Book (1)

Mathematical Physical Development of Quantum Theory

Giuseppe A. Nisticò
Dipartimento di Matematica e Informatica
Università della Calabria

Chapter 0

Conceptual Foundations of Quantum Theory

It is shown that the conceptual basis of classical physical theories is inconsistent with particular physical phenomena. More precisely, the classical idea that at any time each specimen of a physical system has a value for each of its magnitudes as an objective property of the specimen leads to predictions contradicted by particular real experimental results. Therefore, classical theories become empirically invalid. To attain a physical theory empirically consistent it is necessary to identify a conceptual basis that renounces to such an idea. Such a consistent conceptual basis is identified within the von Neumann approach to quantum mechanics. The basic concepts are the concept of observable and the concept of expectation value. The very physical meaning of these concepts compels the set of observables and the set of expectation values to satisfy precise mathematical conditions.

1. An experimental paradox

Let us consider the following experimental setup, consisting of three apparatuses (Fig.1). The first apparatus is a source that emits identical particles one at once, under identical

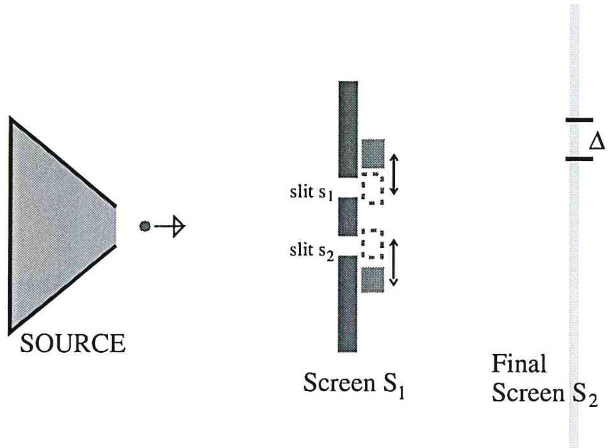

Figure 1. Experimental setup

conditions of the source, in such a way that each particle travels towards the second apparatus.

The second apparatus is an impenetrable screen S_1 that supports two parallel slits, we call slit s_1 and slit s_2, through which the particles can pass. One of the slits can be chosen to be closed, so that the crossing of S_1 is allowed only through the other slit.

A screen S_2, at a certain distance behind S_1, ascertains

i) whether each particle emitted from the source passes beyond S_1, simply by revealing if it hits S_2 itself, and also
ii) whether it hits or does not hit a fixed region Δ of S_2.
This is the third apparatus of the setup.

By making use of this experimental setup three different experiments E(1), E(2), E(1,2) can be performed.

Experiment E(1). Experiment E(1) consists in performing a large number N of particle emissions with s_2 closed and s_1 open. A particle emitted from the source at time t_0, will reach S_1 at time $t_1 > t_0$. The condition of the particle between the source and S_1 is that of free particle. The particle goes beyond S_1 and hence hits S_2 only if its position at time t_1 is in front the aperture of slit s_1.

The condition of the particle between S_1 and S_2 is again that of free particle. Let n_1 and $n_1(\Delta)$ be the number of the particles that pass beyond S_1 and hence hit S_2, and the number of the particles that are revealed to hit Δ after N runs, respectively. Of course $n_1 \geq n_1(\Delta)$. The number n_1 coincides with the number of particles localized in front the aperture of slit s_1 when S_1 is reached.

Experiment E(2). Experiment E(2) is almost identical to E(1); the only difference is that slit s_1 is closed at time t_1, while slit s_2 is left (7) open. By n_2 and $n_2(\Delta)$ we denote the number of the particles that pass beyond S_1 through s_2 to hit S_2, and the number of particles that are revealed to hit Δ after N runs, respectively.

Experiment E(1,2). In experiment E(1,2) both slit s_1 and slit s_2 remain always open and a number M$\gg$2N of emissions are performed. By n and $n(\Delta)$ we denote the number of the particles that pass beyond S_1 and hence hit S_2, and the number of particles that are revealed to hit Δ after the M runs, respectively. By $n(1)$ and $n(2)$ we denote the number of particles that hit S_2 passing through s_1 and s_2 respectively; of course $n(1,2) = n(1) + n(2)$. By $n(\Delta \mid 1)$ and $n(\Delta \mid 2)$ we denote the number of particles that hit Δ passing through s_1 or s_2, respectively; of course $n(\Delta \mid 1) + n(\Delta \mid 2) = n(\Delta)$.

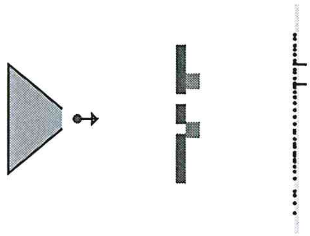

Figure 2. Experiments E(1) with a setting of the source such that $n_1(\Delta \mid 1) > 0$

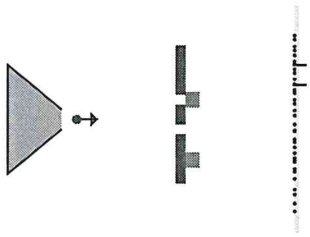

Figure 3. With the same setting of the source of experiment E(1), experiment E(2) yields $n_2(\Delta \mid 2) > 0$

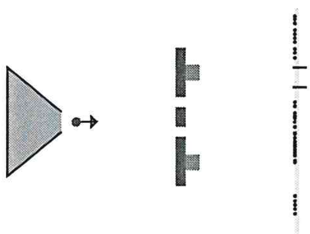

Figure 4. With the same setting of the source of experiments E(1) and E(2), in experiment E(1,2) $n(\Delta) = 0$ has been found.

By *setting* of the setup we mean
i) the concrete choice of the source and of all its controllable conditions,

ii) the concrete choice of the screens S_1 and S_2,
iii) the concrete choice of the distances source-S_1, S_1-S_2 and of the region Δ.

Remark 1.1. The physical condition of the particles that pass through slit s_1 in experiment E(1) is the same of those that pass through slit s_1 in E(1,2): in both cases they are free particles outcoming from s_1 with the same past history (they come from the source under identical conditions and go towards s_1). Therefore, there is no reason for which the behavior of the particles of the first set should be different from the behavior of the particles of the second set. In other words, the fact that s_2 is closed in no way affects the condition of the particles that pass through s_1 in E(1) with respect the particles that passes through s_1 with s_2 open in E(1,2).

Let us suppose that for a given setting of the apparatus experiment $E(1)$ is performed and $n_1(\Delta) > 0$ is found as a result. This implies in particular that the number n_1 of the particles that pass beyond S_1 is positive, that is to say n_1 particles had a position in front the aperture of slit s_1 at time t_1.

Let experiment $E(1,2)$ be performed with the same setting. Since according to remark 1.1 the behavior of the particles between the source and S_1 mus be indistinguishable from that of the particles in the same region in E(1), if M is large enough then a number $n(1) > n_1$ of particles will have a position in front the aperture of s_1 at time t_1 and hence they go in the region between S_1 and S_2. Here the behavior of these $n(1)$ particles is indistinguishable from that of the n_1 particles between S_1 and S_2 in E(1); since $n_1(\Delta) > 0$ in E(1), we must infer that $n(\Delta \mid 1) > 0$ and hence $n(\Delta) > 0$. Thus, the following statement should hold.

(St.1) *Whatever the setting of the apparatus, if experiment E(1) yields $n_1(\Delta) > 0$, then a number M exists such that an execution of experiment E(1,2) with the same setting yields $n(\Delta) > 0$.*

This statement holds also if we replace E(1) with E(2) and $n_1(\Delta)$ with $n_2(\Delta)$, because the argument leading to (St.1) holds also with these changes.

In fact, particular settings of the source have been realized (Figs. 2-4) such that in actual performances of the experiments E(1), E(2), E(1,2) turned out to be found that

$$n_1(\Delta) > 0, \; n_2(\Delta) > 0, \quad \text{but } n(\Delta) = 0 \text{ for every } M > 0.$$

This experimental result contradicts statement (St.1). So the question arises: given that the condition of the particles that pass beyond S_1 is the same in E(1) and E(1,2), why does a difference arise in their behavior? The answer is that something must be wrong in our reasoning that leads to $(\mathcal{S}_0)$

2. What's wrong

What is wrong arises clearly if we analyze another experimental procedure characterized by analogous paradoxical fatures. This new experimental procedure concerns with a particular kind of physical system, a particle P that after a given time from its emission by a suitable source S divide into two particles P_L and P_R that travel in opposite directions, say left h.s. and right h.s., respectively.

Once P_L and P_R are separated they can simultaneously undergo separated measurement procedures of a physical magnitude $\mathcal{L}$ of P_L and of another physical magnitude $\mathcal{R}$ of P_R. Every magnitude considered in this experiment has two possibile values, 0 or 1; so the value $v(\mathcal{L})$ obtained by measuring $\mathcal{L}$ on P_L, as well as the value $v(\mathcal{R})$ measured on P_R, can be 0 or 1.

The question can be asked whether for a particular pair $(\mathcal{L}, \mathcal{R})$ of such magnitudes a setting of the experimental apparatus exists such that the values $v(\mathcal{L})$ and $v(\mathcal{R})$ obtained by the respective measurements are constrained to each other, for instance in such a way that the correlation

$$v(\mathcal{L}) = 1 \quad \text{implies} \quad v(\mathcal{R}) = 1$$

holds between the actually measured values, with that given setting.

In fact, a setting has been concretely found such that the following correlations hold for three particular pairs $(\mathcal{L}_1, \mathcal{R}_1)$, $(\mathcal{L}_2, \mathcal{R}_1)$, $(\mathcal{L}_2, \mathcal{R}_2)$ of magnitudes:

(1) if $\mathcal{L}_1$ and $\mathcal{R}_1$ are measured then $v(\mathcal{L}_1) = 1$ implies $v(\mathcal{R}_1) = 1$;

(2) if $\mathcal{L}_2$ and $\mathcal{R}_1$ are measured then $v(\mathcal{R}_1) = 1$ implies $v(\mathcal{L}_2) = 1$;

(3) if $\mathcal{L}_2$ and $\mathcal{R}_2$ are measured then $v(\mathcal{L}_2) = 1$ implies $v(\mathcal{R}_2) = 1$.

These correlations are indisputable facts in that setting.

Now, in the case that the pair $(\mathcal{L}_1, \mathcal{R}_2)$ is measured with the same setting for which (1-3) hold, and $v(\mathcal{L}_1) = 1$ is obtained, what can we state about $v(\mathcal{R}_2)$? The answer is straightforward in the case that all magnitudes $\mathcal{L}_1, \mathcal{L}_2, \mathcal{R}_1, \mathcal{R}_2$ are measured together at time t_0 on the same specimen, i.e. the pair P_L, P_R, of the physical system in the setting for which (1-3) hold: the statement

$$v(\mathcal{L}_1) = 1 \textit{ implies } v(\mathcal{R}_2) = 1,$$

is an unavoidable implication of (1-3).

Let us now suppose that $\mathcal{L}_2, \mathcal{R}_1$ cannot be measured together with $\mathcal{L}_1, \mathcal{R}_2$ that are actually performed. We can argue according the following steps.

Step 1. Let the outcome of an actually performed measurement of $\mathcal{L}_1$ at time t_0 be $v(\mathcal{L}_1) = 1$. Though correlation (1) implying $v(\mathcal{R}_1) = 1$ refers to actually measured outcomes, such assignment can be extended to the (not measured) value of $\mathcal{R}_1$ according to the following statement.

Value Assignment Assumption. The outcome of a measurement of a magnitude at any time t on a specimen of the physical system is nothing else but the revealation of the value of the magnitude at that time.

Step 2. By making use of this assumption and of (2), we can state that for this specimen if $\mathcal{L}_2$ were measured instead of $\mathcal{L}_1$, then the outcome would be $v(\mathcal{L}_2) = 1$ with certainty. Indeed, the outcome of $\mathcal{R}_1$ if measured instead of $\mathcal{R}_2$ at t_0 would be 1 with certainty, and the outcome of an eventual measurement of $\mathcal{L}_2$ at t_0 cannot depend on the fact that $\mathcal{R}_1$ is measured or not, because the measurements occur at the same time t_0 in spatially separated places; therefore we can state that

$$v(\mathcal{L}_2) = 1 \textit{ if } \mathcal{L}_2 \textit{ were measured instead of } \mathcal{L}_1.$$

Step 3. From this statement, by making use once again of the Value Assignment Assumption and of (3), we infer that if $\mathcal{R}_2$ were measured , then the outcome is $v(\mathcal{R}_2) = 1$ with certainty.

Thus, the conclusion of our conceptually compelling reasoning is the following statement.

($\mathcal{S}$) *If $\mathcal{L}_1$ and $\mathcal{R}_2$ are measured at time t_0, then $v(\mathcal{L}_1) = 1$ implies $v(\mathcal{R}_2) = 1$.*

In fact, analogously to the experiment of section 1, a setting for this new experimental procedure has been concretely found for which (1-3) hold, but in some actually performed measurements of $(\mathcal{L}_1, \mathcal{R}_2)$ the outcome turned to be

$$(\mathcal{ER}) \qquad v(\mathcal{L}_1) = 1 \quad \text{and} \quad v(\mathcal{R}_2) = 0,$$

in contradiction with statement ($\mathcal{S}$). Therefore, there must be a mistake in the reasoning that leads to ($\mathcal{S}$) from (1-3). Since (1-3) are indisputable facts, the unique possible conclusion is that the Value Assignment Assumption cannot be maintained. This assumption is an unavoidable consequence of the following apparently obvious principle.

($\mathcal{BP}$) *Every specimen of the physical system has a precise value of each of its magnitudes at any time.*

This principle establishes the existence of these values as objective properties of the specimen, independently of its experimental revealation. Also if ($\mathcal{BP}$) is not explicitly formulated, the development of all classical theories makes continuously use of it, and their formulation cannot be carried out if its validity is removed.

The experimental result ($\mathcal{ER}$) implies that the principle ($\mathcal{BP}$) does not always hold in Physics. So, classical theories are not able to explain these physical phenomena, because they assume a principle that turns out to be empirically invalid.

3. Conceptual re-foundation for empirically consistent physical theories

Since all classical physical theories have been developed complying with ($\mathcal{BP}$), and therefore are empirically unadequate, the physical theories must be re-founded. A physical theory of a specific system is a specific formal system whose role is to establish the relationships among the phenomena of that physical system. These physical phenomena include the occurences of the outcomes of measurements performed on specimens of the physical system, of course. In our re-foundation of the physical theories, however, the existence of the value of every magnitude at any time cannot be assumed as physical phenomenon. Coherently, the classical concept of physical magnitude is replaced by the following concept of *observable.*

Observables. By observable we mean any physical magnitude measurable, by means of a concrete apparatus, on individual specimens of the physical system under investigation, which has real numbers as outcomes; the coordinates of the position of a particle are examples of observables. An observable is not assigned an objective value if such a value is not the outcome of an actually performed measurement.

The set of all observables is denoted by $\mathcal{O}$. The set of all possible outcomes of the measurement of an observable is denoted by $\tilde{\sigma}(\mathcal{A})$, and it is called *physical spectrum of the observable* $\mathcal{A}$.

Remark 1. Coherently with the new basic concept, the performance of a measurement of an observable $\mathcal{R}$ on a specimen of the physical system assigns that specimen the objective value of $\mathcal{R}$ that is necessarily consistent with the ascertained physical circumstances. Therefore, if any value assignment to a set of oservables is not consistent with the ascertained conditions, then the measurement of these observables together on the same specimen is not possible – otherwise the assignment would be consistent.

In the experimental procedure of section 2 the assignment of values to $\mathcal{L}_2$ and $\mathcal{R}_1$ is proved to be inconsistent with the ascertained conditions (1-3) and the ascertained outcomes of the measurements of $\mathcal{L}_1$ and $\mathcal{R}_2$. Thus, $\mathcal{L}_1$, $\mathcal{R}_1$, $\mathcal{L}_2$ and $\mathcal{R}_2$ cannot be measured together. In the experiment of section 1 we have shown that in experiment E(1,2) the outcomes of the measurement of the position at the time of the final impact on S_2 makes inconsistent the assignment of position at time t_1; therefore, according to the implication *assignment not consistent* $\Rightarrow$ *measurement not possible* above, we have to conclude that in E(1,2) the measurement of the position at the time of the final impact cannot be performed together to the measurement of the position at time t_1 on the same specimen of the physical system.

Hence, a theory with the present conceptual basis allows for the existence of observables that cannot be measured together

3.1. Functional Principle for observables and co-measurability

The very meaning of the concepts of observable allows to establish mathematical conditions to be satisfied by the set of observables, which are expressed by the following principle.

Functional Principle. Let $\mathcal{R}$ be an observable. In correspondence with each function

$f : \tilde{\sigma}_{\mathcal{R}} \to \mathbb{R}$, there is another observable, denoted by $f(\mathcal{R})$, whose outcomes can be obtained by applying the function f to the outcomes of $\mathcal{R}$; we notice that, if f is injective, $\mathcal{R}$ and $f(\mathcal{R})$ measure the same magnitude, by using two different scales.

Now we show how co-measurability can be characterized. Given two observables $\mathcal{A}$ and $\mathcal{B}$, let us suppose that a third observable $\mathcal{C}$ and two functions f and g exist such that $\mathcal{A} = f(\mathcal{C})$ and $\mathcal{B} = g(\mathcal{C})$ according to the functional principle. If a measurement of $\mathcal{C}$ is performed with outcome c, then the values of a and b can be obtained, according to the functional principle, simply as $a = f(c)$ and $b = g(c)$. Thus, the following statement holds.

(Co.1) If $\mathcal{A} = f(\mathcal{C})$ and $\mathcal{B} = g(\mathcal{C})$ then $\mathcal{A}$ and $\mathcal{B}$ are measurable together.

Conversely, let $\mathcal{A}$ and $\mathcal{B}$ be two observables that are measurable together on a same specimen of the physical system. In the case that their physical spectra are finite, say

$$\tilde{\sigma}(\mathcal{A}) = \{\lambda_1, \lambda_2, ..., \lambda_N\} \quad \text{and} \quad \tilde{\sigma}(\mathcal{B}) = \{\mu_1, \mu_2, ..., \mu_M\},$$

the existence can be proved of a third observable $\mathcal{C}$ and of two functions f, g such that $\mathcal{A} = f(\mathcal{C})$ and $\mathcal{B} = g(\mathcal{C})$ according to the following procedure:

(1) Fix a bijection:

$$\varphi : \tilde{\sigma}(\mathcal{A}) \times \tilde{\sigma}(\mathcal{B}) \to \{\gamma_1, \gamma_2, ..., \gamma_{N\times M}\} = \Gamma \subseteq \mathbb{R}, \quad (\lambda_j, \mu_k) \to \varphi(\lambda_j, \mu_k) = \gamma_n.$$

A unique pair $\tilde{j} : \{1, 2, ..., N \times M\} \to \{1, 2, ..., N\}$, $\tilde{k} : \{1, 2, ..., N \times M\} \to \{1, 2, ..., M\}$ of mappings exists such that $\varphi(\lambda_{\tilde{j}(n)}, \mu_{\tilde{k}(n)}) = \gamma_n$.

(2) Now, let us define $\mathcal{C}$ as the observable that can be measured by measuring together $\mathcal{A}$ and $\mathcal{B}$, being λ_j, μ_k the respective outcomes, and assigning $\mathcal{C}$ the outcome $\varphi(\lambda_j, \mu_k) = \gamma_n$. Of course, $\tilde{j}(n) = j$ and $\tilde{k}(n) = k$.
Accordingly, the spectrum of $\mathcal{C}$ is $\tilde{\sigma}(\mathcal{C}) = \Gamma$.

(3) Finally, define: $f : \tilde{\sigma}(\mathcal{C}) \to \tilde{\sigma}(\mathcal{A})$, $f(\gamma_n) = \lambda_{\tilde{j}(n)}$ and $g : \tilde{\sigma}(\mathcal{C}) \to \tilde{\sigma}(\mathcal{B})$, $g(\gamma_n) = \lambda_{\tilde{k}(n)}$.
If $\mathcal{A}$ and $\mathcal{B}$ are measured together with outcomes λ_{j0} and μ_{j0} respectively, then the outcome of $\mathcal{C}$ is $c = \varphi(\lambda_{j0}, \mu_{k0}) \equiv \gamma_{n_0}$, so that $f(c) = \lambda_{j0}$ and $g(c) = \mu_{j0}$.
Thus, $\mathcal{A} = f(\mathcal{C})$, $\mathcal{B} = g(\mathcal{C})$.

We can conclude that

(Co.2) in the case that $\tilde{\sigma}(\mathcal{A})$ and $\tilde{\sigma}(\mathcal{B})$ are finite, if $\mathcal{A}$ and $\mathcal{B}$ are measurable together then $\mathcal{C}, f, g$ exists such that $\mathcal{A} = f(\mathcal{C})$ and $\mathcal{B} = g(\mathcal{C})$.

This statement could be immediately extended to the case of general spectra $\tilde{\sigma}(\mathcal{A})$, $\tilde{\sigma}(\mathcal{B})$ if a bijection $\varphi : \tilde{\sigma}(\mathcal{A}) \times \tilde{\sigma}(\mathcal{B}) \to \Gamma \subseteq \mathbb{R}$ existed, where $\tilde{\sigma}(\mathcal{A})$ and $\tilde{\sigma}(\mathcal{B})$ are not necessarily finite closed subsets of $\mathbb{R}$. In general, continuous such bijections do not exist: in other words, it is not possible to transform a plane in a line bijectively by means of a continuous function. However, this kind of transformation is possible if the continuity condition is removed, keeping measurability. Since the continuity condition is not required by our argument, it can be conclude that:

(Co) *Two observables $\mathcal{A}$ and $\mathcal{B}$ are measurable together if and only if a third observable $\mathcal{C}$ and two functions f, g exist such that $\mathcal{A} = f(\mathcal{C})$ and $\mathcal{B} = g(\mathcal{C})$.*

3.2. Expectation Value of observables

By *expectation value* we mean a function $\mathsf{E}_v : \mathcal{O}_{E_v} \to \mathbb{R}$ assigning a numerical value $\mathsf{E}_v(\mathcal{R})$ to every observable $\mathcal{R}$ in a suitable subset $\mathcal{O}_{E_v}$ of observables, which is to be interpreted as the expectation value of the measurement of the observable $\mathcal{R}$, in the sense of statististical probability theory. Hence, any expectation value E_v refers to a *population* $\mathcal{N}$ of specimens of the physical systems such that if the measurement of an observable $\mathcal{R}$ is performed on each specimen of a concrete sample $\mathcal{N}_1 \subseteq \mathcal{N}$ of N_1 specimens, anyway extracted from $\mathcal{N}$, with actual outcomes outcomes $a_1, a_2, \ldots, a_{N_1}$ and mean value $\langle\mathcal{R}\rangle_{\mathcal{N}_1} = \sum_{i=1}^{N_1} a_i/N_1$, then $\langle\mathcal{R}\rangle_{\mathcal{N}_1}$ converges to the expectation value $\mathsf{E}_v(\mathcal{R})$ as $N_1 \to \infty$.

The population corresponding to an expectation value E_v is to be identified with the process, natural or laboratorial or of any nature, that selects the specimens of the physical system that belongs to the population. Two different such selection processes are physically equivalent if they yields the same expectation values. In the experiments of sections 1 and 2 such a selection is operated by the source in a fixed setting.

This peculiar possibility is not in contradiction with the existence of the expectation values of two different observables, also in the case that they cannot be measured together. Indeed, the two expectation values can be determined as the limits of the mean values of the two observables measured on two *different* sequences of samples for the two observables such that every sample of the first sequence has empty set theoretic intersection with any sample of the other sequence.

The theory based on these basic concepts does not make predictions about a single measurement, but rather it establishes the probability of each possible outcome; then, in general, its prediction can be verified by repeating the selection and the measurement many times, and then constructing the statistical distribution of the results.

4. Quantum Theory

The *quantum theory* of a specific physical system is the physical theory of that physical system developed coherently with the basic concepts of observable and expectation value defined in this chapter. The formulation of complete specific quantum theories requires the mathematical formalism of particlar classes of operators in Hilbert space.