Pergamon

Stud. Hist. Phil. Mod. Phys., Vol. 28, No. 3, pp. 307–323, 1997

Printed in Great Britain
1355-2198/97 $17.00 + .00

Gauge Theory, Anomalies and Global Geometry: The Interplay of Physics And Mathematics

Dana Fine † *and Arthur Fine* *

1. Introduction

Whether today's physics will lead to a grand theory of everything or to a more modest layer cake of effective theories of different orders (or perhaps to something else), certainly the next theoretical turnings will evolve out of the successes and failures of quantum field theory. In its postwar re-incarnation, field theory has been at the centre of developments not only in the physics of fundamental particles (quantum electrodynamics, chromodynamics and the standard model), but also in condensed-matter physics and cosmology. One of the most striking aspects of field theory in this period has been its alliance with the frontiers of research in 'pure' mathematics, especially in the areas of differential geometry and topology. At least since the advent of relativity, the prominent role of abstract mathematics, although not well understood (witness Wigner's (1967, p. 222) 'unreasonable effectiveness of mathematics'), has nevertheless been accepted as a fact of life in theoretical physics. What is perhaps new in the context of field theory is the reciprocal feedback of physics on pure mathematics, not just in the promotion of mathematical work in fields and topics allied to physics but also in concrete suggestions, deriving from the physics, for the solution to outstanding mathematical problems. The present work is a small contribution to tracing some of the interplay of physics and mathematics in field theory. Our particular subject is the treatment of anomalies in gauge theory, specifically the global chiral (i.e. left–right) anomalies. In general, the term 'anomaly' signifies the breakdown, upon quantisation, of a symmetry present in the classical action where that breakdown does not depend on special features of the vacuum state (as it does in the case of spontaneously broken symmetry).

(*Received 12 August 1996; revised 2 January 1997*)
* Department of Philosophy, Northwestern University, Evanston, Illinois 60208, U.S.A.
† Department of Mathematics, University of Massachusetts Dartmouth, North Dartmouth, MA 02747, U.S.A.
PII:S1355-2198(97)00011-7

Since symmetries are associated with conservation laws, one can regard the question of anomalies as a question of what remedy to apply when a classical conservation law seems to fail, or actually does fail, in the quantum treatment of the phenomena. A standard example of the former would be the seeming failure of energy conservation in beta decay, whose remedy was the neutrino hypothesis. The chiral anomaly is an example of the latter; that is, chirality is a classical symmetry that actually disappears after (second) quantisation. Originally the chiral anomaly arose in perturbative calculations relating to models of neutral pion dissociation. Using Feynman rules, the calculations were seen to involve a fermion triangle diagram with one axial and two vector currents. Imposing current conservation and Bose symmetry in the vector channels leads to non-conservation of the axial current. This breaks chiral symmetry and results in neutral pion decay. Our interest is in showing how this chapter in physics and perturbation theory connects with the introduction of global mathematical methods, and the ramifications of that. The story begins in the laboratory.

2. Penetrating Showers in Lead

This is the title of a paper by W. B. Fretter (1948) that provides perhaps the first experimental evidence for the existence of neutral mesons. Fretter's experiment involves a cloud chamber containing eight 1/2-inch thick lead plates, eight inches wide, used to track the production of penetrating showers by cosmic rays. Discussing the interpretation of these showers, Fretter refers to J. R. Oppenheimer, who appears to have been the first to propose the γ-instability of neutral mesons when coupled to nucleons: 'Recently J. R. Oppenheimer has suggested that in these nuclear events not only charged mesons may be produced, but also uncharged mesons. The neutral mesons are calculated to have an extremely short life (10^{-15} s). Thus, even if the neutral mesons were given a large amount of energy in the initial event they would decay almost immediately and produce a pair of γ-rays, which would then create the electron shower' (pp. 45–46). If it had mattered, Fretter could have improved on the estimate of 10^{-15} s, for Oppenheimer had already assigned the problem of calculating the lifetime to R. J. Finkelstein. Finkelstein's result (1947) gave an order of magnitude estimate of 1×10^{-16} s for the decay of neutral 'pseudo-scalar' mesons into two photons. This turns out to be in good agreement with the experimental decay rate of $(1.12\pm0.202)\times10^{16}\,\mathrm{s}^{-1}$ derived from subsequent measurements. [1] The physical picture employed by Finkelstein was that of 'dressed' nucleons; i.e. of nucleons surrounded by a virtual meson cloud. Coupling a neutral meson to the electromagnetic field permits γ-decay: the neutral pion dissociates into a proton–antiproton pair which then radiates the photons. Finkelstein's calculations, however, encountered divergent integrals of a sort reminiscent of the self-energy divergences in electrodynamics. Finkelstein's resolution was to employ a rather *ad hoc* method of subtraction for cancelling the divergences.

[1] For example, see Rosenfeld (1968).

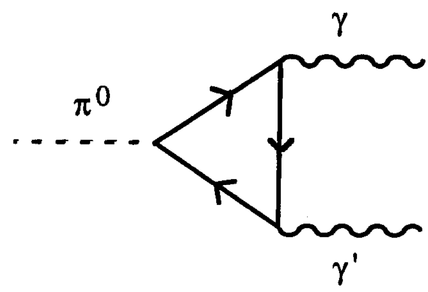

Fig. 1. *The triangle diagram.*

3. Perturbations

Follow-up studies by J. Steinberger (1949) made use of the then new technique of Feynman diagrams to effect the perturbative calculation.[2] Steinberger represented neutral pion decay by the graph in Fig. 2, where the central triangular 'loop' corresponds to a linearly divergent integral: 'All infinities in field theory are similar to that of this example. Somewhere in the Feynman diagram there is a closed loop which gives rise to the infinite integral' (p. 1181). To deal with this Steinberger borrowed a regularisation technique (Pauli–Villars regularisation) and wrote down general consistency conditions for a formal 'method of subtraction fields' (p. 1180), applying it first to the $\pi_0 \rightarrow \gamma\gamma$ decay. Two years later Schwinger (1951) treated the same problem as an application of his general 'proper-time method' (p. 672). Using Green's functions without the aid of the Feynman diagrams. Schwinger's treatment depended on formulating the perturbation expansion in such a way that a so-called proper-time integration was reserved to the last. Both Schwinger and Steinberger based their perturbation calculations for the neutral pion decay rate on approximations to an explicit Lagrangian function with interaction $g\overline{\psi}\gamma_5\psi\phi$ (where ψ, $\overline{\psi}$ are the proton fields, ϕ is the pion field and g is the coupling constant.) The concordance of these methods for handling the divergences suggested that something was right about the renormalisation techniques. But clearly some things, if not actually wrong, were certainly not well understood.[3]

The problem of neutral pion decay arose again in the context of the 'current algebra' approach to symmetry introduced by M. Gell-Mann (1962) and the exploration of the assumption (called PCAC) that axial-vector currents are partially conserved: explicitly, that the divergence of the current is proportional to the particle mass. Arguments by D. G. Sutherland (1967) and M. Veltman (1967) implied that under this assumption (given gauge invariance) the am-

[2] Steinberger's paper appeared in the October issue of *Physical Review*. The September issue contained Feynman's (1949) now famous paper, 'Space–Time Approach to Quantum Electrodynamics'.

[3] Concerning the cogency of Steinberger's calculation R. Jackiw remarked: 'It is presumably an accident that this completely implausible calculation gives a result in excellent agreement with experiment' (1972, p. 167).

plitude for $\pi_0 \to \gamma\gamma$ vanishes in the limit where the pion mass goes to zero. In this limit exact chiral symmetry holds, and J. C. Taylor (1958) had already shown that chiral symmetry is incompatible with the beta decay of pions. J. S. Bell and R. Jackiw (1969) formulated the problem this way: perturbation theory calculations involving methods for subtracting infinities yield a non-zero amplitude for neutral pion decay in the limit where the pion mass is zero. PCAC, however, which is formally satisfied in the Lagrangian models on which the perturbative calculations are based, implies that the decay amplitude vanishes in that limit. How can we reconcile the two? The answer provided by Bell and Jackiw was to suggest an unorthodox regularisation scheme. Their approach modifies the customary Pauli–Villars regularization. Bell and Jackiw require that the coupling constants g_i for the auxiliary fermion fields used to 'subtract off' the infinities in the triangle diagram co-vary with the regulator masses m_i of those fields; i.e. that $m_i/g_i = m/g =$ constant (where g is the original coupling constant and m the fermion mass). This scheme preserves gauge invariance and the PCAC conditions (along with Lorentz covariance and Bose symmetry). The price for preserving these symmetries in the regularisation is the Sutherland–Veltman theorem: the neutral pion decay rate vanishes. In addition to this awkward conflict with experiment, S. Adler (1969) points out that for strong interactions the Bell and Jackiw regularisation would result in unrenormalisable infinities. Adler concludes that in order to preserve the other symmetries and also the possibility of renormalisation the PCAC condition must go. Adler's analysis pinpoints the problem. The PCAC condition arises by calculating the divergence of the axial-vector current formally from the equations of motion inherent in the Lagrangian of the interaction model. These formal equations lead to the conservation of the axial-vector current (hence to chiral symmetry) in the zero mass limit. Translated into a condition on amplitudes this symmetry is expressed by a so-called Ward identity. If the corresponding calculation for the Ward identity is performed in perturbation theory, however, one encounters divergent path integrals, linearly divergent for a single triangle (or loop) graph. Thus one must introduce a regularisation technique. Because of contributions from the regulator terms, which are non-vanishing and finite even as the regulation is removed, the Ward identity acquires an extra ('anomalous') term and global chiral symmetry is broken. Similarly, regularisation modifies the PCAC condition. These modifications deny the critical assumptions of the Sutherland–Veltman theorem and allow the perturbative calculation for the $\pi_0 \to \gamma\gamma$ decay rate to proceed essentially as in Finkelstein, Steinberger and Schwinger; i.e. to yield a non-zero rate within the range of experimental error. W. Bardeen (1969) made a similar analysis for the non-Abelian case (essentially $SU(3)_L \times SU(3)_R$), showing that anomalous terms arise in the Ward identities due to the singular nature of the spinor loops. Bardeen also calculates an expression for the minimal anomalous divergence in the *axial*-vector current on the assumption that the vector currents themselves are conserved.[4]

[4] There is an interesting sidelight to this analysis. As described above, the Finkelstein, Steinberger and Schwinger calculations involve a proton loop (or triangle). If one does the calculation in

4. Clarification and Development

Chiral symmetry is a global property of the Lagrangian. It is incompatible with neutral pion decay. Neutral pion decay is a fact. Upon regularisation, neutral pion decay emerges naturally via the anomalous terms in the Ward identity. Thus chiral symmetry is a property of a classical Lagrangian that does not survive its quantisation via perturbation theory. In view of this, questions arise as to what other classical symmetries might be lost in a perturbative calculation, and how can we tell? Looked at the other way around, how can we be sure of having encoded a desired symmetry in a physical theory? In particular, since renormalisability depends crucially on symmetry considerations, how can we tell which renormalisability arguments are (or will be) safe? Finally, are anomalies such as we encounter in the breaking of chiral symmetry just an artefact of perturbation theory, or do they have a deeper and more significant source? D. Gross and R. Jackiw (1972) address these questions: 'The anomalies of the axial-vector current are thus a consequence of a chirally invariant regulator procedure for fermions. However it must be emphasized that this difficulty is not merely technical; one must not entertain the hope that eventually a proper regulator procedure will be found' (p. 478). No doubt the emphasis here was enhanced by the shortcomings of the Bell and Jackiw attempt at constructing such a 'proper regulator'. More than a decade later Jackiw (1984) wrote again in a similar vein, having in mind both the Abelian Adler–Bell–Jackiw anomaly and the non-Abelian Bardeen anomaly: 'Nevertheless, there is good reason to believe that anomalies are not an obscure consequence of problems with perturbation theory, but reflect a deep fact about Nature' (p. 278). In 1972, however, neither the 'good reasons' nor the nature of this 'deep fact' were apparent. Gross and Jackiw move toward a deeper understanding by studying the interplay between gauge invariance and renormalisation. Looking at massive vector mesons coupled to an axial-vector current constructed from massless fermions they begin by showing that in the presence of gauge invariance, the axial vector anomaly prevents the resulting theory from being renormalisable. As expected, they localise the source of the difficulty in the anomalous triangle graph. They then extend the scope of non-renormalisability, first to an Abelian version of the spontaneously broken gauge theory of weak interactions and then to non-Abelian vector-meson gauge fields with axial-vector coupling to fermions. In this latter case, a more realistic model for weak interactions, they succeed in deriving an explicit formula for the anomalous contribution to the divergence of the current. Their formula involves a trilinear function of the structure constants of the gauge group associated with the action.

This formula is picked up by H. Georgi and S. Glashow (1972) in the context

more modern terms using a quark loop for u, d and s quarks alone, then the experimental rate for $\pi_0 \rightarrow \gamma\gamma$ decay is smaller than the calculated rate by a factor of three. To make up the difference one needs to triple the u, d and s quarks; i.e. to add another quantum number. Colour does the trick and thus the $\pi_0 \rightarrow \gamma\gamma$ decay rate can be used to lend support to the quantum chromodynamic model. See Aitchison (1982, p. 153).

of what became known as the Glashow–Salam–Weinberg electroweak theory; that is, a unitary theory of the weak and electromagnetic fields, where the vector-meson masses are acquired via the Higgs mechanism. Georgi and Glashow note that the theory must be free of anomalies if it is to be renormalisable. The Gross and Jackiw formula enables them to catalogue anomaly-free models. In particular Georgi and Glashow identify classes of 'safe' Lie algebras (or groups) in which the anomalous Gross and Jackiw term vanishes. They show that this includes all the Lie groups representable by generators each of which is unitarily equivalent to its adjoint. Fortunately, $SU(2)$ is one of these groups; it is safe, with the anomalies posing no barrier to renormalisability. The article by Georgi and Glashow moves toward a structural characterisation for the presence of an anomaly; that is, toward detaching the anomaly from explicit perturbative calculations linked to a particular diagram and relating it instead to features of the gauge transformations that characterize a theory. A similar move was also made by Wess and Zumino (1971) who introduce a set of consistency conditions obtained by combining the Ward identities with the structure relations of the gauge group. They formulate things in terms of an effective action, essentially a functional of the meson and gauge fields sufficient to produce the same expectations with respect to these fields as does the full classical action (with the additional fermion fields). The Ward identities state the invariance of this functional. The anomalous Ward identities give a particular form to its variation. Thus anomalous terms in the Ward identity correspond to transformations of the effective action. Such transformations can be described in terms of infinitesimal gauge operators. The Wess and Zumino consistency conditions require that these transformations of the effective action provide a representation of the Lie algebra formed by the infinitesimal gauge operators. Concentrating on the expression that Bardeen had calculated for the minimal anomalous divergence in the axial-vector currents, Wess and Zumino show that if one knows the leading term in the anomalous Ward identity (that given by Adler for the $\pi_0 \to \gamma\gamma$ amplitude) then their consistency conditions suffice to determine all the other anomalous terms. 'In this sense one can say that [the minimal anomaly] is model independent' and determined 'up to an overall constant, by the structure of the gauge group' (p. 95).

5. The Cohomological Description of the Anomaly

In 1975, when C. Becchi, A. Ravet and R. Stora (1975) described their new approach to renormalisation, they were well aware that it also implied a new understanding of anomalies. Indeed, in their introduction, they claim that anomalies 'can be read off on [*sic*] the classical Lagrangian' (p. 128).

This 1975 paper, however, which only treats the anomaly-free Abelian Higgs–Kibble theory, does not provide a primer for how to read off the anomaly. The treatment of the non-Abelian version (Becchi *et al.*, 1976) contains the first explicit discussion. Stora (1977) gives a concise account, which is as follows.

Let $\Gamma(\phi, A)$ denote the effective Lagrangian as a function of matter fields ϕ and the non-Abelian gauge (vector) field A. As in the treatment by Wess and Zumino, here too the Bardeen anomaly $\Delta(\omega, \phi, A)$ is the effect on Γ, denoted $W(\omega)\Gamma(\phi, A)$, of an infinitesimal gauge transformation ω of A. That is,

$$\Delta(\omega, \phi, A) = W(\omega)\Gamma(\phi, A). \tag{1}$$

Writing the anomaly as an integral over the Euclidean spacetime manifold M, one has

$$\Delta(\omega, \phi, A) = \int_M \alpha(\omega, \phi, A). \tag{2}$$

Now think of the action of ω as a new kind of exterior derivative denoted by $\mathbf{s}$ (from A. Slavnov (1972); it later became known as the BRS transformation). Then equation (1) is the statement that the anomaly is an exact differential with respect to $\mathbf{s}$. ($\Delta = -\mathbf{s}\Gamma$, the minus sign being conventional.) To describe $\mathbf{s}$ explicitly, set

$$\mathbf{s}A = -D_A\omega \tag{3}$$

and

$$\mathbf{s}\omega = -\frac{1}{2}[\omega, \omega]. \tag{4}$$

The first equation states that $\mathbf{s}$ acts on A as does an infinitesimal gauge transformation by ω. (Here $D_A\omega = \mathbf{d}\omega + [A, \omega]$ is the covariant exterior derivative.) The second equation ensures that $\mathbf{s}^2 = 0$ when applied to A or ω. Then the fact that the anomaly is $\mathbf{s}$-exact implies that it is closed. That is,

$$\mathbf{s}\Delta = 0. \tag{5}$$

That the algebraic manipulation agrees with successive gauge transformation of Γ is the Wess–Zumino consistency condition on the anomaly. (This identification also requires $\mathbf{s}$ to be an *anti*-derivation that *anti*-commutes with the ordinary exterior derivative $\mathbf{d}$.) In terms of the integral representation of Δ, the above becomes

$$\mathbf{s}\Delta = \int_M \mathbf{s}\alpha = 0. \tag{6}$$

Assuming M to be compact, there must be some 3-form Q for which

$$\mathbf{s}\alpha = -\mathbf{d}Q. \tag{7}$$

The question, 'is there an anomaly?' has thus been rephrased, 'is there a 4-form whose BRS transformation is an exact differential?'. The anomaly has become the integral of such a 4-form (properly normalised). Notice, too, that to phrase the question this way required making explicit an assumption about the topology of M that had been implicit in the calculations of the contribution of the triangle diagram; namely, that M is compact.

To derive the Bardeen anomaly, as the solution to equation (7), Stora adapts the theory of secondary characteristic classes, due to S. S. Chern and J. Simons

(1971). This adaptation requires treating **s** on an equal footing with **d**. Indeed, Stora (following Slavnov, 1972) associates a degree (the 'ghost number') to **s**, and combines the two operators as $(\mathbf{s} + \mathbf{d})$. Notice that with this new notion of degree, α and Q are each forms of total degree 5: α has ordinary degree 4 (to be integrated over M) and ghost number 1 (since Δ is **s** of something with no dependence on ω), while Q must therefore have ordinary degree 3 and ghost number 2. He then applies Chern and Simons' construction to the cohomology of $(\mathbf{s} + \mathbf{d})$. This construction begins with a G-invariant, symmetric trinomial T (more generally, T might be an invariant, symmetric polynomial) on the Lie algebra, and composes it with the curvature two-form of a connection. The result is an ordinary 6-form, which is $(\mathbf{s} + \mathbf{d})$-exact. That is, one obtains forms Δ_i^{5-i} of ordinary degree i and ghost number $5 - i$ such that

$$(\mathbf{s} + \mathbf{d}) \left(\Delta_0^5 + \Delta_1^4 + \cdots + \Delta_5^0\right) = T(F, F, F). \qquad (8)$$

The anomaly emerges upon breaking this equation down into several equations distinguished by ghost number. After applying $(\mathbf{s} + \mathbf{d})$ as indicated, the only term on the left-hand side of ghost number 6 is $\mathbf{s}\Delta_0^5$; hence, $\mathbf{s}\Delta_0^5 = 0$. The terms of ghost number 5 must likewise vanish (when added together), so $\mathbf{d}\Delta_0^5 + \mathbf{s}\Delta_1^4 = 0$. Continuing in this fashion produces a sequence of identities, ending with $\mathbf{d}\Delta_5^0 = T(F, F, F)$. One of these identites is $\mathbf{d}\Delta_3^2 + \mathbf{s}\Delta_4^1 = 0$. This is exactly equation (7), with forms of the correct degrees. The anomaly is thus the integral over M of Δ_4^1.

This new characterisation and computation of the anomaly in cohomological terms neatly explained Georgi and Glashow's (1972) connection between a trilinear function on the Lie algebra and the anomaly. It also made explicit the role of the topology of M. Furthermore, it readily generalised to higher dimensions. L. Bonora and P. Cotta-Ramusino (1981) seem to have first taken advantage of this in a paper which standardises Stora's argument, generalises it to higher dimensions, and points out that this argument fails to fix the normalisation constant multiplying the anomaly. This would be a minor point, except that this constant, which depends on the fermion field content, can be zero. Indeed, this is the mechanism by which the appearance of quarks of three colours saved the renormalisability of the electroweak theory of Georgi, Glashow and Weinberg (see Aitchison, 1982, p. 155). The cohomological description of the anomaly also raised a new question: for what manifold is $(\mathbf{s} + \mathbf{d})$ the natural exterior derivative? Although in this study we will not trace developments so far as to present the answer, we will return to this line of thought after a brief look at what was happening in the same period on another track.

6. The Anomaly as the Index of the Dirac Operator

Between 1976 and 1979 several authors identified the Adler–Bell–Jackiw anomaly as the analytic index of the Dirac operator. The latter is the difference between the number of zero modes (eigenfunctions with eigenvalue zero) of positive chirality and the number of zero modes of negative chirality. The chirality of a spinor ψ corresponds to the sign in the equation $\gamma_5\psi = \pm 1$. The pioneering work of G. 't Hooft (1976a) links the Adler–Bell–Jackiw anomaly to what he describes as a 'topological quantum number' (now known as the instanton number or Pontryagin index). Then 't Hooft (1976b) draws a connection between the same anomaly and the zero modes of the Dirac operator. Papers by J. Kiskis (1977), L. Brown, R. Carlitz and C. Lee (1977) and N. Nielsen and B. Schroer (1977), which are all based on the Green's function approach to the anomaly due to Schwinger (1951), focus on these connections. Citations give credit to S. Coleman who apparently outlined an updated version of Schwinger's argument, but did not publish it. This approach does not rely directly on perturbation theory and hence provides *theoretical* evidence for the inescapability of the anomaly.

Coleman's argument, as presented by Kiskis, starts with the usual expression for the expectation of the current in terms of the Green's function, or propagator, of the Fermi field in the presence of a gauge field:

$$\langle \psi(x)^\dagger \gamma^\mu \gamma_5 \psi(x) \rangle = \gamma^\mu \gamma_5 G(x, x; A). \tag{9}$$

Expanding the Green's function in eigenfunctions of the Dirac operator, regulating the infinite expression arising as the result of evaluating a Green's function at two coincident points, and taking the divergence of the resulting equation leads to a Ward identity. This identity involves the usual triangle anomaly and a sum over only the zero-modes of the Dirac operator. Integrating over spacetime, with some assumptions about boundary contributions, leads to an equality; namely, the analytic index of the Dirac operator is equal to the (integrated) anomaly. The question of boundary contributions, and a subtle question about additional contributions from the continuous part of the spectrum of the Dirac operator prevent Kiskis from asserting that this equality is correct. In his acknowledgements, Kiskis credits L. Dolan and K. Macrae with pointing out the connection between his work and the index theorems of M. Atiyah and I. M. Singer (1968a; 1968b), Atiyah and G. Segal (1968) and Atiyah, R. Bott and V. Patodi (1973).

Like Kiskis, so too Brown *et al.* (1977) base their derivation on manipulations of the Green's function for a massive fermion field ψ; however, they do not explicitly employ an eigenfunction expansion. After regularising the formal expression for the Green's function, they relate the spacetime integral of the large mass limit to the integrated anomaly, and they relate the zero mass limit to the index. (The former relation follows from the asymptotic form of the Green's function, while the latter relation is almost immediate from the formal expression for the Green's function.) They then note that the formal expression is independent of the mass (which they point out is consistent with the Pauli–Villars regulation in perturbation theory) and conclude that the index is equal to the anomaly.

Nielsen and Schroer (1977) differ from Kiskis in working with massless Fermi fields and in defining the Green's function in terms of the inverse $G'(x, y; A)$ of the *restriction* of the Dirac operator to the orthogonal complement of its kernel. Moreover, they explicitly work on a compactified spacetime (S^4 in four dimensions and S^2 in two), which ensures that the Dirac operator has a discrete spectrum. This avoids the difficulties Kiskis encountered. Where Kiskis used Pauli–Villars regularisation, Nielsen and Schroer give their own prescription for regulating the right-hand side of equation (9), according to which its divergence is finite and manifestly gauge-invariant. This leads to a current whose divergence is the Adler–Bell–Jackiw anomaly plus a sum over (normalised) zero modes, $\sum_n \phi_{0n}^\dagger \gamma_5 \phi_{0n}$. That is, they obtain essentially the same Ward identity as did Kiskis. This follows from using the definition of G' to compute the divergence of their presciption for regulating the term on the right-hand side. They then extract the short-distance behaviour of G' (that is, $G'(x, y; A)$ for y near to x) to compute the divergence directly, and find that it vanishes. Upon integration over spacetime, they thus obtain Kiskis' equality between the index and the integrated anomaly. (The sum over zero modes becomes the index, because $\gamma_5 \phi_{0n} = \pm 1$.) Having, like Brown *et al.*, read 't Hooft's papers, Nielsen and Schroer were aware that the anomaly is the topological (Pontryagin) index, and, indeed, they present the above arguments as an alternative proof of the Atiyah–Singer index theorem (which asserts the equality of the topological and analytic indices) in the case of the Dirac operator on S^4. They seem to be the first to identify the usual conditions on the large-distance behaviour of the gauge fields explicitly as a compactification of spacetime.

K. Fujikawa (1979) rederived the anomaly directly from the fermion functional integral and, also, without recourse to perturbation theory. Likewise, as a by-product of his calculation, Fujikawa was able to identify the anomaly as the index of the Dirac operator and relate it via the index theorem to the Pontryagin index. Moreover, Fujikawa gave a new interpretation for the anomaly: it is the (unanticipated) variation of the fermionic functional integral 'measure' under chiral transformations. His idea is to consider the effect of an infinitesimal chiral transformation on the generating function for the Euclidean Green's functions,

$$Z(\eta, \bar{\eta}, J) = \frac{1}{N} \int e^{\bar{\psi} \not{\partial}_A \psi + (F_A, F_A) + \bar{\psi}\eta + \bar{\eta}\psi - (J, A)} \mathcal{D}\bar{\psi} \mathcal{D}\psi \mathcal{D}A. \tag{10}$$

The Green's functions are derivatives of Z with respect to the external currents η, $\bar{\eta}$ and J, evaluated where each of these currents are zero. To calculate the effect of a chiral transformation, expand ψ in terms of the (orthogonal) eigenstates ϕ_n, with eigenvalues λ_n of the Dirac operator. In this basis, the first term in the exponent diagonalises, the fermionic measure $\mathcal{D}\bar{\psi}\mathcal{D}\psi$ becomes an infinite product of measures on the spaces of coefficients in the expansion of ψ, and the chiral transformation is given by an infinite-dimensional matrix. The determinant of this matrix is the Jacobian describing the effect of the chiral transformation on the fermionic measure. The logarithm of this Jacobian is simply

$$\sum_n \phi_n^\dagger \gamma_5 \phi_n. \tag{11}$$

Upon regularisation, this yields the usual expression for the chiral anomaly. Fujikawa remarks that because $\not{\partial}_A$ and γ_5 anti-commute, only zero-modes can contribute to the above sum. (For the other modes, inserting $\frac{\not{\partial}_A}{\lambda_n}$ to either side of γ_5 should not affect the term in question, but the anti-commutativity implies that the two choices are opposite in sign.) Hence, this sum reduces to Nielsen and Schroer's sum over zero modes and is precisely the analytic index of the Dirac operator. Fujikawa then uses the Atiyah–Singer index theorem to identify the analytic index with the Pontryagin index, thereby identifying the anomaly with this topological object. He describes this as an independent check on his direct calculation.

7. The Unifying Geometric Framework

To return to the trail of the cohomological developments, the question of the underlying geometry was addressed by J. Thierry-Mieg (1980) who tried to tie the ghost forms directly to the principal fibre bundle. M. Quirós *et al.* (1981) presented a geometrical picture involving product bundles over the principal fibre bundle. In the same year, Bonora and Cotta-Ramusino (1981) gave a superspace formulation and subsequently presented a truly geometric description (1983). T. R. Ramadas (1984) gave a slightly different geometric setting. Indeed, during 1984, there were a flurry of papers elaborating the cohomological constructions, providing a geometric basis and apparently rediscovering the view of the anomaly as the obstruction (identifiable as the Dirac index) to the consistent definition of the determinant of the Dirac operator (e.g. Stora, 1984; Zumino, 1984; Atiyah and Singer, 1984). In addition, E. Witten (1983) gave a topological argument that determined the normalisation of the anomaly (which was also fixed by the interpretation of the anomaly as an index).

It is clear that in the period in which these publications were being developed, most of the authors mentioned (along with many others) were in close, informal communication. The circular flow of ideas and information makes it difficult to track the generation of specific ideas on the basis of published accounts. As an instance of the circular paths of citation in this period, note that Atiyah and Singer (1984) cite Stora (1984), who in turn refers to Atiyah and Singer (1984) and also back to lectures Singer presented in 1982 (in Santa Barbara). Moreover, in his acknowledgements Stora mentions intense correspondence with Singer and Zumino and discussions with six other authors (some of whom Atiyah and Singer similarly acknowledge).

A significant feature of the interchange amongst and between mathematicians and physicists at this time is their insistence on exposing their own colleagues to the other community's perspective, and, quite strikingly, expounding the other community's notation. Stora's 1976 paper, in which he writes his connections and derivatives in the coordinate-free notation of differential forms (maybe after working through Chern and Simons (1974))is perhaps the earliest example. At that time this coordinate-free formulation was still somewhat novel even among differential geometers. In 1984, Zumino took pains to re-write the current (non-)conservation laws in this notation. Bonara and Cotta-Ramusino (1981; 1983) likewise introduce this notation in the spirit of presenting something new and useful. On the mathematical side, Atiyah and Singer devote half a page of their very concise four-page paper to a description of the Yang–Mills functional integral and to a heuristic argument linking the first Chern class of the Dirac determinant bundle with the anomaly (viewed as the obstruction to defining a gauge-invariant effective action).

We see here a willingness in both communites to consider topics each would have dismissed a decade before as 'irrelevantly abstract' (physicist to mathematician) or as 'hopelessly vague' (mathematician to physicist). To get some sense of the excitement generated by the interplay at that time we will describe Atiyah and Singer's (1984) elaboration and practical application of the geometric picture underlying the BRS cohomology. The space of interest is a bundle $\mathcal{Q}$ over $M \times \mathcal{A}/G$, with group G. Here $\mathcal{A}$ is the space of connections on a fixed principal fibre bundle P over spacetime M, and $\mathcal{G}$ is the group of gauge transformations (restricted to be the identity on the fiber over a fixed point of M). One of the sources of mathematical interest in the bundle $\mathcal{Q}$ is the fact that $\mathcal{A}$, $\mathcal{G}$ and the quotient are infinite-dimensional spaces. To construct $\mathcal{Q}$, begin with the space $P \times \mathcal{A}$ and quotient by the action of $\mathcal{G}$ on each factor. Now the action of G on P defines $\mathcal{Q}$. The latter has a natural metric (determined by a metric on G and a metric on M) which defines a connection w on $\mathcal{Q}$ (in the $\mathcal{A}/G$-directions) as the orthogonal complement of the G-orbits.

Atiyah and Singer (1984) connect the bundle $\mathcal{Q}$ with the anomaly by two different arguments. One is a functional integral argument. Starting with the fermionic functional integral for the Euclidean Green's function,

$$\mathcal{I}_r = \int e^{\overline{\psi}\partial\!\!\!/_A\psi}\overline{\psi}(y_1)\psi(x_1)\cdots\overline{\psi}(y_r)\psi(x_r), \tag{12}$$

they argue that to make sense of $\mathcal{I}_r$ in the presence of zero modes of $\partial\!\!\!/_A$ requires that the Dirac determinant line bundle be trivial (equivalently, its first Chern class must vanish). The Dirac determinant line bundle has the same Chern characters as the index bundle Ind $\partial\!\!\!/$. The latter is the vector bundle over $\mathcal{A}/G$ defined by the (formal) difference between the kernel and cokernel of the Dirac operator $\partial\!\!\!/_A$ mapping right-handed to left-handed spinor fields. The connection between Ind $\partial\!\!\!/$ and $\mathcal{Q}$ is given by the index theorem, which computes the Chern character of Ind $\partial\!\!\!/$ in terms of the characteristic classes of the vector bundle associated to $\mathcal{Q}$ by a representation of G. These latter classes can be expressed directly in terms of polynomials in the curvature of the connection w. Integrating over M the form constructed from the curvature gives a form on $\mathcal{A}/G$, which, in turn lifts to a form on $\mathcal{A}$. Restriction to the orbits defines a form on $\mathcal{G}$, which represents a cohomology class. In particular, for $M = S^4$, $G = SU(N)$ and the identity representation, the computation of the first Chern class of Ind $\partial\!\!\!/$ and the above-mentioned integration, lift and restriction reproduces exactly Stora's cohomological computation. Thus the Bardeen anomaly appears as that element of the cohomology of $\mathcal{G}$ which corresponds to the first Chern class of Ind $\partial\!\!\!/$, and the significance of the trinomial in the Lie algebra is that it serves to represent the relevant Chern class of the vector bundle associated to the principal bundle

$\mathcal{Q}$. Moreover, the (integrated) Adler–Bell–Jackiw anomaly is the zeroth Chern class of Ind $\partial\!\!\!/$.[5] In this picture, these anomalies are two of a family of Chern classes of Ind $\partial\!\!\!/$, all but the zeroeth of which correpond to elements of the cohomology of $\mathcal{G}$.

The other way to connect Ind $\partial\!\!\!/$ with the anomaly is to consider the operator $T_\phi = \partial\!\!\!/_A^* \; \partial\!\!\!/_{\phi \bullet A}$. Clearly, T_ϕ varies with ϕ (or not) just as does $\partial\!\!\!/_A$. Since, T_ϕ maps right-handed spinors to right-handed spinors, its determinant is relatively straightforward to interpret. In the absence of zero modes, zeta-function regularisation gives a well-defined interpretation of $\det T_\phi$ for a given ϕ. The existence of a gauge-invariant definition of the determinant of the Dirac operator, and hence of a gauge-invariant effective action, thus requires that $\mathbf{d}\,(\det T_\phi) = 0$. (Here $\det T_\phi$ is function from $\mathcal{G}$ to the non-zero complex numbers.) Less stringently, one might ask whether

$$\frac{\mathbf{d}\,(\det T_\phi)}{\det T_\phi} = 0. \tag{13}$$

The left-hand side represents a generator of the first cohomology of $\mathcal{G}$ (the pullback of the generator of the first cohomology of the complex plane minus the origin). The direct computation of the regularised determinant shows that this generator is the same as that obtained from the first Chern character of Ind $\partial\!\!\!/$; that is, it is the Bardeen anomaly.

Atiyah and Singer's geometric construction and index computation answered some of the most pressing questions surrounding the anomaly. The topological origin of the anomaly was made precise by identifying it as an element of the first cohomology of $\mathcal{G}$. The geometric space underlying the BRS cohomolgy was seen to be the vector bundle over $M \times \mathcal{A}/G$ associated to $\mathcal{Q}$. Furthermore, this construction was quite general: M, G, P could in principle be arbitrary. Finally, this setting raised the question of whether there was physical significance to the higher Chern characters of Ind $\partial\!\!\!/$.

8. Summary

Figure 5 diagrams the evolution of the anomaly from its origins as the triangle diagram's anomalous contribution to neutral pion decay to its realisation as a Chern character of a line bundle over an infinite-dimensional manifold. The elegance and completeness of the latter picture inspired physicists to tackle the formidable task of gaining control over the machinery of global differential geometry.[6]

[5] The *un*integrated Adler–Bell–Jackiw anomaly appears as the second Chern character of the vector bundle associated to $\mathcal{Q}$.

[6] Physicists were of course not unanimous in this. L. Alvarez-Gaumé and P. Ginsparg (1984) extend Fujikawa's fermionic integral approach to treat the Bardeen anomaly, which they interpret directly in terms of the cohomology of the space of gauge transformations, without referring to an underlying geometric picture. Though they are well-versed in the work of Atiyah and Singer, and adopt the language of differential forms, their explicit aim is to by-pass the families index

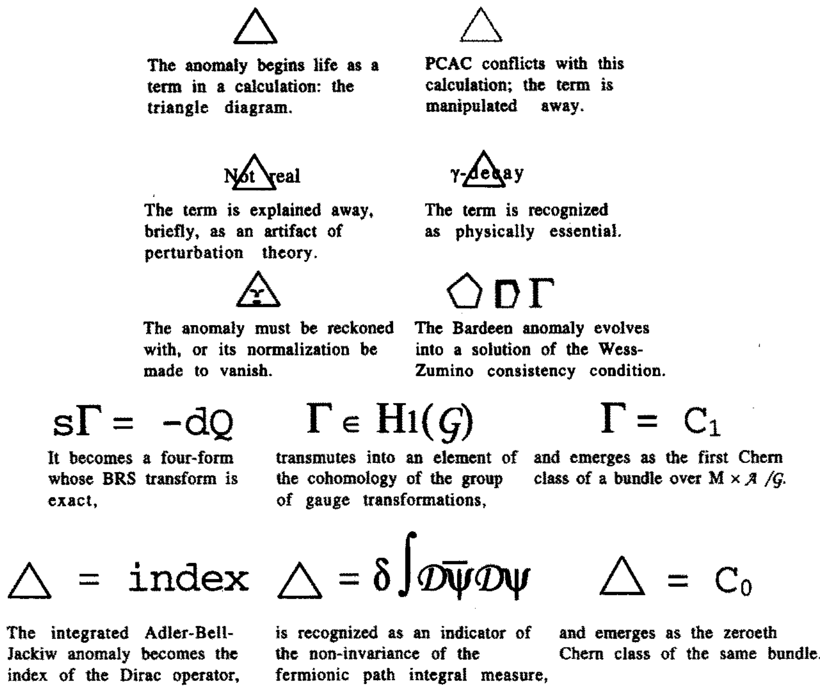

Fig. 2. *The evolution of the anomaly.*

Stora, Zumino and Jackiw (among many others) were already on track to do so; each wrote primarily expository papers between 1984 and 1986 on their perceptions of the intimate relation between anomalies and topology (Stora, 1984; Zumino, 1984; Jackiw, 1985). Mathematicians saw this picture not only as a striking application of intrinsically important mathematics, but as an indication of the potential of quantum field theory as a source of intriguing and tractable problems in differential geometry and topology. It is hard to imagine, for instance, that the topology of infinite-dimensional spaces, such as $\mathcal{Q}$ or $\mathcal{G}$ would have been addressed for internal, mathematical reasons. Note, too, that the quantum-field-theoretic Feynman diagrams and functional integrals compute the Chern characters of infinite-dimensional bundles in a way that is not obviously equivalent to Chern and Simons' construction.

The achievement of the geometrical picture of the anomaly, and of the BRS cohomology, marks a reasonable point to end the story of penetrating showers in lead; that is, the story of the development, understanding, and successive use of the chiral anomaly. Later chapters while filling in details and extending its range of applicability, preserve this basic picture.[7] This endpoint, however,

theorem: 'We hope, in particular, that the pedestrian perspective on the families index theorem implicit in our approach will be useful for those physicists to whom K-theory proves anathema' (p. 452).

[7] These developments include Moore and Nelson's (1985) work on the σ-model anomaly, where

also marks the start of a period of intense interaction between the physics and mathematics communities. The interaction has become so pervasive that critics in each field complain that it has fundamentally distorted their discipline. [8] Nevertheless, the interaction has brought to each field new ideas, new techniques and new results which neither could have developed in isolation.

The global geometric approach to anomalies led directly to new physical constructs, Chern–Simons field theory and the Wess–Zumino–Witten model, which in turn led to more general topological (and conformal) quantum field theories and has played an important role in the development of string theory. [9] Some of these theories have direct implications for physical phenomena ranging from the rather hypothetical ground state of the universe to the observed fractional quantum Hall effect. Moreover, these physical theories play back into mathematics in such diverse realms as Donaldson theory, knot invariants, and infinite-dimensional Lie algebras. Topologists and analysts are struggling to keep up with the pace at which results (to be made rigorous, to be re-derived conventionally, and to be applied) are suggested by these quantum field theories and their functional integrals. There is no way the triangle diagram in meson decay could explain the fractional quantum Hall effect without the topological perspective on the anomaly, nor could the use of secondary characteristic classes have led to a generalisation of the Jones knot polynomial without the intervention of quantum field theory. This remarkable interplay between physics and mathematics is the legacy of the decay of the neutral pion.

Acknowledgements—DF would like to acknowledge the hospitality of the Department of Mathematics at M. I. T. AF would like to thank the Dibner Institute for the History of Science and Technology at M. I. T. for a visiting fellowship in support of this research. The authors thank S. Axelrod, S. Sawin, and, most especially, I. M. Singer for useful suggestions in response to a draft of this paper.

References

Adler, S. L. (1969) 'Axial-Vector Vertex in Spinor Electrodynamics', *Physical Review* **177**, 2426–2438.

the global geometric picture is arguably essential to the computation of the anomaly, and the development, beginning with Bonora and Cotta-Ramusino (1983), of the novel mathematical category of local cohomology.

[8] In the mathematics community this view has been expressed, for example, in the very existence of the workshop *Proof and Progress in Mathematics* held at Boston University on 12 February 1996. The papers presented there by A. Jaffe and S. Maclane addressed this concern directly. The physics community's concerns in this respect have been communicated privately to one of the authors by former graduate and post-doctoral students in theoretical physics, most of whom are no longer in the field.

[9] Indeed, in the early 1980s, it was the cancellation of anomalies that sparked renewed interest in string theory. Most recently, D. Freed, in a seminar at M. I. T. (29 October 1996) described a result of P. Hořava and E. Witten (1996) as applying Atiyah and Singer's construction to 11-dimensional supergravity. Physically, this provides evidence for a conjectured duality with a certain 10-dimensional string theory. Freed notes that an extension of the families index theorem, which he proves directly, underlies this result. Here the connections between the mathematics and the physics are very tight.

Aitchison, I. (1982) *An Informal Introduction to Gauge Field Theories* (Cambridge: Cambridge University Press).
Alvarez-Gaumé, L. and Ginsparg, P. (1984) 'The Topological Meaning of Non-Abelian Anomalies', *Nuclear Physics* **B243,** 449–474.
Atiyah, M. and Segal, G. (1968) 'The Index of Elliptic Operators: II', *Annals of Mathematics* **87,** 531–545.
Atiyah, M. and Singer, I. M. (1968a) 'The Index of Elliptic Operators: I', *Annals of Mathematics* **87,** 484–530.
Atiyah, M. and Singer, I. M. (1968b) 'The Index of Elliptic Operators: III', *Annals of Mathematics* **87,** 546–604.
Atiyah, M. and Singer, I. M. (1984) 'Dirac Operators Coupled to Vector Potentials', *Proceedings of the National Academy of Science* **81,** 2597–2600.
Atiyah, M., Bott, R. and Patodi, V. K. (1973) 'On the Heat Equation and the Index Theorem', *Inventiones Mathematicae* **19,** 279–330.
Bardeen, W. (1969) 'Anomalous Ward Identities in Spinor Field Theories', *Physical Review* **184,** 1848–1859.
Bell, J. and Jackiw, R. (1969) 'A PCAC puzzle: $\pi_0 \rightarrow \gamma\gamma$ in the σ-Model', *Nuovo Cimento* **60A,** 47–61.
Becchi, C., Rouet, A. and Stora, R. (1975) 'Renormalization of the Abelian Higgs–Kibble Model', *Communications in Mathematical Physics* **42,** 127–162.
Becchi, C., Rouet, A. and Stora, R. (1976) 'Renormalization of Gauge Theories', *Annals of Physics* **98,** 287–321.
Bonora, L. and Cotta-Ramusino, P. (1981) 'ABJ Anomalies and Superfield Formalism in Gauge Theories', *Physics Letters* **B107,** 87–90.
Bonora, L. and Cotta-Ramusino, P. (1983) 'Some Remarks on BRS Transformations, Anomalies and the Cohomology of the Lie Algebra of the Group of Gauge Transformations', *Communications in Mathematical Physics* **87,** 589–603.
Brown, L., Carlitz, R. and Lee, C. (1977) 'Massless Excitations in Pseudoparticle Fields', *Physical Review* **D16,** 417–422.
Chern, S. S. and Simons, J. (1971) 'Some Cohomology Classes in Principal Fiber Bundles and Their Applications to Riemannian Geometry', *Proceedings of The National Academy of Science (USA)* **68,** 791–794.
Chern, S. S. and Simons, J. (1974) 'Characteristic Forms and Geometrical Invariants', *Annals of Mathematics* **99,** 48–69.
Feynman, R. (1949) 'Space–Time Approach to Quantum Electrodynamics', *Physical Review* **76,** 769–789.
Finkelstein, R. J. (1947) 'The γ-Instability of Mesons', *Physical Review* **72,** 415–422.
Fretter, W. B. (1948) 'Penetrating Showers in Lead', *Physical Review* **73,** 41–46.
Fujikawa, F. (1979) 'Path-integral Measure of Gauge-Invariant Fermion Theories', *Physical Review Letters* **42,** 1195–1198.
Gell-Mann, M. (1962) 'Symmetries of Baryons and Mesons', *Physical Review* **125,** 1067–1084.
Georgi, H. and Glashow, S. (1972) 'Gauge Theories Without Anomalies', *Physical Review* **D6,** 429–431.
Ginsparg, P. (1986) 'Applications of Topological and Differential Geometric Methods to Anomalies in Quantum Field Theory', in J. Abad, M. Asorey and A. Cruz (eds), *New Perspectives in Quantum Field Theories* (Singapore: World Scientific), pp. 29–81.
Gross, D. and Jackiw, R. (1972) 'Effect of Anomalies on Quasi-Renormalizable Theories', *Physical Review* **D6,** 477–493.
Hŏrava, P. and Witten, E. (1996) 'Eleven-dimensional Supergravity on a Manifold with Boundary', hep-th/9603142.
Jackiw, R. (1972) 'Field Theoretic Investigations into Current Algebra' in S. Treiman, R. Jackiw and D. Gross, *Lectures on Current Algebra and its Applications* (Princeton: Princeton University Press), pp. 97–254.

Jackiw, R. (1984) 'Topological Investigations of Quantized Gauge Theories' in B. DeWitt and R. Stora (eds), *Relativity, Groups & Topology II* (Amsterdam: North Holland), pp. 221–331.
Jackiw, R. (1985) 'Topological Investigations of Quantized Gauge Theories', in S. Triemann, R. Jackiw, B. Zumino and E. Witten (eds), *Current Algebra and Anomalies* (Princeton: Princeton University Press/World Scientific), pp. 211–359.
Kiskis, J. (1977) 'Fermions in a Pseudoparticle Field', *Physical Review* **D15,** 2329–2336.
Moore, G. and Nelson, P. (1985) 'The Aetiology of Sigma Model Anomalies', *Communications in Mathematical Physics* **100,** pp. 83–132.
Nash, C. (1986) 'Contemporary Gravitational and Yang–Mills Theories—from Anomalies to Non-Commutative Geometry', in J. Hietarinta and J. Westerholm (eds), *Topological and Geometrical Methods in Field Theory (Espoo, Finland June 1986)* (Singapore: World Scientific), pp. 355–372.
Nielsen, N. and Schroer, B. (1977) 'Axial Anomaly and Atiyah–Singer Theorem', *Nuclear Physics* **B127,** pp. 493–508.
Quirós, M. *et al.* (1981) 'Geometrical Structure of Faddeev–Popov Fields and Invariance Properties of Gauge Theories', *Journal of Mathematical Physics* **22,** pp. 1767–1774.
Ramadas, T. R. (1984) 'The Wess–Zumino Term and Fermionic Solitons', *Communications in Mathematical Physics* **93,** 355–365.
Rosenfeld, H. A. *et al.* (1968) 'Data on Particles and Resonant States', *Reviews of Modern Physics* **40,** 77–128.
Schwinger, J. (1951) 'On Gauge Invariance and Vacuum Polarization', *Physical Review* **82,** 664–679.
Slavnov, A. (1972) 'The Ward Identities in Gauge Theories', *Teoreticheskaia I Matematicheskaia Fizika* **10,** 153–161.
Steinberger, J. (1949) 'On the Use of Subtraction Fields and the Lifetimes of Some Types of Meson Decay', *Physical Review* **76,** 1180–1186.
Stora, R. (1977) 'Continuum Gauge Theories', in M. M. L. Lévy and P. Mitter (eds), *New Developments in Quantum Field Theory and Statistical Mechanics: Cargèse 1976* (New York: Plenum Press), pp. 201–224.
Stora, R. (1984) 'Algebraic Structure and Topological Origin of Anomalies', in G. t'Hooft *et al.* (eds), *Progress in Gauge Field Theory* (New York: Plenum Press), pp. 543–562.
Sutherland, G. D. (1967) 'Current Algebra and Some Non-Strong Mesonic Decays', *Nuclear Physics* **B2,** 433–440.
Taylor, C. J. (1958) 'Beta Decay of The Pion', *Physical Review* **110,** 1216.
Thierry-Mieg, J. (1980) 'Geometrical Reinterpretation of Faddeev–Popov Ghost Particles and BRS Transformations', *Journal of Mathematical Physics* **21,** 2834–2838.
't Hooft, G. (1976a) 'Symmetry Breaking through Bell–Jackiw Anomalies', *Physical Review Letters* **37,** 8–11.
't Hooft, G. (1976b) 'Computation of the Quantum Effects due to a Four-dimensional Pseudoparticle', *Physical Review* **D14,** 3432–3450.
Veltman, M. (1967) 'Theoretical Aspects of High Enery Neutrino Interactions', *Proceedings of The Royal Society* **A301,** 107–112.
Wess, J. and Zumino, B. (1971) 'Consequences of Anomalous Ward Identities', *Physics Letters* **37B,** 95–97.
Wigner, E. P. (1967) *Symmetries and Reflections* (Bloomington: Indiana University Press).
Witten, E. (1983) 'Global Aspects of Current Algebra', *Nuclear Physics* **B223,** 422–432.
Zumino, B. (1984) 'Chiral Anomalies and Differential Geometry', in B. DeWitt and R. Stora (eds), *Relativity, Groups & Topology II* (Amsterdam: North Holland Press), pp. 1291–1322.