International Journal of Theoretical Physics, Vol. 32, No. 8, 1993

Reflections on the Evolution of Physical Theories

Henri Bacry[1]

Received February 27, 1993

> Most physicists are very satisfied with this situation. They argue that if the results agree with observation, that is all that one requires. [...] For these reasons I find the present Q.E.D. quite unsatisfactory.
>
> P. A. M. Dirac

> We conclude that a convergent theory cannot be formulated consistently within the framework of present spacetime concepts.
>
> J. Schwinger

> I evidence that Landau insisted upon the inconsistency of the local Lagrangian approach to the relativistic theory up to the last day of his active life. That was the reason why he evaded the title *Quantum Field Theory* in his course since it referred to fields in space-time.
>
> M. Marinov

1. INTRODUCTION

My aim in this article is not an investigation of the historical or the philosophical aspects of physical theories, even if some of my remarks may have some interest for historians or philosophers. I am a physicist and, as a physicist, I have the deep conviction that trying to discover in the historical evolution of physical theories missed opportunities, contradictions, or paradoxes is not only a cultural exercise permitting us to evaluate the value of science, but could also provide an important help in the search for better theories. The period I am interested in starts at the birth of statistical mechanics (Boltzmann, Gibbs), particle physics (the discovery of the elec-

[1]Centre de Physique Théorique, Case 907, F 13288 Marseille Cedex 9, France.

902/32/8-1

tron by Thomson), special relativity (Einstein), and the appearance of h (Planck constant). At first sight, three questions arise: (1) the ability of each theory to interpret known phenomena, (2) the consistency of each theory, and (3) their mutual compatibility. However, the inconsistencies it is possible to discover in physical theories oblige us to ask a fourth question, namely, what are the axioms which must be modified if we want to remove these inconsistencies? My aim is not to propose specific changes. It is to show in which part of the theory the changes would have an important influence.

Just a word to conclude this introduction. The kind of reflection I am reporting in the present article is not new. The reader is referred in particular to my previous works (e.g., Bacry, 1981, 1988a, b, 1989, 1991, 1992).

2. MISSED OPPORTUNITIES

Some years ago, Dyson (1972) wrote an article entitled "Missed opportunities." He wanted to give some examples where scientists were, in principle, able to make interesting remarks but did not. Dyson mentions his personal experience about the sequence 3, 8, 10, 14, 15, 21, 24, 28,..., which appears in two distinct chapters of mathematics, namely number theory and Lie algebras. Having worked in these two domains, Dyson knew the existence of these sequences, but he never realized that they were identical. The link between them was discovered by MacDonald. It is possible to imagine a change in the history of mathematics if Dyson had been puzzled by the coincidence. Building this kind of *historical fiction* is a (difficult) exercise which has no historical or philosophical interest. My feeling is that it is of scientific interest because it is a way to ask questions which could be fruitful.

Let me give an example taken from the classical kinetic theory of gases. A perfect gas obeys the equation $PV = NkT$. It is well known that the entropy is given by the expression $S = (PV/T)\,\mathrm{Log}(V^{5/2}P^{3/2}) + \mathrm{const.}$ This can be put in the form

$$S = \frac{PV}{T}\,\mathrm{Log}(aV^{5/2}P^{3/2})$$

where a is an unknown constant. Although this constant plays no role in thermodynamics, such a formula is unsatisfactory for two reasons. First, it is known in dimensional analysis that every physical quantity has "dimensions." In order to satisfy this condition, we must require that the argument of the logarithm, namely $aV^{5/2}P^{3/2}$, is dimensionless. Second, we know that entropy is an extensive variable. This requires that the quantity $aV^{5/2}P^{3/2}$ is intensive. Therefore, the constant a can only depend on N, on m (the mass of a molecule), and some fundamental constants. Condition 2 implies that a is of the form $bN^{-5/2}$, where b is a function of m.[2] It is known that this result suffices to solve the famous Gibbs paradox and to introduce the concept of indistinguishability of identical molecules. Condition 1 is satisfied if there exists a universal constant having the dimension of an action.[3]

This example is interesting in that it tells us that the classical kinetic theory of gases (a combination of Newtonian mechanics and the hypothesis of pointlike molecules) is not consistent, a statement which seems to be ignored in textbooks. We must underline that the proof of it is obtained with the aid of general principles which were well known at the end of the 19th century. However, the only contradiction which is mentioned about this period is the incompatibility between Newtonian mechanics and Maxwell's theory of electrodynamics.

I will give in the present article examples of missed opportunities of a special kind. Very often, a physicist has the intuition of some principle but, for obscure reasons, does not state it explicitly. Similarly, a physicist has the idea of a thought experiment but does not explore it completely. The reader will see that the word *implicit* appears many times in this article.

3. MAXWELL THEORY AND THE ARROW OF TIME

It is often said that Maxwell's theory is invariant under time reversal. Let us examine this point. At the end of the 19th century, Newtonian mechanics was considered as a *consistent* and *accepted* theory. This does not mean that it was acceptable. In particular, Newtonian gravitation theory is unable to explain the exact value of the precession of Mercury's perihelion. In a sense, we can say that idea which solves this difficulty is the rejection of action at a distance, an idea already accepted in Maxwell's theory of electromagnetic interactions.

It is interesting to underline the following fact. In 1919, physicists recognized that the Mercury difficulty was a failure of Newtonian theory which can be cured by a theory of where the gravitational field propagates.

[2] m is an intensive variable.
[3] The velocity of light is useless.

At that time, two kinds of forces were known and understood:

1. The gravitational forces (source: energy-momentum)
2. The electromagnetic forces (source: moving electric charges[4])

In both cases, we need *retarded fields*. This condition has important consequences for the second principle of thermodynamics. Suppose, for instance, that a hot body is placed in a cold container. We know that the hot body radiates more than the container itself. That explains the temperature variations of the body and the container. If we reverse time, the retarded fields are transformed into advanced fields and the hot body absorbs more radiation *because it is hotter* than the container. This is also true for each atom. It follows that it is wrong to say that, microscopically, time is reversible. Why do modern textbooks say it? To my knowledge, nobody has tried to incorporate retarded fields in the kinetic theory of gases.

Let us mention another incompatibility. It is impossible to consider a perfect gas of point molecules obeying special relativity. In such a gas, the molecules do not interact, except with the walls of the container. In the approximation where the container is a rigid cube, the momentum component which is orthogonal to the wall changes sign. So does the corresponding speed component. The independence of the x, y, z components provides us with Maxwell-type distribution laws, laws of the exponential form $\exp[-\beta(v_x^2+v_y^2+v_z^2)]$ and $\exp[-\beta(p_x^2+p_y^2+p_z^2)]$. These expressions are incompatible if $\mathbf{p} \neq m\mathbf{v}$.

4. EINSTEIN UNIFYING PRINCIPLE (1905)

This unifying principle was not explicitly stated by Einstein. However, it was used by him in the elaboration of special relativity. It can be set out in the following way.

A continuous set of experiments must be interpreted by a single theory.

I will try to convince the reader that the fact that this obvious principle was never stated had important consequences for the history of physics. Suppose that the set we have in mind is parametrized by a real number t taking all the values between zero and one. Saying that we have as many theories as we have values of t is equivalent to saying that we have a single theory which involves t as an observable. It would be stupid to build two distinct theories, one for $0 \leqslant t \leqslant 1/2$ and one for $1/2 < t \leqslant 1$, because the continuity assumption requires that the theories coincide for the limit value $t = 1/2$.

[4]At that time, a model was needed to explain magnetic matter.

According to Einstein himself, the guiding remark which led him to special relativity was the following one (see, e.g., Holton, 1973). Electromagnetic induction has two interpretations, depending on which of the two elements, namely the magnet and the electric wire, is moving. The phenomenon itself only depends on the relative motion. Therefore, there exists a single explanation of that, a *relative theory*. Clearly, Einstein had in mind a *continuous* set of experiments where the speeds of the magnet and the electric wire are arbitrary. That is the reason I consider that the unifying principle is due to Einstein. Clearly, I have no explanation why he did not state it explicitly. It is natural to recall here the two interpretations Einstein was referring to.

1. The magnet is fixed, the wire is moving. The explanation makes use of the Lorentz force

$$\mathbf{F} = e\mathbf{v} \times \mathbf{B}$$

2. The wire is fixed, the magnet is moving. The explanation is based on a flux argument

$$\text{e.m.f.} = -\frac{\partial \Phi}{\partial t}$$

5. BOHR'S VIOLATION OF THE EINSTEIN UNIFYING PRINCIPLE (1927)

We must underline that the notion of continuity is essential in our statement. Nothing obliges us to have a unique theory for explaining a discrete number of distinct experiments. When Einstein suggested in 1905 to consider a light beam as composed of quanta, the problem was to understand why light would be both a continuous phenomenon and a discrete phenomenon. There were physicists who were trying to build a new theory which would be a compromise between the two aspects (Einstein, de Broglie), physicists who were in favor of a corpuscular interpretation (Heisenberg), and those who thought that the wave concept was essential (Schrödinger). They were all, more or less unconsciously, in favor of a unique theory of light. Bohr adopted a different attitude. His attempt to officialize the particle–wave duality in his *complementary principle* would be acceptable if there were actually two kinds of experiments with light, ones where light behaves as a wave, and ones where it behaves as particles. Unfortunately, this is not the case. There exists at least one continuous set of experiments for which we cannot choose clearly between the wave and the photon interpretations. This set was *implicitly* described by Feynman.

6. THE FEYNMAN CONTINUOUS SET OF EXPERIMENTS

Among the set of experiments I have in mind, only the two extreme cases are described by Feynman in his famous lectures. It concerns the electron two-slit experiment, when a light source is used in order to localize the electrons. Suppose that the light source has a fixed power W, the continuous parameter being the frequency ν. We suppose that W is not very large and we want to explain why we have still an interference pattern, although this pattern is not very neat.

The two extreme cases investigated by Feynman are the following ones. Suppose first that ν is very large, which implies that the photon rate is very small. The rareness of photons implies that the probability for an electron to be scattered by one of them is small and therefore there are electrons which contribute to the interference pattern. This is a *corpuscular* interpretation, since photons are involved. Now, suppose that ν is very small, that is, the wavelength λ very large. As we know from *wave* theory, the image of an electron is not a sharp point, it is a spot with a width proportional to λ. Therefore we cannot say each time which slit the electron went through. Only some of them contribute to the interference pattern.

7. THE DE BROGLIE PRINCIPLE (1923) AND ITS VIOLATION BY BORN (1926)

When de Broglie proposed to associate with each particle of momentum $\mathbf{p}$ a wavelength $\lambda = h/p$ he wanted to put the electron and the photon on the same footing. However, he did not state explicitly the following symmetry principle:

All particles must be put on the same footing.

It is well known that Schrödinger found a way to associate a wave with the electron. At the time, the Schrödinger equation was considered to play for the electron the role played by the Maxwell equations for the photon. Surprisingly, nobody protested when Born violated de Broglie's principle in proposing his probabilistic interpretation of the wave function. Such an interpretation was not acceptable for the photon! The simplest way to check it is to show that neither the electric field $\mathbf{E}$ nor the potential vector $\mathbf{A}$ can be used to build a probability density in space. Dimensional analysis arguments again show that, even if we try to call the fundamental constants h and c.

This difficulty was confirmed by Newton and Wigner (1949) when they proved that the photon has no localized states. However, Wigner (1939) proved that the de Broglie principle was right, since he stated implicitly that:

The Hilbert space of one-particle states is always an irreducible representation space of the Poincaré group.

Is de Broglie's symmetry principle right?

At this stage, I have three important remarks to make: (1) the difficulties we have met in the Einstein and de Broglie principles are related to the problem of *localization* (Feynman experiment in the first case, non-localized states in the second case), (2) the large success of the Poincaré group representations concerns the energy-momentum space and the angular momentum, not the position, (3) Newton and Wigner failed in their attempt to discover worldlines in irreducible representations of the Poincaré group.[5]

The energy-momentum is also involved in classical special relativity and the success of energy-momentum conservation is not questionable. It is natural to examine the problem of localization in special relativity, that is, to examine the role of the Minkowski space at the particle scale.

8. MINKOWSKI SPACE AND ENERGY-MOMENTUM SPACE

In almost all textbooks, the Minkowski space is implicitly identified with the energy-momentum space. They are however distinct. Their main properties are given in Table I.

9. EINSTEIN AND SPACE

Everybody knows how Einstein modified Newtonian mechanics in order to subject it to Lorentz invariance. The paradox I want to explore came when he decided to explain how to measure distances with the aid of rulers and *light* signals. When one knows that the wavelength is not an invariant in special relativity, one must be surprised that Einstein used *light* signals instead of *electromagnetic* signals. From the special relativity point of view, a light signal must be as useful as a radio signal or even a signal with a light-year wavelength. However, stated in this way, it is obvious that the measurement of a distance of 1 cm cannot be performed with the aid of a signal of 1 km wavelength. Why not?

[5]In particular, if an electron is localized for an observer, it is not for another one. This result contradicts the well-known position of Bohr, according to which any measurement gives a result which can be described classically. A wordline is a classical concept...

Table I

Minkowski space-time	Energy-momentum space
1. From the Lorentz group point of view	
Invariant: $x^\mu x_\mu$	Invariant: $p^\mu p_\mu$
An infinite number of light cones	One light cone
Lorentz transformation formulas are the same	
2. From the Poincaré group point of view: $\mathscr{P} = \mathscr{T} \cdot \mathscr{L}$	
Homogeneous space (one orbit)	Many orbits and state (timelike, spacelike, etc)
Affine space	Vector space
$\mathscr{P}/\mathscr{L}$	Dual of $\mathscr{T}$ or dual of the Lie algebra of $\mathscr{T}$
Representation theory ignores it	Energy-momentum is involved in the little group approach
Newton–Wigner failed to find a covariant position operator	
3. Dimensional analysis ($c = 1$, $\hbar = 1$)	
Length	Length^{-1}
4. Meaning in QED	
No interpretation (dummy indices)	Energy-momentum (conserved quantities)
Locality of the Lagrangian density	Space for Feynman diagrams
	Cutoffs are involved

In order to know where the signal is coming from, we must use directed signals. If we want to have a relativistic measurement, the signal must be a Maxwell signal. Such a signal has a width and must propagate at the speed of light. Is it possible? Such a question was asked by Poincaré (1892) and his answer was *no*.

To propagate at speed c, a beam must have a width of about ten times its wavelength!

As a consequence, if we really want to use *light* signals (say $\lambda = 0.6 \pm 0.1\ \mu$m) to measure distance between "points," we know that the Doppler effect forbids us to perform large boosts; a simple calculation shows that *the ratio v/c of the boost must be less than 0.324.* It follows that we are not authorized to build a relativistic space-time with such signals.

To be more precise, let us adopt the point of view of Poincaré. For him a reasonable signal can be described approximately with the aid of a circular function of the type

$$\Psi(t, \rho, y, z) = NJ_0\left(\frac{2\pi\rho}{\lambda}\left(1 - \frac{V^2}{c^2}\right)^{1/2}\right)\exp\left[2\pi i\left(\frac{ct}{\lambda} - \frac{Vz}{c\lambda}\right)\right]$$

It is a cylindrical beam of radius ρ_0 and group velocity V[6]. Let us give some numerical values:

V	ρ_0
c	∞
$0.9999999c$	552λ
$0.999999c$	174λ
$0.99999c$	55λ
$0.9999c$	17λ
$0.999c$	5.5λ
$0.99c$	1.7λ
$0.87c$	0.55λ
$0.5c$	0.285λ
0	0.2468λ

One verifies that the notion of a light *ray* propagating at a speed close to c is not realistic. A corollary is that Minkowski spacetime can be only considered as a macroscopic object. The paradox is that Minkowski spacetime is used in QED to require the local character of the Lagrangian density, as if Minkowski spacetime was made of Euclidean points! Another paradoxical fact is that QED does not really interpret the x_μ as spacetime coordinates.

10. WHAT TO KEEP FROM THE POINCARÉ GROUP

When a theory has successes and an inconsistency, the question arises of how to perform a separation between the part of the theory which is responsible for the success and the part which contains contradictions. Let us try to do that. The construction of the unitary irreducible representations of the Poincaré group is probably the most successful part of special relativity (in particle physics, not in gravitation theory, for which it is a disaster). It permits us to classify all kinds of particles and implies the main conservation laws (energy-momentum and angular momentum). We insist on the fact that it is in agreement with the de Broglie symmetry principle.

[6]The expression *group velocity* is unknown in Poincaré.

The commutation relations of the generators $M_{\mu\nu}$ and P_μ are well known. They are

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(g_{\nu\rho} M_{\mu\sigma} - g_{\mu\rho} M_{\nu\sigma} - g_{\nu\sigma} M_{\mu\rho} + g_{\mu\sigma} M_{\nu\rho})$$
$$[M_{\mu\nu}, P_\rho] = i(g_{\nu\rho} P_\mu - g_{\mu\rho} P_\sigma)$$
$$[P_\mu, P_\nu] = 0$$

Let us consider a unitary irreducible representation describing a spinless particle of mass m; such a representation satisfies the conditions

$$P_\mu P^\mu = m^2, \qquad W_\mu = \tfrac{1}{2} \varepsilon_{\mu\nu\rho\sigma} M^{\rho\sigma} P^\nu = 0$$

(the vanishing of the Pauli–Lubansky vector W_μ guarantees that the spin of the particle is zero).

If we define the position operator as

$$X^i = \frac{1}{2}\left(\frac{1}{P_0} M^{0i} + M^{0i} \frac{1}{P_0}\right) \tag{1}$$

we obtain the following commutation relations:

$$[X^i, X^j] = 0, \qquad [X^i, P^j] = i\delta^{ij}$$

If we add the relation $[P^i, P^j] = 0$ to these, we obtain the well-known set of Heisenberg commutation relations for a spinless massive particle.

The problem is to define a position operator for an arbitrary particle (spinning or massless). If we want to preserve the validity of the de Broglie symmetry principle, the most natural choice consists in adopting (1) as the general definition.[7] There exist many arguments in favor of that. The only "difficulty" is that the relation $[X^i, X^j] = 0$ is no longer valid, a result which is hard to accept, but which has many advantages, especially the one of preserving the Einstein unifying principle.

11. A STRANGE AFFAIR IN THE THIRTIES

It is clear that the relation $[X^i, X^j] \neq 0$ is a departure from the usual commutation relations obeyed by the so-called Schrödinger position operator. When spin was discovered for the electron, physicists decided to replace the Schrödinger equation by a couple of them. In doing so, they admitted implicitly the fact that the X^i were commuting with the spin

[7]The reader will find other arguments in favor of this in my previous works and in the works of Grigore (1989) and Duval *et al.* (1990).

operator. We note that the relation $[X^i, X^j]=0$ is necessary to have a wave function $\psi(x, y, z)$, but the condition $[X^i, X^j]\neq 0$ does not contradict Dirac's axioms of quantum mechanics. The paradox is that Schrödinger, the passionate defender of wave theory, was the first to propose a position operator with noncommuting components! The reason he did it was to cure some diseases of the Dirac equation—considered as a relativistic generalization of the Schrödinger equation. It is strange that almost all quantum textbooks describe a part of his proposal, the famous *zitterbewegung*, related to the difference of the two Schrödinger position operators. It is also strange that the Schrödinger position operator for the Dirac equation is exactly the one defined by equation (1), when the representation of the Poincaré group is the one associated with the Dirac equation (a reducible representation).

12. FROM HEISENBERG TO CONNES

For Heisenberg, the commutation relations for a spinless particle with mass related noncommuting coordinates of phase space. He never said that he invented a *noncommutative space*. If he had said such a thing, physicists would have tried to define statistical mechanics with the aid of this noncommutative phase space. However, in inventing noncommutative geometry, Connes (1986, 1990) is referring explicitly to Heisenberg.

With the above analysis, I am tempted to say that the noncommutativity of the X^i for spinning particles only makes the noncommutative phase space a little bit more noncommutative and, in that sense, it is not a big revolution. However, it must have consequences about our space structure and my belief is that we have to understand that in order to guess which theory will replace QED.

13. CONCLUSION

Let me come back to the successes of the Poincaré group in particle physics. This is a group with ten generators. The translation generators are responsible for the energy-momentum conservation laws, the rotation generators of the conservation of angular momentum, and the boost generators of the conservation of *initial position*. If positions are slightly different from the ones described by Minkowski space, it means that we have to change slightly the notion of boosts. If we remember that boosts were questionable in Minkowski space (see Section 9), we are not surprised. We are naturally led to a deformation of the Poincaré group which would preserve translations and rotations [such a deformation has been proposed by Lukierski *et al.* (n.d.)]. By duality, small changes at short distances must correspond to small changes in large momenta. The fact that cutoffs for momenta are involved in QED is perhaps related to a noncommutative structure for our space. With such a structure, making the size of an electron go to zero is meaningless and consequently the difficulty of an electron with infinite energy also becomes meaningless. A noncommutative space is probably a way to solve the difficulties mentioned in the epigraphs to this paper.

REFERENCES

Bacry, H. (1981). *Journal of Physics A*, **14**, L73.

Bacry, H. (1988a). *Annales de l'Institut Henri Poincaré*, **49**, 245.

Bacry, H. (1988b). *Localizability and Space in Quantum Physics*, Springer-Verlag, Berlin.

Bacry, H. (1989). *Nuclear Physics B (Proceedings Supplement)*, **6**, 222.

Bacry, H. (1991). in *Lecture Notes in Physics*, Vol. 382, *Springer*-Verlag, Berlin.

Bacry, H. (1991). *Annales de l'Institut Henri* Poincaré, **56**, 345.

Connes, A. (1986). *Publications Mathematiques* IHES, **62**, 257.

Connes, A. (1990). *Géométrie non commutative*, InterEditions.

Duval, C., Elhadad, J., and Tuynman, G. M. (1990). Pukanszky's condition and symplectic induction, preprint CPT-90/P.2398, Centre de physique théorique, Marseille, France.

Dyson, F. (1972). *Bulletin of the American Mathematical Society*, **78**, 635.

Feynman, R. P., Leighton, R. B., and Sands, M. (1965). *The Feynman Lectures in Physics*, Vol. 3, Chapter 1, Addison-Wesley, Reading, Massachusetts.

Grigore, D. R. (1989). *Journal of Mathematical Physics*, **30**, 2646.

Holton, G. (1973). *Thematic Origins of Scientific Thought*, Harvard University Press, Cambridge, Massachusetts.

Lukierski, J., Nowicki, A., and Ruegg, H. (n.d.) New quantum Poincaré algebra and κ-deformed field theory, *Physics Letters B*, in press.

Newton, T. D., and Wigner, E. P. (1949). *Review of Modern Physics*, **21**, 400.

Poincaré, H. (1892). *Théorie mathématique de la lumière*, Georges Carré, Paris.

Schrödinger, E. (1984). *Collected Papers*, Vol. 3, Austrian Academy of Science, Vienna.

Wigner, E. P. (1939). *Annals of Mathematics*, **40**, 149.