \tag{24}$$

where $\tilde{L}$ is the transpose of L.

The matrix relation of Eq. (24) leads to the Lorentz invariance of the quantity

$$(B_u, B_v) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} A_u \\ A_v \end{pmatrix}. \tag{25}$$

This invariance is identical to the cross-product invariance of Eq. (15).

Because the Lorentz transformation matrix of Eq. (22) is diagonal and does not mix the upper and lower components of the vector, we can write the Lorentz transformation of A_u as

$$\begin{pmatrix} A'_u \\ 0 \end{pmatrix} = \begin{pmatrix} e^{-\eta} & 0 \\ 0 & e^{+\eta} \end{pmatrix} \begin{pmatrix} A_u \\ 0 \end{pmatrix}, \tag{26}$$

without the participation of the A_v component. We can write a similar expression for the transformation of A_v. From this, we can conclude that both $A_u B_v$ and $A_v B_u$ remain invariant under Lorentz transformations. This will also lead to the Lorentz invariance given in Eq. (14), as well as that of Eq. (15).

Another interesting property of the metric matrix J of Eq. (23) is that the rotation matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{27}$$

also satisfies the relation

$$\tilde{R} J R = J. \tag{28}$$

Thus the boost matrix in the "rotated" coordinate system[17]

$$L'(\theta,\eta) = R(\theta)\tilde{L}(\theta)R(\theta) \tag{29}$$

also satisfies the condition

$$\tilde{L}' J L' = J. \tag{30}$$

Because the light-cone coordinate system is a 45° rotation of the Cartesian zt system, we can go through the above calculation with the conventional z and t components. In this case, the Lorentz transformation matrix is not diagonal, but is

$$L'(45°,\eta) = R(-45°)L(\eta)R(45°).$$
$$= \begin{pmatrix} \cosh\eta & -\sinh\eta \\ -\sinh\eta & \cosh\eta \end{pmatrix}. \tag{31}$$

This form is of course the familiar Lorentz transformation matrix in the coordinate system of z and t.

The metric matrix J of Eq. (23) should not be confused with the familiar matrix

$$G = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{32}$$

applicable to the vector (A_t, A_z). As is well known, this matrix leads to the Lorentz invariance of $(A_t B_t - A_z B_z)$. However, it is important to note that the G matrix is fundamentally different from the J matrix given in Eq. (23). The G matrix does not commute with J, and takes the form

$$G' = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{33}$$

after the 45° rotation, in the (u,v) coordinate system.

The Lorentz transformation matrix $L(\eta)$ in the (u,v) coordinate system also satisfies

$$\tilde{L}(\eta)G'L(\eta) = G'. \tag{34}$$

For this reason, the matrices

$$K_{\pm} = (J \pm G')/2$$

should also satisfy the relation

$$\tilde{L}(\eta)K_{\pm}\ L(\eta) = K_{\pm}. \tag{35}$$

Since the K matrices take the form

$$K_{+} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$K_{-} = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}. \tag{36}$$

The relation of Eq. (35) leads to the boost invariance of the quantities given in Eq. (19).

VI. SYMPLECTIC TRANSFORMATION

The mathematics presented in Secs. II–V is not well known among physicists, but is surprisingly simple. Thus we are led to inquire whether mathematicians have a specific name for this kind of construction. The transformation which preserves the metric of Eq. (23) is called the "symplectic" transformation, or more specifically the transformation of the group Sp(2). The word "symplectic group" was invented by Weyl in 1938,[18] and is therefore relatively new even in mathematics. This explains why the use of this group has not yet been common in physics research and teaching.

However, it is very important to realize that the use of the symplectic group is one of the new elements in many theoretical research lines. This is due to the fact that this group preserves skew symmetric products, and that there are many skew symmetric products in physics. For instance, the Poisson brackets in classical mechanics are skew symmetric and are therefore symplectic.[12] Consequently, Heisenberg's uncertainty relation in quantum mechanics is a symplectic construction.[14]

We have shown in the preceding sections that the Lorentz boost in a given direction is also a symplectic transformation. The use of the symplectic group in relativity does not end here. There are many skew symmetric products in differential geometry needed in studying general relativity.[15]

The symplectic group has been recently shown also to be very useful in geometrical optics,[13] and even in calculations of lens aberrations.[19] This new trend indeed encourages us to look at all branches of physics with the symplectic transformation in mind. It would be interesting to see how the symplectic group can be used in studying Maxwell's equations which are partly based on cross products.[6]

Because the symplectic property is common among different branches of physics, it is not unreasonable to ask whether it can serve as a unifying agent. It has now been firmly established that it serves as a unifying language for classical mechanics and geometrical optics.[13] Because both quantum mechanics and special relativity contain symplectic ingredients, this group may play an important role in constructing relativistic quantum mechanics.[16]

Throughout this paper, we ignored the transverse components which are not affected by Lorentz boosts along the z axis. As is well known, the representation of the full Lorentz group requires one boost operator and two rotation operators to produce its six generators. It is important to note that the symplectic transformation discussed in the present paper is applicable only to boosts along a given direction. We should realize also that including the rotations to produce all six generators is not equivalent to enlarging the dimension of the symplectic group.[20]

VII. CONCLUDING REMARKS

When we teach/learn transformations in which the dot product is preserved, we ask ourselves what transformation preserves the cross product.[10] This question is well justified in view of the present paper. The task of answering this question indeed constitutes a subject with rich content.

The mathematics required to answer this question is not difficult, and can start with a simple example of Lorentz boost,[8] if not the stress tensor.[10] One of the authors (YSK) in fact has used the contents of Secs. II, III, and IV in teaching special relativity in a freshman physics course for the physics majors at the University of Maryland, with encouraging results.

1. P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
2. E. F. Taylor and J. A. Wheeler, *Spacetime Physics* (Freeman, San Francisco, 1963).
3. A. Shadowitz, *Special Relativity* (Saunders, Philadelphia, 1968).
4. H. Amar, Am. J. Phys. **23**, 487 (1948).
5. R. W. Brehme, Am. J. Phys. **30**, 489 (1962); **36**, 896 (1968).
6. L. P. Parker and G. M. Schmieg, Am. J. Phys. **38**, 218, 1298 (1970).
7. J. S. Rigden, Am. J. Phys. **36**, 896 (1972).
8. E. N. Glass, Am. J. Phys. **43**, 1013 (1975).
9. C. Leubner, Am. J. Phys. **49**, 368 (1981).
10. R. P. Feynman, R. B. Leighton, and M. Sands, *Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1964), Vol. II, Sec. 39.
11. S. D. Drell and T. M. Yan, Ann. Phys. **60**, 578 (1971); Y. S. Kim and M. E. Noz, Phys. Rev. D **15**, 335 (1977); P. E. Hussar, Phys. Rev. D **23**, 2781 (1981).
12. V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, Berlin and New York, 1978).
13. R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978); S. Sternberg, Bull. Am. Math. Soc. **2** (2), 378 (1978).
14. M. Moshinsky and P. Winternitz, J. Math. Phys. **21**, 1667 (1980).
15. L. D. Faddeev, Actes Congr. Int. Math. **3**, 35 (1970).
16. Y. S. Kim and M. E. Noz, J. Math. Phys. **22**, 2289 (1981).
17. This rotation is strictly in the coordinate system of z and t or u and v, and is not to be confused with a rotation in the three-dimensional Euclidian space consisting of three spatial coordinates.
18. H. Weyl, *Classical Groups* (Princeton University, Princeton, NJ, 1946).
19. A. J. Dragt, J. Opt. Soc. Am. **72**, 372 (1982).
20. R. Gilmore, *Lie Groups and Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).

We note here that the transition from Eq.(18) to Eq.(19) is a squeeze transformation. The wave function of Eq.(18) is distributed within a circular region in the uv plane, and thus in the zt plane. On the other hand, the wave function of Eq.(19) is distributed in an elliptic region. This is how the wave function is Lorentz-boosted.

5 Feynman's Parton Picture

It is safe to believe that hadrons are quantum bound states of quarks having localized probability distribution. As in all bound-state cases, this localization condition is responsible for the existence of discrete mass spectra. The most convincing evidence for this bound-state picture is the hadronic mass spectra which are observed in high-energy laboratories [2, 15]. However, this picture of bound states is applicable only to observers in the Lorentz frame in which the hadron is at rest. How would the hadrons appear to observers in other Lorentz frames?

In 1969, Feynman observed that a fast-moving hadron can be regarded as a collection of many "partons" whose properties do not appear to be identical to those of quarks [19]. For example, the number of quarks inside a static proton is three, while the number of partons in a rapidly moving proton appears to be infinite. The question then is how the proton looking like a bound state of quarks to one observer can appear different to an observer in a different Lorentz frame? Feynman made the following systematic observations.

a). The picture is valid only for hadrons moving with velocity close to that of light.

b). The interaction time between the quarks becomes dilated, and partons behave as free independent particles.

c). The momentum distribution of partons becomes widespread as the hadron moves very fast.

d). The number of partons seems to be infinite or much larger than that of quarks.

Because the hadron is believed to be a bound state of two or three quarks, each of the above phenomena appears as a paradox, particularly b) and c) together. We would like to resolve this paradox using the covariant harmonic oscillator formalism.

For this purpose, we need a momentum-energy wave function. If the quarks have the four-momenta p_a and p_b, we can construct two independent four-momentum variables [15]

$$P = p_a + p_b, \qquad q = \sqrt{2}(p_a - p_b). \tag{20}$$

The four-momentum P is the total four-momentum and is thus the hadronic four-momentum. q measures the four-momentum separation between the quarks.

We expect to get the momentum-energy wave function by taking the Fourier transformation of Eq.(19):

$$\phi_\eta(q_z, q_0) = \left(\frac{1}{2\pi}\right) \int \psi_\eta(z,t) \exp\{-i(q_z z - q_0 t)\} dx dt. \tag{21}$$

Let us now define the momentum-energy variables in the light-cone coordinate system as

$$q_u = (q_0 - q_z)/\sqrt{2}, \qquad q_v = (q_0 + q_z)/\sqrt{2}. \tag{22}$$

In terms of these variables, the Fourier transformation of Eq.(21) can be written as

$$\phi_\eta(q_z, q_0) = \left(\frac{1}{2\pi}\right) \int \psi_\eta(z,t) \exp\{-i(q_u u + q_v v)\} du dv. \tag{23}$$

The resulting momentum-energy wave function is

$$\phi_\eta(q_z, q_0) = \left(\frac{1}{\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}\left(e^{-2\eta} q_u^2 + e^{2\eta} q_v^2\right)\right\}. \tag{24}$$

Since we are using the harmonic oscillator, the mathematical form of the above momentum-energy wave function is identical to that of the space-time wave function. The Lorentz squeeze properties of these wave functions are also the same, as are indicated in Fig. 1. These squeeze transformations perfectly consistent with the algorithms of the Poincaré group [20].

When the hadron is at rest with $\eta = 0$, both wave functions behave like those for the static bound state of quarks. As η increases, the wave functions become continuously squeezed until they become concentrated along their respective positive light-cone axes. Let us look at the z-axis projection of the space-time wave function. Indeed, the width of the quark distribution increases as the hadronic speed approaches that of the speed of light. The position of each quark appears widespread to the observer in the laboratory frame, and the quarks appear like free particles.

Furthermore, interaction time of the quarks among themselves become dilated. Because the wave function becomes wide-spread, the distance between one end of the harmonic oscillator well and the other end increases as is indicated in Fig. 1. This effect, first noted by Feynman [19], is universally observed in high-energy hadronic experiments. The period is oscillation is increases like e^{η}. On the other hand, the interaction time with the external signal, since it is moving in the direction opposite to the direction of the hadron, it travels along the negative light-cone axis. If the hadron contracts along the negative light-cone axis, the interaction time decreases by $e^{-\eta}$. The ratio of the interaction time to the oscillator period becomes $e^{-2\eta}$. The energy of each proton coming out of the Fermilab accelerator is $900 GeV$. This leads the ratio to 10^{-6}. This is indeed a small number. The external signal is not able to sense the interaction of the quarks among themselves inside the hadron. This is the reason why the partons appear to be incoherent to external signals. Indeed, Feynman's decoherence is an effect of the Lorentz covariance.

Concluding Remarks

Due to Einstein, this world, at least the physics world, became Lorentz-covariant. The lack of coherence in Feynman's parton picture is the most puzzling question in covariance. It is a pleasure to report that Wigner's formulation of the internal space-time symmetries of relativistic particles provide a resolution to this problem.

In this report, we discussed Wigner's 1939 paper on the representations of the Poincaré group. Wigner wrote many other papers. They were also discussed at this conference. We are grateful to Professors Joszef Janszky and Peter Adam for organizing this historical conference. The author would like to thank Jiri Kvita for pointing out an typographical error in the original version.

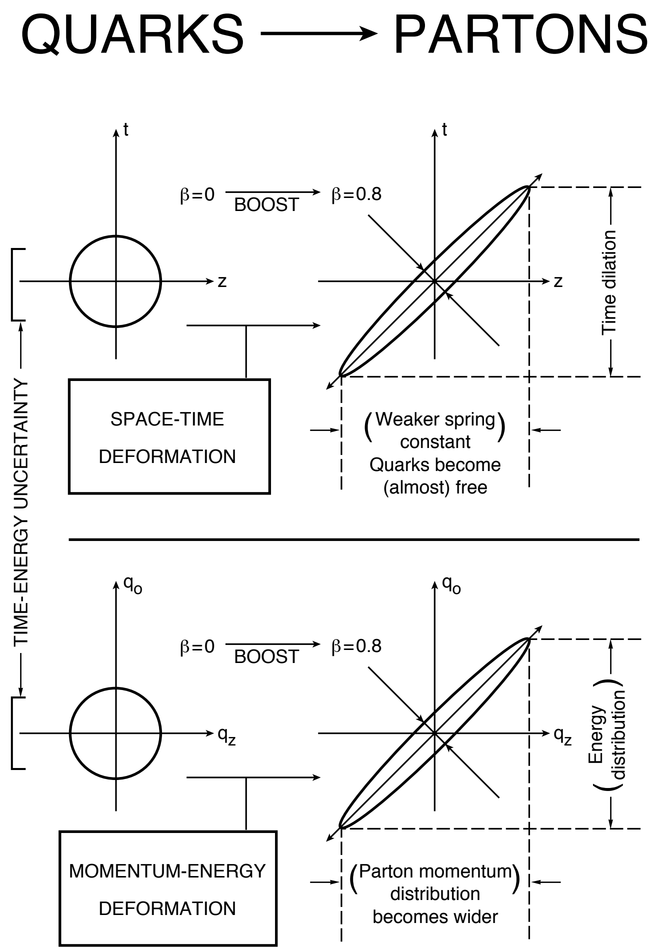


Figure 1: Lorentz-squeezed space-time and momentum-energy wave functions. As the hadron's speed approaches that of light, both wave functions become concentrated along their respective positive light-cone axes. These light-cone concentrations lead to Feynman's parton picture.

From Stokes

to

Bourbaki

Stokes' Theorem

“The general theorem is due to Nicolas Bourbaki ... and vice-versa !”

http://home.att.net/~numericana/answer/forms.htm

One of the primary motivations for the creation of Nicolas Bourbaki was to develop the mathematics required to prove the Generalized Stokes Theorem.

The generalized Stokes Theorem in 1, 2, and 3 dimensions

GAUSS

Surface encloses volume

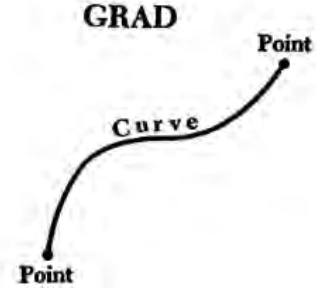
$$\int_{\text{surface}} \mathbf{F}\cdot d\mathbf{a} = \int_{\text{volume}} \operatorname{div}\mathbf{F}\, dv$$

STOKES

Curve encloses surface

$$\int_{\text{curve}} \mathbf{A}\cdot d\mathbf{s} = \int_{\text{surface}} \operatorname{curl}\mathbf{A}\cdot d\mathbf{a}$$

GRAD



Points enclose curve

$$\varphi_2 - \varphi_1 = \int_{\text{curve}} \operatorname{grad}\varphi\cdot d\mathbf{s}$$

IN CARTESIAN COORDINATES

$$\operatorname{div}\mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} = \nabla\cdot\mathbf{F}$$

$$\begin{aligned}\operatorname{curl}\mathbf{A} &= \hat{\mathbf{x}}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) \\ &+ \hat{\mathbf{y}}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) \\ &+ \hat{\mathbf{z}}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \\ &= \nabla\times\mathbf{A}\end{aligned}$$

$$\operatorname{grad}\varphi = \hat{\mathbf{x}}\frac{\partial\varphi}{\partial x} + \hat{\mathbf{y}}\frac{\partial\varphi}{\partial y} + \hat{\mathbf{z}}\frac{\partial\varphi}{\partial z} = \nabla\varphi$$

$$\nabla = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}$$

From Purcell

The History of Stokes' Theorem

Let us give credit where credit is due: Theorems of Green, Gauss and Stokes appeared unheralded in earlier work.

VICTOR J. KATZ
University of the District of Columbia
Washington, D.C. 20005

Most current American textbooks in advanced calculus devote several sections to the theorems of Green, Gauss, and Stokes. Unfortunately, the theorems referred to were not original to these men. It is the purpose of this paper to present a detailed history of these results from their origins to their generalization and unification into what is today called the generalized Stokes' theorem.

Origins of the theorems

The three theorems in question each relate a k-dimensional integral to a $k-1$-dimensional integral; since the proof of each depends on the fundamental theorem of calculus, it is clear that their origins can be traced back to the late 17th century. Toward the end of the 18th century, both Lagrange and Laplace actually used the fundamental theorem and iteration to reduce k-dimensional integrals to those of one dimension less. However, the theorems as we know them today did not appear explicitly until the 19th century.

The first of these theorems to be stated and proved in essentially its present form was the one known today as Gauss' theorem or the divergence theorem. In three special cases it occurs in an 1813 paper of Gauss **[8]**. Gauss considers a surface (superficies) in space bounding a solid body (corpus). He denotes by PQ the exterior normal vector to the surface at a point P in an infinitesimal element of surface ds and by QX, QY, QZ the angles this vector makes with the positive x-axis, y-axis, and z-axis respectively. Gauss then denotes by $d\Sigma$ an infinitesimal element of the y-z plane and erects a cylinder above it, this cylinder intersecting the surface in an even number of infinitesimal surface elements $ds_1, ds_2, \dots, ds_{2n}$. For each j, $d\Sigma = \pm ds_j \cos QX_j$ where the positive sign is used when the angle is acute, the negative when the angle is obtuse. Since if the cylinder enters the surface where QX is obtuse, it will exit where QX is acute (see FIGURE 1), Gauss obtains $d\Sigma = -ds_1 \cos QX_1 = ds_2 \cos QX_2 = \dots$ and concludes by summation that "The integral $\int ds \cos QX$ extended to the entire surface of the body is 0."

He notes further that if T, U, V are rational functions of only y,z, only x,z, and only x,y respectively, then "$\int (T\cos QX + U\cos QY + V\cos QZ)\,ds = 0$." Gauss then approximates the volume of the body by taking cylinders of length x and cross sectional area $d\Sigma$ and concludes in a similar way his next theorem: "The entire volume of the body is expressed by the integral $\int ds\, x(\cos QX)$ extended to the entire surface." We will see below how these results are special cases of the divergence theorem.

Finally, Cartan defines the "derived expression" of a first degree differential expression $\omega = A_1 dx_1 + A_2 dx_2 + \cdots + A_n dx_n$ to be the second degree expression $\omega' = dA_1 dx_1 + dA_2 dx_2 + \cdots + dA_n dx_n$, where, of course,

$$dA_i = \Sigma_j \frac{\partial A_i}{\partial x_j} dx_j.$$

For the case $n=3$ one can calculate by using the above rules that if $\omega = A_1 dx + A_2 dy + A_3 dz$, then

$$\omega' = \left(\frac{\partial A_3}{\partial y} - \frac{\partial A_2}{\partial z} \right) dy\, dz + \left(\frac{\partial A_1}{\partial z} - \frac{\partial A_3}{\partial x} \right) dz\, dx + \left(\frac{\partial A_2}{\partial x} - \frac{\partial A_1}{\partial y} \right) dx\, dy.$$

Comparing this with the example we gave in discussing Volterra's work, it is clear that Volterra's M_{23}, M_{31}, and M_{12} are precisely the coefficients of Cartan's ω'.

Cartan in **[2]** did not discuss the relationship of his differential expressions to Stokes' theorem; nevertheless, by the early years of the twentieth century the generalized Stokes' theorem in essentially the form given by Poincaré was known and used by many authors, although proofs seem not to have been published.

By 1922, Cartan had extended his work on differential expressions in **[3]**. It is here that he first uses the current terminology of "exterior differential form" and "exterior derivative." He works out specifically the derivative of a 1-form (as we did above) and notes that for $n=3$ Stokes' theorem states that $\int_C \omega = \int\int_S \omega'$ where C is the boundary curve of the surface S. (This is, of course, exactly Volterra's result in the same special case.) Then, defining the exterior derivative of any differential form $\omega = \Sigma A\, dx_i dx_j \ldots dx_l$ to be $\omega' = \Sigma dA\, dx_i dx_j \ldots dx_l$ (with dA as above), he works out the derivative of a 2-form Ω in the special case $n=3$ and shows that for a parallelepiped P with boundary S, $\int\int_S \Omega = \int\int\int_P \Omega'$. One can easily calculate that this is the divergence theorem, and we must assume that Cartan realized its truth in more general cases. He was, however, not yet ready to state the most general result.

The "d" notation for exterior derivative was used in 1902 by Theodore DeDonder in **[6]**, but not again until Erich Kähler reintroduced it in his 1934 book *Einführung in die Theorie der Systeme von Differentialgleichungen* **[11]**. His notation is slightly different from ours, but in a form closer to ours it was adopted by Cartan for a course he gave in Paris in 1936–37 (published as *Les Systèmes Différentiels Extérieurs et leurs Applications Géométriques* **[4]** in 1945). Here, after discussing the definitions of the differential form ω and its derivative $d\omega$, Cartan notes that all of our three theorems (which he attributes to Ostrogradsky, Cauchy-Green, and Stokes, respectively) are special cases of $\int_C \omega = \int_A d\omega$ where C is the boundary of A. To be more specific, Green's theorem is the special case where ω is a 1-form in 2-space; Stokes' theorem is the special case where ω is a 1-form in 3-space; and the divergence theorem is the special case where ω is a 2-form in 3-space. Finally, Cartan states that for any $p+1$-dimensional domain A with p-dimensional boundary C one could demonstrate the general Stokes' formula:

$$\int_C \omega = \int_A d\omega$$

(For examples of the use of these theorems, see any advanced calculus text, e.g., **[1]** or **[24]**. For more information on differential forms, one can consult **[7]**.)

Appearance in texts

A final interesting point about these theorems is their appearance in textbooks. By the 1890's all three theorems were appearing in the analysis texts of many different authors. The third of our theorems was always attributed to Stokes. The French and Russian authors tended to attribute the first theorem to Ostrogradsky, while others generally attributed it to Green or Gauss; this is still the case today. Similarly, Riemann is generally credited with the second theorem by the French, while Green is named by most others. Before Cartan's 1945 book, about the only author to attribute that result to Cauchy was H. Vogt in **[26]**.

The generalized Stokes' theorem, first published, as we have seen, in 1945, has only been appearing in textbooks in the past twenty years, the first occurrence probably being in the 1959 volume of Nickerson, Spencer, and Steenrod **[14]**.

References

[1] R. C. Buck, Advanced Calculus, 2nd ed., McGraw-Hill, New York, 1965.
[2] E. Cartan, Sur certaines expressions différentielles et sur le problème de Pfaff, Annales École Normale, v. 16, 1899, pp. 239–332; Oeuvres, Part II, Vol. I, pp. 303–397.
[3] E. Cartan, Leçons sur les invariants intégraux, Chap. VII, Hermann, Paris, 1922.
[4] E. Cartan, Les Systèmes Différentiels Extérieurs et leurs Applications Géométriques, Hermann, Paris, 1945.
[5] A Cauchy, Sur les intégrales qui s'étendent à tous les points d'une courbe fermée, Comptes Rendus, v. 23, 1846, pp. 251–55; Oeuvres, lre Série, Tome X, pp. 70–74.
[6] T. DeDonder, Étude sur les invariants intégraux, Rendiconti del Circolo Matematico di Palermo, v. 16, 1902, pp. 155–179.
[7] H. Flanders, Differential Forms, Academic Press, New York, 1963.
[8] C. F. Gauss, Theoria Attractionis Corporum Sphaeroidicorum Ellipticorum Homogeneorum Methodo nova tractata, Commentationes societatis regiae scientiarium Gottingensis recentiores, v. II, 1813; Werke, v. 5, pp. 1–22.
[9] G. Green, An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism, London, 1828; Green's Mathematical Papers, pp. 3–115.
[10] H. Hankel, Zur allgemeinen Theorie der Bewegung der Flüssigkeiten, Dieterische Univ. Buchdruckerei, Göttingen, 1861.
[11] E. Kähler, Einführung in die Theorie der Systeme von Differentialgleichungen, Leipzig, 1934.
[12] C. Maxwell, Letter to Stokes in G. Stokes, Memoir and Scientific Correspondence, ed. by J. Larmor, Cambridge, 1907, v. II, p. 31.
[13] C. Maxwell, A Treatise on Electricity and Magnetism, Oxford, 1873, v. I.
[14] H. Nickerson, D. Spencer, N. Steenrod, Advanced Calculus, Van Nostrand, Princeton, 1959.
[15] M. Ostrogradsky, Démonstration d'un théorème du calcul intégral, pub. in Russian in Istoriko-Matematicheskie Issledovania, v. XVI, 1965, pp. 49–96.
[16] M. Ostrogradsky, Note sur la Théorie de la Chaleur, Mémoires de L'Acad. Imp. des Sciences de St. Petersburg, ser. 6, v. 1, 1831, pp. 129–133.
[17] M. Ostrogradsky, Sur le calcul des variations des intégrales multiples, Journal fur die Reine und angewandte Mathematik, v. XV, 1836, pp. 332–354.
[18] H. Poincaré, Les Méthodes Nouvelles de la Mécaniques Céleste, v. III, p. 10, Gauthier-Villars, Paris, 1899, Dover, N.Y. 1957.
[19] S. Poisson, Mémoire sur l'Équilibre et le Mouvement des Corps Élastiques, Mémoires de l'Académie Royale des Sciences de l'Institut de France, v. VIII, 1829, pp. 357–571.
[20] B Riemann, Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse, Gottingen, 1851; Werke, pp. 3–48.
[21] F. Sarrus, Mémoire sur les oscillations des corps flottans, Annales de mathématiques pures et appliquées (Nismes), v. XIX, 1828, pp. 185-211.
[22] G. Stokes, Mathematical and Physical Papers, v. 5, p. 320, Cambridge Univ. Press, Cambridge, 1905.
[23] P. Tait, On Green's and other allied Theorems, Transactions of the Royal Society of Edinburgh, 1869–70, pp. 69–84.
[24] A. Taylor, W. Mann, Advanced Calculus, 2nd. ed., Xerox, Lexington, 1972.
[25] W. Thomson, P. Tait, Treatise on Natural Philosophy, Vol. 1, Oxford, 1867.
[26] H. Vogt, Éléments de Mathématiques Supérieures, Paris, 1925.
[27] V. Volterra, Delle Variabili Complesse Negli Iperspazi, Rendiconti della R. Accad. der Lincei, ser. IV, v. V, 1889, pp. 158–165; Opere, v. I, pp. 403–411.
[28] A. Yushkevich, Istoriya Matematiki v Rossii do 1917 goda, Moscow, 1968, p. 290.

Matthew effect (sociology)

From Wikipedia, the free encyclopedia

In sociology, the **Matthew effect** (or **accumulated advantage**) is the phenomenon where "the rich get richer and the poor get poorer".[1][2] Those who possess power and economic or social capital can leverage those resources to gain more power or capital. The term was first coined by sociologist Robert K. Merton in 1968 and takes its name from a line in the biblical Gospel of Matthew:

> For to all those who have, more will be given, and they will have an abundance; but from those who have nothing, even what they have will be taken away.
> —Matthew 25:29, New Revised Standard Version.

Contents [hide]

Sociology of science [edit]

In the sociology of science, "Matthew effect" was a term coined by Robert K. Merton to describe how, among other things, eminent scientists will often get more credit than a comparatively unknown researcher, even if their work is similar; it also means that credit will usually be given to researchers who are already famous.[3][4] For example, a prize will almost always be awarded to the most senior researcher involved in a project, even if all the work was done by a graduate student.

Could Feynman Have Said This?

N. David Mermin

Fifteen years ago, I mused in a Reference Frame column on how different generations of physicists differed in the degree to which they thought that the interpretation of quantum mechanics remains a serious problem (PHYSICS TODAY, April 1989, page 9). I declared myself to be among those who feel uncomfortable when asked to articulate what we really think about the quantum theory, adding that "If I were forced to sum up in one sentence what the Copenhagen interpretation says to me, it would be 'Shut up and calculate!'"

In the intervening years, I've come to hold a milder and more nuanced opinion of the Copenhagen view, but that should be the subject of another column. The subject of this one is the habit of misquotation or misattribution that afflicts our profession, a rather different example of which I pointed out a few months ago (February 2004, page 10).

Given my capacity for intellectual development ("inconsistency," in the terminology deployed in the current American political season), it's fortunate that I've now reached an age at which I tend to forget about things I've written more than a few years ago. Indeed, I find it downright irritating when somebody asks me questions about papers I wrote a mere half dozen years ago, naively identifying me with the author of those ancient texts. Until quite recently, I had no memory of ever having written such a childishly brusque dismissal of such an exquisitely subtle point of view, much less of having published it in so widely read a venue.

This amnesia, combined with the evolution in my thinking that had distanced me from my long forgotten words, may explain why I was initially somewhat puzzled by the slight sensation of discomfort that passed over me when, browsing the e-print archive earlier this year, I read a characterization of Max Born's probability rule as "the favorite ingredient of what has been nicknamed, after Feynman's famous dictum, the shut up and calculate interpretation of quantum mechanics."

I yield to nobody in my admiration for Richard Feynman's aphorisms on the nature of quantum mechanics. Indeed, long ago I published a poem (PHYSICS TODAY, April 1985, page 47) consisting of nothing more than a resetting as verse of a paragraph Feynman had written about his own attitude toward the quantum theory, in his now (but not then) famous article that launched the whole field of quantum computation. I like to think I have devoured everything Feynman ever wrote on the character of quantum mechanics.

But while "shut up and calculate" sounded dimly familiar to me as a characterization of a certain interpretive stance, I couldn't recall where Feynman had written it. Mulling this over, a terrible thought began to dawn on me. Could it be that I myself had once used the phrase? If so, then it would appear that I had picked it up from something by Feynman, forgotten the source, and presented it as my own. Devastating!

It was devastating because I have a horror of writing or uttering any witticism that is not original with me, unless I make it absolutely clear where and (if known) from whom I got it. I don't even like to tell jokes unless I've made them up myself.

(I digress to offer you my favorite:

> Question: What is the difference between theoretical physics and mathematical physics?
>
> Answer: Theoretical physics is done by physicists who lack the necessary skills to do real experiments; mathematical physics is done by mathematicians who lack the necessary skills to do real mathematics.

Mathematical physicists tend not to like this joke, but other physicists seem to. Nonphysicists, of course, are entirely immune to its charms.)

So with growing trepidation, I searched through my past writings on quantum mechanics. I was dismayed when I came upon my 1989 column, which confirmed my worst fears. Not only had I appropriated without proper attribution a Feynman quote, but it appeared to be a famous one. How humiliating! I was afflicted with visions of knowledgeable PHYSICS TODAY readers shaking their heads 15 years ago at what must have struck them as my shameless attempt to ride to literary glory on the unacknowledged shoulders of Feynman.

So I went to the Web to find the source, hoping I could then salvage my reputation by persuading PHYSICS TODAY to print an addendum or erratum. Google gives more than 130 hits containing both "shut up and calculate" and "feynman." Most of these do not directly link the two, but about a dozen do. Here are a few:

> "Shut up and calculate" was a motto of Richard Feynman.
>
> For example, there's Feynman's "shut up and calculate."
>
> My personal philosophy is that of the famous physicist Richard Feynman, who said: "Shut up and calculate."
>
> When asked which interpretation of QM he favored, Feynman replied "Shut up and calculate."
>
> Richard Feynman foreslog ligefrem en "shut up and calculate" fortolkning af kvantemekanikken.
>
> Shut up and calculate—Richard Feynman.

Just to make sure, I also searched for "shut up and calculate" and "mermin." I found only 10 hits, all of them mentioning me in ways that had nothing to do with their use, elsewhere, of "shut up and calculate." So it would have been clear to the world that I had

N. David Mermin *(ndm4@cornell.edu) is a semi-retired professor of physics at Cornell University, where he nervously awaits your evidence that Feynman said it.*

indeed passed off Feynman's words as my own.

Or had I . . . ?

I noticed that not a single one of the Web sites attributing the phrase to Feynman cited a source or hinted at the circumstances under which he had said it. A ray of hope flickered on: Could I once again have become a victim of the Matthew effect?

The Matthew effect was enunciated by the great sociologist of science, Robert Merton.[1] Merton worked in those innocent days when sociologists were interested only in the behavior of scientists and not in the content of their science. (To be fair to contemporary sociologists of science, I should modify that last phrase to "and not in the manifestations of that behavior in the content of their science.") I first learned of the Matthew effect more than 20 years ago, on the occasion of my first and, perhaps until now, only, victimization at the hands of the *New York Times*.

I learned the name for what the *Times* had done to me when I received a very nice note from P. W. Anderson in which he expressed his regret that the newspaper had given him exclusive credit for a nomenclatural advance that was entirely due to me. "A depressingly typical example of the Matthew effect" was how he characterized the misattribution. (I reported the entire history of this contretemps in these pages back in those dark ages [April 1981, page 46] before there were Reference Frame columns.) When I wrote back asking him what the Matthew effect was, he referred me to Merton.

It was Merton who identified and named the tendency always to assign exclusive scientific credit to the most eminent among all the plausible candidates. At least I hope it was he, though I'm sure Merton, who invented many wonderful jokes himself, would have been delighted if the credit for it turned out to be misattributed to him. Merton named the effect after the Gospel According to Matthew, because there it is written,

> For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath.
>
> —Matthew 25:29.

Could the widespread attribution of my wretched witticism to Feynman be another instance of this same deplorable practice? Had I once again been matthewed?

Although I didn't say so in my old Reference-Frame article, what inspired this not so terribly bon mot were vivid memories of the responses my conceptual inquiries elicited from my professors—whom I viewed as agents of Copenhagen—when I was first learning quantum mechanics as a graduate student at Harvard, a mere 30 years after the birth of the subject. "You'll never get a PhD if you allow yourself to be distracted by such frivolities," they kept advising me, "so get back to serious business and produce some results." "Shut up," in other words, "and calculate." And so I did, and probably turned out much the better for it. At Harvard, they knew how to administer tough love in those olden days.

This bit of history is relevant to the question of whether Feynman's abundance might have been augmented by a portion of the little that I had. Can you imagine the young Feynman ever having had a similar experience that seared "shut up and calculate" into his tender consciousness? No, of course you can't! Nobody could ever have had the slightest reason to direct the best human calculator that ever was to shut up and calculate.

But perhaps Feynman was offering such advice to others who were searching for a better understanding of the quantum mechanical formalism. I can't believe that. He said that he "always had a great deal of difficulty understanding the world view that quantum mechanics represents," and added, "I still get nervous with it."[2] Nobody who felt that way would ever respond with "shut up and calculate" to conceptual inquiries from the perplexed.

Well maybe Feynman, like me, was merely dismissing an interpretive position of others by lampooning it as a "shut up and calculate interpretation." I find this unlikely. For one thing, his strong preference for working things out for himself and, of course, his well-known disdain for philosophy make me doubt that he ever paid much attention to the interpretive positions of others. For another, would one for whom calculation was so effortless and understanding so important be likely to translate anybody's admonition against fruitless speculation into such terms?

In short, I suspect that it is only Feynman's habitual irreverence that has linked him in the minds of many to the phrase "shut up and calculate." Who else among the high and mighty—and Merton has taught us that it is only among the high and mighty that people tend to look—could have said it? Albert Einstein? Don't be silly. Erwin Schrödinger? Of course not. Niels Bohr? Don't make me laugh. None of them besides Feynman could have said it. Does that mean that Feynman said it? No!

Broaden the search to embrace the low and powerless. Among them am I, who hereby put forth the hypothesis that I was the first to use "shut up and calculate" in the context of quantum foundations. I'm not proud of having said it. It's not a beautiful phrase. It's not very clever. It's snide and mindlessly dismissive. But, damn it, if I'm the one who said it first, then that means I did not, even unconsciously, appropriate the words of Richard Feynman and pass them off as my own. So I have nothing to be ashamed of other than having characterized the Copenhagen interpretation in such foolish terms—a lesser offense than unconscious plagiarism, in my moral bookkeeping.

So, dear reader, if you have evidence that Feynman really did say "shut up and calculate," please send it to me. I will not be happy to receive it. I'd rather be a Matthew victim than a plagiarist. But I'd like to know the truth.

References

1. R. K. Merton, *Science* **159**, 56 (1968).
2. R. P. Feynman, *Int. J. Theor. Phys.* **21**, 471 (1982). ■

6.10 The generalized Stokes's theorem

We worked hard to define the exterior derivative and to define orientation of manifolds and of boundaries. Now we are going to reap some rewards for our labor: we are going to see that there is a higher-dimensional analogue of the fundamental theorem of calculus, Stokes's theorem. It covers in one statement the four integral theorems of vector calculus, which are explored in section 6.11.

Recall the fundamental theorem of calculus:

Theorem 6.10.1 (Fundamental theorem of calculus). *If f is a C^1 function on a neighborhood of $[a,b]$, then*

$$\int_a^b f'(t)\,dt = f(b) - f(a). \qquad 6.10.1$$

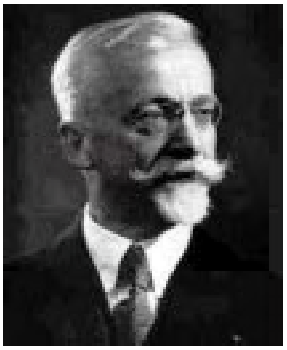


Figure 6.10.1.
Elie Cartan (1869–1951) formalized the theory of differential forms in the early twentieth century. Other names associated with the generalized Stokes's theorem include Henri Poincaré, Vito Volterra, and Luitzen Brouwer.

One of Cartan's four children, Henri, became a renowned mathematician; another, a physicist, was arrested by the Germans in 1942 and executed 15 months later.

Theorem 6.10.2 is probably the best tool mathematicians have for deducing global properties from local properties. It is a wonderful theorem.

It is often called the generalized Stokes's theorem, to distinguish it from the special case (surfaces in $\mathbb{R}^3$) also known as Stokes's theorem. Special cases of the generalized Stokes's theorem are discussed in section 6.11.

To lighten notation, in theorem 6.10.2 we write ∂X. However, we are actually integrating φ over $\partial_M^s X$, the smooth part of the boundary that sets off $X \subset M$ from M.

Restate this as

$$\int_{[a,b]} df = \int_{\partial[a,b]} f, \qquad 6.10.2$$

i.e., the integral of df over an oriented interval is equal to the integral of f over the oriented boundary of the interval. In this form, the statement generalizes to higher dimensions:

Theorem 6.10.2 (Generalized Stokes's theorem). *Let X be a compact piece-with-boundary of a $(k+1)$-dimensional oriented manifold $M \subset \mathbb{R}^n$. Give the boundary ∂X of X the boundary orientation, and let φ be a k-form defined on an open set containing X. Then*

$$\int_{\partial X} \varphi = \int_X d\varphi. \qquad 6.10.3$$

This beautiful, short statement is the main result of the theory of forms. Note that the dimensions in equation 6.10.3 make sense: if X is $(k+1)$-dimensional, ∂X is k-dimensional, and if φ is a k form, $d\varphi$ is a $(k+1)$-form, so $d\varphi$ can be integrated over X, and φ can be integrated over ∂X.

Example 6.10.3 (Integrating over the boundary of a square). You apply Stokes's theorem every time you use antiderivatives to compute an integral: to compute the integral of the 1-form $f\,dx$ over the oriented line segment $[a,b]$, you begin by finding a function g such that $dg = f\,dx$, and then say

$$\int_a^b f\,dx = \int_{[a,b]} dg = \int_{\partial[a,b]} g = g(b) - g(a). \qquad 6.10.4$$

This isn't quite the way Stokes's theorem is usually used in higher dimensions, where "looking for antiderivatives" has a different flavor.

Michael Spivak

CALCULUS ON MANIFOLDS

The letter from Thomson to Stokes on the cover of Spivak's Calculus on Manifolds

63 KELVIN to STOKES, 2 July 1850
Stokes Collection, K39

9 Barton Street,
Westminster
July 2, 1850

MY DEAR STOKES

As I have not a copy of your paper on the Equil. & Motion of Elastic Solids,[1] nor any other work of reference for the purpose, by me, I shall be much obliged by your sending me the *equations of equil*^{m} of ⟨an⟩ a non crystalline elastic solid under the action of any forces, and the formulae for the mutual actions betw. any two contiguous portions of the body. I have been trying but as yet without success, to make out something about the integration of the equations for the case of a solid of any form, with each point of its surface displaced to a given extent & in a given direction from its natural position. I think I see how it can be done when the solid is a rectangular parallelepiped, but not in a very inviting way. It was reading your paper on diffraction[2] on my way from Cambridge that made me take up the subject again.

Do you know that the condition that $\alpha dx + \beta dy + \gamma dz$ may be the diff[l] of a function of two indep[t] variables for all points of a surface is

$$l\left(\frac{d\beta}{dz}-\frac{d\gamma}{dy}\right)+m\left(\frac{d\gamma}{dx}-\frac{d\alpha}{dz}\right)+n\left(\frac{d\alpha}{dy}-\frac{d\beta}{dx}\right)=0?$$

I made this out some weeks ago with ref[ce] to electromagnetism. With ref[ce] to an elastic solid, the cond[n] may be expressed thus – the resultant axis of rotation at any point of the surface must be perp[r] to the normal.

Your's very truly
WILLIAM THOMSON

P.S. The following is also interesting, & is of importance with reference to both physical subjects.

$$\int(\alpha dx+\beta dy+\gamma dz)=\pm\iint\left\{l\left(\frac{d\beta}{dz}-\frac{d\gamma}{dy}\right)+m\left(\frac{d\gamma}{dx}-\frac{d\alpha}{dz}\right)+n\left(\frac{d\alpha}{dy}-\frac{d\beta}{dx}\right)\right\}dS$$

where l, m, n denote the dir[n] cosines of a normal through any el[t] dS of a (limited) surface; & the integ[n] in the sec[d] member is performed over a portion of this surface bounded by a curve round w[h] the int[n] in the 1[st] member is performed.[3]

1 Stokes (11).
2 Stokes (40).
3 Stokes included the equation in this postscript on the Smith's prize examination for 1854 (the year Maxwell took the examination), and it has become known as Stokes's Theorem. (See Larmor's footnote in Stokes's *MPP*, v, 320–1.)

The Preface to Spivak's Calculus on Manifolds

Preface

This little book is especially concerned with those portions of "advanced calculus" in which the subtlety of the concepts and methods makes rigor difficult to attain at an elementary level. The approach taken here uses elementary versions of modern methods found in sophisticated mathematics. The formal prerequisites include only a term of linear algebra, a nodding acquaintance with the notation of set theory, and a respectable first-year calculus course (one which at least mentions the least upper bound (sup) and greatest lower bound (inf) of a set of real numbers). Beyond this a certain (perhaps latent) rapport with abstract mathematics will be found almost essential.

The first half of the book covers that simple part of advanced calculus which generalizes elementary calculus to higher dimensions. Chapter 1 contains preliminaries, and Chapters 2 and 3 treat differentiation and integration.

The remainder of the book is devoted to the study of curves, surfaces, and higher-dimensional analogues. Here the modern and classical treatments pursue quite different routes; there are, of course, many points of contact, and a significant encounter occurs in the last section. The very classical equation reproduced on the cover appears also as the last theorem of the book. This theorem (Stokes' Theorem) has had a curious history and has undergone a striking metamorphosis.

The first statement of the Theorem appears as a postscript to a letter, dated July 2, 1850, from Sir William Thomson (Lord Kelvin) to Stokes. It appeared publicly as question 8 on the Smith's Prize Examination for 1854. This competitive examination, which was taken annually by the best mathematics students at Cambridge University, was set from 1849 to 1882 by Professor Stokes; by the time of his death the result was known universally as Stokes' Theorem. At least three proofs were given by his contemporaries: Thomson published one, another appeared in Thomson and Tait's *Treatise on Natural Philosophy*, and Maxwell provided another in *Electricity and Magnetism* [13]. Since this time the name of Stokes has been applied to much more general results, which have figured so prominently in the development of certain parts of mathematics that Stokes' Theorem may be considered a case study in the value of generalization.

In this book there are three forms of Stokes' Theorem. The version known to Stokes appears in the last section, along with its inseparable companions, Green's Theorem and the Divergence Theorem. These three theorems, the classical theorems of the subtitle, are derived quite easily from a modern Stokes' Theorem which appears earlier in Chapter 5. What the classical theorems state for curves and surfaces, this theorem states for the higher-dimensional analogues (manifolds) which are studied thoroughly in the first part of Chapter 5. This study of manifolds, which could be justified solely on the basis of their importance in modern mathematics, actually involves no more effort than a careful study of curves and surfaces alone would require.

The reader probably suspects that the modern Stokes' Theorem is at least as difficult as the classical theorems derived from it. On the contrary, it is a very simple consequence of yet another version of Stokes' Theorem; this very abstract version is the final and main result of Chapter 4.

It is entirely reasonable to suppose that the difficulties so far avoided must be hidden here. Yet the proof of this theorem is, in the mathematician's sense, an utter triviality—a straightforward computation. On the other hand, even the statement of this triviality cannot be understood without a horde of difficult definitions from Chapter 4. There are good reasons why the theorems should all be easy and the definitions hard. As the evolution of Stokes' Theorem revealed, a single simple principle can masquerade as several difficult results; the proofs of many theorems involve merely stripping away the disguise. The definitions, on the other hand, serve a twofold purpose: they are rigorous replacements for vague notions, and machinery for elegant proofs. The first two sections of Chapter 4 define precisely, and prove the rules for manipulating, what are classically described as "expressions of the form" $P\,dx + Q\,dy + R\,dz$, or $P\,dx\,dy + Q\,dy\,dz + R\,dz\,dx$. Chains, defined in the third section, and partitions of unity (already introduced in Chapter 3) free our proofs from the necessity of chopping manifolds up into small pieces; they reduce questions about manifolds, where everything seems hard, to questions about Euclidean space, where everything is easy.

Concentrating the depth of a subject in the definitions is undeniably economical, but it is bound to produce some difficulties for the student. I hope the reader will be encouraged to learn Chapter 4 thoroughly by the assurance that the results will justify the effort: the classical theorems of the last section represent only a few, and by no means the most important, applications of Chapter 4; many others appear as problems, and further developments will be found by exploring the bibliography.

The problems and the bibliography both deserve a few words. Problems appear after every section and are numbered (like the theorems) within chapters. I have starred those problems whose results are used in the text, but this precaution should be unnecessary—the problems are the most important part of the book, and the reader should at least attempt them all. It was necessary to make the bibliography either very incomplete or unwieldy, since half the major branches of mathematics could legitimately be recommended as reasonable continuations of the material in the book. I have tried to make it incomplete but tempting.

Many criticisms and suggestions were offered during the writing of this book. I am particularly grateful to Richard Palais, Hugo Rossi, Robert Seeley, and Charles Stenard for their many helpful comments.

I have used this printing as an opportunity to correct many misprints and minor errors pointed out to me by indulgent readers. In addition, the material following Theorem 3-11 has been completely revised and corrected. Other important changes, which could not be incorporated in the text without excessive alteration, are listed in the Addenda at the end of the book.

Michael Spivak

Waltham, Massachusetts
March 1968

Suppose now that ω is an arbitrary k-form on M. There is an open cover $\mathcal{O}$ of M such that for each $U \in \mathcal{O}$ there is an orientation-preserving singular k-cube c with $U \subset c([0,1]^k)$. Let Φ be a partition of unity for M subordinate to this cover. We define

$$\int_M \omega = \sum_{\varphi \in \Phi} \int_M \varphi \cdot \omega$$

provided the sum converges as described in the discussion preceding Theorem 3-12 (this is certainly true if M is compact). An argument similar to that in Theorem 3-12 shows that $\int_M \omega$ does not depend on the cover $\mathcal{O}$ or on Φ.

All our definitions could have been given for a k-dimensional manifold-with-boundary M with orientation μ. Let ∂M have the induced orientation $\partial\mu$. Let c be an orientation-preserving k-cube in M such that $c_{(k,0)}$ lies in ∂M and is the only face which has any interior points in ∂M. As the remarks after the definition of $\partial\mu$ show, $c_{(k,0)}$ is orientation-preserving if k is even, but not if k is odd. Thus, if ω is a $(k-1)$-form on M which is 0 outside of $c([0,1]^k)$, we have

$$\int_{c_{(k,0)}} \omega = (-1)^k \int_{\partial M} \omega.$$

On the other hand, $c_{(k,0)}$ appears with coefficient $(-1)^k$ in ∂c. Therefore

$$\int_{\partial c} \omega = \int_{(-1)^k c_{(k,0)}} \omega = (-1)^k \int_{c_{(k,0)}} \omega = \int_{\partial M} \omega.$$

Our choice of $\partial\mu$ was made to eliminate any minus signs in this equation, and in the following theorem.

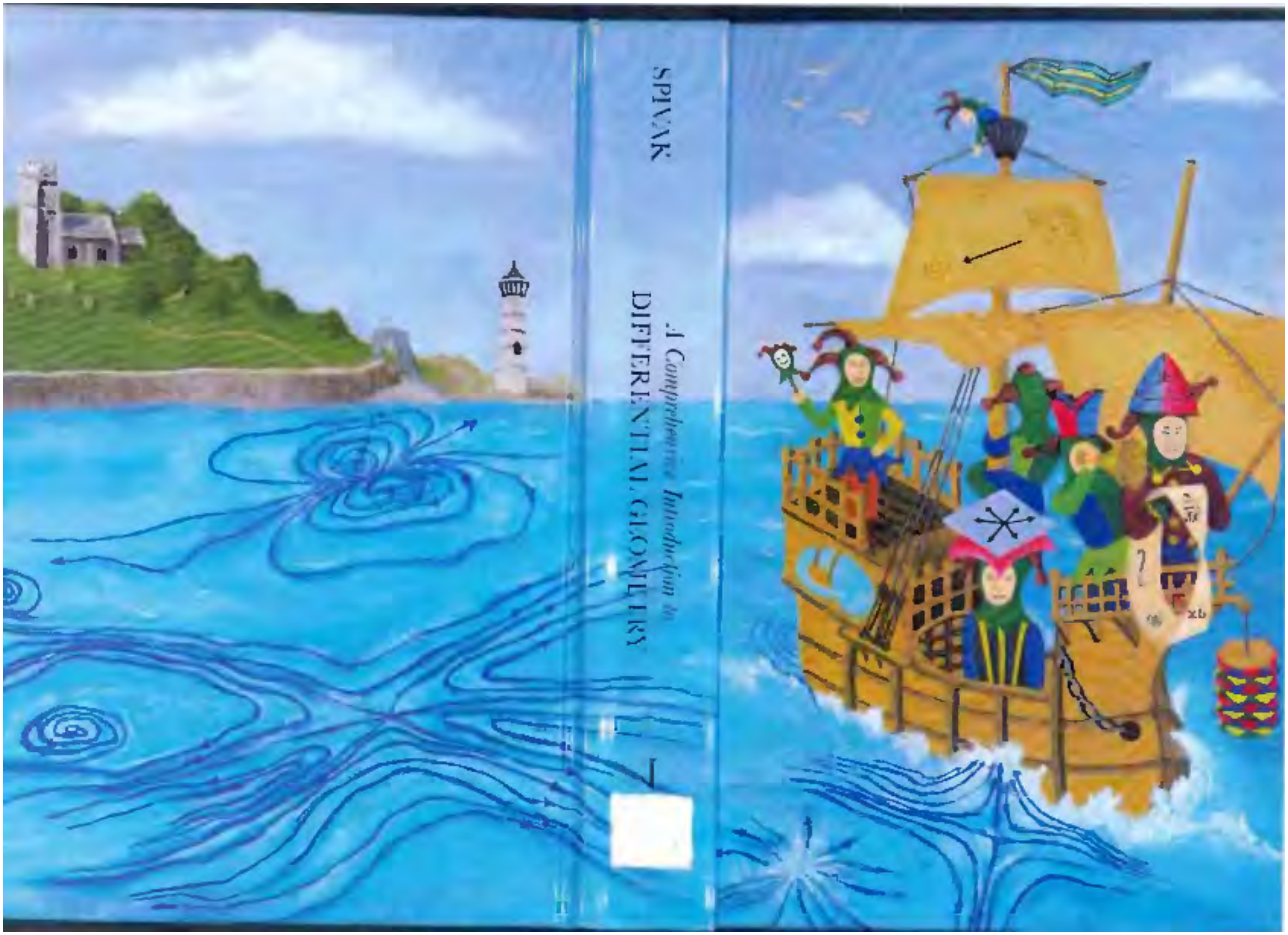
5-5 Theorem (Stokes' Theorem). *If M is a compact oriented k-dimensional manifold-with-boundary and ω is a $(k-1)$-form on M, then*

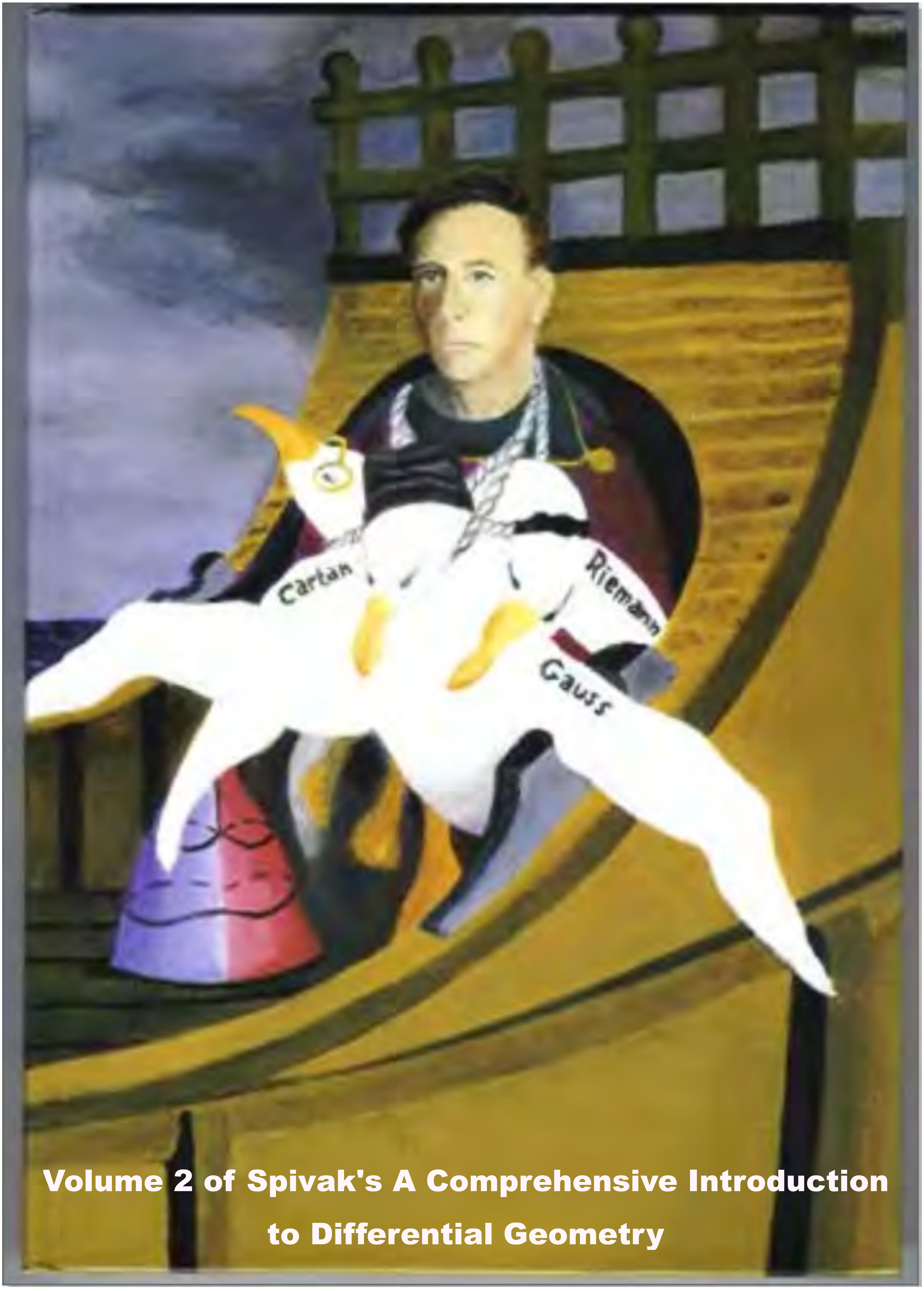
$$\int_M d\omega = \int_{\partial M} \omega.$$

(Here ∂M is given the induced orientation.)

Proof. Suppose first that there is an orientation-preserving singular k-cube in $M - \partial M$ such that $\omega = 0$ outside of

Volume 1 of Spivak's A Comprehensive Introduction to Differential Geometry





Volume 2 of Spivak's A Comprehensive Introduction to Differential Geometry

QUESTIONS AND ANSWERS

Contributions to this section, both Questions and Answers, are welcomed. Please submit four copies to the editorial office. Please include a *title* for each submission, include name and address at the end, and put references in the standard format used in the American Journal of Physics. For further suggestions, sample Questions and Answers, and requested form for both Questions and Answers, see Robert H. Romer, "Editorial: 'Questions and Answers,' a new section of the American Journal of Physics," Am. J. Phys. **62** (6) 487–489 (1994).

Questions at any level and on any appropriate AJP topic, including the "quick and curious" question, are encouraged.

Editorial Note on Answers to Question #55. Are there pictorial examples that distinguish covariant and contravariant vectors?

We reviewed over a dozen responses to Neuenschwander's question[1] that were sent in by readers. Although many of them make similar and overlapping points, we have chosen to print the following three answers as ones which cover most clearly the several aspects of the question posed. Napolitano and Lichtenstein,[2] as well as Schmidt,[3] make the point that at a conceptual level, covariant and contravariant vectors are different kinds of geometric objects, but, given a metric, there is a natural identification between them. As a result, in a space with a metric, one may speak of covariant and contravariant components of either type of vector.

While the conceptual distinction between a (contravariant) vector and a co(variant)-vector is sometimes important, because physical situations almost always involve metric spaces, and because our intuitions are so deeply encoded with notions of distances and angles, it is actually harder than one might first suppose to communicate to students why one must distinguish the two types of geometric objects. The gradient of a function (the prototype of a covector), and the flow velocity of a particle [the prototype of a (contravariant) vector] are good places to start. With the oblique, rectilinear axes in the flat plane, and with the usual metric, one may be able to provide the student some intuition in this regard. For example, if one defines a function on this plane by $f_1(x^1,x^2)=x^1$, then the lines of constant f_1 are parallel to the x^2 axis, and the gradient "vector" is perpendicular to that axis. It is easy to imagine extending this type of example to get at the more general conceptual distinction between co- and contra-variant vectors.

Neuenschwander's question also asked about pictorial illustrations. The most popular (and appropriate) response we got involved the use of oblique axes in the Euclidean plane with the usual metric. As Evans's[4] answer points out, this is a good starting example for clarifying the distinction between co- and contra-variant components. However, all the answers we got were either incomplete or too cryptic regarding the full set of circumstances where such distinctions may be usefully maintained. The metric, in the case of flat space with oblique axes, is nondiagonal. But an orthogonal, curvilinear co-ordinate system (polar co-ordinates, for example) in flat space also gives rise to a distinction between covariant and contravariant vector components. Indeed, as discussed in Mary Boas's popular text, in such cases "any vector has three kinds of components: contravariant, covariant and what we might call ordinary components."[5] Here, the metric is diagonal, but is not the identity matrix as it would be in orthogonal, rectilinear co-ordinates (aka Cartesian co-ordinates). Of course, when the space is not flat, but smoothly curved, since it is impossible to introduce Cartesian co-ordinates in a whole neighborhood, the distinction between the two kinds of vector components becomes mandatory.

[1]Dwight E. Neuenschwander, "Question #55. Are there pictorial examples that distinguish covariant and contravariant vectors?," Am. J. Phys. **65** (1), 11 (1997).

[2]J. Napolitano and R. Lichtenstein, "Answer to Question #55," Am. J. Phys. **65** (11), 1037–1038 (1997).

[3]Hans-Jürgen Schmidt, "Answer to Question #55," Am. J. Phys. **65** (11), 1038 (1997).

[4]James Evans, "Answer to Question #55," Am. J. Phys. **65** (11), 1039 (1997).

[5]Mary L. Boas, *Mathematical Methods in the Physical Sciences* (Wiley, New York, 1983), 2nd ed., Chap. 10, Sec. 13, pp. 447–449.

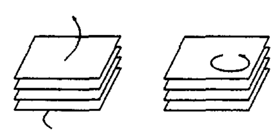
Kannan Jagannathan
Assistant Editor
American Journal of Physics
Amherst College
P.O. Box 2262
Amherst, Massachusetts 01002

Answer to Question #55. Are there pictorial examples that distinguish covariant and contravariant vectors?

Pictorial examples as called for by Neuenschwander's question[1] indeed provide a very useful device for making the distinction between covectors (i.e., covariant vectors) and vectors (i.e., contravariant vectors), and they are used in some textbooks. For example, see the popular textbook by Misner, Thorne, and Wheeler.[2] We prefer using examples that are a bit tongue-in-cheek.

Consider first a vector. This is, in fact, the object most familiar to students, which is drawn simply as a "stick" (e.g., "→"). We refer to such an object as a "stick vector." One example is the displacement vector between two points in coordinate space. The magnitude of a stick vector is simply proportional to the length of the stick drawn on the blackboard. The arrow demonstrates the sense of direction of the stick vector.

Now consider a covector. This should be familiar to most students in terms of a gradient. We can picture a gradient best in terms of the equipotential surfaces to which it refers, and this is the basis of the pictorial representation. That is, draw the surfaces themselves, along with some sense of direction, which might be indicated by a wavy line with an arrow at the end, or with a whorl on one of the sheets:



Note that, in any case, the magnitude of the covector is proportional to the *density* of sheets.

We refer to this pictorial representation of a covector as a "lasagna vector," the sheets reminiscent of the noodles in a pan of lasagna. Students have no trouble remembering this analogy. It is also handy because if there are many noodles packed closely together in the pan, the lasagna is certainly worth more. That is, it has a larger magnitude.

Next we point out that the inner product can only be taken between a stick vector and a lasagna vector, but never between two of the same kind. The inner product is given, pictorially, by placing the stick vector into the pan of lasagna (of course, maintaining its orientation) and counting the number of noodles pierced by the stick. Clearly, the value of the inner product is both proportional to the length of the stick (i.e., the magnitude of the vector) and the density of the lasagna noodles (i.e., the magnitude of the covector).

Now a student may ask, as happened in our class, why there is any distinction between vectors and covectors since one can easily draw stick vectors for gradients by attaching them at right angles to the equipotentials, or contour lines in two dimensions. This is an excellent question and it strikes to the heart of the meaning of the metric tensor. Given our representation of vector and covector, there is not yet any way to define an angle!

An angle is defined by the inner product between two stick vectors, but as we have defined it, this operation is not possible. We need some mechanism for turning a stick vector into a lasagna vector. We would then take the inner product between this transformed stick vector and the untransformed one that remains. The object (or function) which maps a stick vector into the corresponding lasagna vector is called the metric tensor.

[1]Dwight E. Neuenschwander, "Question #55. Are there pictorial examples that distinguish covariant and contravariant vectors?," Am. J. Phys. **65** (1), 11 (1997).

[2]Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, *Gravitation* (W. H. Freeman and Company, San Francisco, 1973), pp. 53–59, with examples throughout the book.

J. Napolitano and R. Lichtenstein
Department of Physics, Applied Physics, and Astronomy
Rensselaer Polytechnic Institute
110 8th Street
Troy, New York 12180-3590

Answer to Question #55. Are there pictorial examples that distinguish covariant and contravariant vectors?

Neuenschwander[1] asked how to visualize the distinction between co- and contravariant vectors. Most textbooks introduce this distinction on an abstract level; the only exception I know is that of Stephani,[2] and below I will show how I present it in my lectures "Introduction to Differential Geometry" at Potsdam University.

If *no metric exists* at all, then covariant vectors and contravariant vectors are different types of objects.

If *a metric exists*, then there is a canonical isomorphism between them; so we introduce *vectors*, and after fixing a coordinate system, we speak about their covariant and their contravariant components.

In the following, we will deal with the second case only, because it is easier to visualize: The chalkboard has a canonical metric which makes it a flat two-dimensional Riemannian manifold.

Neuenschwander[1] wrote that the mentioned distinction is necessary when dealing with curved spaces. This is not wrong, but it is a little bit misleading, and I prefer to say: "... is necessary when dealing with a non-Cartesian coordinate system." Example: We fix a point (the "origin" O) in the Euclidean plane; then there is a one-to-one correspondence between points and vectors. (The point P is related to the vector $\overline{OP}$.) First, we use rectangular coordinates. We might call them x and y; however, as we are interested in seeing how the situation is changed by introducing nonrectangular coordinates, we call them x^i with $i \in \{1,2\}$. So the point P has coordinates (x^1, x^2); cf. diagram 1.

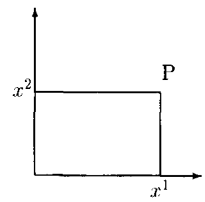


diagram 1

The coordinate system is a rectangular one, and so the component x^1 can be equivalently described as the perpendicular projection to the x^1 axis or as the projection parallel to the x^2 axis.

Let us now consider the case of an inclined system (see diagram 2). Let the angle between the axes be α with

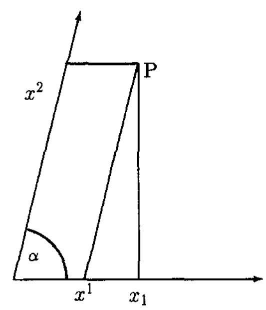


diagram 2

$0 < \alpha < \pi$.

Here, x^1 is the projection parallel to the x^2 axis, and x_1 is the perpendicular projection to the x^1 axis. We get $x_1 = x^1 + x^2 \cos \alpha$, i.e., $x_1 = x^1$ if and only if $\alpha = \pi/2$. In general, we get the following linear relation:

$$x_i = g_{ij} x^j$$

by the use of the metric g_{ij}, where $g_{12} = g_{21} = \cos \alpha$, $g_{11} = g_{22} = 1$, and summation over $j \in \{1,2\}$ is automatically assumed.

[1]D. Neuenschwander, Am. J. Phys. **65** (1), 11 (1997).

[2]H. Stephani, *General Relativity* (Cambridge U. P., Cambridge, England, 1990), 2nd ed., p. 26. (In the first German edition, which appeared in Berlin in 1997, this distinction is on p. 35.)

Hans-Jürgen Schmidt
Universität Potsdam, Institut f. Mathematik
D-14415 POTSDAM, PF 601553,
Am Neuen Palais 10, Germany

Answer to Question #55. Are there pictorial examples that distinguish covariant and contravariant vectors?

It is important to distinguish[1] between covariant and contravariant components of a vector whenever we deal with nondiagonal metric tensors, in fact, whenever the metric tensor is not the identity matrix. It is possible to construct informative pictorial examples even in the context of Euclidean plane geometry, if we choose to work with nonorthogonal axes.

Consider two axes, x and y, inclined at an angle α. An arbitrary vector $\mathbf{V}$ in the plane may be resolved into x and y components. But since the axes are not orthogonal, it is not so clear what we want to mean by "components of a vector." Two different definitions might be sensible. First, we might mean that $\mathbf{V}$ can be written in terms of line segments directed parallel to the unit vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{V}=V^x\mathbf{x}+V^y\mathbf{y}. \quad (1)$$

V^x and V^y are the contravariant components of $\mathbf{V}$. Second, we might mean that we should be able to pick off components by taking inner products of $\mathbf{V}$ with $\mathbf{x}$ and $\mathbf{y}$:

$$V_x=\mathbf{V}\cdot\mathbf{x}=V\cos\theta_x, \quad V_y=\mathbf{V}\cdot\mathbf{y}=V\cos\theta_y \quad (2)$$

(where θ_x and θ_y are the angles $\mathbf{V}$ makes with the x and y axes). V_x and V_y are the covariant components of $\mathbf{V}$. V^x, V^y and V_x, V_y are shown in diagram 1.

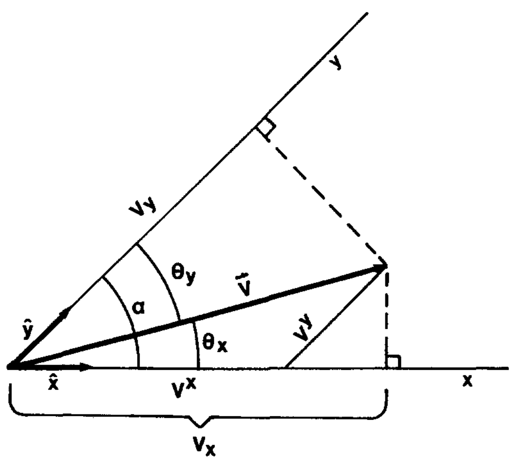


diagram 1

Let us construct the metric tensor. Taking the inner product of $\mathbf{x}$ with Eq. (1) we have

$$V_x=V^x+V^y\cos\alpha,$$

since $\mathbf{x}\cdot\mathbf{y}=\cos\alpha$. Similarly, taking the inner product of $\mathbf{y}$ with Eq. (1) gives

$$V_y=V^x\cos\alpha+V^y.$$

Thus the covariant components may be obtained from the contravariant components by the rule

$$V_i=g_{ij}V^j,$$

where i and j can each take on the values x and y, and where the repeated index is summed. The elements of the metric tensor are then

$$g_{ij}=\begin{pmatrix} 1 & \cos\alpha \\ \cos\alpha & 1 \end{pmatrix}.$$

It is easy to show that the inverse transformation is

$$V^i=g^{ij}V_j,$$

with

$$g^{ij}=\begin{pmatrix} \sin^{-2}\alpha & -\cos\alpha\sin^{-2}\alpha \\ -\cos\alpha\sin^{-2}\alpha & \sin^{-2}\alpha \end{pmatrix}.$$

The metric tensor has the usual properties. For example, $V^2=\mathbf{V}\cdot\mathbf{V}$ is an invariant, independent of the angle α between the axes. It may be calculated from

$$V^2=V^iV_i=V^ig_{ij}V^j=V_ig^{ij}V_j.$$

An easy way to see this is to begin from diagram 2,

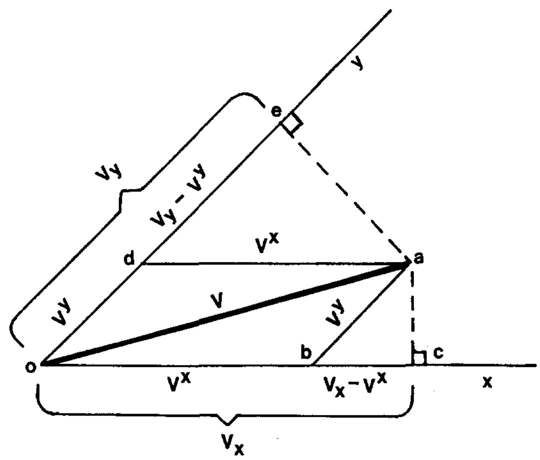


diagram 2

which is merely a more symmetric version of diagram 1. For the sake of having an expression of symmetric form, let us write

$$V^2=\tfrac{1}{2}[(oc)^2+(ca)^2+(oe)^2+(ea)^2].$$

Now,

$$oc=V_x$$

$$(ca)^2=(ab)^2-(bc)^2=(V^y)^2-(V_x-V^x)^2$$

$$oe=V_y$$

$$(ea)^2=(ad)^2-(de)^2=(V^x)^2-(V_y-V^y)^2.$$

With these substitutions, we find immediately

$$V^2=V^xV_x+V^yV_y.$$

Similarly, if $\mathbf{A}$ and $\mathbf{B}$ are two vectors, it may be shown that the invariant inner product (independent of α) is given by

$$\mathbf{A}\cdot\mathbf{B}=AB\cos\theta_{AB}=A^iB_i=A_iB^i,$$

where θ_{AB} is the angle between $\mathbf{A}$ and $\mathbf{B}$.

[1]D. Neuenschwander, Am. J. Phys. **65**(1), 11 (1997).

James Evans
Department of Physics
University of Puget Sound
Tacoma, Washington 98416
JCEvans@UPS.edu

$a = \frac{dH}{dy} - \frac{dG}{dz}$ $b = \frac{dF}{dz} - \frac{dH}{dx}$ $c = \frac{dG}{dx} - \frac{dF}{dy}$	(A)	$\mathbf{B} = \nabla \times \mathbf{A}$
$P = c\frac{dy}{dt} - b\frac{dz}{dt} - \frac{dF}{dt} - \frac{d\psi}{dx}$ $Q = a\frac{dz}{dt} - c\frac{dx}{dt} - \frac{dG}{dt} - \frac{d\psi}{dy}$ $R = b\frac{dx}{dt} - a\frac{dy}{dt} - \frac{dH}{dt} - \frac{d\psi}{dz}$	(B)	$\mathbf{E} = \mathbf{v} \times \mathbf{B} - \frac{\partial \mathbf{A}}{\partial t} - \nabla \phi$
$X = vc - wb$ $Y = wa - uc$ $Z = ub - va$	(C)	$\mathbf{F} = \mathbf{J} \times \mathbf{B}$
$a = \alpha + 4\pi A$ $b = \beta + 4\pi B$ $c = \gamma + 4\pi C$	(D)	$\mathbf{B} = \mu_o \mathbf{H} + \mathbf{M}$
$4\pi u = \frac{d\gamma}{dy} - \frac{d\beta}{dz}$ $4\pi v = \frac{d\alpha}{dz} - \frac{d\gamma}{dx}$ $4\pi w = \frac{d\beta}{dx} - \frac{d\alpha}{dy}$	(E)	$\mathbf{J} = \nabla \times \mathbf{H}$
$\mathfrak{D} = \frac{1}{4\pi} K \mathfrak{E}$	(F)	$\mathbf{D} = \epsilon \mathbf{E}$
$\mathfrak{K} = C\mathfrak{E}$	(G)	$\mathbf{J}_c = \sigma \mathbf{E}$
$\mathfrak{C} = \mathfrak{K} + \dot{\mathfrak{D}}$	(H)	$\mathbf{J} = \mathbf{J}_c + \frac{\partial \mathbf{D}}{\partial t}$
$u = p + \frac{df}{dt}$ $v = q + \frac{dg}{dt}$ $w = r + \frac{dh}{dt}$	(H*)	$\mathbf{J} = \mathbf{J}_c + \frac{\partial \mathbf{D}}{\partial t}$
$\mathfrak{C} = (C + \frac{1}{4\pi} K \frac{d}{dt})\mathfrak{E}$	(I)	$\mathbf{J} = \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t}$
$u = CP + \frac{1}{4\pi} K \frac{dP}{dt}$ $v = CQ + \frac{1}{4\pi} K \frac{dQ}{dt}$ $w = CR + \frac{1}{4\pi} K \frac{dR}{dt}$	(I*)	$\mathbf{J} = \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t}$
$\rho = \frac{df}{dx} + \frac{dg}{dy} + \frac{dh}{dz}$	(J)	$\varrho = \nabla \cdot \mathbf{D}$
$\sigma = lf + mg + nh + l'f' + m'g' + n'h'$	(K)	$\varrho_s = \mathbf{n} \cdot (\mathbf{D}_1 - \mathbf{D}_2)$
$\mathfrak{B} = \mu \mathfrak{H}$	(L)	$\mathbf{B} = \mu \mathbf{H}$

Fig. 1.1 *The original set of equations (A)–(L) as labeled by Maxwell in his Treatise (1873), with their interpretation in modern Gibbsian vector notation. The simplest equations were also written in vector form.*

$$\begin{aligned} \mathbf{d}\wedge\mathbf{\Psi} &= \gamma \\ \mathbf{d}\wedge\mathbf{\Phi} &= 0 \\ \mathbf{\Psi} &= \overline{\overline{\mathsf{M}}}|\mathbf{\Phi} \end{aligned}$$

Fig. 1.2 *The two Maxwell equations and the medium equation in differential-form formalism. Symbols will be explained in Chapter 4.*

Grassmann had hoped that the second edition of *Ausdehnungslehre* would raise interest in his contemporaries. Fearing that this, too, would be of no avail, his final sentences in the foreword were addressed to future generations [15, 75]:

> ... But I know and feel obliged to state (though I run the risk of seeming arrogant) that even if this work should again remain unused for another seventeen years or even longer, without entering into actual development of science, still that time will come when it will be brought forth from the dust of oblivion, and when ideas now dormant will bring forth fruit. I know that if I also fail to gather around me in a position (which I have up to now desired in vain) a circle of scholars, whom I could fructify with these ideas, and whom I could stimulate to develop and enrich further these ideas, nevertheless there will come a time when these ideas, perhaps in a new form, will rise anew and will enter into living communication with contemporary developments. For truth is eternal and divine, and no phase in the development of the truth divine, and no phase in the development of truth, however small may be region encompassed, can pass on without leaving a trace; truth remains, even though the garments in which poor mortals clothe it may fall to dust.
>
> Stettin, 29 August 1861

1.2 VECTORS AND DUAL VECTORS

1.2.1 Basic definitions

Vectors are elements of an n-dimensional vector space denoted by $\mathbb{E}_1(n)$, and they are in general denoted by boldface lowercase Latin letters $\mathbf{a}$, $\mathbf{b}$,... Most of the analysis is applicable to any dimension n but special attention is given to three-dimensional Euclidean (Eu3) and four-dimensional Minkowskian (Mi4) spaces (these concepts will be explained in terms of metric dyadics in Section 2.5). A set of linearly independent vectors $\{\mathbf{e}_i\} = \mathbf{e}_1, \mathbf{e}_2,..., \mathbf{e}_n$ forms a basis if any vector $\mathbf{a}$ can be uniquely expressed in terms of the basis vectors as

$$\mathbf{a} = \sum_{i=1}^{n} a_i \mathbf{e}_i, \tag{1.1}$$

where the a_i are scalar coefficients (real or complex numbers).

TABLE I
DIFFERENTIAL FORMS OF EACH DEGREE

Degree	Region of Integration	Example	General Form
0-form	Point	$3x$	$f(x, y, z, \ldots)$
1-form	Path	$y^2\, dx + z\, dy$	$\alpha_1\, dx + \alpha_2\, dy + \alpha_3\, dz$
2-form	Surface	$e^y\, dy\, dz + e^x g\, dz\, dx$	$\beta_1\, dy\, dz + \beta_2\, dz\, dx + \beta_3\, dx\, dy$
3-form	Volume	$(x + y)\, dx\, dy\, dz$	$g\, dx\, dy\, dz$

TABLE II
THE DIFFERENTIAL FORMS THAT REPRESENT FIELDS AND SOURCES

Quantity	Form	Degree	Units	Vector/Scalar
Electric Field Intensity	E	1-form	V	**E**
Magnetic Field Intensity	H	1-form	A	**H**
Electric Flux Density	D	2-form	C	**D**
Magnetic Flux Density	B	2-form	Wb	**B**
Electric Current Density	J	2-form	A	**J**
Electric Charge Density	ρ	3-form	C	q

double integral over a surface, so its degree is two. A 3-form is integrated by a triple integral over a volume. 0-forms are functions, ''integrated'' by evaluation at a point. Table I gives examples of forms of various degrees. The coefficients of the forms can be functions of position, time, and other variables.

A. Representing the Electromagnetic Field with Differential Forms

From Maxwell's laws in integral form, we can readily determine the degrees of the differential forms that will represent the various field quantities. In vector notation,

$$\oint_P \boldsymbol{E} \cdot d\boldsymbol{l} = -\frac{d}{dt} \int_A \boldsymbol{B} \cdot d\boldsymbol{A}$$

$$\oint_P \boldsymbol{H} \cdot d\boldsymbol{l} = \frac{d}{dt} \int_A \boldsymbol{D} \cdot d\boldsymbol{A} + \int_A \boldsymbol{J} \cdot d\boldsymbol{A}$$

$$\oint_S \boldsymbol{D} \cdot d\boldsymbol{S} = \int_V q dv$$

$$\oint_S \boldsymbol{B} \cdot d\boldsymbol{S} = 0$$

where A is a surface bounded by a path P, V is a volume bounded by a surface S, q is volume charge density, and the other quantities are defined as usual. The electric field intensity is integrated over a path, so that it becomes a 1-form. The magnetic field intensity is also integrated over a path, and becomes a 1-form as well. The electric and magnetic flux densities are integrated over surfaces, and so are 2-forms. The sources are electric current density, which is a 2-form, since it falls under a surface integral, and the volume charge density, which is a 3-form, as it is integrated over a volume. Table II summarizes these forms.

B. 1-Forms: Field Intensity

The usual physical motivation for electric field intensity is the force experienced by a small test charge placed in the field. This leads naturally to the vector representation of the electric field, which might be called the "force picture." Another physical viewpoint for the electric field is the change in potential experienced by a charge as it moves through the field. This leads naturally to the equipotential representation of the field, or the "energy picture." The energy picture shifts emphasis from the local concept of force experienced by a test charge to the global behavior of the field as manifested by change in energy of a test charge as it moves along a path.

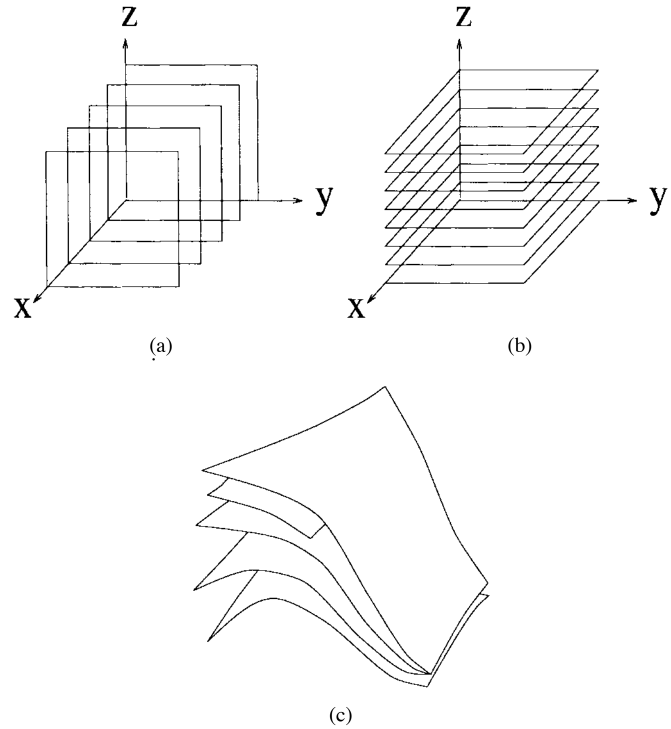


Fig. 1. (a) The 1-form dx, with surfaces perpendicular to the x-axis and infinite in the y and z directions. (b) The 1-form $2\, dz$, with surfaces perpendicular to the z-axis and spaced two per unit distance in the z direction. (c) A general 1-form, with curved surfaces and surfaces that end or meet each other.

Differential forms lead to the "energy picture" of field intensity. A 1-form is represented graphically as surfaces in space [1], [3]. For a conservative field, the surfaces of the associated 1-form are equipotentials. The differential dx produces surfaces perpendicular to the x-axis, as shown in Fig. 1(a). Likewise, dy has surfaces perpendicular to the y-axis and the surfaces of dz are perpendicular to the z-axis. A linear combination of these differentials has surfaces that are skew to the coordinate axes. The coefficients of a 1-form determine the spacing of the surfaces per unit length; the greater the magnitude of the coefficients, the more closely spaced are the surfaces. The 1-form $2\, dz$, shown in Fig. 1(b), has surfaces spaced twice as closely as those of dx in Fig. 1(a).

In general, the surfaces of a 1-form can curve, end, or meet each other, depending on the behavior of the coefficients

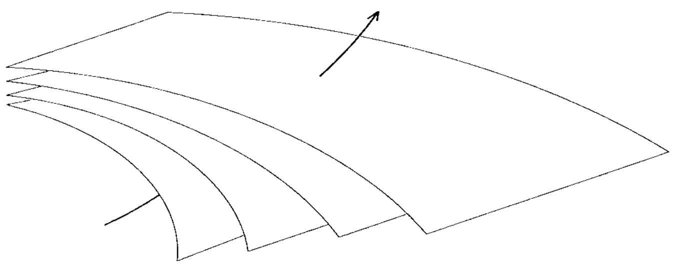


Fig. 2. A path piercing four surfaces of a 1-form. The integral of the 1-form over the path is four.

of the form. If surfaces of a 1-form do not meet or end, the field represented by the form is conservative. The field corresponding to the 1-form in Fig. 1(a) is conservative; the field in Fig. 1(c) is nonconservative.

Just as a line representing the magnitude of a vector has two possible orientations, the surfaces of a 1-form are oriented as well. This is done by specifying one of the two normal directions to the surfaces of the form. The surfaces of $3\,dx$ are oriented in the $+x$ direction, and those of $-3\,dx$ in the $-x$ direction. The orientation of a form is usually clear from context and is omitted from figures.

Differential forms are by definition the quantities that can be integrated, so it is natural that the surfaces of a 1-form are a graphical representation of path integration. The integral of a 1-form along a path is the number of surfaces pierced by the path (Fig. 2), taking into account the relative orientations of the surfaces and the path. This simple picture of path integration will provide in the next section a means for visualizing Ampere's and Faraday's laws.

The 1-form $E_1\,dx + E_2\,dy + E_3\,dz$ is said to be *dual* to the vector field $E_1\hat{\boldsymbol{x}} + E_2\hat{\boldsymbol{y}} + E_3\hat{\boldsymbol{z}}$. The field intensity 1-forms E and H are dual to the vectors $\boldsymbol{E}$ and $\boldsymbol{H}$.

Following Deschamps, we take the units of the electric and magnetic field intensity 1-forms to be volts and amperes, as shown in Table II. The differentials are considered to have units of length. Other field and source quantities are assigned units according to this same convention. A disadvantage of Deschamps' system is that it implies in a sense that the metric of space carries units. Alternative conventions are available; Bamberg and Sternberg [5] and others take the units of the electric and magnetic field intensity 1-forms to be volts per meter and amperes per meter, the same as their vector counterparts, so that the differentials carry no units and the integration process itself is considered to provide a factor of length. If this convention is chosen, the basis differentials of curvilinear coordinate systems (see Section IV) must also be taken to carry no units. This leads to confusion for students, since these basis differentials can include factors of distance. The advantages of this alternative convention are that it is more consistent with the mathematical point of view, in which basis vectors and forms are abstract objects not associated with a particular system of units, and that a field quantity has the same units whether represented by a vector or a differential form. Furthermore, a general differential form may include differentials of functions that do not represent

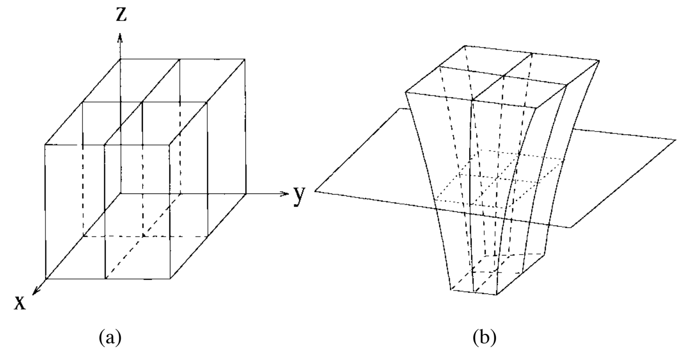


Fig. 3. (a) The 2-form $dx\,dy$, with tubes in the z direction. (b) Four tubes of a 2-form pass through a surface, so that the integral of the 2-form over the surface is four.

position and so cannot be assigned units of length. The possibility of confusion when using curvilinear coordinates seems to outweigh these considerations, and so we have chosen Deschamps' convention.

With this convention, the electric field intensity 1-form can be taken to have units of energy per charge, or joules per coulomb. This supports the "energy picture," in which the electric field represents the change in energy experienced by a charge as it moves through the field. One might argue that this motivation of field intensity is less intuitive than the concept of force experienced by a test charge at a point. While this may be true, the graphical representations of Ampere's and Faraday's laws that will be outlined in Section III favor the differential form point of view. Furthermore, the simple correspondence between vectors and forms allows both to be introduced with little additional effort, providing students a more solid understanding of field intensity than they could obtain from one representation alone.

C. 2-Forms: Flux Density and Current Density

Flux density or flow of current can be thought of as tubes that connect sources of flux or current. This is the natural graphical representation of a 2-form, which is drawn as sets of surfaces that intersect to form tubes. The differential $dx\,dy$ is represented by the surfaces of dx and dy superimposed. The surfaces of dx perpendicular to the x-axis and those of dy perpendicular to the y-axis intersect to produce tubes in the z direction, as illustrated by Fig. 3(a). (To be precise, the tubes of a 2-form have no definite shape: tubes of $dxdy$ have the same density those of $[0.5\,dx][2\,dy]$.) The coefficients of a 2-form give the spacing of the tubes. The greater the coefficients, the more dense the tubes. An arbitrary 2-form has tubes that may curve or converge at a point.

The direction of flow or flux along the tubes of a 2-form is given by the right-hand rule applied to the orientations of the surfaces making up the walls of a tube. The orientation of dx is in the $+x$ direction, and dy in the $+y$ direction, so the flux due to $dx\,dy$ is in the $+z$ direction.

As with 1-forms, the graphical representation of a 2-form is fundamentally related to the integration process. The integral of a 2-form over a surface is the number of tubes passing through the surface, where each tube is weighted positively if always rank 2, and the dual metric tensor is a rank 2 symmetric contravariant tensor [335]. In the literature the process of converting a contravariant vector to a covariant vector is referred to as *index lowering* while converting in the opposite direction from covariant to contravariant is referred to as *index raising*, most often in the context of tensor analysis.[6]

In some applications, a metric is provided without explicit identification as to whether the metric applies to contravariant quantities or covariant quantities, and a decision must be made to interpret the metric as primal or dual. We shall consider this issue further below as it arises often in the discrete context of graph theory.

For completeness, we define a **norm** on vectors using the inner product defined by the metric tensor. The norm of any 1-vector $\bar{v}$ is given by $\|\bar{v}\| \equiv \sqrt{\langle \bar{v}, \bar{v} \rangle}$. This norm can be shown to satisfy the triangle inequality and it, as well as the underlying inner product, satisfies the Schwartz inequality. Finally a **metric** can be defined that establishes the distance between two vectors at a point q: for a pair of vectors $\bar{u}$ and $\bar{v}$ in $T\mathcal{M}^n_q$, the metric is defined as $\rho(\bar{u}, \bar{v}) = \|\bar{u} - \bar{v}\|$.

The metric tensor provides a unique mapping between 1-forms and 1-vectors, but requires the specification of a metric on the underlying vector space in the form of a bilinear pairing. This pairing is fundamental in that once it is established, lengths of vectors can be computed, from which one can then compute lengths of curves, angles, and perform parallel translation. In the next section we will consider integration, which is another such pairing that is purely topological and thus does not require a metric.

2.2.2 *Differentiation and Integration of Forms*

So far we have considered the exterior algebra of both p-vectors and differential p-forms and the mechanism by which the metric tensor can provide a one-to-one mapping between forms and vectors. We now consider the operations of differentiation and integration defined for forms in order to assemble the tools that are needed to use differential forms for calculus in arbitrary dimensions.

2.2.2.1 The Exterior Derivative

The exterior derivative operator extends the notion of a derivative to differential forms in a way that is invariant under coordinate transformations and does not require the specification of a metric. We define the **exterior derivative** as the operator d that maps p-forms to $(p+1)$-forms, $\mathrm{d} : \bigwedge^p(T^*\mathcal{M}^n) \to \bigwedge^{p+1}(T^*\mathcal{M}^n)$. Just as

[6] In the differential forms literature this same process is sometimes described by the so-called *musical isomorphisms*, $\sharp$ and $\flat$, where $\tilde{\alpha}^\sharp$ is the contravariant version of the form $\tilde{\alpha}$, and $\bar{v}^\flat$ is the covariant version of the vector $\bar{v}$.

Who was Nicolas Bourbaki?

Book Review

Bourbaki, A Secret Society of Mathematicians and *The Artist and the Mathematician*

Reviewed by Michael Atiyah

Bourbaki, A Secret Society of Mathematicians
Maurice Mashaal
AMS, June 2006
US$29.00, 260 pages
ISBN-13: 978-0821839676

The Artist and the Mathematician: The Story of Nicolas Bourbaki, the Genius Mathematician Who Never Existed
Amir D. Aczel
Thunder's Mouth Press, August 2006
US$23.95, 272 pages
ISBN-13: 978-1560259312

All mathematicians of my generation, and even those of subsequent decades, were aware of Nicolas Bourbaki, the Napoleonic general whose reincarnation as a radical group of young French mathematicians was to make such a mark on the mathematical world. His memory may now have faded, the books are old and yellowed, but his influence lives on. Many of us were enthusiastic disciples of Bourbaki, believing that he had reinvigorated the mathematics of the twentieth century and given it direction. But others believed that Bourbaki's influence had been pernicious and narrow, confining mathematics behind walls of rigour, and cutting off its external sources of inspiration.

Sir Michael Atiyah is Honorary Professor of Mathematics at the University of Edinburgh. His email address is `M.Atiyah@ed.ac.uk`.

Now that we are in the twenty-first century it is perhaps the right time to look back and try to assess the overall impact of Bourbaki, before all the principal players leave the scene. The basic historical facts are well known and are set out in both the books under review. France had lost a whole generation of intellectuals in the 1914–18 war, and the young mathematicians of Paris, in the inter-war period of the 1920s and 1930s, were looking for new guidance and inspiration. Only Hadamard and Élie Cartan of the older generation still commanded respect. Talented youth, unconstrained by higher authority, is a powerful force and, whatever one's views about Bourbaki, there is no doubt that the talent was quite exceptional. The list of the early members of Bourbaki is truly impressive: André Weil, Henri Cartan, Claude Chevalley, Jean Dieudonné, Laurent Schwartz… Later recruits were of similar calibre: Jean-Pierre Serre, Armand Borel, Alexandre Grothendieck… Harnessing the powers of such a formidable group was not an easy task. There were fierce debates, some serious quarrels, and much passion. The remarkable fact is that the group, by and large, stayed together and kept Bourbaki alive and active over several decades. This was a tribute to the idealistic vision that they shared, that of remoulding the shape of mathematics in the twentieth century.

Much of the atmosphere of the early days is brought vividly to life by the many informal photographs in the Mashaal book. It is fascinating to see pictures of the young André Weil, relaxing in a deck chair, though Henri Cartan was always impeccably dressed in jacket and tie, resisting trendy fashion.

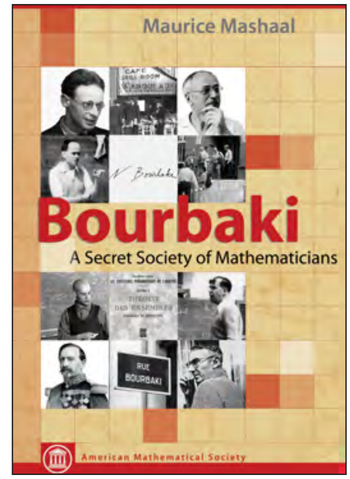
I myself attended a Bourbaki conference in my youth and can attest to the lively experience of debating vigorously (and usually critically) the latest version of the next book. Summer sunshine in the south of France and the friendly and casual atmosphere did much to prevent arguments developing into armed conflict. To paraphrase Winston Churchill, "never in the course of human argument has so much been spoken by so many on so little." It appeared a miracle that books, many of them, actually emerged from this process, a result undoubtedly due to the diligence and energy of Dieudonné. If Weil was the prime inspiration behind Bourbaki, it was Dieudonné who carried it to fruition.

So what were the basic aims of Bourbaki, and how much was achieved? Perhaps one can pick out two central objectives. One was that mathematics needed new and broad foundations, embodied in a series of books that would replace the old-fashioned textbooks. The other was that the key idea of the new foundations lay in the notion of "structure", illustrated by the now common word "isomorphism".

There is no doubt that, with its clear emphasis on "structure", Bourbaki produced the right idea at the right time and changed the way most of us thought. Of course it fitted in well with Hilbert's approach to mathematics and the subsequent development of abstract algebra. But structure was not confined to algebra, and it was particularly fruitful in topology and associated areas of geometry, all of which were to see spectacular developments in the period following World War II. Here the impact of Bourbaki was decisive, and, in the hands of Serre and Grothendieck, algebraic geometry rose to incredible heights.

Laying universal foundations is another matter. Each time it is tried it inevitably gets bogged down by the sheer scale and ambition of the operation. The "ne plus ultra" in this direction was the *Éléments de Géométrie Algébrique* of Grothendieck and Dieudonné, which expanded voluminously both forward and backward and was in danger of sinking under its own weight.

Laying ambitious foundations is not only a dangerous delusion, it can also be a didactic disaster. Encyclopaedias are not textbooks, and much of the critique directed against Bourbaki is that it was used, or perhaps misused, to reform school education. This may be unfair, since many of the great mathematicians in Bourbaki were excellent lecturers and knew well the difference between formal exposition and the conveying of ideas. But, as so often happens, the disciples are more extreme and fanatical than their masters, and education in France and elsewhere suffered from a dogmatic and ill-informed attempt at reform. Jesus Christ is not responsible for the excesses perpetrated in the name of Christianity.

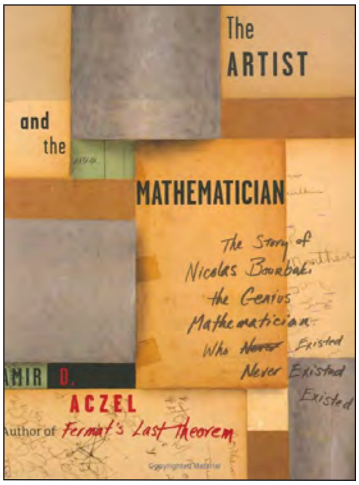


Bourbaki was to some extent the victim of its own success. The original aim had been the modest one of writing a modern replacement for Goursat's *Cours d'Analyse* but, buoyed up by enthusiasm and the success of recruiting many of the leading mathematicians of the time, horizons broadened. All of mathematics was to be included, analysis, algebra, and geometry. For obvious reasons algebra lent itself best to the Bourbaki treatment. The volumes on commutative algebra and particularly on Lie groups were excellent and became standard references, due in large part to the personal contribution of Serre, whose influence and taste guided this whole area.

The formal aspects of analysis, as exemplified in functional analysis, also had success, though Bourbaki's treatment of probability came in for severe criticism from the experts who argued that important parts of the theory were excluded by the restriction to locally compact spaces. A concern for elegance had led to too great a price being paid.

But this little battle over probability was a mere sideshow in the Bourbaki approach to analysis, a subject too varied, complex, and untidy to be taken over by Bourbaki. Glimmerings of these problems already appear in differential geometry, a subject at the interface between analysis and geometry, where structure, though present, is a less dominating concept. Though Riemann surface theory, after a century of active development, could conceivably be given a coherent Bourbaki treatment, the same could hardly be said for the current work of Thurston-Perelman in three dimensions. Another severe limitation of Bourbaki, no doubt conscious, was the restriction to pure mathematics. Applied mathematics is too messy and disparate to be included, and theoretical physics hovers on an uncertain borderline. One distinguishing feature of Bourbaki was the emphasis on clear and unambiguous definitions and on rigorous proofs. This was, as in algebraic geometry, a reaction against some sloppy treatments of the past, and it served a purpose in creating a firm platform for the future. Unfortunately, when taken to extremes, the requirement for total rigour excludes large areas of mathematics which are in their early creative stages. Had Euler worried too much about rigour, mathematics would have suffered.

Over the past thirty years, arguably in the declining years of Bourbaki, some of the most exciting developments in mathematics have arisen from the interface with physics and particularly quantum



field theory. New concepts and explicit results have emerged from this interaction, notably Donaldson's work on four-manifolds, mirror symmetry in algebraic geometry, and quantum cohomology. Much of this came directly from very heuristic work by physicists such as Edward Witten. Most of it, though by no means all, has now been given a cloak of respectability involving rigorous proofs.

Clarity and rigour have a vital place in mathematics but they must not be used as a barrier to new ideas from other fields. Free trade is a benefit to us all and should not be inhibited by excessive attachment to national sovereignty.

Although Bourbaki recruited most of the famous French mathematicians of the time (and several from outside France), there were some notable exceptions, the most obvious being Jean Leray (who left very early) and René Thom. In retrospect it is clear that neither fitted the Bourbaki role. The fact that they were also two of the most original mathematicians of the time does perhaps suggest that such originality has difficulty flourishing in a constrained atmosphere. Both were also closer to applied mathematics than their colleagues.

Of the two books under review, the first by Maurice Mashaal might be described as "authorized". It has the sanction of the AMS and was first published several years ago in French. It seems clear that the author knew many of the French mathematicians personally and derived his information and in particular the photographs from this source. It is reliable on the history, the personalities, and the mathematics. It is also highly readable and noncontroversial.

The other book by Amir Aczel is totally different. It has a more ambitious aim, which is to examine the Bourbaki influence on "structure" in the social sciences. It is also highly controversial in its extensive treatment of the Grothendieck story. I was not convinced of the total reliability of its sources, nor of its philosophical credentials.

Although written in English this book is permeated by French intellectual ideas and will probably seem strange to those not part of that scene. A slightly tenuous link between André Weil and the sociologist Claude Levi-Strauss is used to claim that Bourbaki made a major impact on sociology and related fields such as psychology, anthropology, and linguistics. This grand aim is clearly set out by the title, and I have no expertise in any of these fields. It may be that the author is a polymath, an intellectual colossus, who straddles the entire scene from mathematics to the social sciences. The only place where I can examine the evidence for this and make an informed comment, is in his treatment of mathematics and the people in it. Here I have profound misgivings, which relate mainly to Grothendieck, who occupies a central place in the author's pantheon.

There is no doubt that Grothendieck was an exceptional figure in the mathematical world and that he deserves a scholarly full-length biography, preferably written by a mathematician who knew him personally. I believe such a book is in preparation, and I look forward to reading it. Aczel's book does not measure up to the level of the subject, because of his uncritical acceptance of Grothendieck as the great prophet, spurned eventually by his people (including Bourbaki).

I knew Grothendieck well when he was in his prime. I greatly admired his mathematics, his prodigious energy and drive, and his generosity with ideas, which attracted a horde of disciples. But his main characteristic, both in his mathematics and in social life, was his uncompromising nature. This was, at the same time, the cause both of his success and of his downfall. No one but Grothendieck could have taken on algebraic geometry in the full generality he adopted and seen it through to success. It required courage, even daring, total self-confidence and immense powers of concentration and hard work. Grothendieck was a phenomenon.

But he had his weaknesses. He could navigate like no one else in the stratosphere, but he was not sure of his ground on earth—examples did not appeal to him and had to be supplied by his colleagues.

Aczel is right when he identifies Grothendieck as someone who took the new Bourbaki philosophy seriously and made a tremendous success of it. Where I part company with Aczel is in his assertion that Bourbaki made a fatal mistake in not taking Grothendieck's advice and rewriting its foundations in the new language of category theory. Aczel believes that Bourbaki had turned its face away from the future in not following Grothendieck. I doubt whether history will come to this verdict. Grothendieck's own *EGA*, as well as the general fate of over-confident universalists, might suggest otherwise. Moreover, given Grothendieck's uncompromising nature and supreme self-confidence, it is difficult to see how, with him at the helm, Bourbaki could have continued as a collegial enterprise.

Aczel's total endorsement of Grothendieck leads him to make such fatuous statements as: "Weil was a somewhat jealous person who clearly saw that Grothendieck was a far better mathematician than he was." Subtle balanced judgement is clearly not Aczel's forte, and it hardly encourages the reader to take seriously his confident and sweeping assertions in the social sciences.

Bourbaki and Algebraic Topology

by John McCleary

The principal aim of the Bourbaki group (L'Association des Collaborateurs de Nicolas Bourbaki) is to provide a solid foundation for the whole body of modern mathematics. The method of exposition is axiomatic and abstract, logically coherent and rigorous, proceeding normally from the general to the particular, a style found to be not altogether congenial to many readers. The ongoing series of books began with Éléments de Mathématiques in 1939, and other books on algebra, set theory, topology, and other topics have followed. Many books in the series have become standard references, though some mathematicians are critical of their austerely abstract point of view.

from http://www.encyclopedia.com/html/B/BourbakiN1.asp, Dec. 3, 2004

It is now more than 70 years ago that the founders of *Le Comité de rédaction du traité d'analyse* met in Paris at the *Café A. Capoulade*, 63 boulevard Saint-Michel, to discuss the drafting of a textbook on analysis. This meeting included (recent centenarian) Henri Cartan (1904–), Claude Chevalley (1909–1984), Jean Delsarte (1903–1968), Jean Dieudonné (1906–1992), René de Possel (1905–1974), and André Weil (1906–1998). The fate of this project is the story of the *Bourbaki*, or should I say, the story of the character Nicolas Bourbaki, author of *Éléments de mathématique*, a series of influential expositions of the basic notions of modern mathematics.

This talk is based on a wild goose chase after a document. The project was supported by the Gabriel Snyder Beck Fund at Vassar College that funds research on anything French. In early 2000 I learned at a meeting in Oberwohlfach that an archive of papers and internal documents of the Bourbaki was soon to be opened to scholars in Paris. The Beck fund provided me the means to visit the archive. The managers of this archive, Liliane Beaulieu and Christian Houzel, showed me great hospitality during my visit to Paris in July 2003, and made it possible for me to rummage through the Bourbaki papers.

Historical research poses questions, to which various methods may be applied. My interests include the history of algebraic topology, a subject whose development during the twentieth century influenced a great deal of that century's mathematics. The years following the Second World War represent a high point in this story, and several important members of Bourbaki contributed to this development. However, algebraic topology does not appear among the topics treated in *Éléments*—admittedly many other important topics were also omitted. The involvement of so many pioneering topologists makes this omission stand out.

While a graduate student, I collected a rumor that there was a manuscript, 200 pages long, prepared for *Élements* by Cartan, Koszul, Eilenberg, and Chevalley, treating algebraic topology. Furthermore, this document was based on the use of differential forms, that is, algebraic topology chez Élie Cartan (1869–1951) (*le pere d'Henri*). According to the rumor, the manuscript was abandoned when the doctoral theses of Jean-Pierre Serre (1926–) and Armand Borel (1923–2003) were published. Serre's and Borel's subsequent papers did change the focus of research in topology, away from differential geometric methods to more algebraic methods, principally the spectral sequence and the Steenrod algebra, making the manuscript obsolete. So what was in this manuscript? Could

I get a look at it? The historian salivates at the chance to look at the state of affairs before and after a key event.

Well, the manuscript wasn't there, if, in fact, it exists at all. The archival work I was able to do, however, offered many insights into the workings and spirit of Bourbaki and I will relate some findings in this report. As my story unfurls, I want to consider the allure of the axiomatic method before and after Bourbaki, one of the features of their exposition that has received criticism.

Who is Bourbaki?

> *His name is Greek, his nationality is French and his history is curious. He is one of the most influential mathematicians of the 20th century. The legends about him are many, and they are growing every day. ... The strangest fact about him, however, is that he doesn't exist.*
>
> Paul Halmos, 1957

André Weil was on the faculty at the University of Strasbourg in 1934, together with Henri Cartan. They were responsible for the course on the differential and integral calculus, one of three standard courses required for the *license de mathématiques*, along with general physics and rational mechanics. The standard text was *Cours d'Analyse mathématique* by ÉDUOARD GOURSAT (1858–1936), written before the First World War. Cartan found it wanting, incomplete where generalizations were known, and simply not the best way to present these topics. An explicit example, one with a story of its own, is the formulation of Stokes's Theorem. It may be written

$$\int_{\partial X} \omega = \int_X d\omega,$$

where ω is a differential form, $d\omega$ its exterior derivative, X the domain of integration and ∂X the boundary of X. When everything in sight is smooth, the proof is clear, but the importance of this formula in the case of more general domains of integration is the content of the celebrated theorem of GEORGES DE RHAM (1903–1990), proved in 1931 to answer a question of Elie Cartan relating invariant integrals on Lie groups to the topology of such manifolds.

Persistent badgering by Cartan led Weil to suggest that they write a textbook that they could be satisfied with. Weil writes that he told Cartan, "Why don't we get together and settle such matters once and for all, and you won't plague me with your questions any more?"

The first meeting on 10 December 1934 in Paris to plan the book occurred after a meeting of *le Seminaire Julia*, another of Weil's and Cartan's efforts to fill the gap left in French mathematics after World War I, which Weil called "hectatomb of 1914–1918 which had slaughtered virtually an entire generation" of French mathematicians. The seminar, organized by these young turcs in imitation of the seminars in Germany, needed a sponsor in order to get a room at the Sorbonne. GASTON JULIA (1893–1978) had been the youngest of their teachers at the *École Normale Supérieure* and he stepped up to sponsor them. The seminar treated a topic a year, beginning in 1933-34 with groups and algebras, going on to

Hilbert spaces, then topology. The seminar continued until 1939 when it was superseded by the Seminar Bourbaki.

The committee's first plans were for a text in analysis, that would, according to Weil, "*fix the curriculum for 25 years for differential and integral calculus.*" This text should be *aussi moderne que possible*, *un traité utile à tous*, and finally, *aussi robustes et aussi universels que possible.* Weil already knew a potential publisher in his friend Enriques Freymann, a Mexican diplomat who married the daughter of the founder of *Maisson Hermann*, a scientific publisher. Freymann became the chief editor and manager of the publishing house.

Among the innovations of this text was the suggestion, insisted on by Delsarte, that it be written collectively without *expert leadership.* The initial expectation was that the text would run to 1000–1200 pages and be done in about six months. The initial group of six was expanded to nine members in January 1935, with Paul Dubreil (1904–1994), Jean Leray (1906–1998) and Szolem Mandelbrojt (1899-1983) added. Dubreil and Leray were replaced by Jean Coulomb and Charles Ehresmann (1905–1979) before the first summer workshop in July, 1935.

The first Bourbaki congress was held in Besse-en-Chandesse in the Vosges mountains. At this workshop, the proposal was made to expand the project to add a *paquet abstrait*, treating abstract (new and modern) notions that would support analysis. These included abstract set theory, algebra, especially differential forms, and topology, with particular emphasis on existence theorems (Leray).

The *paquet* eventually became the *Fascicule de Résultats*, a summary of useful results presented in such a way that a competent mathematician could see where a desired result might be found, and provide the result themselves if they needed it. In fact, the last publication, *Fascicule XXXVI*, part two of *Variétés différentielles et analytiques*, is such a summary. By the way, it is in *Fascicule XXXVI* that the statement of Stokes's Theorem found its place.

During the first conference, with a group of young, eager, and able mathematicians in one place, a new result on measures on a topological space was proved. A note was written up to submit to *Comptes-Rendues.* The name of *Bourbaki* for the group was based on a story out of school: In 1923, Delsarte, Cartan, and Weil were members of the newly matriculated class at *École Normale Superieure*, when they received a lecture notice by a professor with a vaguely Scandinavian name, for which attendance was strongly recommended. The speaker was a prankster, Raoul Husson, wearing a false beard and speaking with an undefinable accent. Taking off from classical function theory, the talk had its climax in *Bourbaki's Theorem* leaving the audience "speechless with amazement." (This Bourbaki was the general who traveled with Napoleon.) Weil recalled this story and the family name was adopted. But why Nicolas? For the submission of the paper, the author needed a *prenom.* It was Weil's wife Eveline who christened the new Bourbaki Nicolas. The note was handled at the *Académie des Sciences* by Élie Cartan who stood up for the unfortunate Poldevian mathematician. The note was accepted and published.

To produce the constituent parts of *les Éléments*, the method of editing adopted by the Bourbaki emphasized communal involvement. A text was brought before a meeting and presented, page by page, line by line, to the group who then expressed any and all criticism. A revision was handed over to another member of the group and the process repeated when a new draft was available. After enough iterations to obtain unanimous approval—either for the strength of the text or the fatigue of the group with the topic—the text would be finalized (usually by Dieudonné) and sent to the publisher.

Digression: The Axiomatic Method

> *In spite of the high pedagogic value of the genetic method, the axiomatic method has the advantage of providing a conclusive exposition and full logical confidence to the contents of our knowledge.*
>
> David Hilbert, 1900

During his 'apprenticeship' (documented in [Weil]), Weil traveled extensively, spending time in Germany while the rise of National Socialism to power took place. As he was interested in number theory, he admired the mathematics of the German schools, especially the axiomatic approach led by the work of David HIlbert (1862–1943) and the Göttingen school. French mathematics through the nineteenth century and into the twentieth was dominated by analysis. Even results of a number-theoretic nature were proved through analytic means. The success of Hilbert's ideas in many fields attracted mathematicians everywhere and so, when looking for a model to shape their project, the members of Bourbaki turned to the axiomatic method.

This phenomenon was not without precedent. When E.H. Moore (1862–1932) came to lead the University of Chicago mathematics department around 1900, he consciously adopted the style of Hilbert's *Grundlagen der Geometrie* as modern, precise, and a model to be imitated. His earliest students at Chicago included Oswald Veblen (1880–1960), Frederick Owens, and R.L. Moore (1882–1974) whose PhD theses concerned the foundations of geometry, axiom systems, and the economy of expression Hilbert achieved. The goals of some of this work were to tighten the systems of axioms describing geometry, to root out redundancy and present the least one needs to assume to achieve Euclid's bounty. These goals, however, though laudable, do not exhaust the depth of the axiomatic method.

Roughly speaking, the *axiomatic method* is an approach to producing mathematics that presents, after some analysis, a set of axioms from which a set of theorems is deduced. The goal in presenting the *right set of axioms* is to avoid deception by intuition. Hilbert's experience with algebraic number theory (his *Zahlbericht*) and invariant theory led him to tread a path leading to more abstract generalization.

When he turned to elementary geometry in his lectures of 1898–99, students in Göttingen were surprised. Already in his early career, Hilbert had remarked of geometry, "One must be able to say at all times—instead of points, straight lines, and planes—tables, chairs, and beer mugs." His stated goal in the *Grundlagen* was " to attempt to choose for geometry a *simple* and *complete* set of *independent* axioms and to deduce from these the most important geometrical theorems in such a manner as to bring out as clearly as possible the significance of the different groups of axioms and the scope of the conclusions to be derived from the individual axioms."

The *Grundlagen* was an immediate success, drawing the following reaction from Henri Poincaré (1858–1912): "The logical point of view alone appears to interest Professor

Hilbert. Being given a sequence of propositions, he finds that all follow logically from the first. With the foundation of this first proposition, with its psychological origin, he does not concern himself .... The axioms are postulated; we do not know from whence they come; it is then as easy to postulate A as C .... His work is thus incomplete, but this is not a criticism I make against him. Incomplete one must indeed resign oneself to be. It is enough that he has made the philosophy of mathematics take a step forward ...."

The philosophical and foundational aspects of Hilbert's efforts are clear. However, the mathematical aspects are not the focus of most discussions of the *Grundlagen*. Among the exercises in independence he has introduced new objects—in particular, non-Archimedean geometries. By isolating the relations among axiom groups, one can discover how the failure of one or more of the assumptions produces new results, the model of this activity being non-Euclidean geometry. His experience in algebra and number theory also supported this view, that the axiomatic method sharpened one's tools with which to craft new arguments, discover new phenomena, and retain the past in a tidy manner to boot.

Another Göttingen product of importance to Bourbaki is in the same spirit: *Moderne Algebra* by B.L. van der Waerden (1903–1996) first appeared in 1930, giving an organized account of algebra based on axioms that revealed the similarity in approaches to certain results. The notion of isomorphism plays an important role in algebra and later surfaces as a leitmotif for Bourbaki.

It is important to see that Hilbert and van der Waerden, though formal in presentation, really sought mathematical goals that were not about the past, to recover a complete description of a known theory, but were forward-looking, providing the mathematician with a slim but firm scaffolding on which many new results could be built. The degree to which this view became part of the manner in which modern mathematics was done can be measured by the natural feel we have today for this sort of presentation. It was not always so.

Algebraic Topology chez Bourbaki

> *A côté des structures algébriques (groupes, anneaux, corps, etc.) ... dans toutes les parties de l'Analyse, des structures d'une autre sorte : ce sont celles où l'on donne un sens mathématique aux notions intuitives de limite, de continuité et de voisinage.*
>
> Bourbaki, Topologie 1965

The goal of producing a modern, robust, and universal text led to the most characteristic quality of Bourbaki—a topic was discussed repeatedly in an effort to "digest mathematics, to go to the essential points, and reformulate the math in a more comprehensive and conceptual way [Borel]." The sessions were animated to achieve this goal; after the war, there is a record in *La Tribu* of the rebirth of what were considered classic duels between Cartan and Dieudonné. With their work style and clear goal, "whatever was accepted would be incorporated without any credit to the author. Altogether, a truly unselfish, anonymous, demanding work by people striving to give the best possible exposition of basic mathematics, moved by their belief in its unity and ultimate simplicity [Borel]."

From the first 1935 summer meeting we have the earliest list of topics dates and those responsible for a write-up:

```
Abstract sets (HC)
Algebra (Delsarte)
Real numbers (Dieudonné)
Topology.  Theorems of existence (AW, deP)
Integration
Real functions, series, infinite products
Inequalities:  O and o
Calculus of differential forms
Geometry
Analytic functions:  general part
```

The subject of topology appears in the list and there was a discussion in the spring of 1935 of possible texts that would support their presentation. The classic books by Kerekjarto, Seifert and Threlfall, and Kuratowski were mentioned (none in French). In the first issues of the *Journal de Bourbaki* (later to become *La Tribu*), edited by Delsarte, it was reported that Weil was reading the newly published *Topologie I* of Alexandroff and Hopf, and this text was expected to help them avoid any errors in their presentations. The team writing the topology section, Weil, de Possel, and Henri Cartan are reported in 1936 to be reading (Weil), sleeping (de Possel), or to have written nothing but still thinking about it (Cartan).

The earliest references to 'algebraic topology' in the reports to Bourbaki use the term to refer to duality in topological groups—a discussion later to become 'topological algebra.' In the 1930's the essential points of combinatorial topology was discussed among the Bourbaki: already at the summer conference of 1935, an outline by Weil includes dimension, intersection, linking, degree of mappings, and the index of fixed points among the combinatorial topics. The fundamental group (*groupe de Poincaré*) and covering surfaces were also included. By 1938, Weil made a report on degree and combinatorial topology.

By 1937 there was a plan for the first volumes together with a target date—completion of the first volume by 1.I.1938. The *paquet abstrait* had grown to include the topics of set theory, algebra, set-theoretic topology and abstract integration. In fact, in keeping with the goal of producing a toolbox for mathematicians, the first publication was not a textbook but a list of results (*un fascicule de résultats sans demonstations*) on set theory. Beginning the march toward analysis, it was agreed that set theory served as a basis for future volumes.

Plans for the future volumes were discussed in the *Journal de Bourbaki* until 1940 when the *Journal* was replaced by *La Tribu (Bullétin, apériodique et bourbachique)*. By the time of *La Tribu* the use of the notion of *structure* dominated the formulation of the publishing project. As described later in Bourbaki's entry in Le Lionnais's *Les grands courants de la pensée mathématique*, there were 'mother-structures,' simplest and shared by many mathematical activities; beyond this, one finds 'multiple structures' which blend some number of the mother-structures, for example, topological groups blend the group structure with continuity, while order structures together with algebraic structures give rise to the study of ideals and to integration.

It is this organization by structures that is Bourbaki's lasting legacy. The influence of this notion was far-reaching, even including a psychological discussion of development by

Jean Piaget that cites a correspondence between the mother-structures and a child's first forms of interaction with the world.

Based on the hierarchy of structures, the plan for the *Élements de Mathématique* broke into parts. Part I dealt with the fundamental structures of analysis. In *La Tribu* of 3–15.IX.1940, Part II treated linear analysis, Part III algebraic analysis (to include elliptic functions, the theory of numbers), and Part IV differential topology. We find algebraic topology (that is, combinatorial topology) in this scheme in Part I.

```
Book 1.  Set theory
Book 2.  Algebra
Book 3.  General topology
Book 4.  Topological vector spaces
Book 5.  Elementary techniques of infinitesimal calculus
Book 6.  Integration
Book 7.  Combinatorial topology
Book 8.  Differentials
Book 9.  Calculus of variation
Book 10.  Analytic functions
```

A 25 page report on the shape of books 3 and 7 was titled *Topologia Bourbachica* in which the main topics were I. general topology, 2. topological degree, 3. covering spaces and the Poincaré group, and 4. combinatorial topology (surfaces, Betti groups, Euler-Poincaré formula, indices of vector fields).

Weil was reported to be 'meditating' on the subject of Books 7 and 8, while Ehresmann was working on parts 3 and 4 of Book 7. In late 1941, these books were listed as urgently in need of work, *"la rédaction a le regret ... que ces livres brillent toujours par leur inexistence."* *

The summer meeting of 1942 (in Clermont) presented a new organization of Part I:

```
1.  Sets
2.  Algebra
3.  General Topology
4.  Functions of a real variable (elementary theory)
5.  Combinatorial topology
6.  Topological vector spaces
7.  Differential calculus and manifolds
8.  Integral calculus and differential forms
9.  Analytic functions
```

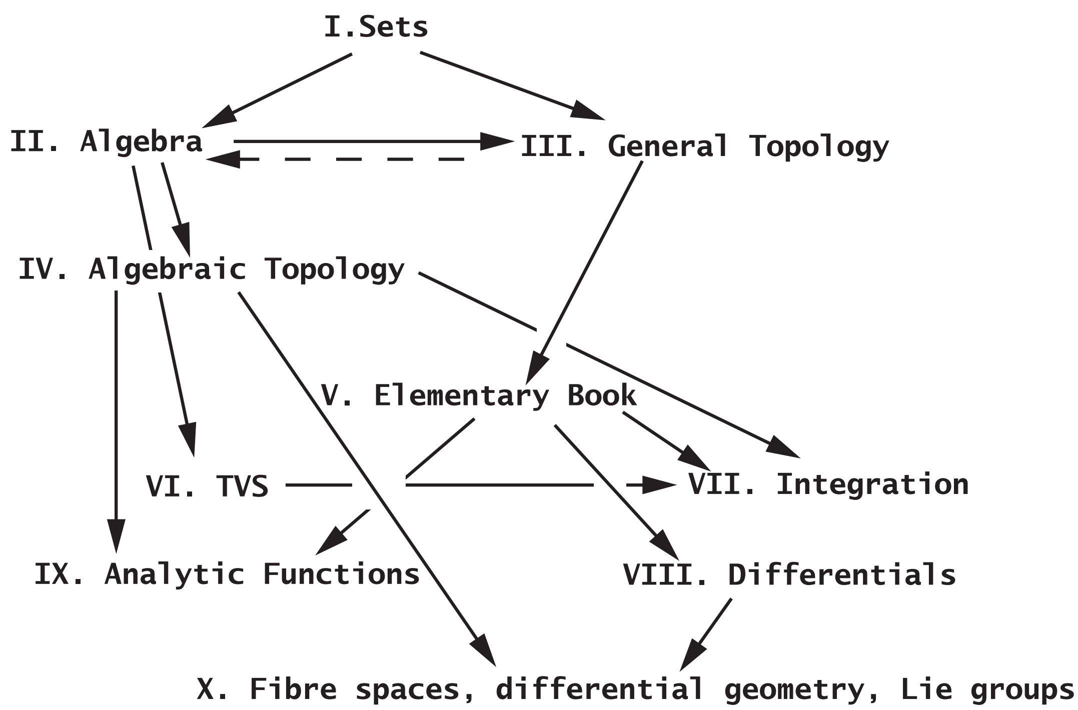
On this plan little progress on algebraic topology took place. In *La Tribu* no. 10 of 10–15.IV.1944, it is reported that *"le récent Congrés Bourbaki que s'est tenu à Paris du 6*

* "the editors regret ... that these books are conspicuous by their nonexistence."

*au 8 Avril 1944 n'on a pas moins réalisé au progrés important et depuis longtemps souhaité par la rédaction: le demarrage de la Topologie algébrique."**

A description of the core of the subject at the time was given, however: a) there should be no Menger theory of curves, no graphs, no Peano continua, no continua; b) a chapter on knots; c) higher homotopy groups and fibre spaces, which they deemed interesting, having a future, but at present in a state *"trop larvaire."* The development of this topic took place during the war with the work of Ehresmann, Feldbau, Cartan, and Leray in France, Steenrod and Whitney in the US, and Hopf and Eckmann in Switzerland (see [McCleary]).

La Tribu of 11–15 July 1945 contains a picture of the dependencies among topics in Part I, once again featuring algebraic topology near the foundations.



11-15.IV.1945 Congress in Paris, from La Tribu no. 8

The 1947 organization of the general plan changed again—the basics now broke up into blocs:

```
                         General Plan
I. Sets, II. Algebra, III. General Topology
Linear bloc:  IV. Functions of a real variable,
    V. Topological vector spaces, VI. Integration, VII. Local differentials
Topologico-differential bloc:  VIII. Algebraic topology, IX. Manifolds,
    X. Lie groups
```

* "the recent Bourbaki Congress that was held in Paris from the 6th to the 8th of April 1944 nevertheless realized important progress, long wished for by the editors: the beginning of algebraic topology."

In 1946, with the end of World War II, and travel easier, Samuel Eilenberg (1913–1998) was drafted as a member, explicitly to prepare a report on algebraic topology. By 1949 there was an 82-page document, `Rapport SEAW sur la topologie préhomologique`, by Eilenberg and Weil, treating the important aspects of the topology of fibre spaces. This densely written report developed the point-set properties of fibre spaces, including some new ideas. For example, they defined the *épiderme* of a space (with the parenthetical remark, *pourquoi pas*); this "skin" is a covering of the space with good properties of extension.

It is the 1950 Grand Plan that gives the familiar list of topics to be treated:

```
Part I.
1.  Sets
2.  Algebra
3.  General topology
3bis.  Geometric topology
4.  Functions of a real variable
5.  Topological vector spaces
6.  Integration
7.  Manifolds
8.  Analytic functions
9.  Lie groups
```

Part II treated Commutative Algebra, Part III Algebraic Topology and its applications, and Part IV Functional Analysis.

The new topic, Geometric topology, was named by Serre to treat topics like coverings, fibre spaces, homotopy, polyhedra, retracts, and the fundamental group. This term went on in the literature, but it did not sit well with the Bourbaki who coined other terms to mock it.

So What Happened?

> *Furthermore, in a time in which indiscriminate use of science and technology threatens the future of the human race, or at least the future of what we now call civilization, it is surely essential that a well integrated report about our mathematical endeavors be written and kept for the use of a later day "Renaissance."*
>
> Pierre Samuel, 1972

Another French enterprise was born about this time that affected the efforts to bring a text on algebraic topology together. In 1948/49, the *Séminaire Henri Cartan* began in Paris. Cartan had just come from Harvard in 1948, having spoken on topological notions, especially what later became sheaves. From its inception the seminar treated topological themes, beginning with basic notions in 48/49 and going on to treat fibre spaces, spectral sequences, sheaves, homology of groups and Eilenberg-Mac Lane spaces, in later years. The level of exposition of these lectures was consistent with the expectations of the Bourbaki, and many of the lectures were given by then current members of Bourbaki.

The discussions of algebraic topology in the earliest plans for *Élements de mathématique* and its appearance among the basic tools for the intended audience of Bourbaki make it clear what status the topic had for the group. However, the development of the subject was so rapid in the post-war years that it could not be understood in the manner that the Bourbaki set as a standard for their published work—that the essential concepts be identified, and the axiomatic basis presented in such a way that the main theorems would be smoothly proven from first principles. The collateral development of homological algebra, which would provide a tool for algebraic topology was finally taken up by Bourbaki, but only in recent times (1980). It is significant that some of this development was carried out by members of Bourbaki itself—Cartan, Eilenberg, Serre, Borel, and others. The press of new discoveries caused Bourbaki to wait.

The published work of Bourbaki does not make for easy reading. The austere style is associated with a monolithic view of the unity of mathematics that is precisely and properly presented in their work. The philosophical cadre of "structure" as guidepost and goal makes for a good explanation of the finished product. However, the record of the archives tells a different story. The austerity is a result of group editing. The course of a document was almost chaotic from first presentation to final publication, spiced by the lively interchanges of mathematicians of the first order, committed to an extraordinary standard.

From the point of view of an enterprise, Bourbaki's *Élements* stands out as an effort to rebuild a mathematical culture, based on a method (the axiomatic method) that was seen to be fruitful, by a collective of gifted mathematicians whose anonymity in their work was offset by the *joie de vivre* the process involved. We should all be so moved to do the same. And I wonder what kind of report on algebraic topology we would produce today.

Bibliography

1. Beaulieu, L., A Parisian café and ten proto-Bourbaki meetings (1934–35), Math. Intell. **15**(1993), 27–35.
2. Borel, A., Twenty-five years with Nicolas Bourbaki, 1949–1973, Notices A.M.S. **45** (1998), 373–380.
3. Bourbaki, N., L'architecture des Mathématiques, in F. Le Lionnais, *Les grands courants de la pensée mathématique*, Cahiers du Sud, Paris, 1948. (Translated by Arnold Dresden in Amer. Math. Monthly **57**(1950), 221–232.)
4. Bourbaki, N., Eléments de Mathématique, Hermann et cie, Paris, 1939–1998.
5. Chouchan, Michèle, Nicolas Bourbaki: Faits et légendes, Éditions du Choix, Agrenteuil Cedex, 1995.
6. Corry, L., Nicolas Bourbaki and the concept of mathematical structure, Synthese **92**(1992), 315–384.
7. Dieudonné, J., The work of Nicholas Bourbaki, Amer. Math. Monthly **77**(1970), 134–145.
8. Halmos, Paul, Nicolas Bourbaki, Scientific American, **196**(1957), 88–99.
9. Hilbert, David, Grundlagen der Geometrie, Teubner, Leipzig, 1899.
10. Houzel, C., Bourbaki und danach, Math. Semesterber. **49**(2002), 1–10.

11. Mashaal, Maurice, Bourbaki, une société secrète de mathématiciens, Belin, Pour la Science, 2002, Paris.
12. McCleary, John, Tortoises and hares: a history of manifolds and bundles, Rend. Circ. Mat. di Palermo Series II, Suppl. **72**(2004), 9–29.
13. Parshall, K. H., Entering the International Arena: E. H. Moore, The University of Chicago, and Hilbert's Grundlagen der Geometrie, in Il Sogno di Galois: Scritti di Storia della matematica dedicati a Laura Toti Rigatelli per il suo 60o compleanno, ed. Raffaella Franci et al. (Siena: Centro Studi della Matematica Medioevale, 2003), pp. 199-227.
14. Samuel, Pierre, Review of Bourbaki (1939–), Book II, Chapters 1–3 (1970), Math Reviews, **43**(1972), #2.
15. Weil, André, The apprenticeship of a mathematician; translated from the French by Jennifer Gage, Basel; Boston: Birkhäuser Verlag, 1992.

Maurice Mashaal

Bourbaki

A Secret Society of Mathematicians

American Mathematical Society

The name Bourbaki is known to every mathematician. Many also know something of the origins of Bourbaki, yet few know the full story. In 1935, a small group of young mathematicians in France decided to write a fundamental treatise on analysis to replace the standard texts of the time. They ended up writing the most influential and sweeping mathematical treatise of the twentieth century, *Les élements de mathématique*.

Maurice Mashaal lifts the veil from this secret society, showing us how heated debates, schoolboy humor, and the devotion and hard work of the members produced the ten books that took them over sixty years to write. The book has many first-hand accounts of the origins of Bourbaki, their meetings, their seminars, and the members themselves. He also discusses the lasting influence that Bourbaki has had on mathematics, through both the *Élements* and the *Seminaires*. The book is illustrated with numerous remarkable photographs.

Preface

Let's say it straightaway: Nicolas Bourbaki is not an individual but a group of mathematicians, almost all French. Formed in 1935, this colorful group consists of a dozen members and remains active today. Despite being unknown to most of the general public, these mathematicians changed the face of mathematics during the sixties and seventies.

Bourbaki neither invented revolutionary techniques nor proved grandiose theorems—and neither did it try to do so. What the group did bring, primarily through its imposing treatise *Éléments de mathématique*, was a new vision of mathematics, a profound reorganization and clarification of its components, lucid terminology and notation, and a distinctive style. It seduced many mathematicians, to such an extent that Bourbaki's philosophy has pervaded the international mathematics community. This has increased the influence of French mathematics throughout the world.

It was not only *Éléments de mathématique* that contributed to Bourbaki's fame. The exceptional quality of the group's members also contributed fundamentally to the quality of the group. André Weil, a key figure in Bourbaki since its creation, was one of the greatest mathematicians of the century. The other mathematicians who helped found the group, including Henri Cartan and Claude Chevalley, were also of international stature. More recent members include prestigious figures of mathematics such as Laurent Schwartz, Alexandre Grothendieck, and Jean-Pierre Serre. The members of Bourbaki all carried out individual mathematical work, earning them a range of high international honors. Several of them, notably Claude Chevalley, Laurent Schwartz, Alexandre Grothendieck, and Roger Godement, also committed much of their time to philosophy and politics. The group's philosophy also found its way into the New Math revolution of the seventies. While Bourbaki regretted this extension of their work, like Antigone they watched in fear as their views broke free from the group and started to lead their own existence.

In addition, Bourbaki constructed an extensive folklore around the group, building on the group's secrecy, its name, its humor and schoolboy pranks, its structure and method of working. Bourbaki is indebted to this folkore for much of the group's success. Yet the use of the word "success" must be qualified. Bourbaki did not complete its mission, and the evolution of mathematics will surely prevent it from ever being completed. Bourbaki and its philosophy had some unwelcome idiosyncracies that their critics readily denounced. And the future survival of this multi-brained mathematician is questionable.

Bourbaki's story is one of shadows and dazzling light. I hope I have shown both equally in this book.

—*Maurice Mashaal*

1

A Group Forms

On December 10, 1934, a handful of young mathematicians gathered in a cafe in the Latin Quarter of Paris. Their goal: to write a textbook on analysis. This is the beginning of a legendary enterprise that would change mathematics forever.

"André Weil and I were both at the University of Strasbourg in 1934. I often talked with him about the course on differential and integral calculus that I was teaching. At that time, a degree in mathematics was composed of certificates in general physics, differential and integral calculus, and classical mechanics. In other words, there was only one certificate in mathematics. [...] Therefore, I had to put as much as possible into what there was. I often wondered about the best way to teach this course because the existing textbooks were not satisfactory, especially when it came to multiple integrals and Stokes' Theorem. I discussed my concerns several times with André Weil. One beautiful day he told me, 'I've had it, we need to fix this for good. We need to write a good textbook on analysis. Then you'll stop complaining!'" This is how Henri Cartan described the origins of Bourbaki to Marian Schmidt in 1982, and André Weil confirms it in his *Souvenirs d'apprentissage* (translated as *The Apprenticeship of a Mathematician*), published in 1991: "One winter's day, in late 1934, I came up with a terrific idea for putting an end to my friend's persistent interrogation. I told him [Henri Cartan], 'There are five or six of us teaching the same course in various universities. Let's get together, fix the problems, and after that I'll finally be free from your questions.' I didn't realize that Bourbaki was born at this moment."

It's impossible to guarantee that human memory accurately recalls events from more than fifty years in the past. Yet even though this conversation from 1934 only gives a hint of the profound reasons for forming the group and of what would happen later, these two quotations do summarize the birth of the Bourbaki group.

The University of Strasbourg, where André Weil and Henri Cartan taught at the beginning of their careers.

The prehistory of Bourbaki's unique mathematical adventure began in the 1920s at l'École Normale Supérieure in Paris. Almost all the future Bourbaki members would study in this institution, a fact that remains true today. The ENS, as it is usually abbreviated, is a member of the Grandes Écoles, the elitist system of institutions of higher education so particular to France. Created in 1794, the ENS originally focused on producing professors of secondary education (including high school and the preparatory classes for entry into the Grandes Écoles). By the end of the nineteenth century, however, the Normaliens (as students of the ENS are called) increasingly turned to jobs in higher education and research. The ENS is composed of two faculties: humanities (with about thirty students per year in the 1920s) and sciences (with about twenty students per year). Admission is based on a highly selective exam following two or three years of preparatory classes. In the 1920s, the admitted students stayed at the ENS for three years. The first two were devoted to university classes and the third was devoted to preparing for the final examinations for teacher certification.

The entrance of l'École Normale Supérieure in 1924.

Bourbaki's Cradle, l'École Normale Supérieure

L'École Normale Supérieure has produced numerous authors, intellectuals, and politicians, including Raymond Aron, Jean-Paul Sartre, and Georges Pompidou. In the sciences as well, it is the ENS (along with l'École Polytechnique) that produces most of the elite of French research. At the beginning of the nineteenth century, most French mathematicians had studied at the École Polytechnique; by the end of that century, most had studied at the ENS. To cite a few examples, Gaston Darboux, Émile Picard, Paul Painlevé, Jacques Hadamard, Élie Cartan, René Baire, and Henri Lebesgue were all Normaliens; however, Henri Poincaré was a Polytechnicien.

It is at the prestigious ENS on the rue d'Ulm that the five future founding members of Bourbaki, Henri Cartan, Claude Chevalley, Jean Delsarte, Jean Dieudonné, and André Weil, met and became friends. Delsarte and Weil arrived in 1922, Cartan in 1923, Dieudonné in 1924, and Chevalley in 1926. These mathematicians took part in the adventures of Bourbaki from the beginning in 1934 and didn't leave until age forced them to. However, they were not the only mathematicians who participated in founding the group. Other mathematicians of the same generation were involved in Bourbaki's first steps to various degrees.

After the founding plenary conference held in July 1935, where the collective pseudonym Bourbaki was adopted, the nine official members of the group were Jean Coulomb, Charles Ehresmann, Szolem Mandelbrojt (the uncle of Benoit Mandelbrot, the father of fractals), René de Possel, and the five founding members already named. Of these nine, only Mandelbrojt was not a Normalien. Born in Poland, Mandelbrojt had completed his doctorate in Paris and had been teaching at the University of Clermont-Ferrand since 1929. Coulomb, who was more of a geophysicist than a mathematician, left the group relatively early, in 1937.

These young mathematicians, each at most thirty years old, held their first work meeting on Monday, December 10, 1934. Initially, their project was to write an analysis textbook to replace the unsatisfactory exisiting books—Édouard Goursat's book in particular—for use in French higher education. The meeting was held over lunch at the cafe A. Capoulade in the Latin Quarter of Paris, near the Pantheon. This cafe, which was located at 63 boulevard Saint-Michel on the corner with rue Soufflot, has since been replaced by one of the many American-style fast-food havens. Gathered around a basement table were Henri Cartan, Claude Chevalley, Jean Delsarte, Jean Dieudonné, René de Possel, and André Weil. Almost all of these

The students of "l'École" in 1924:
1. Henri Cartan,
2. Georges Canguilhem,
3. Jean-Paul Sartre,
4. Jean Dieudonné,
5. Raymond Aron,
6. René de Possel,
7. Charles Ehresmann,
8. Paul Nizan,
9. Louis Néel.

mathematicians held positions at univerisities outside of Paris: Cartan and Weil were at Strasbourg, Delsarte at Nancy, Dieudonné at Rennes, and de Possel at Clermont-Ferrand. They were in Paris for the Julia Seminar, held at 4:30 pm at the newly founded Henri Poincaré Institute on the second and fourth Mondays of each month of the academic year.

What did they say during this first meeting? We can answer this thanks to the work of Liliane Beaulieu, a mathematical historian from Quebec, who had access to the minutes of Bourbaki's meetings. For Weil, who already seemed to have precise ideas about what he and his friends should set out to accomplish together, it was necessary to "establish the content of the differential and integral calculus certificate for twenty-five years by writing a treatise on analysis" and that this treatise would be "as modern as possible." Weil anticipated that the publisher would be Hermann, where he knew the director, Enrique Freymann. Delsarte strongly supported the idea that the book, covering a wide range of material, would be written collectively. Animated discussion about the size of the project, the

The cafe A. Capoulade, where our secret society of mathematicians gathered in 1934.

schedule of publication, and the organization of the work followed. Cartan proposed that the work would be at most 1000 or 1200 pages; Delsarte proposed that the first volumes would be published in at most six months; Weil proposed that, after a few preliminary meetings, the group would create subcommittees in charge of designing each section of the book, and that the group would decide on a precise and final plan during a plenary meeting during the following summer vacation. They also started to discuss the nature and content of the future book, although of course one meeting was not enough to answer all their questions.

The Committee for Writing a Treatise on Analysis, as the group called itself at the time, met ten times between December 1934 and May 1935—about once every two weeks and, like the first meeting, on Mondays at the cafe A. Capoulade. During the second meeting, the group decided that the Committee would include at most nine members. Although this rule was only enforced briefly following the birth of Bourbaki, the number of members never increased to more than about a dozen. In January 1935, even before the founding plenary conference that would take place that summer, Paul Dubreil, Jean Leray, and Szolem Mandelbrojt joined the group; however, Dubreil and Leray quit before the summer. Jean Coulomb replaced Dubreil in April, and Charles Ehresmann replaced Leray in the summer. Arrivals

At the first Bourbaki conference, held in Besse-en-Chandesse in 1935. Standing (left to right): Henri Cartan, René de Possel, Jean Dieudonné, André Weil, and a university lab technician. Seated (left to right): Miriès (a guinea pig, or potential future Bourbaki), Claude Chevalley, Szolem Mandelbrojt.

and departures, which were marked by no ceremony, remained common throughout the history of Bourbaki, and the composition of the group varied continually.

A Modest Project Turns Ambitious

As for the project's nature and content, the initial intention of creating a book used for teaching differential and integral calculus quickly turned into a much more ambitious task. Already during the second meeting, Weil stated that "we must write a book that can be used by anyone: by researchers (both students and professors), by future teachers, by physicists, and by engineers." To accomplish this, the book would need to provide the readers with a collection of mathematical tools "as robust and universal as possible," and the group would need to develop a detailed outline for the book before deciding what tools to present. At the same time, the book would need to simplify the tools as much as possible; that is, to determine the real substance of these tools and to present the most general, and therefore universal, versions of them. This is not the case of most textbooks, where one of the weaknesses is often that fundamental theorems "are presented with a pretty incredible restraint: far

The small village of Besse-en-Chandesse, where the first Bourbaki conference was held.

more hypotheses than necessary are used."

Forming a detailed outline of the book turned out to be a long process, certainly much longer and arduous than the members of Bourbaki initially anticipated. The planning took several years and was very productive. Gradually, the group's extensive reflection and lively discussions led to a new vision of mathematics, a modern way of teaching it and even of doing it. The volumes of the book presented this new vision and had a great influence on the mathematical community, both in France and throughout the world.

The first complete outline was established during the founding plenary conference, which took place from July 10 through July 17, 1935, in Besse-en-Chandesse, a small village about twenty-five miles from Clermont-Ferrand. The main subjects to be included in the book (such as functions of real and complex variables, integrals, differential equations and partial differential equations, and special functions) were approximately the same as in earlier textbooks on analysis. In addition, the group planned to add a small number of more abstract and innovative chapters providing basic concepts of abstract algebra, set theory, and topology, which the group deemed necessary for the coherent presentation of the main material. The entire work was expected to cover 3200 pages, which is three times more than the Committee had anticipated during its first meeting, held just a few months earlier!

Szolem Mandelbrojt, Claude Chevalley, René de Possel, and André Weil (from left to right) during the 1935 conference.

Bourbaki, as the group was named at the Besse-en-Chandesse conference (the first name Nicolas was added later), planned to complete a first draft of the book within a year. While this goal was far from achieved, it did set the group's plans in motion. Of course there were still debates, questions, hesitation, and backtracking, all of which resulted in modifications and refinements of the outline for the book. Over the course of the following years, the "abstract packet" (Bourbaki's name for the sections dealing with abstract algebra and topology, which were to serve as a foundation for the rest of the book) grew relentlessly, while the other chapters were delayed and and even deemphasized. Initially conceived as auxiliary to the other chapters and intended to be as short as possible, the abstract packet became one of the main parts of the project. In fact, Bourbaki's project had become so broad and ambitious that the phrase "treatise on analysis" no longer applied. Thus it was under the title *Éléments de mathématique*—the resemblance to Euclid's *Elements* is no coincidence—that Bourbaki's first volume appeared in 1939-1940. The first volume to appear was *Fascicule des résultats de théorie des ensembles*, a summary of set theoretical results presented without proof.

During World War II, despite the scattering of the Bourbaki members, the group managed to publish three more volumes of *Éléments de mathématique*. Numerous other volumes followed in the decades

René de Possel
The portaits ringed in blue show the future members of Bourbaki at their entrance to l'École Normale Supérieure.

Bourbaki's first work: a summary of results in set theory.

between 1940 and 1970, with the production rate slowing down markedly after this. The most recent volume was published in 1998, while the one before that had been published fifteen years earlier, in 1983.

Éléments de mathématique (which currently contains about 7000 pages dense with definitions, axioms, lemmas, corollaries, and theorems)—and especially a few particular volumes—was a worldwide success and made Bourbaki famous. But the success and fame of the group stem equally from the distinctive way in which the members lived and worked. Such a huge and influential treatise succeeded not only because of the mathematical talent of its authors, but also because of the group's enthusiasm, friendship, schoolboy solidarity, and belief in a common goal.

The Bourbaki group is devoid of any hierarchy. All decisions must be unanimous, and while there is no official voting, anyone can veto a proposal. This system particularly applies to decisions about the treatise. The final version must be accepted by every member, which requires, in most cases, many years of work. The writing process itself is distinctive as well. The group gives the task of writing the rough draft of a given section to one or two members. Once the section is written, it is read out loud to all the members, after which the other members mercilessly criticize the draft. The group then assigns the task of rewriting the section to a different member. The process continues until it converges—sometimes wearily—to a manuscript that the members unanimously accept as ready for publication.

However, the lack of hierarchy doesn't imply that all the members have equal weight in the group. Some members put more into the group; some have more influence. André Weil, who can be considered as the group's first leader, is *primum inter pares*, the first among equals. He was also the target of jokes and mockeries less often than the other members. Even Jean Dieudonné, a vocal member who worked intensely for the group, was aware of Weil's role. Henri Cartan tells how "one day, Dieudonné said [metaphorically], 'I won't drink my *café au lait* before Weil.'" In the more recent history of Bourbaki, key members of the group include Jean-Pierre Serre, Michel Demazure, Pierre Cartier, and Jean-Louis Verdier.

Laurent Schwartz

André Weil, Armand Borel, and Jean-Pierre Serre on the terrace of the hotel in Pelvoux-le-Poët (in the Alps), July 1951.

Conferences in the Country

Bourbaki holds three conferences each year to keep track of progress and make decisions about the future. These meetings usually take place in the countryside, in calm and agreeable places. In recent years, each conference lasted one week; in the past, when summer vacations were longer, each conference lasted two weeks. These conferences, gathering a dozen mathematicians working seven or eight hours a day, seemed to proceed amidst a cheerful hubbub. There were always many people talking loudly at the same time. Jokes rang out as frequently as insults, which were delivered more or less purposefully but always quickly forgotten. The group enjoyed an animated atmosphere of comradeship away from work as well, and

André and Eveline Weil in 1939. André Weil (1906–1998), the brother of the philosopher Simone Weil, was one of the great mathematicians of the twentieth century. He played a central role in Bourbaki, even after he left France in 1941.

many people who spotted the group during one of their conferences called them "a bunch of lunatics"!

On the other hand, this bunch of lunatics behaved (and behaves) much more discreetly outside their conferences. One of the most striking peculiarities of the Bourbaki enterprise is that no outside person is supposed to know who the current members are, what the group is working on, or when and where the conferences take place. During a series of interviews with former Bourbaki members, led by the reporter Michèle Chouchan and aired on France-Culture in 1988, Laurent Schwartz, an early Bourbaki recruit, said that "whenever I was asked if I was a member of Bourbaki, I had to say no. If I wasn't a member, I was speaking the truth, and if I was a member, I was required to say that I was not." Also, at the Bourbaki secretariat, housed in an office at the École Normale Supérieure, one is politely told that Bourbaki doesn't help newspapers or accept interviews, and that it "neither confirms nor denies any information circulating about the group." In short, it is difficult to obtain direct information. Only former members agree to break the silence.

There are several reasons for this tradition (or mania?) of secrecy. According to Bourbaki, the group preserves its secrecy to preserve the collective nature of their enterprise. Bourbaki writes its books as a collective effort, and no member must be allowed to put himself ahead of others, be it for scientific merit or for the collection of royalties received from sales of the group's books. Of course, the secrecy also allows the group to work productively without being disturbed; the fact that the group was most secretive during its golden years from 1950 to 1970 supports this view. This secrecy may also protect the group's members from influential scientists who were skeptical or hostile towards the project, and such people did exist from the very beginning. Withholding the names of the members also strengthens the authority of the group's treatise: the contents of the books appear as the expression of a consensus, with any disagreements among the members of the group remaining invisible. Finally, the secrecy served a social function: strengthening the cohesion of the group and helping to create the myth surrounding it. This is not the least among the charms of Bourbaki.

It appears that Bourbaki was initially less secretive than it became in later years. An example of this lack of secrecy is a letter written on November 17, 1936 by Mandelbrojt, Delsarte, Cartan, Weil, Dieudonné, de Possel, Coulomb, Chevalley, and Ehresmann to the physicist Jean Perrin (at the time the State Undersecretary for Scientific Research). This letter, a grant request, summarizes what Bourbaki was and describes the financial difficulties of the project, which at the time was far from producing royalties. This is what Bourbaki wrote:

"Dear Mr. Perrin:

"Perhaps you have already heard that the mathematicians named above are devoting a large amount of their time to writing a Treatise on Mathematical Analysis, which—we hope—will be the basis of analysis education for the next decade.

"We have adopted a new method of collaboration. We have not restricted ourselves by distributing the task of writing distinct pieces of the whole work; instead, each chapter is assigned to one of us only after we have discussed and prepared it at great length, and the resulting draft is again discussed in detail after every member has studied it. Each chapter is revised at least once, and often several times. Through this method, we hope to obtain a truly collective work, one with a profoundly unified character.

"It is clear that the method we have chosen is not a lazy one, and that it requires numerous meetings and frequent travel. Furthermore, much of our time is spent on reproducing and distributing various drafts and manuscripts. For the past two years, we ourselves have taken responsibility for financing these activities. Now that the State officially sponsors scientific research, we thought that it might be able to help us. It is this financial help, Mr. Perrin,

André Weil and Armand Borel in 1955. André Weil was formidably caustic. To a mathematician who asked him, "Can I ask you a stupid question?" André Weil replied, "You just did..."

Samuel Eilenberg (1913–1998), one of the fathers of category theory, at a Bourbaki conference. Eilenberg was one of the few non-French members of the group.

that we request of you with all respect.

"Following is a summary of the funds we are requesting. Seven of us live outside of Paris and we hold at least four meetings each year; allowing on average 250 francs per person per meeting to cover travel and living expenses yields a total of 7000 francs. In addition, we estimate that 3000 francs are needed for various expenses, including correspondance, office supplies, typing, duplication, and especially the payment of assistants who copy formulas by hand onto the duplicated manuscripts, a service which is as vital as it is expensive.

"A grant of 10000 francs per year for four or five years would allow us to complete our project successfully.

"Mr. Perrin, we hope that you will accept our profound deference and respectful admiration."

The letter closed with the signature of Mandelbrojt, who was the oldest member of the group; the pseudonym Bourbaki was never mentioned. The grant was awarded for a year and subsequently renewed.

A Guinea Pig Must Prove Himself

Bourbaki quickly gained new members through a unique recruitment procedure. Bourbaki often invited one or two additional people to the group's conferences. In some cases, these were "guinea pigs," potential future recruits who were being tested. Once a promising young mathematician had been spotted, Dieudonné explained at a conference in Romania in 1968, "the procedure was to invite him to a conference as a 'guinea pig,' it's a traditional method. You all know about laboratory guinea pigs, who are exposed to all sorts of viruses. Well, it's the same sort of thing, the poor guy is thrown into the Bourbaki discussion sessions, and not only must he understand, he must also participate. If he's silent, he simply won't be invited again."

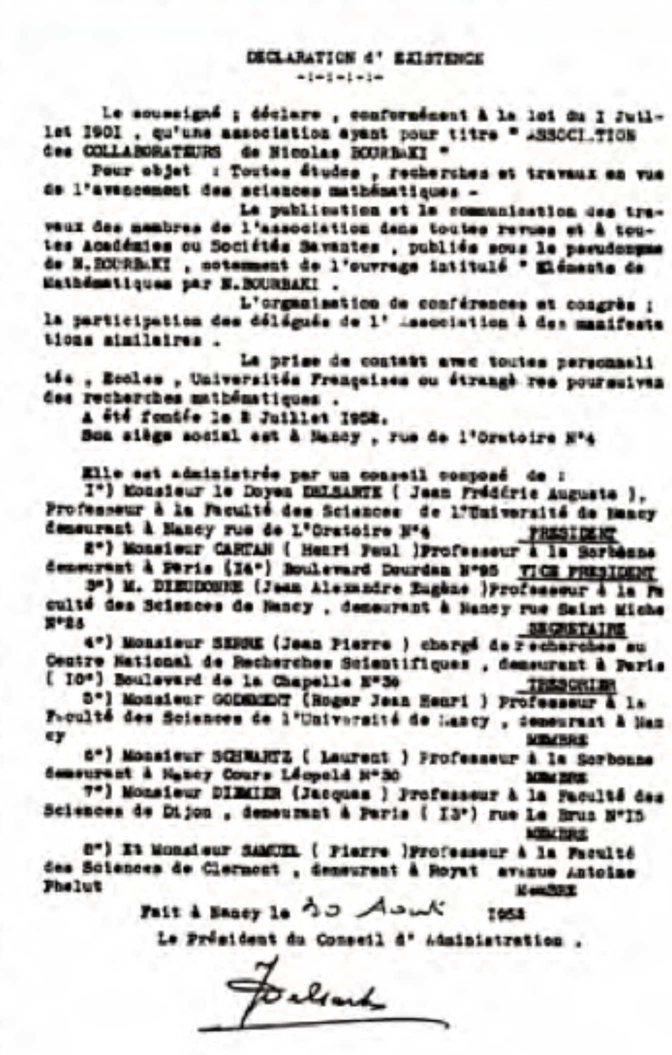
If he is recruited, it means that other members must leave, since Bourbaki hardly ever grew beyond a dozen members. Sometimes, departures stem from more or less significant disagreements about the group's goals and methods of working. For example, Paul Dubreil left even before the founding plenary conference of 1935, in part because of unavailability (his wife had been hired at Rennes, while he was at Nancy) and in part because, as he confided to Liliane Beaulieu, he disliked the disorderly discussions and preferred working with one or two people on precise problems than to working in a large group. Jean Leray, an eminent French mathematician, also left early because his proposals had been impertinently shot down and also because, according to Beaulieu, he opposed the group's principle of starting from the very basics of mathematics. Also, as in any social group, there were discords of a more personal nature. This appears to be the case of René de Possel: in 1937, his wife Éveline became Éveline Weil.

The Association of Collaborators of Nicolas Bourbaki

To manage the group's various practical issues, and especially its financial affairs, Bourbaki made itself official by creating a not-for-profit organization. This "Association of Collaborators of Nicolas Bourbaki" was registered on August 20, 1952 at the prefecture in Nancy. Its headquarters was Jean Delsarte's house, at 4 rue de l'Oratoire in Nancy. In 1966, the headquarters was moved to Jean-Pierre Serre's address in Paris, and it has been at l'École Normale Supérieure, at 45 rue d'Ulm in Paris, since 1972.

One should differentiate the Bourbaki group from the Assocation of Collaborators of Nicolas Bourbaki. Although the members are the same, the association is only an administrative body dealing with the outside world. Therefore, the statutes of the association don't mention the specific rules about secrecy, unanimity, retirement at fifty, the recruitment process, and so forth.

The first board of directors consisted of Jean Delsarte (president), Henri Cartan (vice-president), Jean Dieudonné (secretary), Jean-Pierre Serre (treasurer), Roger Godement, Laurent Schwartz, Jacques Dixmier, and Pierre Samuel. In 1995, the general assembly of October 20 was presided by Bernard Tessier, with the other members present being Arnaud Beauville, Gérard Ben Arous, Daniel Bennequin, Patrick Gérard, Guy Henniart, Pierre Julg, Olivier Mathieu, Joseph Oesterlé, Marc Rosso, Georges Skandalis, and Jean-Christophe Yoccoz.



DECLARATION d' EXISTENCE
-!-!-!-!-

Le soussigné ; déclare , conformément à la loi du I Juillet I90I , qu'une association ayant pour titre " ASSOCIATION des COLLABORATEURS de Nicolas BOURBAKI "

Pour objet : Toutes études , recherches et travaux en vue de l'avancement des sciences mathématiques -

La publication et la communication des travaux des membres de l'association dans toutes revues et à toutes Académies ou Sociétés Savantes , publiés sous le pseudonyme de N.BOURBAKI , notamment de l'ouvrage intitulé " Eléments de Mathématiques par N.BOURBAKI .

L'organisation de conférences et congrès ; la participation des délégués de l' Association à des manifestations similaires .

La prise de contact avec toutes personnalités , Ecoles , Universités Françaises ou étrangères poursuivant des recherches mathématiques .

A été fondée le 8 Juillet 1952.

Son siège social est à Nancy , rue de l'Oratoire N°4

Elle est administrée par un conseil composé de :

I°) Monsieur le Doyen DELSARTE (Jean Frédéric Auguste), Professeur à la Faculté des Sciences de l'Université de Nancy demeurant à Nancy rue de L'Oratoire N°4 PRESIDENT

2°) Monsieur CARTAN (Henri Paul)Professeur à la Sorbonne demeurant à Paris (I4°) Boulevard Jourdan N°95 VICE PRESIDENT

3°) M. DIEUDONNE (Jean Alexandre Eugène)Professeur à la Faculté des Sciences de Nancy , demeurant à Nancy rue Saint Michel N°25 SECRETAIRE

4°) Monsieur SERRE (Jean Pierre) chargé de recherches au Centre National de Recherches Scientifiques , demeurant à Paris (I0°) Boulevard de la Chapelle N°30 TRESORIER

5°) Monsieur GODEMENT (Roger Jean Henri) Professeur à la Faculté des Sciences de l'Université de Nancy , demeurant à Nancy MEMBRE

6°) Monsieur SCHWARTZ (Laurent) Professeur à la Sorbonne demeurant à Nancy Cours Léopold N°30 MEMBRE

7°) Monsieur DIXMIER (Jacques) Professeur à la Faculté des Sciences de Dijon , demeurant à Paris (I3°) rue Le Brun N°I5 MEMBRE

8°) Et Monsieur SAMUEL (Pierre)Professeur à la Faculté des Sciences de Clermont , demeurant à Royat avenue Antoine Phelut MEMBRE

Fait à Nancy le 20 Août 1952

Le Président du Conseil d' Administration .

Jean-Pierre Serre

Retirement at Fifty

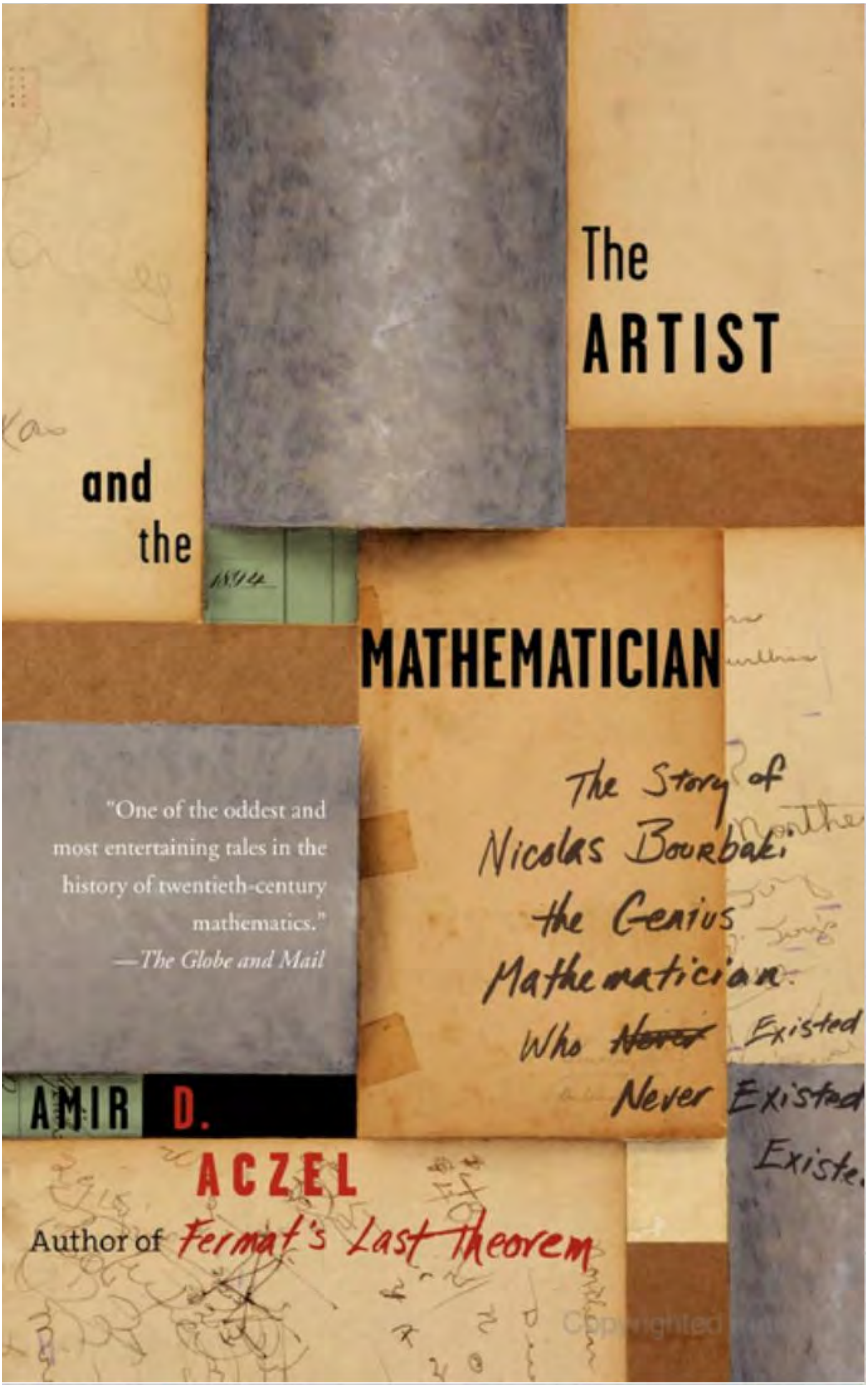
Another reason for departure from Bourbaki was age, since members must retire at fifty. This rule was initiated by Weil when the founding members were approaching this age. At the end of Dieudonné's birthday lunch during the summer conference of 1956, held at Sallières-les-Bains in the department of Drôme, Henri Cartan read a letter from Weil (who had been living in the United States since 1947 and only came to one in three conferences), which proposed a "gradual disappearance of the founding members."

According to Weil, two considerations motivated his proposal. The first was that "the number of conference participants is sometimes too high to allow fruitful work," and the second was that "the founding members are 'more equal than others,' and thus the more recent members 'don't feel obliged to take on full responsibility' in group discussions." But surely an additional motivating factor was the widely held belief that a mathematician is most brilliant and productive in his youth. As Dieudonné said in 1968, "[...] a mathematician above fifty can still be a very good mathematician, still very productive, but he rarely can adapt to new ideas, ideas developed by people 25 or 30 years younger than him. And an enterprise like Bourbaki would like to be eternal [...]."

Thus most of the founding members left around 1956–1958, and the fifty year rule was retained in the following years. In some sense, Bourbaki is eternally young! However, the retirement of a member does not mean that he is entirely cut off from the active members, since for example the retired members continue to receive *La Tribu* ("The Tribe"), the group's internal newsletter describing the proceedings of its conferences.

In Bourbaki's sixty-five years of existence, the group has involved about forty members, almost all of which were French and students of the ENS. Among the exceptions (who all spoke French) are Samuel Eilenberg, an American of Polish origin who founded category theory with Saunders MacLane around 1942 and collaborated with Bourbaki for fifteen years until 1966; Armand Borel, a Swiss living in the United States and a member for twenty years until 1973; and Serge Lang, an American of French origin.

In addition to the mathematicians already cited, notable members of Bourbaki include Arnaud Beauville (b. 1947), Claude Chabauty (1910–1990), Alain Connes (b. 1947), Jacques Dixmier (b. 1924), Adrien Douady (b. 1935), Roger Godement (b. 1921), Alexandre Grothendieck (b. 1928), Jean-Louis Koszul (b. 1921), Charles Pisot (1909–1984), Pierre Samuel, and Bernard Teissier. It would be diffi-


The
ARTIST
and
the
MATHEMATICIAN
The Story of
Nicolas Bourbaki
the Genius
Mathematician
Who ~~Never~~ Existed
Never Existed
"One of the oddest and
most entertaining tales in the
history of twentieth-century
mathematics."
—The Globe and Mail
AMIR D.
ACZEL
Author of Fermat's Last Theorem

A real-life mathematical mystery

Nicolas Bourbaki was perhaps the greatest mathematician of the twentieth century. Responsible for the emergence of the "new math" that swept through American and foreign education systems in the middle of the century, Bourbaki originated the modern concept of the mathematical proof and is credited with the introduction of rigor into the discipline. It can be said that no working mathematician in the world today is free of the influence of Nicolas Bourbaki's seminal work.

This is the story of both Bourbaki and the world that created him. And it is the story of an elaborate intellectual joke—because Bourbaki, the author of dozens of acclaimed papers and one of the foremost mathematicians of his day—never existed.

one

THE DISAPPEARANCE

IN AUGUST 1991, Alexandre Grothendieck, widely viewed as the most visionary mathematician of the twentieth century, a man with insight so deep and a mind so penetrating that he has often been compared with Albert Einstein, suddenly burned 25,000 pages of his original mathematical writings. Then, without telling a soul, he left his house and disappeared into the Pyrenees.

Twice during the mid-1990s Grothendieck briefly met with a couple of mathematicians who had discovered his hiding place high in these rugged and heavily wooded mountains separating France from Spain. But soon he severed even these new ties with the outside world and disappeared again into the wilderness. And for ten years now, no one has reported seeing him. His mail keeps piling up uncollected at the mathematics department of the University of Montpellier in southern France, the last academic institution with which he had been associated. The few individuals whom he had once trusted to forward him the select pieces of mail he did want to receive no longer have any way of making contact with him. His children have not heard from him in many years, and two of his relatives who live in southwest France—not far from the Pyrenees—and with whom he had had limited, sporadic contact, have not had a word from him in years. They do not even know whether he is still alive. It seems as if Alexandre Grothendieck has simply vanished off the face of the earth.

During his most active period as a mathematician, from the 1950s to around 1970, a period when he completely reworked important areas of modern mathematics, lectured extensively on his pathbreaking research, organized leading seminars, and interacted with the most important mathematicians from around the world, Alexandre Grothendieck had been closely associated with the work of Nicolas Bourbaki. And some have surmised that Grothendieck's inexplicable disappearance into the Pyrenees was somehow connected with his relationship with Bourbaki.

Nicolas Bourbaki was the greatest mathematician of the twentieth century. Since his appearance on the world stage in the 1930s, and until his declining years as the century drew to a close, Bourbaki has changed the way we think about mathematics and, through it, about the world around us. Nicolas Bourbaki is responsible for the emergence of the "New Math" that swept through American education in the middle of the century as well as the educational systems of other nations; he is credited with the introduction of rigor into mathematics; and he was the originator for the modern concept of a mathematical proof. Furthermore, the many volumes of Bourbaki's published treatise on "the elements of mathematics" form a towering foundation for much of the modern mathematics we do today. It can be said that no working mathematician in the world today is free of the influence of the seminal work of Nicolas Bourbaki.

But what was the nature of the relationship between Bourbaki and Alexandre Grothendieck, and who is Nicolas Bourbaki?

A mathematical project may entail work on certain aspects of a theory by a French mathematician, a Japanese one, an Englishwoman, a Dutchman, and so on. Who, in fact, even looks at the national origin of a mathematician working in a group on some topic? So the age of a French—and completely male—group of mathematicians is simply over. Who wants to belong to a national or gender-specific group doing anything nowadays?

IF HE IS alive, Grothendieck is still hiding in the Pyrenees. He is hiding very well now, since attempts to find him have failed. Apparently, this is what he wants: to be alone, to write and destroy his own writing, and not to have any connection with people other than grocery store clerks or laborers who might do occasional work for him.

His disappearance and his anger with the world symbolize the demise of Bourbaki. For Grothendieck alone held the great hope for the future of Bourbaki. Grothendieck and his work were the next stage in the program of abstraction and generalization in mathematics that Bourbaki had embarked upon. Alexandre Grothendieck was the human incarnation of the essence of Bourbaki, of the ideals that Bourbaki strove for in mathematics, for here was a man who actually *thought* in great generalities, and for whom axiomatic thinking was natural. Grothendieck's oeuvre was, in fact, all about *structure*, so that the structuralist idealism of Bourbaki found in Grothendieck's work its finest manifestation.

But the man who left mathematics for political causes came up completely empty-handed. Fellow mathematicians were disappointed with his loss of interest in the discipline, and they saw that his political actions were ineffectual and completely useless. Realizing his own failure, Grothendieck drew further apart from the world around him. Perhaps reflecting his own parents' disillusionment with revolution following the defeat of the Republicans in the Spanish Civil War, Grothendieck realized that he could not change society with his political activity. He therefore chose to withdraw from society.

The man who could bring the most beautiful structures into mathematics could not bring structure to political reality. And perhaps politics, unlike other disciplines, is not amenable to the methods of structuralism. At any rate, political change was not to be brought on through the work of Grothendieck, and as the need for political change diminished, the man withdrew from society altogether. It is hoped that he enjoys the peaceful and pristine environment in which he now lives.

French universities, who were now taking positions at universities, and Cartan and Weil were only two members of this group. This was the post-World War I generation of French mathematicians.

The Great War had decimated France's intelligentsia and academia. The numbers of French university graduates who died in that war was staggering. About half the graduating class of the years 1910–1916 died during the First World War.[2] There followed a period of stagnation when new mathematical and educational ideas were sorely needed.

Weil had had excellent connections with German mathematicians, forged during his numerous travels in Germany, and he knew that, despite its difficulties, German mathematics was forging forward: new ideas abounded there, and groups of mathematicians were working together at German universities, such as the excellent research groups Weil had visited in Göttingen and Berlin a few years earlier.

There was no reason why something similar should not be possible in France. The stark inadequacy of prewar textbooks and syllabi in all French universities—along with Cartan's ceaseless questions—were the catalysts that gave Weil the idea of organizing all his colleagues and taking common action. Together, they would do away with the past and start afresh.

Weil called a meeting for noon on December 10, 1934, to be held at the *Café Grill-Room A. Capoulade* (which doesn't exist anymore; today it is the site of a fast-food restaurant) at 63 Boulevard Saint-Michel, at the corner of the Boulevard Saint-Michel and rue Soufflot, across from the beautiful Luxembourg Gardens and near the Panthéon, in the heart of the university area in the Latin Quarter of Paris.

In the below-ground level of this café came together the young mathematicians Henri Cartan, Claude Chevalley, Jean Delsarte, Jean Dieudonné, René de Possel, and André Weil. Among them, they represented the universities of Strasbourg, Nancy, Rennes, and Clermont-Ferrand. All of them happened to be in the capital for a mathematics conference held at the newly opened Institut Henri Poincaré. They embarked on the ambitious project of setting the mathematics curriculum for courses of calculus and mathematical analysis offered in all the universities in France.

This group, collectively, would become Nicolas Bourbaki. So Nicolas Bourbaki was never a single person, even though there had once been a general by the name of Bourbaki. These six young mathematicians meeting in Paris—and others who would join them, while some of them would leave—would in a way continue the pranks of Paris student life by inventing a person and by founding a secret society. They would create for their invented person a family and a family history—hence the visiting cards found in Weil's room in Finland, and the invitation to Bourbaki's daughter's wedding. All of this, including a baptism and a baptismal certificate, godparents, and so on, would be fabricated in order to create a persona—an amalgam of the identities of the individual members of the group. But all this was still in the future.

That December day in 1934 the group lunched on cabbage soup and grilled meats served with *endives braisées* or *pommes soufflées*.[3] André Weil opened the meeting by saying that the goal of the common undertaking was "to define for the next twenty-five years the syllabus for the certificate in differential and integral calculus, by writing collectively a treatise on analysis."

Answering to the fact that such a syllabus may already have existed, he quickly added, "Of course, this treatise will be as *fifteen*

THE DEATH OF BOURBAKI AND HIS LEGACY

THE OULIPO WRITER and mathematician Raymond Queneau had this to say about Bourbaki in 1962[1]

> *Il a nécessairement vieilli, votre fictif mathematician, il doit avoir pris du retard. Eh bien! Non, Bourbaki n'a pas vieilli parce qu'il ne* peut pas *vieiilir.*
>
> Translation: "He must have gotten old, your fictitious mathematician, he had to have surrendered to time. Well, no. Bourbaki has not gotten old because he *cannot* get old."

Bourbaki can't get old because he never existed. But many mathematicians believe that Bourbaki is not old—he is dead. The reasons for this belief are that Bourbaki has not published in many years; he exerted his influence on mathematics in the past, but does not count as a player in the field today; and finally, of the official members in Bourbaki today, none is among the highest ranked French mathematicians. Bourbaki, for all intents and purposes, is dead.

eight

THE ACHIEVEMENTS OF BOURBAKI

ACCORDING TO THE French historian of mathematics Denis Guedj:

> Animated by a profound faith in the unity of mathematics, and wishing to be 'universal mathematicians', Bourbaki undertook to derive the whole of the mathematical universe from a single starting point.

That starting point was the theory of sets.

Twenty-five centuries earlier, Euclid based his entire *Elements* on the elementary notions of point, line, circle, and other geometrical concepts, and on this foundation he constructed his entire system of mathematics. Bourbaki wanted to found their own system of mathematics in a similar way to that of Euclid. Their stated aim was to produce a new set of Euclid's *elements*—"to last for the next 2,000 years." In that spirit, they named their own series of books they planned to publish, "Elements de mathématique." They purposely spelled mathematics in the singular, in French: mathematique, rather than the usual way, mathematiques. They chose this innovative new word to stress the *unity* of mathematics. They were striving under their own names. They were being awarded prizes and medals, the same medals they themselves had fought so strongly against in the early days of the group. The Bourbaki organization had become too powerful; and yet it now lacked its main purpose since all its goals had by now been achieved.

Bourbaki lost an incredibly important opportunity to remake its oeuvre in the new form of the theory of categories, something that would have better suited the study of structures than did the old theory of sets with its myriad problems and inadequacies. In addition, new theories in mathematics, such as chaos and fractals, as well as René Thom's catastrophe theory, emerged and demonstrated that structuralism is not absolutely necessary for doing good mathematics.*

Bourbaki had a chance, through the work of Grothendieck and his students, to refound modern mathematics on the theory of categories, but Bourbaki missed that chance. In part, this missed opportunity led to the demise of Bourbaki. For mathematics remained based on a flawed system, set theory, rather than something that would have been much more appropriate. Bourbaki had a possibility in the late 1960s to redirect itself toward a more ambitious goal. Mathematics had evolved further and reached a place in which new foundations could be laid for the discipline. This direction would have been possible because of the work of one man: Alexandre Grothendieck.

But a rift materialized in the midst of the group, and the members of Bourbaki could not agree to follow in the new direction. Grothendieck left the group in anger. Eventually, as we have seen, he would completely withdraw from society.

There also followed a serious battle for copyright with the group's publisher, a fight that sapped all of Bourbaki's energies over several years and left it ineffective and ineffectual. The legal battle began in 1975, when Bourbaki sued its publisher, Hermann, for the copyright of its books and the right to translate and publish them abroad. The group hired an excellent attorney, Blausteil, who had been the attorney for the heirs of Picasso. The matter was legally settled in 1980, and Bourbaki won. The group was now allowed to sign a publication deal with a new publisher. The members of Bourbaki could now republish all their older texts, and much work went into revising these texts, which left little time and energy for writing new books. Bourbaki had thus won the battle, but lost the war. This legal battle left it weary, exhausted, low on funds, and internally divided. It was left with no aim and no steam.[5]

Thus Bourbaki's pace was slowed down considerably, precisely because the group did win the legal battle for the older books. There followed several failed projects, such as a treatise on functions of several complex variables. Having lost its momentum, the group began its slide toward oblivion.

MOST MATHEMATICIANS AGREE that Bourbaki is dead. None of its present members are among the top forty or so mathematicians in France; the group does not publish; and mathematicians in general do not consider the group "alive." The questions that remain are when Bourbaki died, and why.

One theory is that the student revolt that swept France in 1968, affecting all of society, politicizing people whose political agendas had until then remained latent, and bringing the entire economy to a standstill, spelled the end of Bourbaki. According to this theory, Grothendieck and many other mathematicians in France took to politics as a more important occupation than doing mathematics and thereafter, mathematics was not the same.

Pierre Cartier, on the other hand, even though he took part in the political upheavals of the 1960s and 1970s and even went to Vietnam, as did Grothendieck, to protest against the war, believes that Bourbaki lived throughout most of the twentieth century. And, as he puts it, the twentieth century can be viewed as starting and ending with Sarajevo; so Bourbaki died with the war in Yugoslavia. Certainly Bourbaki was dead when the Berlin Wall came tumbling down toward the end of the century.[6]

These views tie together Bourbaki's work in mathematics and the group's political activity and agendas. This, perhaps, is a good description of Bourbaki's life and death, since the group has always been politically active, as individuals as well as collectively. But what else contributed to its demise? Personalities and their interactions played a major role in the dissolving of the Bourbaki group. André Weil was the main founder of Bourbaki, and its most active member. But Weil left France permanently during the war, and his non-presence in France made things difficult for the group. Following his departure, Bourbaki lacked a leader and a key figure.

It is true that Weil returned to France sporadically, and attended most of the congresses after the war, but he was not in Europe day-to-day, hence he was unable to participate in many activities that took place outside of the formal meetings. In those days, the 1960s and 1970s, communication across the ocean was not as easy and free-flowing as it is today, with fax machines and the Internet connecting people in real time. Long-distance phone calls, the only real-time connection in those days, were very expensive. In addition, there was the psychological factor of Weil's separation from the group.

Bourbaki tried to bridge this gap by listing Nicolas Bourbaki's official affiliation as the University of Nancago. Nancago was the fusion of Nancy, where some of the members were on the faculty, and Chicago, where André Weil was at that time. But this didn't help very much.

Additionally, as is the case with every organization, there was a danger that the members would pursue personal goals rather than the common goals of the group. Or rather, in the case of Bourbaki, the pertinent question is: How much effort does a member devote to the common goals, as compared with the efforts he devotes to personal goals? This question is especially important when applied to the leader of the group. In the case of André Weil, it is very likely that personal goals dictated the agenda. Weil remained in the United States for personal reasons even when the threat of war was over and while clearly Bourbaki needed him in France.

This was in stark contrast with the actions of another member, Claude Chevalley, who indeed returned to France and resumed his activities within the group. From his own memoirs, one gets the impression that Weil cared about his personal life more than he did the common good.

The conflict between personal goals and common goals clearly had its effect and swept with it the entire group of Bourbaki. We know, for example, that Jean Dieudonné did most of the writing of the final drafts of the books. Such an effort required great discipline and putting aside personal goals. For, however we may view the situation, a group that takes only collective credit for its works cannot last very long.

Everyone, at some point, wants to go it alone and get all the glory that's there to be had. Perhaps a comparison with the breakup of the Beatles is not out of place here. And the conflict between personal goals and common ones is most evident works, it often seems that the writers have turned abstraction into a goal, and rigor takes over and leaves absolutely no place for intuition or even general understanding. Bourbaki does not, in general, use pictures or other visual aids: thus it completely discourages any understanding of the material that is "human." For how many people can see what is going on in a proof simply by following very technical details? Most mathematicians rely, at some point, on some kind of a mental picture of what is going on in a theorem or a proof.

It is this blind reliance on technical details, strict adherence to rigorous procedures, and an over appreciation of generalities at the expense of the specific case, the picture, the intuition, the human idea of a mathematical problem which have made Bourbaki disliked by some mathematicians. Having brought us their ideas of abstraction and generality and structure, Bourbaki lost its lead as the world of mathematics moved forward.

THE IDEAS OF French philosopher Michel Serres demonstrate the problems with Bourbaki's structures. In 1961, Serres realized, as did Piaget, that the idea of structure emanated directly from the mathematics of Bourbaki.[8] He defined structure mathematically as follows: "A structure is an operational set with an undefined meaning . . . grouping any number of elements, whose content is not specified, and a finite number of relations whose nature is not specified."[9] Serres noted that in algebra this definition of structure needs no explanation, and that in mathematics, therefore, the definition of structure is the truest.

But soon Serres diverged from the point of view that structures are important. Philosophy required methods and

A mathematical project may entail work on certain aspects of a theory by a French mathematician, a Japanese one, an Englishwoman, a Dutchman, and so on. Who, in fact, even looks at the national origin of a mathematician working in a group on some topic? So the age of a French—and completely male—group of mathematicians is simply over. Who wants to belong to a national or gender-specific group doing anything nowadays?

IF HE IS alive, Grothendieck is still hiding in the Pyrenees. He is hiding very well now, since attempts to find him have failed. Apparently, this is what he wants: to be alone, to write and destroy his own writing, and not to have any connection with people other than grocery store clerks or laborers who might do occasional work for him.

His disappearance and his anger with the world symbolize the demise of Bourbaki. For Grothendieck alone held the great hope for the future of Bourbaki. Grothendieck and his work were the next stage in the program of abstraction and generalization in mathematics that Bourbaki had embarked upon. Alexandre Grothendieck was the human incarnation of the essence of Bourbaki, of the ideals that Bourbaki strove for in mathematics, for here was a man who actually *thought* in great generalities, and for whom axiomatic thinking was natural. Grothendieck's oeuvre was, in fact, all about *structure*, so that the structuralist idealism of Bourbaki found in Grothendieck's work its finest manifestation.

But the man who left mathematics for political causes came up completely empty-handed. Fellow mathematicians were disappointed with his loss of interest in the discipline, and they saw that his political actions were ineffectual and completely useless. Realizing his own failure, Grothendieck drew further apart from the world around him. Perhaps reflecting his own parents' disillusionment with revolution following the defeat of the Republicans in the Spanish Civil War, Grothendieck realized that he could not change society with his political activity. He therefore chose to withdraw from society.

The man who could bring the most beautiful structures into mathematics could not bring structure to political reality. And perhaps politics, unlike other disciplines, is not amenable to the methods of structuralism. At any rate, political change was not to be brought on through the work of Grothendieck, and as the need for political change diminished, the man withdrew from society altogether. It is hoped that he enjoys the peaceful and pristine environment in which he now lives.

“She shaves herself”

The Barber paradox is attributed to the British philosopher Bertrand Russell. It highlights a fundamental problem in mathematics, exposing an inconsistency in the basic principles on which mathematics is founded.

The barber paradox asks us to consider the following situation:

In a village, the barber shaves everyone who does not shave himself, but no one else.

The question that prompts the paradox is this:

Who shaves the barber?

No matter how we try to answer this question, we get into trouble.

If we say that the barber shaves himself, then we get into trouble. The barber shaves only those who do not shave themselves, so if he shaves himself then he doesn’t shave himself, which is self-contradictory.

If we say that the barber does not shave himself, then problems also arise. The barber shaves everyone who does not shave himself, so if he doesn’t shave himself then he shaves himself, which is again absurd.

Even if we try to get clever, saying that the barber is a woman, we do not evade the paradox. If the barber is a woman, then she either shaves herself (and so is one of the people not shaved by the barber), or does not shave herself (and so is one of the people shaved by the barber).

Both cases, then, are impossible; the barber can neither shave himself nor not shave himself. The question ‘Who shaves the barber?’ is unanswerable.

Grothendieck: To ask the question is to commit a category error.

Bourbaki should be redone based on category theory--at least until something more fundamental is discovered

Barber paradox

From Wikipedia, the free encyclopedia

This article is about a paradox of self-reference. For an unrelated paradox in the theory of logical conditionals with a similar name, introduced by Lewis Carroll, see the Barbershop paradox.

The **Barber paradox** is a puzzle derived from Russell's paradox. It was used by Bertrand Russell himself as an illustration of the paradox, though he attributes it to an unnamed person who suggested it to him.[1] It shows that an apparently plausible scenario is logically impossible.

Contents

The paradox

Suppose there is a town with just one male barber; and that every man in the town keeps himself clean-shaven: some by shaving themselves, some by attending the barber. It seems reasonable to imagine that the barber obeys the following rule: He shaves all and *only* those men in town who do *not* shave themselves.

Under this scenario, we can ask the following question: Does the barber shave himself?

Asking this, however, we discover that the situation presented is in fact impossible:

- If the barber does not shave himself, he must abide by the rule and shave himself.
- If he does shave himself, according to the rule he will not shave himself.

History

This paradox is often attributed to Bertrand Russell (e.g., by Martin Gardner in *Aha!*). It was suggested to him as an alternate form of Russell's paradox,[1] which he had devised to show that set theory as it was used by Georg Cantor and Gottlob Frege contained contradictions. Of the Barber paradox, Russell said the following:

> That contradiction [Russell's paradox] is extremely interesting. You can modify its form; some forms of modification are valid and some are not. I once had a form suggested to me which was not valid, namely the question whether the barber shaves himself or not. You can define the barber as "one who shaves all those, and those only, who do not shave themselves." The question is, does the barber shave himself? In this form the contradiction is not very difficult to solve. But in our previous form I think it is clear that you can only get around it by observing that the whole question whether a class is or is not a member of itself is nonsense, i.e. that no class either is or is not a member of itself, and that it is not even true to say that, because the whole form of words is just noise without meaning.
>
> – Bertrand Russell, *The Philosophy of Logical Atomism*

This point is elaborated further under Applied versions of Russell's paradox.

In prolog

In Prolog, one aspect of the Barber paradox can be expressed by a self-referencing clause:

```
shaves(barber, X) :- male(X), not shaves(X,X).
male(barber).
```

where negation as failure is assumed. If we apply the stratification test known from Datalog, the predicate shaves is exposed as unstratifiable since it is defined recursively over its negation.

In first-order logic

$$(\exists x)(barber(x) \land (\forall y)(\neg shaves(y,y) \Leftrightarrow shaves(x,y)))$$

This sentence is unsatisfiable (a contradiction) because of the universal quantifier. The universal quantifier y will include every single element in the domain, including our infamous barber x. So when the value x is assigned to y, the sentence can be rewritten to $\neg shaves(x,x) \Leftrightarrow shaves(x,x)$, which simplifies to $shaves(x,x) \land \neg shaves(x,x)$, a contradiction.

In literature

In his book *Alice in Puzzleland*, the logician Raymond Smullyan had the character Humpty Dumpty explain the apparent paradox to Alice. Smullyan argues that the paradox is akin to the statement "I know a man who is both five feet tall and six feet tall," in effect claiming that the "paradox" is merely a contradiction, not a true paradox at all, as the two axioms above are mutually exclusive.

A paradox is supposed to arise from plausible and apparently consistent statements; Smullyan suggests that the "rule" the barber is supposed to be following is too absurd to seem plausible.

Non-paradoxical variations

A modified version of the Barber paradox is frequently encountered in the form of a brainteaser puzzle or joke. The joke is phrased nearly identically to the standard paradox, but omitting a detail that allows an answer to escape the paradox entirely. For example, the puzzle can be stated as occurring in a small town whose barber claims: I shave *all* and *only* the men in our town who do not shave themselves. This version omits the gender of the barber, so a simple solution is that *the barber is a woman*. The barber's claim applies to only "men in our town," so there is no paradox if the barber is a woman (or a gorilla, or a child, or a man from some other town--or anything other than a "man in our town"). Such a variation is not considered to be a paradox at all: The true Barber paradox requires the contradiction arising from the situation where the barber's claim applies to himself.

Notice that the paradox still occurs if we claim that the barber is a man in our town with a beard. In this case, the barber does not shave himself (because he has a beard); but then according to his claim (that he shaves all men who do not shave themselves), he must shave himself.

In a similar way, the paradox still occurs if the barber is a man in our town who *cannot* grow a beard. (Perhaps he lost all facial hair follicles in a painful accident). Once again, he does not shave himself (because he has no hair on his face), but that implies that he does shave himself.

In music

- Chip Hop (rap) artist MC Plus+ refers to the Barber paradox in his song "Man Vs Machine" from the album Chip Hop. He uses it to defeat his own fictional AI opponent, Max Flow, in a rap-battle.
- Dub legend King Tubby claims that he himself shaves the Barber, in the tune "I Trim The Barber".

References

1. ^ *a b* *The Philosophy of Logical Atomism*, reprinted in *The Collected Papers of Bertrand Russell, 1914-19*, Vol 8., p. 228

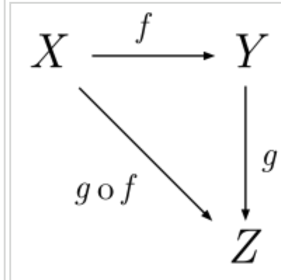
External links

- Proposition of the Barber's Paradox (http://www.umsl.edu/~siegel/SetTheoryandTopology/TheBarber.html)
- Joyce, Helen. "Mathematical mysteries: The Barber's Paradox." (http://plus.maths.org/issue20/xfile/index.html) *Plus,* May 2002.
- Edsger Dijkstra's take on the problem (http://www.cs.utexas.edu/users/EWD/transcriptions/EWD09xx/EWD923a.html)

Category theory

From Wikipedia, the free encyclopedia

In mathematics, **category theory** deals in an abstract way with mathematical structures and relationships between them: it abstracts from *sets* and *functions* to *objects* linked in diagrams by *morphisms* or *arrows*.



A category with objects X, Y, Z and morphisms *f* and *g*

One of the simplest examples of a category (which is a very important concept in topology) is that of groupoid, defined as a category whose arrows or morphisms are all invertible. Categories now appear in most branches of mathematics and also in some areas of theoretical computer science where they correspond to types and mathematical physics where they can be used to describe vector spaces. Category theory provides both with a unifying notion and terminology. Categories were first introduced by Samuel Eilenberg and Saunders Mac Lane in 1942–45, in connection with algebraic topology.

Category theory has several faces known not just to specialists, but to other mathematicians. A term dating from the 1940s, "general abstract nonsense", refers to its high level of abstraction, compared to more classical branches of mathematics. Homological algebra is category theory in its aspect of organising and suggesting manipulations in abstract algebra. Diagram chasing is a visual method of arguing with abstract "arrows" joined in diagrams. Note that arrows between categories are called functors, subject to specific defining commutativity conditions; moreover, categorical diagrams and sequences can be defined as functors (viz. Mitchell, 1965). An arrow between two functors is a natural transformation when it is subject to certain naturality or commutativity conditions. Both functors and natural transformations are key concepts in category theory, or the " real engines" of category theory. To paraphrase a famous sentence of the mathematicians who founded category theory: 'Categories were introduced to define functors, and functors were introduced to define natural transformations'. Topos theory is a form of abstract sheaf theory, with geometric origins, and leads to ideas such as pointless topology. A topos can also be considered as a specific type of category with two additional topos axioms.

Contents

Background

The study of categories is an attempt to *axiomatically* capture what is commonly found in various classes of related *mathematical structures* by relating them to the *structure-preserving functions* between them. A systematic study of category theory then allows us to prove general results about any of these types of mathematical structures from the axioms of a category.

Consider the following example. The class **Grp** of groups consists of all objects having a "group structure". More precisely, **Grp** consists of all sets *G* endowed with a binary operation satisfying a certain set of axioms. One can proceed to prove theorems about groups by making logical deductions from the set of axioms. For example, it is immediately proved from the axioms that the identity element of a group is unique.

Instead of focusing merely on the individual objects (e.g., groups) possessing a given structure, category theory emphasizes the morphisms – the structure-preserving mappings – *between* these objects; it turns out that by studying these morphisms, we are able to learn more about the structure of the objects. In the case of groups, the morphisms are the group homomorphisms. A group homomorphism between two groups "preserves the group structure" in a precise sense – it is a "process" taking one group to another, in a way that carries along information about the structure of the first group into the second group. The study of group homomorphisms then provides a tool for studying general properties of groups and consequences of the group axioms.

A similar type of investigation occurs in many mathematical theories, such as the study of continuous maps (morphisms) between topological spaces in topology (the associated category is called **Top**), and the study of smooth functions (morphisms) in manifold theory.

If one axiomatizes relations instead of functions, one obtains the theory of allegories.

Functors

Abstracting again, a category is *itself* a type of mathematical structure, so we can look for "processes" which preserve this structure in some sense; such a process is called a functor. A functor associates to every object of one category an object of another category, and to every morphism in the first category a morphism in the second.

In fact, what we have done is define a category *of categories and functors* – the objects are categories, and the morphisms (between categories) are functors.

By studying categories and functors, we are not just studying a class of mathematical structures and the morphisms between them; we are studying the *relationships between various classes of mathematical structures*. This is a fundamental idea, which first surfaced in algebraic topology. Difficult *topological* questions can be translated into *algebraic* questions which are often easier to solve. Basic constructions, such as the fundamental group or fundamental groupoid (http://planetphysics.org/encyclopedia/FundamentalGroupoidFunctor.html) of a topological space, can be expressed as fundamental functors (http://planetphysics.org/encyclopedia/FundamentalGroupoidFunctor.html) to the category of groupoids in this way, and the concept is pervasive in algebra and its applications.

Natural transformation

Abstracting yet again, constructions are often "naturally related" – a vague notion, at first sight. This leads to the clarifying concept of natural transformation, a way to "map" one functor to another. Many important constructions in mathematics can be studied in this context. "Naturality" is a principle, like general covariance in physics, that cuts deeper than is initially apparent.

Historical notes

In 1942–45, Samuel Eilenberg and Saunders Mac Lane were the first to introduce categories, functors, and natural transformations as part of their work in topology, especially algebraic topology. Their work was an important part of the transition from intuitive and geometric homology to axiomatic homology theory. Eilenberg and Mac Lane later wrote that their goal was to understand natural transformations; in order to do that, functors had to be defined, which required categories.

Stanislaw Ulam, and some writing on his behalf, have claimed that related ideas were current in the late 1930s in Poland. Eilenberg was Polish, and studied mathematics in Poland in the 1930s. Category theory is also, in some sense, a continuation of the work of Emmy Noether (one of Mac Lane's teachers) in formalizing abstract processes; Noether realized that in order to understand a type of mathematical structure, one needs to understand the processes preserving that structure. In order to achieve this understanding, Eilenberg and Mac Lane proposed an axiomatic formalization of the relation between structures and the processes preserving them.

The subsequent development of category theory was powered first by the computational needs of homological algebra, and later by the axiomatic needs of algebraic geometry, the field most resistant to being grounded in either axiomatic set theory or the Russell-Whitehead view of united foundations. General category theory, an extension of universal algebra having many new features allowing for semantic flexibility and higher-order logic, came later; it is now applied throughout mathematics.

Certain categories called topoi (singular *topos*) can even serve as an alternative to axiomatic set theory as a foundation of mathematics. These foundational applications of category theory have been worked out in fair detail as a basis for, and justification of, constructive mathematics. More recent efforts to introduce undergraduates to categories as a foundation for mathematics include Lawvere and Rosebrugh (2003) and Lawvere and Schanuel (1997).

Categorical logic is now a well-defined field based on type theory for intuitionistic logics, with applications in functional programming and domain theory, where a cartesian closed category is taken as a non-syntactic description of a lambda calculus. At the very least, category theoretic language clarifies what exactly these related areas have in common (in some abstract sense).

Categories, objects and morphisms

A *category* C consists of the following three mathematical entities:

- A class ob(C), whose elements are called *objects*;
- A class hom(C), whose elements are called morphisms or maps or *arrows*. Each morphism f has a unique *source object* a and *target object* b. We write $f: a \to b$, and we say "f is a morphism from a to b". We write hom(a, b) (or Hom(a, b), or $\hom_C(a, b)$, or Mor(a, b), or $C(a, b)$) to denote the *hom-class* of all morphisms from a to b.
- A binary operation $\circ$, called *composition of morphisms*, such that for any three objects a, b, and c, we have $\hom(a, b) \times \hom(b, c) \to \hom(a, c)$. The composition of $f: a \to b$ and $g: b \to c$ is written as $g \circ f$ or gf (some authors write fg), governed by two axioms:
  - Associativity: If $f : a \to b$, $g : b \to c$ and $h : c \to d$ then $h \circ (g \circ f) = (h \circ g) \circ f$, and
  - Identity: For every object x, there exists a morphism $1_x : x \to x$ called the *identity morphism for x*, such that for every morphism $f : a \to b$, we have $1_b \circ f = f = f \circ 1_a$.

From these axioms, it can be proved that there is exactly one identity morphism for every object. Some authors deviate from the definition just given by identifying each object with its identity morphism.

Relations among morphisms (such as $fg = h$) are often depicted using commutative diagrams, with "points" (corners) representing objects and "arrows" representing morphisms.

Properties of morphisms

Some morphisms have important properties. A morphism $f : a \to b$ is:

- a monomorphism (or *monic*) if $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$ for all morphisms $g_1, g_2 : x \to a$.
- an epimorphism (or *epic*) if $g_1 \circ f = g_2 \circ f$ implies $g_1 = g_2$ for all morphisms $g_1, g_2 : b \to x$.
- an isomorphism if there exists a morphism $g : b \to a$ with $f \circ g = 1_b$ and $g \circ f = 1_a$.[1]
- an endomorphism if $a = b$. end(a) denotes the class of endomorphisms of a.
- an automorphism if f is both an endomorphism and an isomorphism. aut(a) denotes the class of automorphisms of a.

Functors

Functors are structure-preserving maps between categories. They can be thought of as morphisms in the category of all (small) categories.

A (**covariant**) functor F from a category C to a category D, written $F{:}C \to D$, consists of:

- for each object x in C, an object $F(x)$ in D; and
- for each morphism $f : x \to y$ in C, a morphism $F(f) : F(x) \to F(y)$,

such that the following two properties hold:

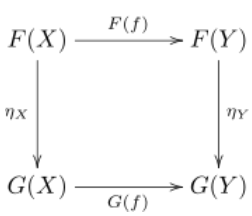
- For every object x in C, $F(1_x) = 1_{F(x)}$;
- For all morphisms $f : x \to y$ and $g : y \to z$, $F(g \circ f) = F(g) \circ F(f)$.

A **contravariant** functor $F: C \to D$, is like a covariant functor, except that it "turns morphisms around" ("reverses all the arrows"). More specifically, every morphism $f : x \to y$ in C must be assigned to a morphism $F(f) : F(y) \to F(x)$ in D. In other words, a contravariant functor is a covariant functor from the opposite category C^{op} to D.

Natural transformations and isomorphisms

A *natural transformation* is a relation between two functors. Functors often describe "natural constructions" and natural transformations then describe "natural homomorphisms" between two such constructions. Sometimes two quite different constructions yield "the same" result; this is expressed by a natural isomorphism between the two functors.

If F and G are (covariant) functors between the categories C and D, then a natural transformation from F to G associates to every object x in C a morphism $\eta_x : F(x) \to G(x)$ in D such that for every morphism $f : x \to y$ in C, we have $\eta_y \circ F(f) = G(f) \circ \eta_x$; this means that the following diagram is commutative:



The two functors F and G are called *naturally isomorphic* if there exists a natural transformation from F to G such that η_x is an isomorphism for every object x in C.

Universal constructions, limits, and colimits

Using the language of category theory, many areas of mathematical study can be cast into appropriate categories, such as the categories of all sets, groups, topologies, and so on. These categories surely have some objects that are "special" in a certain way, such as the empty set or the product of two topologies, yet in the definition of a category, objects are considered to be atomic, i.e., we *do not know* whether an object A is a set, a topology, or any other abstract concept – hence, the challenge is to define special objects without referring to the internal structure of those objects. But how can we define the empty set without referring to elements, or the product topology without referring to open sets?

The solution is to characterize these objects in terms of their relations to other objects, as given by the morphisms of the respective categories. Thus, the task is to find *universal properties* that uniquely determine the objects of interest. Indeed, it turns out that numerous important constructions can be described in a purely categorical way. The central concept which is needed for this purpose is called categorical *limit*, and can be dualized to yield the notion of a *colimit*.

Equivalent categories

It is a natural question to ask: under which conditions can two categories be considered to be "essentially the same", in the sense that theorems about one category can readily be transformed into theorems about the other category? The major tool one employs to describe such a situation is called *equivalence of categories*, which is given by appropriate functors between two categories. Categorical equivalence has found numerous applications in mathematics.

Further concepts and results

The definitions of categories and functors provide only the very basics of categorical algebra; additional important topics are listed below. Although there are strong interrelations between all of these topics, the given order can be considered as a guideline for further reading.

- The functor category D^C has as objects the functors from C to D and as morphisms the natural transformations of such functors. The Yoneda lemma is one of the most famous basic results of category theory; it describes representable functors in functor categories.
- Duality: Every statement, theorem, or definition in category theory has a *dual* which is essentially obtained by "reversing all the arrows". If one statement is true in a category C then its dual will be true in the dual category C^{op}. This duality, which is transparent at the level of category theory, is often obscured in applications and can lead to surprising relationships.
- Adjoint functors: A functor can be left (or right) adjoint to another functor that maps in the opposite direction. Such a pair of adjoint functors typically arises from a construction defined by a universal property; this can be seen as a more abstract and powerful view on universal properties.

Higher-dimensional categories

Many of the above concepts, especially equivalence of categories, adjoint functor pairs, and functor categories, can be situated into the context of *higher-dimensional categories*. Briefly, if we consider a morphism between two objects as a "process taking us from one object to another", then higher-dimensional categories allow us to profitably generalize this by considering "higher-dimensional processes".

For example, a (strict) 2-category is a category together with "morphisms between morphisms", i.e., processes which allow us to transform one morphism into another. We can then "compose" these "bimorphisms" both horizontally and vertically, and we require a 2-dimensional "exchange law" to hold, relating the two composition laws. In this context, the standard example is **Cat**, the 2-category of all (small) categories, and in this example, bimorphisms of morphisms are simply natural transformations of morphisms in the usual sense. Another basic example is to consider a 2-category with a single object; these are essentially monoidal categories. Bicategories are a weaker notion of 2-dimensional categories in which the composition of morphisms is not strictly associative, but only associative "up to" an isomorphism.

This process can be extended for all natural numbers n, and these are called n-categories. There is even a notion of *ω-category* corresponding to the ordinal number ω.

Higher-dimensional categories are part of the broader mathematical field of higher-dimensional algebra,a concept introduced by Ronald Brown. For a conversational introduction to these ideas, see John Baez, 'A Tale of n-categories' (1996). (http://math.ucr.edu/home/baez/week73.html)

See also

- List of category theory topics
- Important publications in category theory
- Glossary of category theory
- Domain theory
- Enriched category theory
- Higher category theory
- Timeline of category theory and related mathematics
- Higher-dimensional algebra

Notes

1. **^** Note that a morphism that is both epic and monic is not necessarily an isomorphism! For example, in the category consisting of two objects A and B, the identity morphisms, and a single morphism f from A to B, f is both epic and monic but is not an isomorphism.

References

Freely available online:

- Adámek, Jiří, Herrlich, Horst, & Strecker, George E. (1990) *Abstract and concrete categories* (http://katmat.math.uni-bremen.de/acc/acc.htm) . John Wiley & Sons. ISBN 0-471-60922-6.
- Freyd, Peter J. (1964) *Abelian Categories. (http://www.tac.mta.ca/tac/reprints/articles/3/tr3abs.html)* New York: Harper and Row.
- Michael Barr and Charles Wells (1999) *Category Theory Lecture Notes. (http://folli.loria.fr/cds/1999/library/pdf/barrwells.pdf)* Based on their book *Category Theory for Computing Science*.
- ——— (2002) *Toposes, triples and theories*. (http://www.cwru.edu/artsci/math/wells/pub/ttt.html) Revised and corrected translation of *Grundlehren der mathematischen Wissenschaften* (Springer-Verlag, 1983).
- Leinster, Tom (2004) *Higher operads, higher categories (http://www.maths.gla.ac.uk/~tl/book.html)* (London Math. Society Lecture Note Series 298). Cambridge Univ. Press.
- Schalk, A. and Simmons, H. (2005) *An introduction to Category Theory in four easy movements. (http://www.cs.man.ac.uk/~hsimmons/BOOKS/CatTheory.pdf)* Notes for a course offered as part of the MSc. in Mathematical Logic, Manchester University.
- Turi, Daniele (1996–2001) *Category Theory Lecture Notes. (http://www.dcs.ed.ac.uk/home/dt/CT/categories.pdf)* Based on Mac Lane (1998).
- Goldblatt, R (1984) *Topoi: the Categorial Analyis of Logic* (http://dlxs2.library.cornell.edu/cgi/t/text/text-idx?c=math;cc=math;view=toc;subview=short;idno=Gold010) A clear introduction to categories, with particular emphasis on the recent applications to logic.
- A. Martini, H. Ehrig, and D. Nunes (1996) *Elements of Basic Category Theory* (http://citeseer.ist.psu.edu/martini96element.html) (Technical Report 96-5, Technical University Berlin)

Other:

- Awodey, Steven (2006). *Category Theory* (Oxford Logic Guides 49). Oxford University Press.
- Borceux, Francis (1994). *Handbook of categorical algebra* (Encyclopedia of Mathematics and its Applications 50-52). Cambridge Univ. Press.
- Freyd, Peter J. & Scedrov, Andre (http://www.cis.upenn.edu/~scedrov/) , (1990). *Categories, allegories* (North Holland Mathematical Library 39). North Holland.
- Hatcher, William S. (1982). *The Logical Foundations of Mathematics*, 2nd ed. Pergamon. Chpt. 8 is an idiosyncratic introduction to category theory, presented as a first order theory.
- Lawvere, William, & Rosebrugh, Robert (2003). *Sets for mathematics*. Cambridge University Press.
- Lawvere, William, & Schanuel, Steve (1997). *Conceptual mathematics: a first introduction to categories*. Cambridge University Press.
- Mac Lane, Saunders (1998). *Categories for the Working Mathematician*. 2nd ed. (Graduate Texts in Mathematics 5). Springer-Verlag.
- ——— and Garrett Birkhoff (1967). *Algebra*. 1999 reprint of the 2nd ed., Chelsea. ISBN 0-8218-1646-2. An introduction to the subject making judicious use of category theoretic concepts, especially commutative diagrams.
- May, Peter (1999). *A Concise Course in Algebraic Topology*. University of Chicago Press, ISBN 0-226-51183-9.
- Pedicchio, Maria Cristina & Tholen, Walter (2004). *Categorical foundations* (Encyclopedia of Mathematics and its Applications 97). Cambridge University Press.
- Taylor, Paul (1999). *Practical Foundations of Mathematics*. Cambridge University Press. An introduction to the connection between category theory and constructive mathematics.
- Pierce, Benjamin (1991). *Basic Category Theory for Computer Scientists*. MIT Press.

External links

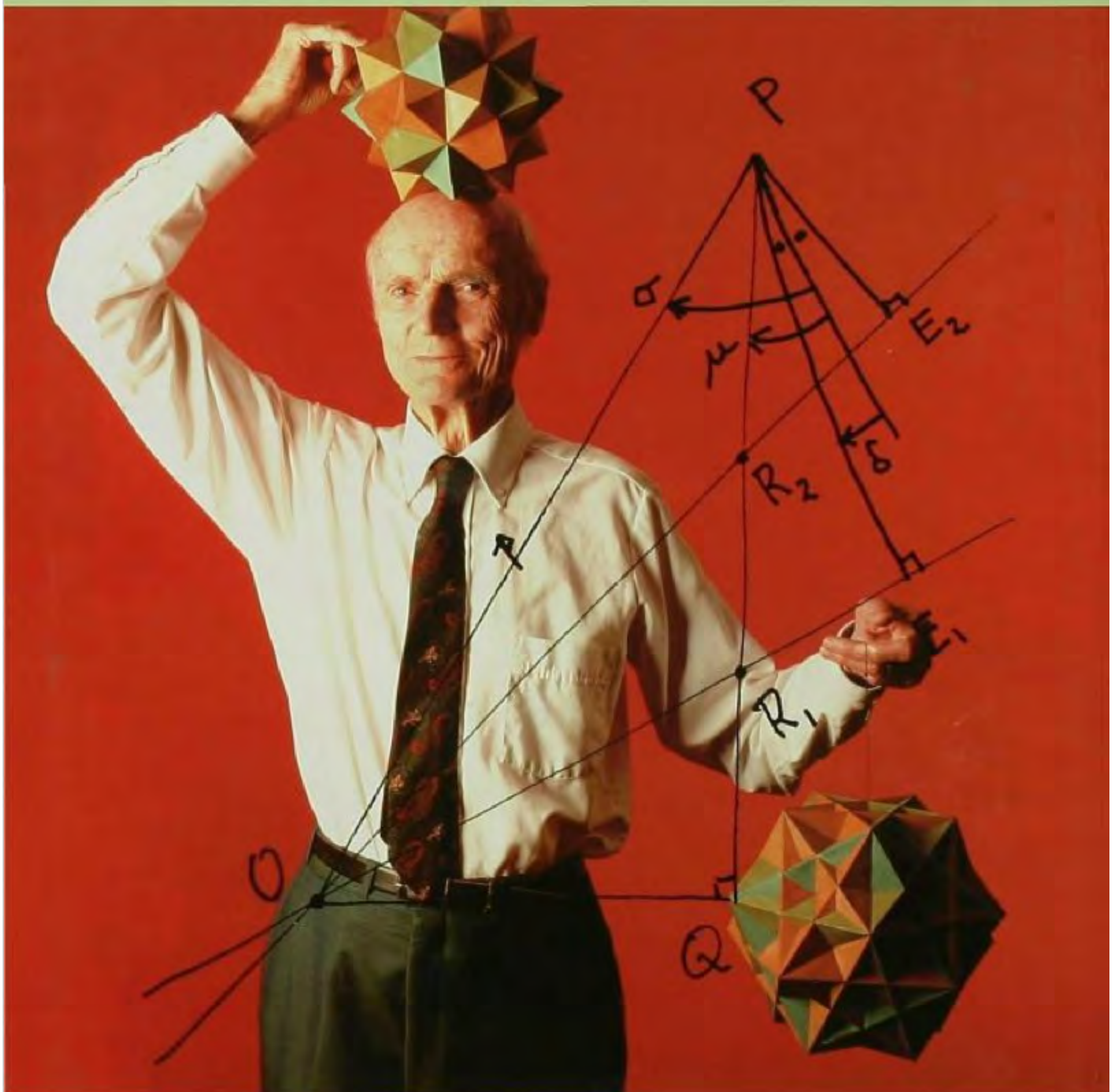
- Chriss Hillman, Categorical primer (http://citeseer.ist.psu.edu/cache/papers/cs/23543/http:zSzzSzwww-aix.gsi.dezSz~appelzSzskriptezSzotherzSzcategories.pdf/hillman01categorical.pdf) , formal introduction to Category Theory.
- J. Adamek, H. Herrlich, G. Stecker, Abstract and Concrete Categories-The Joy of Cats (http://katmat.math.uni-bremen.de/acc/acc.pdf)
- Stanford Encyclopedia of Philosophy: "Category Theory (http://plato.stanford.edu/entries/category-theory/) " -- by Jean-Pierre Marquis. Extensive bibliography.
- Homepage of the Categories mailing list, (http://www.mta.ca/~cat-dist/categories.html) with extensive resource list.
- Baez, John, 1996,"The Tale of *n*-categories. (http://math.ucr.edu/home/baez/week73.html) " An informal introduction to higher order categories.
- The catsters (http://www.youtube.com/user/TheCatsters) " a Youtube channel about category theory.
- Category Theory (http://planetmath.org/?op=getobj&from=objects&id=5622) on PlanetMath
- Categories, Logic and the Foundations of Physics (http://categorieslogicphysics.wikidot.com/) , Webpage dedicated to the use of Categories and Logic in the Foundations of Physics.
- Interactive Web page (http://www.j-paine.org/cgi-bin/webcats/webcats.php) which generates examples of categorical constructions in the category of finite sets. Written by Jocelyn Paine (http://www.j-paine.org/)

Down with Euclid!

Death to Triangles!

KING OF INFINITE SPACE

Donald Coxeter, the Man Who Saved Geometry



Crying `Death to Triangles!' a generation of mathematicians tried to eliminate geometry in favor of algebra. Were it not for Donald Coxeter, they might have succeeded.

FOR A LOT OF PEOPLE, talk of geometry induces flashbacks to high school math class anxieties-fumbling with compasses and protractors and memorizing triangle theorems. So the idea that geometry was once on the brink of extinction as an academic subject does not elicit much regret or nostalgia.

We owe a lot to geometry, however. Geometric algorithms generate the aerodynamic curves of Mercedes-Benz sedans and Boeing aircraft, make possible computer-animated films such as Pixar's ``The Incredibles," and power the data-mining technology used by Amazon.com to find patterns in massive amounts of raw information. Geometry governs in things small (the molecular structure of pharmaceuticals) and large (the shape of our universe).

Yet despite these modern applications, geometry was, for much of the 20th century, a discipline very much in jeopardy. It was deemed by a generation of mathematicians to be old-fashioned, a fine recreation for idling away a lazy afternoon, but in essence little more than a trivial tinkering with toys. Modern mathematics was all about prickly algebraic symbols and undulating equations-impenetrable hieroglyphs with no diagrams, no shapes.

The task of fending off these attacks fell to H.S.M. ``Donald" Coxeter, the greatest classical geometer of the last century. Through his lifelong work as geometry's apostle, Coxeter, who died in 2003 at 96 (prematurely by his measure-his lifelong vegetarianism guaranteed he should live to 100, he figured), became known by his followers around the world as ``the man who saved geometry" in a mathematical era characterized by all things algebraic, abstract, and austere.

Fifty years ago this summer, Coxeter was summoned by the Mathematical Association of America on a roving lecture tour through the United States. He traveled as far north as Fairbanks, Alaska, as far west as Stanford, Calif., and east to New York City, speaking with a missionary's zeal to schoolteachers and any other willing listeners.

Coxeter lectured about ``the beauteous properties of triangles," about circles and spheres, and about the Platonic solids: the tetrahedron, cube, octahedron, icosahedron, and dodecahedron. According to a recent cosmological hypothesis (and a similar theory put forth by Plato) the dodecahedron is a potential model for the shape of the universe-bound by 12 walls, each the shape of a pentagon.

Coxeter had a special affection for the Platonic solids. Educated at Cambridge, in his native England, he spent most of his professional life at the University of Toronto. But before coming to Toronto he did a two-year stint at Princeton. It was there that he launched his career, choosing as his specialty polytopes, an extension of the Platonic solids in higher dimensions.

But just as Coxeter set out upon his career, classical geometry-with its emphasis on shapes and diagrams-was being supplanted by modern mathematicians' penchant for algebra.

A secret society of the créme de la créme of French mathematicians epitomized the shift in the mathematical zeitgeist of the early 20th century. Writing under the pseudonym Nicolas Bourbakis, the collective set out in the 1930s to rewrite the history of mathematics in one grand mathematical treatise, and perhaps the most distinctive feature of their work was the absence of diagrams.

The Bourbakis espoused mathematical rationality and rigor. They believed the subjective and fallible visual sense was easily led astray, falling victim to impressionistic reasoning. In 1959, at a conference in France addressing the need to overhaul the French education system, Jean Dieudonné, a founding member of the Bourbakis and the group's scribe, infamously proclaimed: ``Down with Euclid! Death to Triangles!"

Eventually, the Bourbakis way of mathematics pervaded the Western world, reaching even into grade schools with the Sputnik-motivated New Math reforms of the 1960s, which aimed to improve students' performance and to ensure America was not left in the scientific dust by the Soviet Union. Instead of shapes, children studied axioms and set theory.

As a consequence, mathematical and scientific investigation suffered from what Walter Whiteley, a great admirer of Coxeter and director of applied mathematics at York University in Toronto, calls the ``geometry gap." Whiteley's thesis holds that when the areas of the brain that process visual and geometric concepts fall into disuse, the realms of mathematics and science suffer as well.

So Coxeter set out to make the case for the visual geometric approach, using a number of tactics.

On a popular level, he proselytized for the classical geometric treasures he loved, praising their simple beauty and symmetry. The elegance of his talks and essays gained him an avid following around the world, a fan base of professional and amateur geometers alike who became just as passionate about classical geometry as he was.

Coxeter, for example, was muse to artist M.C. Escher, famous for works like ``Ascending and Descending," a seemingly precarious building of stairs winding in an infinite loop. Coxeter and Escher became friends in the 1950s, and the mathematician's work assisted the artist in his quest to convincingly capture the concept of infinity. (Escher was known to say, ``I'm Coxetering today!") It was a unique collaboration, since Escher, who had no mathematical background, drew entirely from Coxeter's geometric diagrams for inspiration, referring to the accompanying equations as Coxeter's ``hocus-pocus math."

But Coxeter did more than just popularize. He also managed to reinvigorate the discipline through his academic research. He injected a modern relevance, allowing classical geometry to transcend its old-fashioned origins and find far-reaching applications in both mathematics and the sciences.

Specifically, Coxeter classified the symmetries of polytopes, which allowed him to translate these geometric entities into algebra, thus building a powerful bridge between algebra and geometry.

Coxeter also invented mathematical tools-now called Coxeter groups, Coxeter numbers, and Coxeter diagrams-which shed new light on symmetry, broadening and deepening its study. His best-selling book, ``Regular Polytopes," became a classic. ``It's like the bible for me. I refer to it all the time," said John Ratcliffe from Vanderbilt University in Nashville, who has one copy at work and another in his study at home for late-night consultations.

Symmetry underpins all mathematics-an equation being an expression of perfect balance. And it's present throughout nature as well-everything from the smallest spec of a subatomic particle, to a sunflower, to the shape of the universe and the hypothetical parallel universes that mirror our own exhibits symmetry. Applications of Coxeter's work pop up in just about any niche of mathematics or science that explores patterns and symmetry. In fact, Coxeter groups are being used in conjunction with Einstein's equations in the search for supersymmetry, which holds promise to unravel the puzzle of string theory.

The visual and the algebraic perspectives are in constant flux in the mathematical and scientific disciplines. ``The battle between geometry and algebra is like the battle between the sexes," said Sir Michael Atiyah, honorary professor of mathematics at Edinburgh University. ``It's the kind of problem that never disappears. It'll never be dead, and it will never get solved. The question is, `What is the right balance?"'

``It goes back and forth, and not in an accidental way," said Peter Galison, professor of the history of science and physics at Harvard. ``Pushing hard on the visual methods ends up pushing toward the antivisual. Beliefs swing between an almost theological dogma that images are stepping stones to higher knowledge, or that they are deceptive idols that keep us from higher understanding."

Coxeter's legacy is the powerful push he gave the visual geometric method, and the resulting change in perspective that transformed the way mathematicians and scientists create and investigate. ``Coxeter's perspective and ideas are in the air we breathe," said Ravi Vakil, at Stanford. ``It's not that his ideas are used to solve problems, it's that the fundamental problems grow out of his ideas. He's the soil."

"King of Infinite Space can be enjoyed even without a specialized knowledge of geometry or math. (Ms. Roberts's own exposition is admirably clear and conscientiously footnoted.) And the book's narrative is heartening. Too often -- think of A Beautiful Mind or Proof -- mathematicians are portrayed these days as seriously disturbed or weirdly obsessed or burnt out at an early age. Here, by contrast, is the true story of an eminent mathematician, active, alert, acute and ever alive to new ideas over a period of 80 years."
—WALL STREET JOURNAL, by Robert Osserman, special projects director of the Mathematical Sciences Research Institute in Berkeley, Calif.

"Roberts takes readers on a wide-ranging tour of contexts in which Coxeter's beloved symmetries have made themselves known, from geodesic domes to the error-correcting codes that make digital recording possible. As always, what is beautiful has ended up being useful... King of Infinite Space is exhaustive and definitive. Roberts's painstaking research, documented by 73 pages of endnotes, turns up many gems. Especially notable is Roberts's access to Coxeter's diaries, which inject the book with anecdotes of rather startling candor.Invaluable... There is no substitute for Coxeter, and no substitute for this long-overdue treatment of his life."
—WASHINGTON POST, by Jordan Ellenberg, assistant professor of mathematics, University of Wisconsin

"Roberts' book really soars in its description of Coxeter's work and his ability to visualize space, to communicate the poetry of geometry and to inspire other mathematicians, physicists and artists...Through Coxeter, Roberts reminds the reader of the visceral and visual excitement that can be found in the universal alphabet of lines and shapes. Although [Amir] Aczel's book is called "The Artist and the Mathematician," it is Coxeter, and not Bourbaki, who emerges as a true creator of beauty, not just elegance."
—CHICAGO TRIBUNE, Nathan L. Harshman, assistant professor of physics, American University

"[King of Infinite Space] is part biography, part scientific history and part epic.[it] offers poignant looks into Coxeter's soul.Thanks to Roberts's passionate writing, Coxeter the legend lives on."
—GLOBE AND MAIL, Jeffrey Rosenthal, professor in the department of statistics at the University of Toronto

"H.M.S. (Donald) Coxeter (1907-2003) was widely recognized and honored by his peers as the greatest living geometer. He was a prolific writer, publishing 12 books and more than 200 papers while at Cambridge, Princeton, and - for 67 years - the University of Toronto. He influenced prominent researchers, artists, and architects while pursuing theoretical and applied mathematical concepts of space, time, and shape. Canadian journalist Roberts, who won a National Magazine Award for her profile of Coxeter in Toronto Life, uses diaries, interviews, notes, personal vignettes, and stories to depict vividly Coxeter's passion for music, art, mathematics, life in general, and all things of beauty. In addition to successfully crafting a poignant biography, she accurately documents 20th-century mathematical research and scholarship. The author is to be congratulated on the book's simplicity; completeness; excellent use of diagrams, figures, and photographs; appendixes of mathematical notes; and reams of endnotes. A significant work for mathematicians at all levels; recommended for both academic and public libraries."
—STARRED REVIEW in LIBRARY JOURNAL, by Ian D. Gordon, Brock Univ. Lib., St. Catharines, Ont.

"The mathematics of shape and space, geometry was not professionally hip during the career of H. S. M. Coxeter (1907-2003). As Roberts elaborates in this warm but not uncritical portrait, the visual and intuitive aspects of geometry did not attract a field headed in more abstract directions. By the 1950s, a group of French mathematicians mounted the barricades against geometry under the slogan "Death to triangles!" Coxeter took notice but no heed of the radicals, content with his fertile imagination that yielded new geometrical papers up to his nineties. Though keeping geometry vibrant was not Coxeter's intent, it was the effect as, over time, his discoveries came to be useful to architect Buckminster Fuller, string theorists, and Godel, Escher, Bach (1979) author Douglas Hofstadter, who contributes a preface. Roberts accessibly explains the cruxes of Coxeter's discoveries and his place in mathematics history, while her narrative of Coxeter's personal life depicts an aloof but amiable character a bit deficient in the parenting department. With Coxeter appraised by peers as a modern Euclid, Roberts' biography bears inclusion in the popular mathematics collection."
—BOOKLIST, by Gilbert Taylor

“Siobhan Roberts has achieved something extraordinary in this book, a paean to a geometer and all geometry. It tells a brave, compelling story. It comprehends a whole universe — our universe — of kaleidoscopes and crystals, groups and symmetry, bicycles and snowflakes, music and movement. It is lucid, beautiful, and exalting.”
—James Gleick, author of Isaac Newton, Faster, and Chaos

“A biography of Donald Coxeter has long been overdue. Now Siobhan Roberts has provided one, and a marvelous book it is. King of Infinite Space covers all of Coxeter’s major achievements, and in words any reader can understand. Her beautifully written tribute is rich in details about Coxeter’s long life, and his colorful interactions with the world’s top mathematicians. I found it impossible to stop reading.”
—Martin Gardner, longtime "Mathematical Games" columnist in Scientific American, and author of numerous books including The Ambidextrous Universe and most recently Are Universes Thicker Than Blackberries?

“What emerges loud and clear in King of Infinite Space is that Siobhan Roberts understands Coxeter’s spirit very deeply. She understands what drove him, and she knows just how to put into words the fire that always inhabits a great mathematician’s soul. I hope that King of Infinite Space will bring to many people not only a sense for the beauty of mathematics itself, but also a sense for how the very human love of hidden patterns and symmetries can result in a hundred years of exultant exploration.”
—Douglas Hofstadter, author of Gödel, Escher, Bach, from the Foreword of King of Infinite Space

“King of Infinite Space gives us a lively view of the history of mathematics while weaving the story of Donald Coxeter, a broad-minded genius who built an important bridge between two opposite extremes of mathematical creation—the pictorial world of classical geometry, and the ideal world of abstract algebra.”
—Freeman Dyson, Professor of Physics at the Institute for Advanced Study, Princeton University, and author of Disturbing the Universe

“Many mathematicians the world-over are enchanted with the beauty and elegance of Donald Coxeter’s work. Although I never studied with Coxeter, in many ways I consider myself an honorary student of this great geometer. Why is it that Coxeter is affectionately remembered by so many mathematicians? Siobhan Roberts makes the answer quite clear in King of Infinite Space, an elegant biography of an elegant man.”
—John Horton Conway, John von Neumann Professor of Mathematics, Princeton University, and discoverer of Surreal Numbers

“What a wonderful world Siobhan Roberts evokes through this scientific portrait of the inimitable geometer, Donald Coxeter. Geometry: that subject we all learn early and too quickly forget, opens up again to us and what a universe Coxeter made of it. Pure mathematics, of course, but also facets of a pineapple, maps of the early universe, shapes of immunoglobulin, structures of architecture, images within kaleidoscopes. Like the fine and thoughtful sketches of Jeremy Bernstein and James Gleick, Roberts succeeds beautifully in crossing mathematics with the quirky, imaginative, and productive life of one of our greatest modern mathematical thinkers.
—Peter Galison, Professor of History of Science, and of Physics, Harvard University, author of Einstein’s Clocks, Poincaré’s Maps

“From Siobhan Roberts' biography of Donald Coxeter we learn that we have been doing geometry all our lives, for geometry's patterns frame perception. Donald Coxeter, astringent Anglo-Canadian mathematician, was a passionate proponent of geometry; it sustained him for the better part of a century spent 'Voyaging through strange seas of thought, alone' (as Wordsworth said of Newton). Siobhan Roberts must have known him well to write this intimate and engaging account of a life-long devotion to shape, as the key to all creation.”
— John Polanyi, Nobel Laureate

"Little icosahedrons and dodecahedrons often rolled across my dining room table during high school -- in games of dice -- but their complex beauty never really struck me. Donald Coxeter's brilliant geometric vision shows why it should have. Siobhan Roberts has given us a meticulous life of a very special kind of thinker: one who will change the way you experience then world."
—Mark Kingwell, Professor of Philosophy, University of Toronto, author of The World We Want

“A mathematician once wrote Coxeter "I tried very hard not to spend time on your integrals, but the challenge of a definite integral is irresistible." I tried very hard not to spend time reading King of Infinite Space, because I had other work to do, but I found it irresistible. The book shows clearly the degree to which great mathematicians like Coxeter are artists, led by a sense of beauty beyond the fashionable topics of the day into the heart of the deepest and most elegant mysteries.”
—John Mighton, Fields Fellow, Ashoka Fellow, and author of The Myth of Ability

"Donald Coxeter was a remarkable character, and this book is a fine record of his achievements. The author deserves our admiration for having produced such a lively and accessible account of what might at first seem an arcane subject."
—Sir Martin Rees, President of the Royal Society, Master of Trinity College Cambridge, Astronomer Royal

revitalize this sadly neglected subject."[6] *Introduction to Geometry* became widely used as a university textbook.[7] As testament to its popularity, for a time it was the most frequently stolen item in the University of Toronto mathematics library. And it was one of the first textbooks to be built around the concept of symmetry,[8] and of course, it was full of pictures. "I agree with Alice in Wonderland," Coxeter once remarked. "Wasn't it Alice in Wonderland who said, 'What's the good of a book that doesn't have pictures?'"[9]

Coxeter's book received a rave review from the preeminent critic Martin Gardner in his column in *Scientific American*: "Most professional mathematicians enjoy an occasional romp in the playground of mathematics in much the same way that they enjoy an occasional game of chess; it is a form of relaxation that they avoid taking too seriously. On the other hand, many creative, well-informed puzzlists have only the most elementary knowledge of mathematics. H. S. M. Coxeter . . . is one of those rare individuals who are eminent as mathematicians and as authorities on the not-so-serious side of their profession . . . There are many ways in which Coxeter's book is remarkable. Above all, it has an extraordinary range."[10] And while *Introduction to Geometry* is encyclopedic in its scope, like the Bourbaki treatise, it is at once as engaging and awe-inspiring as a curiosity cabinet. In the chapter on hyperbolic geometry, Coxeter prefaced his section on "The Finiteness of Triangles" with a Shakespearean epigram borrowed from *Hamlet*: "I could be bounded in a nutshell and count myself a king of infinite space."[11]

Coxeter's career, considering his two masterpieces—*Regular Polytopes* and *Introduction to Geometry*—can be viewed as two intersecting circles: they overlap, but their circumferences delineate two distinct realms. One realm encompasses Coxeter's role as popularizer and connoisseur of the beauty and fun of classical geometry (symbolized by *Introduction to Geometry*), whereas the other comprises his contribution as a pioneer, an innovator melding classical with modern geometry (as demonstrated in *Regular Polytopes*). The former achievement won him a wide and varied fan base, and the latter cemented his reputation among mathematicians. As observed by Sir Michael Atiyah, "If his fate was just to be a connoisseur of beautiful pictures, he wouldn't have been so widely recognized, he would have been more of a sideline. But you add this extra dimension of symmetries (finite or continuous), and that lifted him up and made him well known and in touch with other aspects of mathematics."[12]

Fields Medal winner David Mumford, who teaches pattern theory and the mathematics of perception at Brown University, felt Coxeter's impact in both realms. He stumbled upon Coxeter's book *Regular Polytopes* as a high school student in the early 1950s. "It was like I had discovered how math was really done," he recalled. "High school mathematics didn't show how deep the subject was. It was a revelation. It made me realize what mathematics was all about."[13] As Mumford continued his studies in mathematics at a scholarly level, Coxeter's work influenced his interests, specifically with the "compactification of modulized spaces"—just as they suggest, these are atlases, of sorts, for algebraic objects—and Coxeter's work fit nicely into the story he and his coauthors told in the book *Smooth Compactification of Locally Symmetric Varieties*. Mumford later met Coxeter in the 1970s, in a replay of the typical scenario. "I had assumed he was dead, and then, 'Oh my god, Coxeter—he's here.'" At the time Mumford was at Harvard, teaching the undergraduate geometry course. He often shook up the syllabus using *Introduction to Geometry* as a text, and he invited Coxeter down to give a lecture.[14]

With the twentieth-century stampede toward modern mathematics, toward all things abstract, algebraic and austere, the Bourbaki enterprise thrived beyond the borders of France—through publication of volumes of its treatise, and via Bourbaki members in the flesh who were installed at various universities on secondments, exposing new guinea pigs to the Bourbaki approach firsthand. Claude Chevalley went to Princeton's Institute for Advanced Study and later to Columbia.[15] Dieudonné spent time in São Paulo, Brazil, as well as at the University of Michigan and Northwestern University.[16]

Over the years the group continued to disguise itself in "mock mystery," and rumors continued to spread about Bourbaki, "the mathematician."[17] In one ruse, Bourbaki applied for an individual rather than group membership in the American Mathematical Society (AMS); the request was refused, twice. The AMS secretary suggested that an application for an institutional membership might meet with more success, but Bourbaki would have none of it.[18]

The Bourbaki group gathered three times per year, once for an extended two-week period, at a youth camp, a monastery, resort, or hotel, where they made major policy decisions, drew up the table of contents for the current volume of the treatise, and delegated research. When their semiregular publication of *Elements of Mathematics* became a commercial success, the royalties paid for travel expenses, wine, and extracurricular activities that enlivened the proceedings.[19] According to *La Tribu*, their internal newsletter, the Bourbaki group played chess, table soccer, volleyball, or Frisbee. They embarked with gusto on mountain hikes, bicycle excursions, or swimming expeditions. They caroused in bumper cars, went butterfly hunting or mushroom picking. They sunbathed, dozed off with text in hand, stuffed themselves with local delicacies and drank until royally drunk—Armagnac, champagne, rum toddies, or wine (wine being the much-needed fuel of Bourbaki's cogitation). Once under the influence, inebriated members sometimes performed a "virile French cancan or a lascivious belly dance . . . The deliberately laid-back attitude . . . gave the impression of insouciant genius."[20]

As far as the tone of the meetings was concerned, Bourbaki's biographer, Liliane Beaulieu, described them as opting "for the humorous and the ribald, on occasion ascending to the heroic contrasted with the loutish"[21]—humor was said to be their second-favorite mind game after mathematics. The mathematical discussions were not nearly as civilized as Henry F. Baker's tea parties at Cambridge. Anybody at any time could interrupt, comment, ask questions, or criticize. Dieudonné observed: "Certain foreigners, invited as spectators to Bourbaki meetings, always come out with the impression that it is a gathering of madmen. They could not imagine how these people, shouting—sometimes three or four at the same time—could ever come up with something intelligent."[22]

As the group's scribe, over the years Dieudonné came to be considered the speaker for the group as well. Between his stentorian voice and propensity for definitive statements and unchallengeable opinions, Dieudonné was known to crank up the decibel level of any conversation. It was Dieudonné who would later declare: "Down with Euclid! Death to Triangles!" He was a giant, a tall, big, and ebullient man, oft times loud and rude. He was flamboyant, with a brutal manner of expression.[23] Pierre Cartier recalled an outing to a concert hall with Dieudonné. "It was fascinating," he said. "He would look at the score in his hand and exclaim with disapproval—'OH!'—if a note was missing from the orchestra."[24] Coxeter, with comparable zeal as a musician, preferred musical scores to books for bedtime reading.[25]

The domineering Dieudonné penned first drafts of Bourbaki chapters, which were referred to as "Dieudonné's monster." From there, each chapter of *Elements of Mathematics* commonly took six, seven, even ten drafts before consensus was reached (unanimity was required, with each member having veto power).[26] And lest the enterprise be misunderstood, Dieudonné clarified: "Bourbaki's treatise was planned as a bag of tools, a *tool kit* for the working mathematician, and this is the key word which I think everybody should keep in mind when talking about Bourbaki or discussing its plan or contents."[27] Cartier agreed: "You can think of the first books of Bourbaki as an encyclopedia. If you consider it as a textbook, it's a disaster."[28]

The popularity of Bourbaki initially brought about something of a revolution in university-level mathematics. Marjorie Senechal was a graduate student in the 1950s at the University of Chicago, a hotbed of Bourbaki in America, under the auspices of Marshall Stone. Stone, strongly influenced by the ideas of Bourbaki, had made the mathematics department at the University of Chicago arguably the best in the country. He recruited the brains of the Bourbaki group, André Weil, and Samuel Eilenberg, who worked

Jean Dieudonné

closely with another Chicago professor, Saunders MacLane. "I suffered under the Bourbaki regime," said Senechal, one of MacLane's students. "Bourbaki was the method taught. I think it cost mathematics a lot of talent—a lot of people who think visually and work visually left the profession, because they felt they didn't have a home there anymore." Coxeter kept the spark alive for people who wanted to continue to do concrete geometry, even if it was unfashionable. "Coxeter was the antithesis to Bourbaki.* He was a lifeline," said Senechal, "a way of salvation from Bourbaki. Because through him I knew there was more to mathematics, I knew there was a whole branch of mathematics I could relate to."[29]

But the Bourbaki revolution shook more than just the universities. Bourbaki trickled down into high schools and public schools, as mathematicians taught by the Bourbaki method became teachers themselves.[30] Bourbaki principles

*Senechal elaborated to say: "If you are thinking about a mathematical idea in the Bourbaki style, you will be working upwards from definitions and axioms and trying to continue through that logical line. If you are working in Coxeter's style, you are also working upward, but you start with some concrete object, asking questions about it, asking how to put that in a more general way and what that leads to. Coxeter's is a visual and hands on approach, as opposed to a strictly logical approach."

infiltrated the "New Math" grade school curriculum reform. The New Math spread throughout the Western Hemisphere, from South America through the American heartland, into Canada, and across the Atlantic to England, Wales, West Germany, Denmark, the Netherlands, and France. New Math overhauled the traditional curriculum, ridding it of trivial problem solving and rote number juggling. Instead, schoolchildren as early as grade one learned the equations of algebra and set theory (the mathematical theory of sets, or collections of abstract objects, and the rules that govern their relationships and manipulations). Dusty and dilapidated Euclidean geometry also was forsaken—like removing Shakespeare from the syllabus and replacing it with grammar, as though one were a minor subset of the other.[31] "This tendency is not only regrettable," said Coxeter, "but unreasonable."[32]

Historically, Euclidean geometry had been under siege ever since its limited scope had been exposed. "Euclid's approach to geometry has been attacked on two grounds—that it is illogical, and that it is boring," Coxeter said in a 1967 report on the state of geometry in primary and secondary school education. "Neither criticism is new," he said, adding: "The objection that Euclid is boring is much more serious than the objection that his logic is imperfect."[33]

If Euclidean geometry was boring, Coxeter argued, this was due to the canned and ossified way it was taught. Like arithmetic, the subject had been reduced to rote learning, with teachers opening a textbook and doing the stultifying "chalk-and-talk" at the front of the classroom. Children mindlessly memorized properties of triangles and their theorems—Side-Angle-Side, Angle-Side-Angle—and regurgitated them on demand to please their teachers. They were robbed of experiencing the beauty and tricks intrinsic to heuristic learning—that is, learning through trial-and-error and making discoveries for oneself.

The French Bourbakis were one influence behind the New Math reforms; the Russians were another. When the Soviets successfully launched the Sputnik satellite into orbit, the Western world got a shock—rudely awakened to the fact that it was falling behind in science, technology, and mathematics. A colorful interpretation of events was articulated in a British report chronicling the New Math:

> It all started on that memorable day in 1957 when the Russians sent their first Sputnik orbiting the earth. Up till then the countries of the West had rather patronizingly regarded the USSR as a backward giant of a nation, hopelessly engaged in trying to educate its largely peasant people to achieve the technological

> advantages of its more favoured European neighbours. The noisy "Bleep-Bleep" of the Sputnik's radio, however, quickly dispelled any notions Westerners might have that the Russians still counted on their hands or that the abacus was the sole piece of educational equipment in Soviet schools. Clearly this formerly retarded people had outstripped Britain and America in finding scientists and mathematicians of a very high caliber indeed. How had this astonishing advance in Russian scientific education come about? No one could supply the answer but it had to be admitted that Soviet schools were obviously producing more and better mathematicians and scientists than were coming from the British system of education.[34]

Following the bleeping Sputnik, the United States Congress released millions of dollars in funding for science education under the "National Defense Education Act." A flurry of international activity led to the formation of the United Nations Educational Scientific and Cultural Organization (UNESCO) and the Organization for European Economic Cooperation (OEEC). Reform of the mathematics curriculum was undertaken with urgency and idealism.[35]

The first forum of debate on the New Math was the 1958 International Congress of Mathematicians, held in Edinburgh. Then came the infamous conference at which Dieudonné whooped his war cry. Held at the Cercle Culturel de Royaumont, Asnières-sur-Oise, France, from November 23 to December 4, 1959, the conference addressed the need for reform in French mathematical education. Here the bombastic Dieudonné rose to his feet and hurled his provocatively planned statement:

"*À bas Euclide! Mort aux triangles!*"[36]

"Down with Euclid! Death to triangles!"

Dieudonné's statement was taken by many as a slap to geometry. Coxeter discussed it with like-minded individuals and was known to now and then unleash a scathingly critical or derogatory comment, though he did not dwell on it, nor did he let the Bourbaki venture as a whole ruffle his feathers. Dieudonné's comment seemed a succinct summary of the Bourbaki agenda—no diagrams—but the interpretation of this event by Bourbaki sympathizers diverges from a geometer's vantage point. Michel Broué, director of the Institut Henri Poincaré, who studied under Bourbaki founder Claude Chevalley in the 1960s, asserted the importance of distinguishing Dieudonné from Bourbaki. Dieudonné, by 1959, was older than fifty and therefore no longer an official member of the Bourbaki group. His "Death to Triangles!" statement is thus disqualified from representing the larger Bourbaki mandate (Broué acknowledged all the same that linking the two has become part of the Bourbaki mythology). "Others in Bourbaki were horrified," Broué said.[37] Especially since Dieudonné stuck with this opinion to his dying day. This was a source of embarrassment for some Bourbakists, such as Cartier. "I was tormented," he said. "The ideology of Bourbaki didn't match with me, it was going too far. Bourbaki was a mathematical priest—pure, pious, rigid. It was a caricature of purity. Purity creates hypocrisy, because if the rule is too strict then life forces you to break it."[38]

Chevalley, for one, espoused the no-pictures dictum, but this belied his closeted use of diagrams. He tried his best to operate by reasoning alone; he earnestly wanted to avoid intuition in mathematics. But he didn't always succeed. Cartier remembered when Chevalley was his professor, and teaching at the front of the classroom he filled the blackboard with symbols and equations. When a student raised their hand with a question, Chevalley dramatically stepped back from the board, crossed his arms, and squinted, contemplating his work through furrowed brow. He was stumped. Then he walked toward the blackboard, standing rather closer than necessary. He huddled in, with hunched shoulders, his arms creating blinders, hiding what he was up to from the class behind. "He drew a picture," said Cartier, "figured out the answer—AHA! he said—and then quickly wiped out the diagram, stood back and continued."[39]

After the "Death to Triangles!" incident at Royaumont, annual conferences on New Math were held in Denmark, Zagreb, Athens, Bologna, elsewhere in Europe, and there was a series of "inter-America" conferences in South America. The first convened in Bogotá, Colombia, in 1961. The ringleader of the New Math reform movement was Marshall Stone, who had been president of the Royaumont conference and led the way internationally.[40] As the mastermind of these international conferences, Stone delivered stirring opening addresses, calling for the modernization of mathematics:

> There are two major factors which require us to examine with fresh eyes the mathematics we propose to teach to young people in the secondary schools and in the first years at the university. One is the extraordinary growth of pure mathematics in modern times. The other is the increasing dependence of scientific thought upon mathematical methods, coinciding in time with a more and more urgent social demand for the services of scientists of every description.

> The forces exerted by these two factors on our educational system are quite clearly on the point of shattering the traditional framework of mathematical instruction and thus preparing the way for an overdue modernization and improvement of our teaching of mathematics. Like the crustacean which has to split and discard its old shell in order to grow, we must at last burst the confines of a curriculum which is plainly no longer suited to our current needs or our current conditions of life.[41]

At Bogotá, Howard Fehr, head of the Department of the Teaching of Mathematics at Columbia University, delivered a lecture titled "Reform of the Teaching of Geometry." "Euclidean geometry," said Fehr, "nowadays . . . is sterile, outside the main course of mathematical advancement, and it can be filed in the archives, without any fear, for the benefit of future historians."[42] Response was mixed. Professor Guillermo Torres, from Mexico, challenged this position and argued that the presentation of mathematics in an exclusively formal aspect "makes it appear to be an inhuman activity and with no sense at all." John Coleman, Coxeter's early student, by then the head of the math department at Queen's University, also expressed doubts during the debate following Fehr's presentation. Based on his experience, he said, students interested in mathematics were first enticed by geometry's intrinsically tactile and visual nature—geometry was the user-friendly interface of mathematics.[43]

New Math sprouted in varied forms internationally. In the United States, the main initiative was the School Mathematics Study Group (SMSG), which produced a new series of textbooks—students renamed it "Some Math Some Garbage."[44] The *American Mathematical Monthly* ran a "Letter of 75 Mathematicians" objecting to the emphasis on abstraction. The leading antagonist was NYU professor Morris Kline, who later sounded the death knell of the New Math in America with his book *Why Johnny Can't Add: The Failure of the New Maths*.[45] And the horrors of it all entered the popular culture via the genius of mathematician cum musical raconteur Tom Lehrer—he documented the debacle on his album *1965: That Was The Year That Was* with the song "New Math." The lyrics poked fun at the fact that the math was so newfangled that parents couldn't make sense of it in helping their children with homework.[46]

In France, of all places—the cradle of Bourbaki—the newspaper *L'Express* ran the headline LE CAUCHEMAR DES MATHS MODERNES (The nightmare of modern maths); "Pornography, drugs, the disintegration of the French language, upheavals in mathematical education all relate to the same process; attacking the central parts of a liberal society," the subtitle continued. And a report to the French Academy of Sciences decried: "The set-theoretic option in the definition of geometry is a dangerous utopia . . . this reform misappreciates the intellectual aptitude and needs of the adolescents who attend our . . . high schools. The reform in progress seriously endangers the economic, technical, and scientific future of the Nation."[47]

In England, a telling cross section of the changes is found in the career of Sir Michael Atiyah. He was a student at Cambridge in the 1950s, when aspects of classical geometry were still hanging on as part of the university curriculum. But by the 1960s, this last bastion had languished, linear algebra having been decreed fundamental and geometry old fashioned and inessential.[48] Sir Michael's 1981 presidential address to the Mathematical Association, titled "What Is Geometry?" bemoaned this unfortunate turn in geometry's history. "Of all the changes that have taken place in the mathematical curriculum, both in schools and universities, nothing is more striking than the decline in the central role of geometry," he said. "Euclidean geometry has been dethroned and in some places almost banished from the scene."[49]

"The battle between geometry and algebra is like the battle between the sexes," said Sir Michael, contemplating the issue recently. "It's perpetual. It's an ongoing battle. And it really is a battle in the sense that these are two sides of the same story, and you've got to have both sides present." Both algebra and geometry are essential, both must be taught properly at all levels, and the resulting interaction in the highest tiers of research move the frontier forward. "It's the kind of problem that never disappears," he said. "It will never be dead and it will never get solved. The dichotomy between algebra, the way you do things with formal manipulations, and geometry, the way you think conceptually, are two main strands in mathematics. The question is what is the right balance."[50]

One outpost that kept the balance weighted toward geometry over algebra was Eastern Europe—Latvia, Hungary, and Russia. The reason is the object of speculation. Perhaps their prophylactic was the Iron Curtain—cut off from the rest of the world, and poor, they continued on with the old-world ways. Russia had a long and fine mathematical tradition all its own. Certainly, the fact that the Russians printed Coxeter's works demonstrated that they liked their classical geometry.[51]

At the 1966 International Congress of Mathematicians in Moscow, Coxeter learned of his considerable popularity in Russia. Prior to the congress he had no idea whether his books had been published there—ostensibly, Russia had agreed to international copyright laws but Russian editions, seldom the products of contractual agreement, were pirated more often than not. If mathematicians wanted royalties, they had to produce proof of publication, which was difficult if one was not in Russia. The International Congress occasioned an olive branch in the form of a book exhibit that allowed mathematicians to peruse a warehouse of all Russian mathematical publications. If they found their books, they were entitled to collect royalties on the spot. Coxeter walked around the warehouse with John Conway, who recalled that Coxeter made a lengthy list of his books and then walked away with his pockets full of rubles.[52]

Coxeter's classical geometry also thrived in Italy. Geometer Emma Castelnuovo was a Coxeter fan from afar, and vice versa. "I have all of Coxeter's books,"[53] said Castelnuovo, now in her nineties. She devoted her life not to higher math like her father, Guido Castelnuovo, but to teaching geometry in grade schools. She worked with children aged eleven to fourteen, in Italy and Africa, doing geometry "by hand," and organizing exhibits of the children's work. She attended all the congresses and commissions on mathematics education, including the "Death to Triangles!" conference in Royaumont, and worked with Piaget on the Commission for the International Study and Improvement of Education in Mathematics. In 1949, she published her first book, *La Geometria Intuitiva*, and wrote many textbooks for students. Coxeter had high praise for Castelnuovo's work and cited her as an example worth following in his report on geometry education. "In Italy today, Emma Castelnuovo has popularized and developed a [new approach to Euclidean geometry]," he said. "Her book, *La Geometria Intuitiva*, describes the teaching of geometry with apparatus resembling Meccano.* The book, beautifully illustrated, shows how geometrical shapes are used in the architecture of Italy."[54]

Another beacon was the Netherlands, where German expatriate Hans Freudenthal was credited with saving Holland from the New Math. In his 1971 article "Geometry Between the Devil and the Deep Sea," Freudenthal cast it all in lyric terms:

> Geometry is endangered by dogmatic ideas on mathematical rigor. They express themselves in two different ways: absorbing geometry in a system of mathematics as linear algebra, or strangulating it by rigid axiomatics. So it is not one devil menacing geometry as I suggested in the title of my paper. There are two. The escape that is left is the deep sea. It is a safe escape if you have learned swimming. In fact, that is the way geometry should be taught, just like swimming.[55]

Coxeter had the same sensibility: "The ability to study, grasp, and master topics in mathematics resembles in some ways the ability to swim or to ride a bicycle," he said in the geometry report, "each of which is, in a static sense, impossible of achievement. There is a trick to it, and strong motivation is needed to learn the trick. Perhaps one difference is that children seldom encounter oppressive authoritarian discipline in connection with the technique of riding a bicycle."[56] Geometry, Freudenthal said, would die of suffocation as a "prefabricated subject." It could be saved if presented as a field of wonderment and activity—folding, cutting, gluing, drawing,

---

*Meccano is the trade name for colorful metal construction toys assembled with nuts and bolts, invented in 1901 by Frank Hornby, of Liverpool, England.

Canadian Mathematical Congress, Fredericton, 1959. Top (left to right): Irving Kaplansky, Alex Rosenberg, Coxeter. Bottom: Werner Fenchel, Philip Wallace, Max Wyman, C. Ambrose Rogers, Hans Freudenthal.

painting, measuring, and fitting. "Coxeter's *Introduction to Geometry*," he said, "is a marvelous demonstration of this attitude. The author knows, in any case, exactly where the horizon is lying."[57]

Introduction to Geometry circulated internationally, with translations into six languages—German, Japanese, Russian, Polish, Spanish, and Hungarian. The first was the German translation, in 1963, which had a title Coxeter was very fond of: *Unvergangliche Geometrie*—Everlasting Geometry, or Geometry which Survives Everything.[58] With the publication of *Introduction to Geometry* in Japan, in 1965, architect, engineer, and geometer Koji Miyazaki became one of Coxeter's biggest fans. Also a professor emeritus at Kyoto University and Teikyo-Heisei University, Miyazaki recalled: "At that time, the name 'Coxeter' suddenly spread out in Japan as the biggest mathematician's name in the world. I am clearly remembering that time. And from that time I was thinking that Prof. Coxeter is the god of the world of geometry."[59]

A counterinsurgency against the geometry-barren New Math curriculum—as against Bourbaki—continued to take shape in all things Coxeter. From the beginning of his career through the heyday of Bourbaki, Coxeter simply averted his eyes from the antivisual antigeometry trend, and went on a crusade to bring his passion for the intuitive methods to any and all willing spectators. He lectured on "the beauteous properties of triangles," on "The Arrangement of Trees in an Orchard," on the Fibonacci numbers (with nine slides and a pineapple as a prop). On a snowy January evening, he took the night train from Toronto to Philadelphia, putting the final touches on his presentation as he traveled. The following day, he noted in his diary: "About 40 broke into spontaneous applause after my 10 min. lecture on 'Close Packing and Froth.'"* The next month he gave a version of the same lecture to seventy schoolteachers in Toronto. Two months later he spoke to a group of forty prize-winning schoolchildren on "Close Packing of Spheres," this time drawing upon an eighteenth-century book with a title he thought his young pupils might find amusing—it was called *Vegetable Statics*, by Dr. Stephen Hales, wherein Hales investigated how many peas, if as many peas as possible were compressed into a large cubic pod, would abut a central pea.[60]

In 1967, Coxeter published two more books that would become classics: *Projective Geometry* and *Geometry Revisited* (the latter with S. L. Greitzer). He churned out papers asking, "Whence Does a Circle Look Like an Ellipse?" and lectures wondering, "Why Do Most People Call a Helix a Spiral?" In another talk he issued "A Plea for Affine Geometry in the School Curriculum," and in yet another he offered simply "Reflections on Reflections," which he delivered in Pittsburgh in 1967.[61]

After his Pittsburgh talk, he traveled to Minneapolis where he was coming to the end of a long-running pet project, working for four years with a group of mathematicians on educational geometry films, *Dihedral Kaleidoscopes* and *Symmetries of the Cube* (two in a series of five films). The project, aiming to improve geometry teaching in high schools and colleges with the introduction of exciting experimental films and an accompanying series of textbooks, was part of the College Geometry Project at the University of Minnesota, well financed with a million-dollar-budget (funded entirely by the National Science Foundation; classical geometry still had its champions). Coxeter laboriously wrote and rewrote the scripts. And in *Dihedral Kaleidoscopes*, he took the role as starring geometer.[62]

The film began with Coxeter scampering across a busy street, dodging traffic, wanting to get to the other side to look at his reflection in a mirrored store window (the narrator explained: "H. S. M. Coxeter, of the University of Toronto, is a geometer. To Professor Coxeter, reflections are of particular

*Two months before his lecture on "Close Packing and Froth," Coxeter noted the seed of inspiration in his diary: "I saw that the no. of bubbles touching one bubble in froth should be [illegible]."

Coxeter starring in the documentary *Dihedral Kaleidoscopes*.

interest because of their implications for geometry and algebra . . ."). With a lively flute soundtrack, the film followed Coxeter as he manipulated mirrors in a darkened studio. He peered into large kaleidoscopes—constructed like tents or pens and illuminated from within,* dropping in colored paper triangles, watching as they fluttered into place, and grinning when they landed and generated pleasing psychedelic patterns on the plane. The films won many awards—in Canada and the United States, at the American Film Festival and the Golden Eagle at the CINE Film Festival, and internationally, in Belgium, Czechoslovakia, France, Italy, Argentina—broadening Coxeter's fan base even further.[63]

*One of the documentary kaleidoscopes was taller than Coxeter himself and equally wide, gaping jaws of mirrors (the mirrors for the kaleidoscopes were produced, after a long search, by Litton Industries at a cost of $5,500). In an outtake, Coxeter inserted his miniature daschund, Nico, into this monster Kaleidoscope to see what would transpire—Nico was puzzled if not petrified, and stood frozen in place until Coxeter rescued him (Nico died later the same year and Coxeter honored him with a dedication in his book *Twelve Geometric Essays*: "In memoriam NICO 1951–1967").

Close-up of Coxeter positioning props in a kaleidoscope.

For the most part, Coxeter's crusade was all rear-guard action. He simply continued to make his contribution in the most hands-on way he could, propagating his passion. He did, however, keep an eye on the land mines of curriculum committees with their mandates for reform.[64] He voiced his opinion and opposition, at times with uncharacteristic volume. Tim Rooney, a colleague of Coxeter's in the math department at the University of Toronto, remembered the only time he ever saw Coxeter angry: when he perceived his geometry was under attack on home turf. Coxeter was graceful and sweet, said Rooney; there existed no easier man to get along with. But when Rooney bumped into him in the hall one day in the 1960s, Coxeter was fuming. He cornered Rooney, pulled him aside, and gave him an earful about a report from a committee studying the department's roster of mathematics courses. Coxeter interpreted the report as disrespectful and denigrating to geometry; it concluded there was an awful lot of geometry on the department's course list and some of it had to go. He interrogated Rooney about it: "What's your committee doing recommending less geometry be taught?"[65]

Coxeter peeking into a kaleidoscope taller than himself.

"He really was angry," Rooney recalled. "I told him, first of all, I wasn't on that committee, at which point he cooled down a little. And then I told him I didn't agree with what the committee said about geometry, which cooled him down further." Coxeter recruited another member of the department and they tackled the chair of the committee and had a furious argument about the report.[66]

This prompted Coxeter to more directly assume the mantle of the curriculum controversy, pulling himself and the dignity of his geometry together by the frayed laces of his well-worn spectator shoes. He sat on the K-13 Geometry Committee, producing the report *Geometry, Kindergarten to Grade Thirteen* in 1967. It baldly stated: "Some recent innovations under the name of 'modern mathematics' are unsatisfactory and ought to be discontinued . . . We have in mind an excessive tendency to abstractness and rigour, a copying of procedures more appropriate to graduate school." The net effect, the report said, was that the "geometric literacy" of society was even lower than its "numeric literacy":

> The ability to visualize geometrically is a basic part of the scientist's mental equipment . . . Thus scientific literacy is founded in part upon geometric abstraction . . . Geometry is perhaps the most elementary of the sciences that enable man, by purely intellectual processes, to make predictions (based on observation) about the physical world. The power of geometry, in the sense of accuracy and utility of these deductions, is impressive, and has been a powerful motivation for the study of logic in geometry. Unfortunately, however, in the teaching of geometry the role of logic is very likely to overshadow the creative and intuitive aspect of the subject. In the past this tendency has been reinforced by the conventional attitude that visual or intuitive "qualitative" pattern work in geometry was a fit subject only for the kindergarten or lower grades.
>
> We wish to emphasize as strongly as possible that we do not accept this view. Visual and intuitive work are indispensable at every level of mathematics and science, both as an aid to clarification of particular problems, and as a source of inspiration, of new "ideas."[67]

Classical geometry, for Coxeter, was one of the arts—the Seven Liberal Arts, as set out by medieval universities, were the Trivium, "the three roads" of grammar, rhetoric, and logic; and the Quadrivium, "the four roads" of arithmetic, geometry, music, and astronomy or cosmology. And so it followed that the justification for studying the liberal arts applied equally to the study of classical geometry—they may seem obsolete, indulgent, and impractical courses of study, but the arts are fertile soil, fostering a freedom and breadth of thinking from which more "modern" achievements grow. A good number of the report's 120 pages contained specific suggestions for reinstating geometry and tips for teaching it in an inspiring way to primary, intermediate, and senior grade levels—complete with practical instructions for nail and plywood constructions, skeletal models made from straws and pipe cleaners, the use of shadows and mirrors, and how to draw a cube from a circular array of dots.[68]

In 1968, in a nice topological twist of history, the proper nouns "Coxeter diagrams" and "Coxeter groups" finally made their debut in—of all places—the Bourbaki volume on Lie algebras, considered by some as the most successful volume in the whole series.[69] Marjorie Senechal delights in recalling how she once looked through all the Bourbaki volumes to see for herself the depressing dearth of diagrams. Apart from the slippery-argument-caution-ahead S-curve, she found only one. It was in Coxeter's volume and it was the Coxeter diagram.[70]

Coxeter came to be included in Bourbaki after his work intrigued a Belgian mathematician by the name of Jacques Tits, now at the Collège de France.[71] Closely affiliated with Bourbaki, Tits drew the group's attention to Coxeter's work, writing the first paper ever on Coxeter groups—"Groupes et Géometries de Coxeter." The paper went unpublished until the Bourbaki volume, which Tits ghostwrote. Two-thirds of the volume is taken up with expositions on Coxeter, baptizing not only the term "Coxeter group," but also "Coxeter graph" (also known as the Coxeter diagram), "Coxeter matrix," and "Coxeter number."[72]

Coxeter was pleased with the Bourbaki nomenclature. It meant his name was writ large into the history of mathematics. With the publication of the Bourbaki volume on groups, nearly ten years had passed since Dieudonné proclaimed "Death to Triangles!" When Dieudonné visited the University of Toronto in 1969, Coxeter and others took him out for a sumptuous dinner at the Park Plaza hotel, its rooftop restaurant offering a glittering view of the city. Dieudonné was there to give two lectures, one on Lie algebras, the other on Bourbaki.[73] "It . . . seems to me," he commented, "that when examining which tools should be included in Bourbaki, a decisive element was whether or not they had been used by great mathematicians, and what degree of importance these mathematicians had attached to these tools."[74] Coxeter had certainly found success by these criteria.[75] And in another address, in 1968 at the Roumanian Institute of Mathematics, in Bucharest, Dieudonné stated, "[O]ne must never speak of anything dead in mathematics because the day after one says it, someone takes this theory, introduces a new idea into it, and it lives again."*[76] Coxeter could hardly have said it better himself.

One decade later again, in 1980, the bright yellow cover of a publication by the Mathematical Association of America showed a hooded skeleton, the ghost of geometry, his bony finger dangling over a ratty scroll with a diagram

*As for Bourbaki's future, after the group's great success its productivity stalled in the 1970s during a clash with the publisher over royalties and translation rights, resulting in a protracted legal dispute, which was settled in 1980. Bourbaki then had a short resurrection, issuing revised editions of old books, and adding a few volumes to the series. "But then silence," said Pierre Cartier. "In a sense Bourbaki is like a dinosaur, the head too far away from the tail," he observed, of the subsequent generations that inevitably strayed further and further from the group's founding ideals and mandate. Just as Bourbaki members were forced to retire at fifty, Cartier joked that Bourbaki—himself or itself—should have retired at the half-century mark. Regardless, for all intents and purposes, his judgment was that "Bourbaki is dead." There is, however, an annual "Bourbaki Seminar" in Paris. And there are rumblings that further publications and revised editions might be in the works.