

the discovery of the violation of P invariance, the CPT theorem came to play a very important part in discussions of the mode of the violation. Again I cannot help wondering what Weyl would have said about the CPT theorem had he lived two years longer into 1957, not only because he had, quite mysteriously in 1930, written about these three symmetries C , P , and T all in one sentence, as I quoted before, but also because the theorem was given a more profound foundation by Jost⁹ and that foundation involved the Lorentz group and the concept of analytic continuation, subjects that were dear to Weyl's heart.

II.

The next contribution of Hermann Weyl to physics that I shall discuss is gauge theory. There were three periods during which Weyl wrote about gauge theory which we shall now discuss separately.

During the first we find three papers, all written in the years 1918–1919^{10,11,12}. The most important of these is the middle one and indeed throughout his life, when he referred to gauge theory, Weyl always referred to this paper. The background of his thinking at that time can be traced through the preface of the various editions of his book *Space, Time, Matter* and through his articles of 1917–1919. It seemed that Weyl, evidently inspired by the work of Einstein on gravity (1916), and also by the work of Hilbert, Lorentz, and F. Klein, was searching for a geometrical theory that would embrace electromagnetism as well as gravity. He was also influenced by Mie who had, in 1912–1913, attempted to formulate a theory of the electron that does not involve divergent field quantities inside of the electron.

In the beginning paragraphs of¹¹ Weyl said that while Einstein's gravity theory depended on a quadratic differential form, electromagnetism depended on a linear differential form $\Sigma \phi_\mu dx_\mu$ (which in today's notations is $\Sigma A_\mu dx^\mu$). The next crucial sentences are,¹³

The later work of Levi-Civita, Hessenberg and the author shows quite plainly that the fundamental conception on which the development of Riemann's geometry must be based if it is to be in agreement with nature, is that of the infinitesimal parallel displacement of a vector. ... *But a truly infinitesimal geometry must recognize only the principle of the transference of a length from one point to another point infinitely near the first.* This forbids us to assume that the problem of the transference of length from one point to another at a finite distance is integrable, more particularly as the problem of

⁹ R. Jost, *Helv. Phys. Acta* 30, 409 (1957).

¹⁰ Weyl, 1918; GA II, p. 1.

¹¹ Weyl, 1918; GA II, p. 29.

¹² Weyl, 1919; GA II, p. 55.

¹³ I quote here from a 1923 translation of¹¹ in *The principle of Relativity* by H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, translated by W. Perrett and G. B. Jeffery, (first published by Methuen and Co. 1923, reproduced by Dover Publications, 1952).

the transference of direction has proved to be non-integrable. Such an assumption being recognized as false, a geometry comes into being, which...explains...also...*the electromagnetic field*. (Italics original) Thus was born¹⁴ the idea of a nonintegrable scale factor which appeared in one Weyl paper, i.e. paper¹⁰, explicitly as

$$(3) \quad e^{\int^{\alpha} d\phi}$$

Weyl then argued that the addition of a gradient $d(\log \lambda)$ to $d\phi = \Sigma \phi_{\mu} dx_{\mu}$ should not change the physical content of the theory, thus concluding that

$$(4) \quad F_{\mu\nu} = \frac{\partial \phi_{\mu}}{\partial x_{\nu}} - \frac{\partial \phi_{\nu}}{\partial x_{\mu}}$$

has "invariant significance". He naturally then identified $F_{\mu\nu}$ with the electromagnetic field and put

$$(5) \quad \phi_{\mu} = (\text{constant}) A_{\mu}$$

where A_{μ} is the electromagnetic potential. Thus electromagnetism was conceptually incorporated in this theory into the geometrical idea (3) of a nonintegrable scale factor.

Invariance of the theory with respect to the addition $d\phi \rightarrow d\phi + d(\log \lambda)$ led Weyl to the name¹¹ "Maßstab-Invarianz" which was translated^{13,15} as "measure-invariance" and "calibration invariance". Later the German term became "Eich Invarianz" and the English term became "gauge invariance",¹⁵.

When Weyl's paper¹¹ was published in the Sitzber. Preuss. Akad. Wiss. in 1918 there were appended to the end of the paper a postscript by Einstein and a reply by Weyl. This unusual development came about, according to Hendry¹⁷, because while at first Einstein was impressed with Weyl's preprint, he later had a strong objection. Nernst and Planck apparently shared Einstein's objection and they demanded on behalf of the Berlin Academy that Einstein's opinion be appended to Weyl's paper as a postscript.

What was the essence of Einstein's objection? Einstein argued that if Weyl's idea of a nonintegrable scale factor is right, then if one takes two clocks and starts them from one point 0 and brings them along different paths back to the same point 0, their scales would have continuously changed. Thus by the time

¹⁴ In 1950 when Weyl reviewed 50 years of relativity (GA IV, p. 421) he referred to the origin of this 1918 idea by saying that if a vector transported around a closed loop back to its original position could change its direction, "Warum nicht auch seine Länge?"

¹⁵ See p. 528 reference ¹⁶.

¹⁶ Chen Ning Yang, *Selected Papers 1945-1980 with Commentary*, (Freeman and Co. 1983).

¹⁷ J. Hendry: *The Creation of Quantum Mechanics and the Bohr-Pauli Dialogue*, (Reidel Publishing Co. 1984).

the
ha
Th
Ein
ow
pa
ca
bu
fee
dis

Ac
we
pla

192
nar
me
In
by

shc

(6)

18
21,
19
20
21
Ch:



Fig. 3. Illustration for Einstein's Gedankenexperiment

they reached back to 0, since they have traced different histories, they would have, in general, different sizes. They would thus keep time at different rates. Therefore, a clock's measure of time depends on its history. If that is the case, Einstein argued, there cannot be physics, because everybody would have his own laws, and there would be chaos. Weyl's reply, also appended to this paper, did not really explain away the difficulty. In the years 1918–1921 he came back^{18,19} to this subject several times. He did not resolve the problem, but his attempts clearly indicated a strong devotion to the original idea. His feelings can perhaps be gleaned from a sentence he wrote²⁰ in 1949 when discussing the events after Einstein's discovery of general relativity:

A lone wolf in Zürich, Hermann Weyl, also busied himself in this field; unfortunately he was all too prone to mix up his mathematics with physical and philosophical speculations.

Pauli also objected to Weyl's theory, but more on philosophical grounds. According to Mehra and Rechenberg²¹, and to Hendry¹⁷, Pauli's objections were of importance to the subsequent emphasis on the "observable" that was to play a key role in the 1925 Heisenberg discovery of quantum mechanics.

Now we come to the second period of Weyl and gauge theory. In 1925 to 1927, quite unrelated to Weyl's gauge theory, a revolution took place in physics, namely quantum mechanics. One of the important points in quantum mechanics was that the momentum p_μ becomes a differential operator $-i\hbar\partial_\mu$. In 1927, Fock and London independently pointed out that if p_μ is to be replaced by $-i\hbar\partial_\mu$, then the quantity

$$p_\mu - \frac{e}{c} A_\mu$$

should be replaced similarly by

$$(6) \quad -i\hbar\partial_\mu - \frac{e}{c} A_\mu = -i\hbar \left(\partial_\mu - \frac{ie}{\hbar c} A_\mu \right).$$

¹⁸ See the record of a discussion between Weyl, Pauli, and Einstein at Bad Nauheim, *Phys. Z.* 21, 649–651 (1920).

¹⁹ Weyl 1921; GA II, p. 260.

²⁰ Weyl 1949; GA IV, p. 394.

²¹ J. Mehra and H. Rechenberg, *The Historical Development of Quantum Theory*, Vol. 2, Chapter 5, (Springer-Verlag, 1982).

(The quantity $p_\mu - \frac{e}{c} A_\mu$ had already been known to be important in the dynamics of a charged particle.) In London's article, which had the title "Quantum Mechanical Meaning of the Theory of Weyl", it was pointed out that the expression $\partial_\mu - \frac{ie}{\hbar c} A_\mu$ in (6) is similar to the expression $(\partial_\mu + \phi_\mu)$ in Weyl's theory. Thus instead of (5) the identification should be

$$(7) \quad \phi_\mu = -\frac{ie}{\hbar c} A_\mu.$$

Now $\frac{e}{\hbar c}$ is a numerical constant. Therefore, (7) is really the same as Weyl's original identification (5) except for the insertion of $-i$ ($i = \sqrt{-1}$).

But this insertion, although trivial formally, has profound physical consequences, because it changes the meaning of the nonintegrable scale factor (3) into

$$(8) \quad \exp\left(-i \frac{e}{\hbar c} \int_p^q A_\mu dx_\mu\right),$$

which is a nonintegrable *phase* factor. Thus Weyl's theory is *the* theory of electromagnetism in quantum mechanics, provided one changes the idea of a scale factor into a phase factor, with the insertion of a $-i$ ²².

Fock and London in 1927 did not explicitly have the concept of *gauge transformation* (i.e., phase transformation.) That concept was for the first time formulated in a decisive paper^{7,24} of Weyl's in 1929. I now quote from a related paper²⁵ of his, also published in 1929:

By this new situation, which introduces an atomic radius into the field equations themselves – but not until this step – my principle of gauge-invariance, with which I had hoped to relate gravitation and electricity, is robbed of its support. But it is now very agreeable to see that this principle has an equivalent in the quantum-theoretical field equations which is

²² For this history see ¹⁶, p. 525. For an analysis of the physical meaning of a nonintegrable phase factor see ³³. The concept of the "nonintegrable" phase factor occurred to me only in 1967–1968 (see ¹⁶, p. 73). [I was not aware until ~1983 that Weyl had, in 1918, *started* conceptually from the nonintegrable scale factor (3) and proceeded to the differential form $\partial_\mu + \phi_\mu$.] Epistemologically this story is interesting and is representative of the style of Weyl's ideas in physics, in contrast to that of the physicists: Weyl started from the integral approach and proceeded to the differential. Mills and I, physicists, learned the differential approach from Pauli²³ and only much later realized that one could also start from the integral approach.

²³ W. Pauli in Handbuch der Physik, 2. Aufl. 24 part 1 (1933).

²⁴ Weyl 1929, GA III, p. 245.

²⁵ Weyl 1929, GA III, p. 229.

e
s:
a
p
e:

fc
P:
cc
te
W
diver
he ha
F:
E:
(tl
ac
tr:
cc
But ir
ness c
local c
exerte
Th
very r
relativ
place
Twen
our m
not aj
Only:
of non
both v
Be
origin
(3) an

²⁶ W. F
²⁷ C. N
²⁸ C. N
²⁹ p. 73
an adv

exactly like it in formal respects; the laws are invariant under the simultaneous replacement of ψ by $e^{i\lambda}\psi$, ϕ_α by $\phi_\alpha - \frac{\partial\lambda}{\partial x_\alpha}$, where λ is an arbitrary real function of position and time. Also the relation of this property of invariance to the law of conservation of electricity remains exactly as before...the law of conservation of electricity

$$\frac{\partial q_\alpha}{\partial x_\alpha} = 0$$

follows from the material as well as from the electromagnetic equations. The principle of gauge-invariance has the character of general relativity since it contains an arbitrary function λ , and can certainly only be understood in terms of it.

Weyl's emphasis in this passage on the *current density* q_α and its divergenceness as basic to the law of conservation of electricity echoes what he had already said in 1918¹³.

For we shall show that as, according to investigations by Hilbert, Lorentz, Einstein, Klein, and the author, the four laws of the conservation of matter (the energy-momentum tensor) are connected with the invariance of the action quantity (containing four arbitrary functions) with respect to transformations of coordinates, so in the same way the law of the conservation of electricity is connected with the "measure-invariance".

But in 1929 he developed further the idea and expressed it as the divergenceness of the current density q_α . In the language of physics today, this is called *local current conservation*. It was elaborated on by Pauli^(23, p. 111, and 26) and exerted a great influence on my own thinking, as we shall discuss later.

The quote above from Weyl's 1929 paper also contains something which is very revealing, namely, his strong association of gauge invariance with general relativity. That was, of course, natural since the idea had originated in the first place with Weyl's attempt in 1918 to unify electromagnetism with gravity. Twenty years later, when Mills and I^{27, 28} worked on non-Abelian gauge fields, our motivation was completely divorced from general relativity and we did not appreciate that gauge fields and general relativity are somehow related. Only in the late 1960's did I recognize the structural similarity mathematically of non-Abelian gauge fields with general relativity and understand that they both were *connections* mathematically²⁹.

Before proceeding further let us ask what has happened to Einstein's original objection after quantum mechanics inserted an $-i$ into the scale factor (3) and made it into a phase factor (8)? Apparently no one had, after 1929,

²⁶ W. Pauli, Rev. Mod. Phys. 13, 203 (1941).

²⁷ C. N. Yang and R. Mills, Phys. Rev. 95, 631 (1954); reprinted in ¹⁶, p. 171.

²⁸ C. N. Yang and R. L. Mills, Phys. Rev. 96, 191 (1954).

²⁹ p. 73 of ¹⁶. The detachment of gauge field concepts from general relativity, in retrospect, was an advantage because it allowed us to concentrate on one problem at a time.



Fig. 4. Aharonov-Bohm experiment. S is a solenoid with magnetic flux perpendicular to plane of paper.

relooked at Einstein's objection until I did in 1983³⁰. The result is interesting and deserves perhaps to be a footnote in the history of science: Let us take Einstein's Gedankenexperiment in Fig. 3. When the two clocks come back, because of the insertion of the factor $-i$, they would not have different scales but different phases. That would not influence their rates of time-keeping. Therefore, Einstein's original objection disappears. But you can ask a further question: Can one measure their phase difference? Well, to measure a phase difference one must do an interference experiment. Nobody knows how to do an interference experiment with big objects like clocks. However, one can do interference experiments with electrons. So let us change Einstein's Gedankenexperiment to one of bringing electrons back along two different paths and ask: Can one measure the phase difference? The answer is yes. That was in fact a most important development in 1959 and 1960 when Aharonov and Bohm³¹ realized – completely independently of Weyl – that electromagnetism has some meaning which was not understood before. They proposed precisely this experiment, with a slight variation of inserting a solenoid which carries a magnetic flux inside. Changing the flux one can manipulate the phase difference between the two paths. The experiment was done by Chambers³² in 1960. For an analysis of its significance and its relationship with the identification of the nonintegrable phase factor (8) as *the essence* of electromagnetism, see reference³³. For discussions of other experiments related to the Aharonov-Bohm effect see³⁰.

The third period during which Weyl wrote about gauge theory covers the years from 1930 to his death in 1955. One finds Weyl referring to gauge theory in many of his papers throughout this period. For example, he referred to it in 1931 in a paper called "Geometrie und Physik." He referred to it again in 1944 in a paper called "How far can one get with a linear field theory of gravitation in flat space-time?". If additional evidence is needed to demonstrate Weyl's deep attachment to the gauge idea, one can look at the postscript (to the 1918 gauge theory paper¹¹) which he wrote, for inclusion in his *Selecta*, six months before

³⁰ Chen Ning Yang in *Proc. Int. Sym. Foundations of Quantum Mechanics* (Tokyo, 1983), Edited by S. Kamefuchi, H. Ezawa, Y. Murayama, M. Namiki, S. Nomura, Y. Ohnuki, and T. Yajima, p. 5 (Phys. Soc. of Japan, 1984).

³¹ Y. Aharonov and D. Bohm 115, 485 (1959). See also W. Ehrenberg and R. E. Siday, *Proc. Phys. Soc. London B62*, 8 (1949).

³² R. G. Chambers, *Phys. Rev. Lett.* 5, 3 (1960). In this connection see the discussion in ³⁰ of other experiments that are related to the Aharonov-Bohm effect.

³³ Tai Tsun Wu and Chen Ning Yang, *Phys. Rev. D12*, 3845 (1975).

his death in 1955. In this postscript one finds, explicitly stated once more, the reason for his devotion to the idea³⁴:

Das stärkste Argument für meine Theorie schien dies zu sein, daß die Eichinvarianz dem Prinzip von der Erhaltung der elektrischen Ladung so entspricht wie die Koordinaten-Invarianz dem Erhaltungssatz von Energie-Impuls.

III.

Weyl's reason, it turns out, was also one of the melodies of gauge theory that had very much appealed to me when as a graduate student I studied field theory by reading Pauli's articles^{23,26}. I made a number of unsuccessful attempts to generalize gauge theory beyond electromagnetism³⁶, leading finally in 1954 to a collaboration with Mills in which we developed a non-Abelian gauge theory^{27,28}. In ²⁷ we stated our motivation as follows:

The conservation of isotopic spin points to the existence of a fundamental invariance law similar to the conservation of electric charge. In the latter case, the electric charge serves as a source of electromagnetic field; an important concept in this case is gauge invariance which is closely connected with (1) the equation of motion of the electro-magnetic field, (2) the existence of a current density, and (3) the possible interactions between a charged field and the electromagnetic field. We have tried to generalize this concept of gauge invariance to apply to isotopic spin conservation. It turns out that a very natural generalization is possible.

Item (2) is the melody referred to above. The other two melodies, (1) and (3), were what had become pressing in the early 1950's when so many new particles had been discovered and physicists had to understand how they interacted with each other.

I had met Weyl in 1949 when I went to the Institute for Advanced Study in Princeton as a young "member". I saw him from time to time in the next years, 1949-1955. He was very approachable, but I don't remember having discussed physics or mathematics with him at any time. His continued interest in the idea of gauge fields was not known among the physicists. Neither Oppenheimer nor Pauli ever mentioned it. I suspect they also did not tell Weyl of the 1954 papers of Mills' and mine. Had they done that, or had Weyl somehow come across our paper, I imagine he would have been pleased and excited, for we had put

³⁴ Weyl, 1955; GA II, p. 42. In sharp contrast, Pauli had a negative attitude about the idea of gauge fields in the last years of his life ³⁵. In 1956 he wrote a series of Supplementary Notes for the English translation of his 1921 article *Relativitätstheorie*. The note on the "Theory of Weyl" was no more positive than his 1921 original article in German.

³⁵ See A. Pais's forthcoming book on the history of elementary particle physics and an article by C. P. Enz on Pauli (reported at the June 1985 Symposium on the Foundation of Physics in Joensuu, Finland.) See also ³⁶.

³⁶ See [16], p. 19.