A proposal for the solution of the paradox of the double-slit experiment

Gerrit Coddens

A proposal for the solution of the paradox of the double-slit experiment

Gerrit Coddens[a]

Laboratoire des Solides Irradiés,
Institut Polytechnique de Paris, UMR 7642,CNRS-CEA-Ecole Polytechnique,
28, Route de Saclay, F-91128-Palaiseau CEDEX, France

26th October 2021

Abstract. We propose a solution for the apparent paradox of the double-slit experiment within the framework of our reconstruction of quantum mechanics (QM), based on the geometrical meaning of spinors. We argue that the double-slit experiment can be understood much better by considering it as an experiment whereby the particles yield information about the set-up rather than an experiment whereby the set-up yields information about the behaviour of the particles. The probabilities of QM are conditional, whereby the conditions are defined by the macroscopic measuring device. Consequently, they are not uniquely defined by the local interaction probabilities in the point of the interaction. They have to be further fine-tuned in order to fit in seamlessly within the macroscopic probability distribution, by complying to its boundary conditions. When a particle interacts incoherently with the set-up the answer to the question through which slit it has moved is experimentally decidable. When it interacts coherently the answer to that question is experimentally undecidable. We provide a rigorous mathematical proof of the expression $\psi_3 = \psi_1 + \psi_2$ for the wave function ψ_3 of the double-slit experiment, whereby ψ_1 and ψ_2 are the wave functions of the two related single-slit experiments. This proof is algebraically perfectly logical and exact, but geometrically flawed and meaningless for wave functions. The reason for this weird-sounding distinction is that the wave functions are representations of symmetry groups and that these groups are curved manifolds instead of vector spaces. The identity $\psi_3 = \psi_1 + \psi_2$ must therefore be replaced in the interference region by the expression $\psi'_1 + \psi'_2$, for which a geometrically correct meaning can be constructed in terms of sets (while this is not possible for $\psi_1 + \psi_2$). This expression has the same numerical value as $\psi_1 + \psi_2$, such that $\psi'_1 + \psi'_2 = \psi_1 + \psi_2$, but with $\psi'_1 = e^{\imath\pi/4}(\psi_1 + \psi_2)/\sqrt{2} \neq \psi_1$ and $\psi'_2 = e^{-\imath\pi/4}(\psi_1 + \psi_2)/\sqrt{2} \neq \psi_2$. Here ψ'_1 and ψ'_2 are the correct (but experimentally unknowable) contributions from the slits to the total wave function $\psi_3 = \psi'_1 + \psi'_2$. We have then $p = |\psi'_1 + \psi'_2|^2 = |\psi'_1|^2 + |\psi'_2|^2 = p'_1 + p'_2$ such that the apparent paradox that quantum mechanics would not follow the traditional rules of probability calculus for mutually exclusive events disappears.

PACS. 02.20.-a, 03.65.Ta, 03.65.Ca Group theory, Quantum Mechanics

1 Introduction

1.1 Context

The present paper is part of a series of three papers which together propose a reconstruction of QM that permits to understand what its algebra means. The reconstruction proper is outlined in [1]. The goal of the two other papers is to illustrate how the insight gained by the reconstruction permits to make sense of two notoriously difficult examples of experiments that seem to defy any attempt of understanding, viz. the Stern-Gerlach experiment [2] and the double-slit experiment [3,4].[1]

[a] Ancien membre du Laboratoire des Solides Irradiés (retraité).

[1] I cannot insist enough that the contents of [1] are a prerequisite for reading the present paper as the very first lines of this Introduction clearly indicate. In fact, as mentioned, the present paper is based on [1] in its very foundations, and it is unfortunately just impossible to copy into the present paper the large parts from [1] that would be needed to make it self-contained. As one may imagine, some of the contents of [1] are truly beyond guessing and there is no royal short-cut to it such that if the reader skips reading [1] he will almost certainly become upset by some "unsettling" statements in the present article, but wrongly so.

1.2 Methodology

The double-slit experiment has been qualified by Feynman [5,6] as the only mystery of quantum mechanics (QM), although this is admittedly an exaggeration (see e.g. [7]). The mystery resides in an apparent paradox between the QM result and what we expect on the basis of our daily-life, common-sense intuition. What we want to explain in this article is that this apparent paradox is a probability paradox. By this we mean that the paradox does not reside in some special novel physical property of e.g. an electron that could act both as a particle and as a wave, but in the fact that we use two different definitions of probability in the intuitive approach and in the calculations. It is the difference between these two definitions which leads to the paradox, because the two definitions are just incompatible.

As this claim may raise some eyebrows, let us point out how we can become sure about it before we even take off. Solving a paradox in mathematics or physics is not so much an exercise of mathematics or physics as rather an exercise of pure logic. In general we have developed two reasonings which contradict one another and we must try to figure out which one is wrong (if not both contain errors). What one must do then to solve this problem is meticulously reconstructing completely the two logical chains of reasoning which have led to the two conflicting results. We must thus clearly delineate all definition domains, investigate all axioms, theorems and assumptions, and then check everything in the logical chains step by step for weak points to see if we can detect some logical error or a questionable link. At each step we must dissect the flow of input and output of the mathematical statements with a bistoury on their truth.

Let me give an example. We all know the famous twin paradox of special relativity. In its simplest form we can consider the twins as in uniform relative motion in free space. Both twins can consider with equal rights that they are at rest and that the other twin is moving. They therefore conclude with equal rights that the other twin is ageing slower. This is all we need to formulate the paradox. This means that the whole paradox relies uniquely on the definition of space-time $\mathbb{R}^4$ and on the definition of the Lorentz boosts. These are our "axiomatics", i.e. the full set of assumptions from which the paradox is derived.

The logical solution of the twin paradox must therefore reside within the use of this set of assumptions, and not be sought for somewhere else. Anything else must be considered as out-of-context and missing the point. Unfortunately some people have thought that it would be more striking to wrap up the paradox in a movie scenario wherein one twin makes a round trip to Alpha Centauri while the other one stays home on Earth. This approach introduces accelerations during the round trip which are not part of the initial formulation of the problem. The vast majority of the solutions proposed for the paradox have focused their attention on these accelerations (See e.g. the detailed reference list in [8]). This is a solution based on physical intuition, not on logical rigour. Logical rigour tells that these accelerations are not part of the initial problem, such that they cannot intervene in its solution. A solution pointing out the rôle played by the accelerations can be physically correct in every single minute detail, but it is then an ersatz solution for another, modified problem, not a solution of the pristine problem. The solution which zooms in onto the accelerations must therefore be considered as a diversion, based on a tacit change of the axiomatics (which are now straying into general relativity). With some mean-spiritedness, the modified problem could be qualified as a straw man. This analysis proves that there must exist a better, more general and fundamental solution for the paradox, which is uniquely based on the definition of the Lorentz transformations and space-time [9].[2]

What we have opted for in the analysis described here is not the inside perspective of a reasoning on the physics but rather an outside perspective of reasoning on types of reasoning and their frameworks, by pointing out that one reasoning relies uniquely on special relativity while the other one blends in also elements of general relativity. It is this kind of "meta" analysis, whereby we change the perspective that we need to solve a paradox. The reasoning followed in this article might therefore from time to time look highly unconventional due to its out-of-the-box character inherent to a "meta" analysis, but it is not speculative, philosophical or epistemological. Everything in it is only pure logic and pure mathematics. We take the liberty to call (by analogy but in a slight *abus de language*) a "meta" analysis which adopts changes of stance as evoked here "metamathematical". Adopting the term "metaphysical" was out of

[2] The true solution consists in introducing a third neutral observer. This observer will have his own, third reference frame and make measurements in this frame to figure out who is right and who is wrong *by defining trips*, as described in [9]. What we learn from this is that all depends on the relative velocity of the Lorentz frame of this neutral observer with respect to the frames of the two twins. He will conclude that the twin who is in a faster relative motion with respect to him will age less. When the two relative velocities are equal, the two twins will age at the same rate. The conclusion is that it is the choice of the third frame associated with the neutral observer and *the way he defines the trips in his frame* which break the symmetry between the twins, such that in general he finds himself in a biased position wherein he willy-nilly can only violate the required neutrality. Everything is determined by the scheme of time intervals, distances, and simultaneity that prevail within the reference frame we have endowed the neutral observer with. When the frame of the neutral observer coincides with the frame of one of the two twins then his story will be identical to the story of that twin. Hence each twin is right in his own scheme of time intervals, distances and simultaneity. It is just that the two twins are imparted with different schemes of time intervals, distances and simultaneity by nature. There is no contradiction because time is not absolute, a point using accelerations to start and end the journeys in the same frame tends to mask.

the question because it means something entirely different. From now on we will drop the quotes when we use the word metamathematical.

In following the methodology of investigating the logical elements that are present within the two conflicting chains of thought, we can develop for the double-slit experiment the following reasoning. As mentioned in Subsection 1.1, we have proposed a reconstruction of QM in [1,2,10]. This reconstruction yields exactly the same algebraic results and the same agreement of these results with the experimental data as the traditional approach, such that it cannot be attacked for departing in some other aspects from the traditional approach. (The differences do not occur in the algebra, but in the geometrical meaning of the algebra, which has been replaced by an "interpretation" in the traditional approach, while there is nothing to interpret, because the meaning of the algebra is already provided by the mathematics). First we have determined the geometrical meaning of spinors [10,11]. This is pure mathematics and it can be easily checked whether it is right or wrong. Using the geometrical meaning of spinors we have been able to derive the Dirac equation from scratch [1,10]. The derivation has been done with the rigour of a mathematical proof. It is also entirely classical and, surprisingly enough, does not require stunning assumptions to account for some conjectured "quantum magic". From the Dirac equation we can derive the Pauli and Schrödinger equations.

Trying to avoid at any price the introduction of flamboyant, novel physical assumptions is the absolute top priority of our approach, because such assumptions instill doubt about the validity of the whole endeavour of making sense of QM. Some attempts to make sense of QM are introducing indeed alienating assumptions, e.g. about parallel worlds, advanced waves, and so on. Such daring assumptions call into question the whole credibility of the "interpretations" based on them. How can we single out one of them among several alternatives, when they are all equally baffling and inaccessible to direct testing? And what if what eludes our understanding were to consist only of a hard nut to be cracked within the mathematics? It would then be doubly insane and detrimental to sidestep solving the purely mathematical problem by postulating it away through the introduction of "new physics" or "quantum magic".

In our approach the meaning of QM is no longer a matter of "interpretation", because the algebra ceases to be a blackbox and its meaning is just provided in a completely natural way by the geometrical meaning of spinors (or group representations in general). This does not leave any elbow room for speculations about jaw-dropping novel physics or add-ons in the form of "interpretations" that raise philosophical issues. An algebraic formalism should never be a subject of epistemology. Its meaning should be provided by the mathematics itself in the form of an isomorphism with a corresponding geometry. E.g. in algebraic geometry $x^2 + y^2 = R^2$ is the equation of a circle and the meaning of this algebra is not open to guesstimate alternative "interpretation". It is immune to parallel science fiction story-telling. In our derivation of the Dirac equation in [1] the isomorphism is provided by the geometrical meaning of spinors, which is beyond the reach of any polemics by physicists, because it is already firmly established long before we even start to consider applying the algebra to physics. The debates are closed.

Within the context of our approach we can formulate the Dirac equation or the Schrödinger equation for the double-slit experiment and solve these equations. This is pure mathematics. The whole theoretical treatment of the double-slit experiment becomes this way just a matter of well-understood, pure mathematics, yielding a result that agrees with the experimental observations (see Section 3). This implies that we have a logical chain of flawless, well-understood mathematics that has been checked step by step all the way from the geometry (of the rotation group in $\mathbb{R}^3$ and the homogeneous Lorentz group in $\mathbb{R}^4$) to the wave function for the double-slit experiment and the probability distribution derived from it using the Born rule, which we also justified in [1]. This actually means that we have understood the double-slit experiment. The reader can appreciate that the chain of reasoning described here hinges crucially on the truth of the pretended contents of [1], which is why we insisted in Footnote 1 on the absolute necessity of reading it.

We have now reached the point where the reader may start scratching his head, because we have all but the feeling that we understand the double-slit experiment, as Feynman's discussion clearly reveals. That is the paradox, *viz.* that classical intuition tells us that we have not understood it at all. Moreover, this classical intuition is also built on an apparently sound mathematical reasoning, but one that (1) is more cursory and far less formal, (2) as it tries to be synthetic rather than algebraic by basing itself on a number of common-sense intuitions about probability calculus and (3) no longer yields a prediction in agreement with experimental data, because it leads to a classical superposition of probability densities rather than an interference pattern. It is thus more than likely that there is something wrong with our common-sense intuitions and it is then important to figure out where this error in our intuition could be hidden. The metamathematical analysis shows that the paradox takes place between two mathematical formulations. The solution of the paradox must therefore be sought for in the mathematics and nowhere else. It is thus not a physical paradox, although we may all be strongly convinced it is, because this is what our gut feelings are telling us.

But we must adopt a cold composure and hold on to the logic. For sure, all this does not yet mean that we would have solved the paradox, but we can have the rock-solid certainty that we should not search for it in terms of novel, counterintuitive quantum principles such as the particle-wave duality or new rules to deal with probabilities that upset the traditional ones. We can exclude the need of introducing mind-boggling new physical principles just as we could exclude the need of introducing arguments from general relativity to solve the twin paradox. The solution must reside within the logic and/or in the probability calculus. And in searching for weak points, some introspection suggests that

it is the rather casual approach we often take to probability calculus that could be built on shaky foundations. In the casual approach it is easier to overlook important subtle details in the way we must define and treat probabilities (see Subsection 1.4).

The argument we will develop is highly arborescent due to a large amount of connected issues that must be explored in depth. In order to help the reader not to get lost in the complexity of the ramifications, the main structure of the reasoning is therefore summarized in Section 10. A very beautiful result of the present work is the proof obtained in Subsection 3.6. In this work we will use the notation $\mathrm{F}(A,B)$ for the set of functions whose domain is the set A and who take values in the set B, while $\mathrm{L}(A,B)$ will be the set of linear mappings whose domain is the set A and who take values in the set B.

1.3 Limitations and clarifications

1.3.1 Use of spinors

As mentioned, we have used the meaning of spinors in the rotation group SO(3) and the homogeneous Lorentz group SO(1,3) to propose a reconstruction of QM in [1,10]. When we speak about wave functions in this paper, we therefore very often refer to spinors. The wave functions of the Dirac and Pauli equations are spinors, and those of the Schrödinger equation can be considered as simplified spinors.

In fact, we can simplify the Pauli equation for a fermion whose spin axis remains all the time parallel to the z-axis by replacing the SU(2) spinor-valued function $\psi \in \mathrm{F}(\mathbb{R}^4,\mathbb{C}^2)$:

$$\psi(\mathbf{r},t) = e^{\imath(Et-\mathbf{p}\cdot\mathbf{r})/\hbar}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{1}$$

by the short-hand $\eta \in \mathrm{F}(\mathbb{R}^4,\mathbb{C})$:

$$\eta(\mathbf{r},t) = e^{\imath(Et-\mathbf{p}\cdot\mathbf{r})/\hbar}. \tag{2}$$

to obtain a scalar Schrödinger equation. This new wave function η obeying the scalar Schrödinger equation is then just a spinor in disguise, showing that the wave functions in the three equations are in reality all spinors, although this fact becomes concealed in the scalar Schrödinger equation by the use of the short-hand notation. With some provisos to be spelled out below, this implies that everything covered by these three equations is written in terms of spinors, whose geometrical interpretation is automatically carried over into QM. A large part of QM is thus written in terms of spinors.[3]

One can describe the double-slit experiment with each of the three equations, but the paradox and the physics will remain basically the same. We therefore present the analysis in this paper within the framework of the Schrödinger equation, with its scalar wave functions, because the Schrödinger equation is known to a much broader public than the other two equations. But the whole paper has been based on [1], where the wave functions used are spinors, rather than scalar wave functions. We will therefore very often refer to results that apply for spinors.

1.3.2 Generality

Based on this statement one could argue that our approach is not general. We have e.g. excluded bosons from the considerations. But that criticism would not be very fair. The derivation of the three equations based on [1] is the only one that is rigorous and deductive. Before our derivation of the Dirac equation from scratch in [1] the status of all three equations was that they had been obtained by successive "educated" guesses. First de Broglie guessed a wave function to be associated with a particle. Building further on this, Schrödinger guessed his equation, and finally Dirac guessed the further relativistic generalization. Also the Maxwell equations have not been derived deductively from mathematical principles but inductively from experimental observation. Such flimsy foundations compromise the immunity of the theory against interpretation. As only spinor-based equations can boast the status of being deductively proved with mathematical rigour and are therefore immune to parallel interpretation, it is only normal that we base our approach on these equations. It is the understanding of spinors which is the game changer in our approach. But of course this entails that when we use in this work the Schrödinger equation for bosons, we cannot consider that it has been rigorously proved, or that its wave function would be a spinor.

It is in this respect that I have stated in [1] that the fact that the Schrödinger equation works for so many different types of particles must be considered as a fluke, because the generality may well turn out to be true *de facto*, but it

[3] To the author's knowledge the procedure of taking the "quantization axis" along the z-axis in QM is systematical, such that there is nothing unusual in what we propose in Eq. 1.

has not been explained. This façade generality is used in textbooks to present a general treatment of the double-slit experiment based on the Schrödinger equation, which is analogous to the treatment for photons. All this apparent generality is uncritically taken for granted, which it is not. The resulting false illusion of obvious generality naturally gives rise to the *a priori* unjustified conviction that the explanation for the double-slit experiment should be "equally general". In reality, the traditional approach suffers from the very same lack of generality as is being taken issue with here. It is therefore sanctimonious to adopt a moralizing attitude and criticize the present work for its alleged lack of generality. This is the pot calling the kettle black and reminds the parable of the mote and the beam. The traditional dogma cannot pretend to be the ultimate touchstone of truth because the authority of a guessed theory cannot compete with that of a theory that has been mathematically proved, a point that should silence all gatekeepers. But rather than in indulging in polemics about fairness, we will provide a scientific rebuttal of the objection, by discussing the point where it matters the most, i.e. when we generalize to all possible types of wave functions (i.e. group representations) the fact that summing or making linear combinations of spinors is *a priori* meaningless in the group representation theory used.

1.3.3 Linearity of the equations and linear combinations of wave functions

This statement is at variance with what one can read in physics textbooks (see e.g. [12], which explicitly claims in its section 1.4 that spinors form a vector space). It may sound heretic, which is one of the reasons why I insisted in Footnote 1 to read [1], where it is developed in Section 2.5 starting from p. 12.[4]

The doctrine that spinors can be summed and would be just vectors in Hilbert space is indeed deeply entrenched, not in the least, perhaps, because there exists even a book written by Dirac, entitled "Spinors in Hilbert space" [16]. But this notion is a historical error based on ignorance (see Appendix, [15] and Footnote 4). When Dirac introduced his equation and spinors with it, he did not know what they were. He therefore assumed that he could treat them as vectors in some abstract vector space of solutions, based on the linearity of his equation. As discussed in the Appendix, this argument of linearity is unsuspectedly deceptive. The consequence of it is a real watershed, viz. that a century later we must engage in an uphill battle to explain to physicists that one cannot sum spinors like vectors as they routinely do.

We invite the reader to get over it. A part of the present article will aim exactly at repairing for the historical error, by showing on a case example that physicists can continue to carry out their algebra on the wave functions as they have done up to now, i.e. by acting as though spinors really are vectors in Hilbert space. However, this applies only to the calculations. When it comes down to trying to make sense of what the calculations mean, one cannot dispense with debugging the historical error. For our true, geometrical understanding of QM, it would be even appropriate to establish similar proofs for several other case examples, using the methods we are introducing here.

The issue that spinors do not form a vector space is conclusively settled in the Appendix and [15]. We will further address it in Subsection 3.4.2 where we will take on the argument of the linearity of the wave equations to give it a second life in terms of sets. In the meantime we invite the reader to ponder over the following: If spinors were really something as trivial as mere vectors in some vector space, why would they then be called spinors? And why would Cartan have written a whole monograph [17] about them?

We can now address the point where the objection about the generality will matter the most, which is that at face value, what we said about summing spinor wave functions is not true for scalar wave functions. It must be obvious that what we have shown in Eq. 1 can only provide a partial answer to this objection, because it does not apply for bosons. But what we learn from the spinor approach is that the phase of the wave function corresponds to an internal clock for some periodic internal dynamics of a particle.[5]

In fact, as explained during the derivation of the Dirac equation in Subsection 4.1 of [1], the spinning motion of the electron in its rest frame is described by a spinor function $\psi : \tau \rightarrow \psi(\mathbf{s}, \tau)$, obtained by replacing the rotation angle φ in the spinor $\psi(\mathbf{s}, \varphi)$ by $\omega_0 \tau$. Here τ is the proper time, and $\mathbf{s}$ is the unit vector along the spin axis.

[4] This point is really beyond discussion and cannot be turned into an issue. The *ad hoc* claim [13] that there would exist definitions of spinors that are different from the one I am using is a fraud. There is only one definition of a spinor in SU(2), *viz.* a 2×1 column matrix which corresponds to the first column of an SU(2) matrix as proved in the Appendix and on p. 7 of [1]. One cannot deny the obvious factual truth that the 2×1 spinor matrix represents the whole 2×2 SU(2) matrix. Overruling the correct definition by brute force to make prevail the "alternative truth" that spinors form a vector space does not improve one's understanding of the algebra. It just obliterates insight and truth by censorship. This has forced me to develop the arguments further in [15] and in the Appendix.

[5] Of course there is no direct proof of this assumption for other particles than electrons, but it is the simplest assumption available (1) given what we have figured out for electrons and (2) given the universality of the phenomenon of interference. It is already hard to find a mechanism that rationally explains the double-slit paradox for electrons, let alone that we would have to invent other mechanisms for other particles. In the absence of information about the detailed mechanism responsible for the phase of other particles, the multiplicity of mechanisms could be immediately attacked invoking Occam's razor.

It does not make sense to add up expressions for the internal dynamics of two different particles. One could e.g. use two time-dependent spinor functions $\psi_j(\mathbf{s}_j, \sqcup) : t \rightarrow \psi_j(\mathbf{s}_j, t)$ to describe the motions of two spinning tops labeled by $j \in \{1, 2\}$. But what could summing two such mathematical expressions for the dynamics of two spinning objects possibly mean! The sum would not represent any meaningful physical reality. It would be like summing the two algebraic expressions $\mathbf{r}_j(t)$ that describe the orbits of two planets labeled by $j \in \{1, 2\}$. Summing or making linear combinations of scalar or any other types of wave functions is therefore as much of a taboo as summing spinor wave functions. We may note that scalar wave functions can also be representations, e.g. of SO(1,1), SO(2) or U(1) and that these groups are also curved manifolds, such that the objection against summing persists, even if the groups are abelian. For other particles like photons, ^{4}He atoms or C_{60} molecules, these internal dynamics (expressed among others by the phase of their appropriate wave functions) must be different, not only from those of electrons, but also mutually. We should therefore determine the nature of these internal dynamics for each type of particle in order to reach a full understanding of QM. This is of course a gigantic task, beyond anything what a single person can achieve.

Despite its elegance and its generality the cherished Hilbert space formalism is reached with the same kind of blissful ignorance as we evoked in revealing the fluke of the generality of the Schrödinger equation. (Such generalizations are heuristic arguments, not theories even after confirmation by experiment, because they generalize the description, not our understanding). It is a rash move away from the physics towards further pseudo-mathematical abstraction. We can qualify it as pseudo-mathematical because in true mathematics a generalization implies that all special cases that are covered by the final construction have been perfectly understood, justified and checked. Already in mathematics, the power of such an increased generality tends to come at a price in terms of the effort required to "crack the code", because the more abstract, general and elegant a formalism becomes, the more difficult it becomes to figure out what is going on behind the scenes. However, in physics, where nothing has been sorted out and where the issue of understanding QM is exactly the problem of "cracking the code" of the formalism, the craze for would-be-scholar mimicking of the mathematical process of abstraction reaches a pinnacle of thwarting any attempt to reach a deeper understanding of QM. It is not by telling people that they should "shut up and calculate" following a set of abstract rules as zombies that the helpful pictures will emerge. What the Hilbert state vectors mean, with a clear picture of the information content of the complex numbers that occur within them is not even spelled out (see Appendix and [15]).

It is just mindless calculus, without any meaning provided, while in our approach a spinor has a clear geometrical meaning. In fact, having a clear picture of the internal dynamics of all the different types of particles would undoubtedly constitute a very instrumental, less abstract basis for understanding that these internal dynamics cannot be summed in a meaningful way and that therefore summing wave functions is (*a priori*) conceptually meaningless. We will come back on the hypothetical issue of lack of generality in Section 6.

The results for spinors show that taking advantage of the fact that summing wave functions in a scalar context apparently makes sense algebraically is not appropriate for discussing the fundamental principles of QM because it is still physically meaningless. Although I am not able to describe the internal dynamics for other particles than electrons, it can be reasonably conjectured that the logic presented is by analogy correct in general and we may therefore jump from one context to another in developing the argument and interchange the words electron and particle rather sloppily. We think the context will clearly show when we are talking about scalar wave functions and when about spinors.

1.4 Possible sources of errors in our intuition about probabilities

1.4.1 "Non-locality"

In general relativity, which uses Riemannian geometry, we must work with local (and instantaneous) Lorentz frames. The definition of the word "local" used here is not the opposite of the term "non-local" as defined in QM. In QM, the term non-locality is used within the context of discussions about entanglement and what Einstein called a "spooky action at the distance". Confronted with the fact that the brand "non-local" has this way been "reserved" for an exclusive specific use, how should we qualify now a frame that is not local in the sense given to the words in general relativity? We cannot use the expression "not local" because playing with words this way is too thorny. We therefore feel ensnared by a game of definitions that deprives us of the language tools we need.

In the present paper we will use the word "non-local" in a sense that is radically different from its standard definition in QM, because we have not found a satisfactory alternative to express what we have in mind, and we will use it in this unique nonstandard sense throughout the paper, such that it can absolutely never be a source of confusion. We have perfectly the right to introduce such a definition for purely domestic use such that it cannot be a pretext for polemics.

What we are perhaps still not sufficiently aware of is the fact that Euclidean geometry permits to define parameters of the set-up which intervene in the expressions for the probabilities, but are defined at a much larger, macroscopic length scale than the microscopic length scale of local physical interactions. We will call such parameters "non-local" and stress their importance in Section 4.

A nice illustration of a "non-local" physical quantity is the angle $\varphi_A - \varphi_B$ which occurs in the expression $\frac{1}{2}\cos^2(\varphi_A - \varphi_B)$ for the probabilities which occur in the experiments of Aspect et al. [18,19] to test the Bell inequalities. The angles φ_A and φ_B correspond to the orientations of two polarizers A and B which can be as far apart as we like [20,21]. They are thus not defined at the microscopic level of interaction points in $\mathbb{R}^3$. The angle $\varphi_A - \varphi_B$ and the probability $\frac{1}{2}\cos^2(\varphi_A - \varphi_B)$ are therefore very obviously defined in a way that we can qualify as "non-local". Changing the distance between the slits in the double-slit experiment modifies the interference pattern, such that it is also such a macroscopic parameter, used to describe macroscopic correlations.

This "non-locality" of parameters based on classical Newtonian notions intervenes also in our definition of a Lorentz frame. This definition takes it for granted that the clocks in the frame are synchronized up to infinite distance, which is just not feasible. Einstein synchronization can only be done at the speed of light. A Lorentz frame is thus also defined "non-locally" and we find this "non-locality" back in the definition of the spinor wave function for an electron. Here the clocks are the virtual spinning electrons which intervene in the definition of the wave function as described in [1, 2]. In the definition of the wave function all these clocks are synchronized up to infinite distance in a rest frame by adopting the same phase angle for all the spinning motions. The result of this convention within the procedure is that the phase velocity of the wave function, the velocity of the signal we would need to perform this synchronization, takes the superluminal value $c^2/v > c$ for a wave function describing electrons in uniform motion with velocity $v < c$. This has not been realized and therefore textbooks introduce wave packets with a group velocity $v_g = \frac{d\omega}{dk} < c$ in order to "repair" for the situation and permit the description of electrons that travel at a speed $v < c$. But there is absolutely no need for doing this because the electrons described by the wave function are already traveling at a uniform speed $v < c$, as explained in [1]. The superluminal phase velocity $c^2/v > c$ does therefore not at all raise a concern that the waves used in QM would violate special relativity. From now on, we will drop the quotes when we use the word non-local.

1.4.2 Contextuality: Bohr's caveat (Conditional probabilities)

Our intuition about probabilities is subject to cognitive bias. Probability calculus is teeming with paradoxes, showing how prone we are to make errors in using it. We fail to conceive that QM probabilities are conditional in the sense that their definition depends on the details of the experimental set-up. These details can be geometrical but also physical. This was stressed by Bohr, who warned us that in QM the instrumental set-up was part of the physics. We will dub this warning "Bohr's caveat". We may have found this caveat bizarre, mysterious or very hard-going. We may have wondered why on Earth we should take it seriously and why it had to be true. But the conditions we refer to in the expression "conditional probabilities" are defined by the experimental set-up, such that Bohr was absolutely right about this very crucial point. The conditions *are* the experimental set-up because the particles are interacting with it at the microscopic level.

These conditions are represented by the boundary conditions when we solve the Schrödinger or Dirac equations because the boundaries we use include boundaries of the set-up. It is known from the Dirichlet problem that the boundary conditions of a differential equation intervene in its solutions and can have a drastic influence on them. Such considerations about the boundary conditions completely justify what Bohr has said about the crucial rôle played by the experimental set-up in QM. Therefore the act of combining the conditional probabilities stemming from two different experimental set-ups should immediately raise a red flag. In fact, conditional probabilities derived from different set-ups should *a priori* not be used together, as this kind of simultaneous use can lead to logical errors. An example of ignoring Bohr's rather counter-intuitive caveat would be to take it for self-evident that we can combine casually the four conditional probabilities $\frac{1}{2}\cos^2(\varphi_{A_j} - \varphi_{B_k})$ defined by four different experimental set-ups, *viz.* the four combinations $(\varphi_{A_j}, \varphi_{B_k})$ of the settings for the polarizers A and B, within the Bell inequality used by Aspect et al. [18,19] in his photon correlation experiments (see [22,23] and the many references therein).

1.4.3 Further unsuspected aspects of the conditional probabilities that can be accounted for by the boundary conditions

As will be discussed in Sections 2 and 8, the probabilities that intervene in the double-slit experiment are not only running contrary to our preconceived notions in being *conditional* and *non-local*, but they are also *undecided* (in a domestic use of the word to be defined). It goes without saying that by overlooking all these counterintuitive ingredients in a casual approach one can brew a truly potent potion of paradoxical quantum magic.

After the derivation of the Dirac equation from scratch in [1] we asked on p.35 where the "quantum magic" could come from. We considered the possibility that the magic could be introduced by the way we use the QM equations. The way we use these equations is that we introduce boundary conditions for them, such that the magic could be smuggled in by the addition of these boundary conditions, because that is all one adds to the formalism. For the example of the double-slit experiment this will indeed be the case. As we will see, the boundary conditions can affect the conditional probabilities in very unexpected ways. This is an observation which points to these boundary conditions

as the culprit of the difference between what we calculate and what we expect on the basis of less sophisticated common-sense intuition. It confirms the idea that the paradox is a probability paradox. However, boundary conditions do not constitute an exhaustive answer to the general question where the quantum magic may come from.

2 Coherent and incoherent nature of the interactions with the set-up

In our discussion of the double-slit experiment we will very heavily rely on the presentations by Feynman [5,6], even though further strange aspects have been pointed out by other authors later on, e.g. in the discussion of the delayed-choice experiment by Wheeler [24,25] and of the quantum eraser experiment [26], which can also be understood based on our discussion. Feynman's lecture in the video [6] is magnificent and an absolute must.

He illustrates the double-slit paradox by comparing tennis balls and electrons. Tennis balls comply with classical intuition, while electrons behave according to the rules of QM. There is however, a small oversimplification in Feynman's discussion. He glosses over a detail, undoubtedly for didactical reasons. When the electron behaves quantum mechanically and only one slit is open, the experiment will give rise to diffraction fringes, which can also not be understood in terms of a classical description in terms of tennis balls. But the hardest part of the mystery is that in the quantum mechanical regime we get a diffraction pattern when only one slit is open, while we get an interference pattern when both slits are open. This means that the two single-slit probabilities even do not add up to an interference pattern when we allow for the quantum nature of the electron in a single-slit experiment. We will therefore compare most of the time the two quantum mechanical situations rather than electrons and tennis balls.

What Feynman describes very accurately is how quantum behaviour corresponds to the idea that the electron does not leave any trace behind in the set-up of its interactions with it. This renders it impossible to reconstruct its history. (We exclude here from our concept of a set-up the detectors that register the electrons at the very end of their history). We cannot tell with what part of the set-up the electron has interacted, because the interaction has been coherent. This corresponds to "wave behaviour". At the very same energy, a particle may also be able to interact incoherently with the set-up and this will then result in classical "particle behaviour". The difference is that when the electron has interacted incoherently we do have the possibility to figure out afterwards its path through the device, because the electron has left evidence behind of its interaction with the atomic constituents of the measuring device.

A nice example of this difference between coherent and incoherent interactions occurs in neutron scattering, as explained by Feynman. His argument runs as follows. In its interaction with the device, the neutron can flip its spin. The conservation of angular momentum implies then that there must be a concomitant change of the spin of a nucleus within an atom of the device. At least in principle the change of the spin of this nucleus could be detected by comparing the situations before and after the passage of the neutron, such that the history of the neutron could be reconstructed. Such an interaction with spin flip corresponds to incoherent neutron scattering. But the neutron can also interact with the atom without flipping its spin. There will be then no trace of the passage of the neutron in the form of a change of spin of a nucleus, and we will never be able to find out the history of the particle from a *post facto* inspection of the measuring device. An interaction without spin flip corresponds to coherent scattering. Note that this discussion only addresses the coherence of the spin interaction. There are other aspects that can intervene in the interaction and in order to have a globally coherent process nothing in the interaction must permit it to leave a mark of the passage of the neutron in the system that could permit us to reconstruct its history. An example of an alternative distinction between coherent and incoherent scattering occurs in the discussion of the recoil of the atoms of the device in response to the scattering of the particle. A crystalline lattice can recoil as a whole (coherent scattering) as e.g. in the Lamb-Mössbauer effect. Alternatively, the recoil can just affect a single atom or a small group of atoms (incoherent scattering).[6]

Of course, also no information will be obtained about the path of the electron, if it can fly through the slits without any interaction. But this is less likely for slow charged electrons and narrow slits. At higher energies the interactions become predominantly incoherent because they will unavoidably have an impact on the apparatus.

In incoherent scattering the electron behaves like a tennis ball. The hardest part of the mystery of the double-slit experiment is thus the paradox which occurs when we compare coherent scattering in the single-slit and in the double-slit experiment. Feynman resumed this mystery by asking: How can the particle "know" if the other slit is open or otherwise? In fact, as the interactions of the electron must be local it is hard to see how they could be influenced by the open/closed status of the other slit (see Section 4).

[6] In the historical Mössbauer effect with a ^{57}Fe nucleus, the nucleus is tagged by the isomeric transition which has taken place in it, such that it is nevertheless an incoherent process. However, in the later developed technique of Mössbauer diffraction using synchrotron radiation, the nucleus first absorbs and then re-emits the 14.37 keV radiation, such that the initial and final states of the nucleus are identical and the nucleus is no longer tagged. The process becomes then coherent, giving rise to Bragg peaks [27].

3 Analysis of the algebra: superposition or Huygens' principle?

3.1 The ideal logical approach

Let us now take a break until Section 4, leaving our intuition for what it is, and turn to QM to analyse the last step of our metamathematical analysis described in Subsection 1.2. To simplify the formulation, we will in general casually use the term probability for what in reality are probability densities. In a purely QM approach we could make the calculations for the three configurations of the experimental set-up. We could solve the wave equations for the single-slit and double-slit experiments:

$$\begin{array}{llll} -\frac{\hbar^2}{2m}\Delta\psi_1 + V_1(\mathbf{r})\,\psi_1 = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi_1, & \mathrm{S}_1 \text{ open}, & \mathrm{S}_2 \text{ closed}, \\ -\frac{\hbar^2}{2m}\Delta\psi_2 + V_2(\mathbf{r})\,\psi_2 = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi_2, & \mathrm{S}_1 \text{ closed}, & \mathrm{S}_2 \text{ open}, \\ -\frac{\hbar^2}{2m}\Delta\psi_3 + V_3(\mathbf{r})\,\psi_3 = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi_3, & \mathrm{S}_1 \text{ open}, & \mathrm{S}_2 \text{ open}. \end{array} \tag{3}$$

Here S_j refer to the slits. (We have written Schrödinger equations in Eqs. 3 and 5 because they are familiar to a much broader public, but we could have written just as well the corresponding Dirac or Pauli equations). Now the ideal case would be that we have closed-form exact analytical solutions for the three configurations, and that applying the Born rule would show that:

- $|\psi_3|^2$ reproduces the interference pattern.
- $|\psi_1|^2 + |\psi_2|^2$ does not reproduce the interference pattern but reproduces the results for tennis balls.

We have found a justification for the use of the Born rule in [1,11,2]. In a description of the electron's spinning motion based on the spinors of SU(2), each electron in the wave function has its own spinor ψ attached to it to describe its spinning motion. That is why the wave function is a spinor field. The spinors in SU(2) satisfy automatically $\psi^\dagger\psi = 1$. Therefore counting electrons must be done with $\psi^\dagger\psi$. For photons, the justification is different.

We would then have established the logical chain of flawless mathematics that leads us from the geometry to the double-slit experiment, mentioned in Subsection 1.2 And we could then search for errors in the second, intuitive chain of mathematical arguments. But it is not at all obvious to find an exact analytical closed-form solution of the Schrödinger equation for ψ_3. Probably such a solution just does not exist. But we can imagine that numerical solutions of the differential equations could be found and this way nevertheless confirm this scenario. However, the fact that we do not have nice analytical expressions at our finger tips that we could show and analyze with surgical precision, makes formulating the arguments in the discussion much more difficult, because it will look more like loose talk. Therefore, at this stage of the development, we will *assume without proof* that the exact solution of the equation for ψ_3 yields the interference pattern. A rigorous proof for it will be given in Subsection 3.6.

3.2 The textbook rules

What we described above is not completely identical to what is taught in textbooks, which tell us that there are two different rules for calculating probabilities for states that are linear combinations of wave functions in QM:

$$\psi = \sum_{j=1}^{n} c_j\psi_j \quad \Rightarrow \quad p = \begin{cases} \sum_{j=1}^{n} |c_j|^2\,|\psi_j|^2 & (\textit{incoherent summing}) \\ \left|\sum_{j=1}^{n} c_j\psi_j\right|^2 & (\textit{coherent summing}) \end{cases} \tag{4}$$

Incoherent summing must be used when the interactions of the particle with the experimental device are incoherent. Coherent summing must be used when the interactions are coherent. Up to this point, there is perhaps not a problem. It only tells how we must perform the calculations.

But textbooks go one better by telling us that the coherent-summing rule teaches us that we should not add up probabilities but probability amplitudes, $|\psi_3|^2 = |\psi_1 + \psi_2|^2$. For the double-slit experiment this conjures up the impression that our rule $p_3 = p_1 + p_2$ for mutually exclusive probabilities p_1 and p_2 may no longer be correct at the quantum level, or alternatively, that p_1 and p_2 are no longer mutually exclusive. In fact, ignoring Bohr's caveat, we expect that the probabilities p_1' and p_2' of traversing the two slits in the double-slit experiment would be just the same as the single-slit probabilities p_1 and p_2, such that we should obtain $p_3 = p_1 + p_2$ which is factually contradicted by the experimental evidence.[7]

[7] We might be mystified here by the paradigm of wave packets, which can be used to argue that the electron must be a wave packet that can travel through both slits at the same time and even interfere with itself. But by explaining the meaning of the

Textbooks further discuss the coherent-sum rule in terms of a "superposition principle". They compare this summing of probability amplitudes to the addition of the amplitudes of waves as can be observed in a water tank and as also discussed by Feynman. To justify the superposition principle the linearity of the wave equations is evoked.

We will have to object fiercely that there is no superposition principle (for spinors or wave functions in general) and that there is no particle-wave duality. The particle-wave duality is a dogma that defies any logic, but we are told to accept it religiously as a quantum mystery in an act of faith. It sounds as the dogma of the divine Trinity in Christian religion. There is a Father, a Son and a Holy Ghost, but there is only one God. This is a mystery and we must accept it in an act of faith. The analogy must be clear. There is a particle and a wave, but there is only one electron. It is a quantum mystery we must accept in an act of faith. Such contradictions provoke a cognitive dissonance and thwart a rational approach to the paradox. It is a brainwash. A conceptual clean-up in the form of a deconstruction of the Copenhagen doctrine is therefore more than timely.

We will therefore now discuss three principles: a fake superposition principle, a true superposition principle and a Huygens' principle.

3.3 The fake superposition principle

What is used in textbooks in the double-slit context is a fake superposition principle that ignores Bohr's caveat. Because the true superposition principle, based on the linearity of the Schrödinger and Dirac equations would be that a linear combination $\psi = \sum_j c_j \chi_j$ is a solution of a Schrödinger equation:

$$-\frac{\hbar^2}{2m}\Delta\psi + V(\mathbf{r})\,\psi = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi. \tag{5}$$

when all wave functions χ_j are solutions of the *same* Schrödinger equation Eq. 5. The same is correct *mutatis mutandis* for the Dirac equation. This is a well-known straightforward mathematical result,

But telling that the solution ψ_1 of a first equation with potential V_1 can be added to the solution ψ_2 of a second equation with a different potential V_2 to yield a solution ψ_3 for a third equation with yet a different potential V_3 can *a priori* not be justified by the mathematics and is not exact. It has nothing to do with the linearity of the equations. It is just infringing the rule that we should not mix up conditional probabilities defined by different experimental setups. This is pure mathematics and it underpins Bohr's caveat. Summing the equations for ψ_1 and ψ_2 does not yield the equation for ψ_3. A solution of the wave equation for the single-slit experiment will not necessarily satisfy all the boundary conditions of the double-slit experiment, and vice versa. These boundary conditions define the conditional probabilities. This fake superposition principle can thus not be used in QM and it has not its place in a discussion of the paradox.

3.4 The true superposition principle?

3.4.1 Fake after all

The true superposition principle looks mathematically sound, but using it to explain the interference boils in reality down to the fake superposition principle. One could think of defining a solution ψ'_j for the double-slit experiment whereby all particles travel uniquely through slit S_j, by imposing the boundary condition $\psi'_j(\mathbf{r}) = 0$, $\forall \mathbf{r} \in S_{1+|j-2|}$. Here $1 + |j-2| = 2$ for $j = 1$, and $1 + |j-2| = 1$ for $j = 2$. The condition must be strengthened by a condition that $\psi'_j(\mathbf{r})$ remains zero in some neighbourhood behind the slit. In fact, ψ'_j could only accidentally be zero, which would not impede its propagation, such that this must be prevented. This can be done by the conditions imposed on the partial derivatives of ψ'_j. This way slit $S_{1+|j-2|}$ would the open but the particles would not pass through it.

However, these boundary conditions for ψ'_j are just those which express in the mathematics that slit $S_{1+|j-2|}$ is closed. Merely assuming at the same time in your head that both slits would be open or making a drawing showing both slits open, without expressing it in the mathematics is just utopian self-delusion. The assumptions must not be

superluminal phase velocity c^2/v in Subsection. 1.4.1, we have shown that there is no logical need to introduce wave packets such that the electron can stay a point particle. Feynman explains in his discussion that electrons always are detected on a detector screen as points, such that they must be particles. Unless we believe in magic, electrons must therefore always be particles and traverse the slits as particles. It is indeed totally irrational to assume that the electron could miraculously adopt itself to the experimental conditions by morphing itself into a wave packet when it passes through the slits and then become a particle when it reaches the detector. This becomes even more gratuitous once we have understood that wave packets are not a logical necessity as explained in Subsection 1.4.1. Nevertheless, textbooks tell us that an electron is both a particle and a wave, which is a *contradictio in terminis*. The introduction of wave packets introduces a whole chain of further unnecessary fundamental complications (e.g. the collapse of the wave function) as discussed in [1].

expressed in your head or on the drawing, but in the algebra. What you have not entered into the algebra, cannot come out of it by divine intervention, whatever you may have in your head. What you have expressed in the mathematics is that $S_{1+|j-2|}$ is closed, which is exactly the opposite of you had in mind, *viz.* that it should be open. One cannot express in the mathematics that $S_{1+|j-2|}$ is open and closed at the same time.

Hence, the whole promising idea falls apart, because you just have laid down the boundary conditions for the two single-slit experiments, and therefore defined the single-slit solutions ψ_j. This just resumes to overriding Born's caveat by writing $\psi_1' = \psi_1$, $\psi_2' = \psi_2$, and $\psi_3 = \psi_1 + \psi_2$. Furthermore, the very argument based on linearity implies that ψ_1' and ψ_2' would be both valid solutions for the double-slit experiment, which they are definitely not, because they would answer the "which-way" question (see Section 8). We will come back on this problem in Subsection 9.1. What we develop there is subtle, but the contents of the present sub-subsection show that it is not artificial because it will only spell out the rigorous implications of the idea formulated here to justify the textbook calculations by invoking the true superposition principle based on the linearity of the wave equations. And before we took a closer look at it, this idea appeared self-evident rather than artificial.

3.4.2 The argument of linearity revisited

Despite the linearity of the wave equations, with spinors the superposition principle can *a priori* not be applied. In the reconstruction of QM based on the geometrical meaning of spinors, it is *a priori* completely meaningless to make linear combinations of spinors [1,10,11]. The spinors are group elements. In the axioms of a group $(G, \circ)$ with composition law $\circ$ we define a composition of two group elements $g_2 \circ g_1$, but not a linear combination $c_1 g_1 + c_2 g_2$. Linear combinations of group elements are just not defined (See Appendix and [15], see also e.g. eq. 1 in [1]).

When we represent the group elements g of a non-abelian group by matrices $\mathbf{D}(g)$, making linear combinations becomes algebraically feasible. But the spinors and the matrices belong to a curved manifold (because the groups SO(3) and SO(1,3) are non-abelian), such that a formal algebraic linear combination $c_1\mathbf{D}(g_1) + c_2\mathbf{D}(g_2)$ will *a priori* not belong to the manifold. This precludes any geometrical interpretation of such mindless and *a priori* meaningless algebra as discussed in Section 2.5 of [1]. This is further worked out in the Appendix and [15]. In fact, linear combinations can only be made in the tangent space to the manifold of the Lie group. That is why one introduces the Lie algebra by defining infinitesimal generators.

Making linear combinations of spinors is taken for granted in the Hilbert space formulation of QM. As accordingly such linear combinations are routinely used with good results despite the mathematical no-go zone, we are obliged to find a mathematical justification for summing spinors, as promised in Subsection 1.3.3. The situation is somewhat reminiscent of how we have been obliged to find a justification for the use of complex numbers. Due to the underlying assumption $\imath^2 = -1$, their use looked mathematically meaningless but they led to a bountifulness of correct and useful results, which one could not continue to reject with contempt in the name of the sacred truth.

An inspection of group representation theory in [1,11] reveals that the purely formal expression $\sum_{j=1}^{n} g_j$ is in reality a definition of the set $\mathscr{S} = \{g_1, g_2, \cdots g_j, \cdots g_n\}$. In fact, when one defines all-commuting operators $s = \sum_{j=1}^{n} g_j$, such that $\forall h \in G : h \circ s = s \circ h$, this identity just stands for $h \circ \mathscr{S} = \mathscr{S} \circ h$. This permits us to give a precise meaning to $\sum_{j=1}^{n} c_j g_j$. But the sum must then be interpreted as a mere juxtaposition just like the group elements g_j are listed in a mere juxtaposition within the set $\mathscr{S}$. The probabilities must then be summed *incoherently* as explained in [1,11]. However, the group theory cannot be used to justify the coherent-sum rule. There is really absolutely no way to make geometrical sense of coherent summing by using the approach based on sets. In fact, in the case of destructive interference, $\psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) = 0$, it would imply that the union of two non-empty sets would be an empty set: $\mathscr{S}_1 = \{\psi_1(\mathbf{r})\} \neq \emptyset$ & $\mathscr{S}_2 = \{\psi_2(\mathbf{r})\} \neq \emptyset$ & $\mathscr{S}_1 \cup \mathscr{S}_2 = \{\psi_1(\mathbf{r}), \psi_2(\mathbf{r})\}$, in contradiction with $\mathscr{S}_1 \cup \mathscr{S}_2 = \emptyset$, as implied by $\psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) = 0$. That is some hell of a contradiction. It would destroy all our mathematics, which is based on set theory. Adding wave functions and probability amplitudes is certainly algebraically feasible, but *a priori* completely incompatible with their geometrical meaning. Despite this absolute mathematical and logical taboo, QM does resort to coherent summing in order to explain interference, and - startlingly - with very convincing results.

3.5 The Huygens' principle

This brings us to the Huygens' principle as a means to justify the coherent-sum rule. It will allow us to respect Bohr's caveat and the geometrical meaning of spinors. The Huygens' principle is a mathematically correct method to find solutions of certain elliptic partial differential equations [28,29]. It is intuitively clear that the Huygens' principle will lead to the solution $\psi_3 = \psi_1 + \psi_2$ for the differential equations that correspond to the double-slit experiment, at least to a very good precision.

The idea is thus that the prescription $\psi_3 = \psi_1 + \psi_2$ will be an excellent but purely numerical algorithm, whereby the summing procedure has absolutely no physical meaning. That the sum cannot have a physical meaning is easily

seen from the example of destructive interference $\psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) = 0$. In fact, when χ_1 is a spinor of SU(2), then also $\chi_2 = -\chi_1$ is a spinor of SU(2), but $\chi_1 + \chi_2 = 0$ is not a spinor of SU(2) because it cannot be normalized to 1. It can also not be given a meaning in terms of sets because it leads to the contradiction between $\mathscr{S}_1 = \{\chi_1\} \neq \emptyset$ & $\mathscr{S}_2 = \{\chi_2\} \neq \emptyset \Rightarrow \mathscr{S}_1 \cup \mathscr{S}_2 = \{\chi_1, \chi_2\} \neq \emptyset$ and $\mathscr{S}_1 \cup \mathscr{S}_2 = \emptyset$, as explained above.

The sum is thus not only meaningless in physics as pointed out in Subsection 1.3.3, it is also meaningless in the group theory. While the summing procedure is meaningless, the end product is given a meaning in QM, viz. that no electrons will occur in that part of space where we have "destructive interference". Other, truly physical reasons why the Huygens' principle cannot have physical meaning will be given in Section 5. Giving the coherent sum a physical interpretation in terms of a superposition principle is therefore *logically and physically flawed*. Of course, agreement with experiment can only validate here the numerical accuracy of this prescription, not the flawed logic that could be used to give it a nonsensical physical interpretation.

The fact that we should consider coherent summing merely as a good numerical procedure rather than an exact physical truth is important. It is the only logical loophole of escape from the paradox. We must then use the following argument. We must solve a partial differential equation for spinors. Summing of wave functions is therefore not allowed. But the differential equation can be formulated in a broader, less restrictive context than that of spinors and wherein vector-like summing is now allowed. We therefore solve the equation within this broader unrestricted context to find the whole pool of possible solutions. Here we can use the Huygens' principle and it is mathematically justified. Such a mathematical solution must then be taken as is and cannot be given a physical interpretation. The spinor solutions will be a subset of the exhaustive set of solutions obtained. If we can find afterwards a rationale to justify a solution from the pool in terms of spinors then we will have succeeded in solving the equation for the restricted context of spinors.

The subtlety of the argument may upset the reader because it does not look physical at all, but as we pointed out in the Introduction, solving a paradox is a demanding exercise of pure logic and mathematical rigour, not an easy-going exercise of formulating *ad hoc* physical guesses! Rejecting conceitedly rigorous mathematics for the sake of personal comfort or fancies is not accepted methodology. It must be further realized that introducing stunning novel physical principles that are not open to direct testing, such as advanced waves (or many worlds), or that are outright contradictory, such as the particle-wave duality, is far more questionable than proposing a rigorous mathematical treatment with some parts that are admittedly subtle. It is not this particular subtlety in the mathematics which is unpalatable, it is the whole way we have been taught QM, which is! Refusing an argument based on its mathematical subtlety reminds of the joke about a person who searches for his spectacles under a street lamp at night, admits he lost them elsewhere in a very remote place, but argues that searching is less fearsome in places that are well lit.

The only thing the reader must keep in mind is that we do this to show that there are no magic axioms beyond human understanding that rule within QM. There must exist a logical explanation for the paradox. Physics is not a matter of witchcraft.

To take the argument developed into account rigorously, we will define that the solution $\psi_1 + \psi_2$ for the wave function ψ_3 of the double-slit wave equation follows a Huygens' principle and note it as $\psi_3 = \psi_1 \boxplus \psi_2$ to remind that it is only numerically accurate, reserving the term superposition principle for the case when the wave functions we combine are not only all solutions of the same linear equation, but also combined in a way that is compatible with the geometrical meaning of spinors.

We make this distinction between the superposition principle (with incoherent summing) and a Huygens' principle (with coherent summing) to make sure that we respect what we can do and what we cannot do with spinors. We think it is illuminating to make this distinction, because it clarifies the axiomatics at stake in our metamathematical analysis. It lays also a mathematical basis for justifying that we have two different rules for calculating probabilities and that both the incoherent rule $p = \sum_j |c_j|^2 |\chi_j|^2$ and the coherent rule $p = |\psi|^2 = |\sum_j c_j \chi_j|^2$ are "correct" within their respective domains of validity. This is the mathematical essence of the problem. QM just tells us that once we have an *exact* pure-state solution of a wave equation, we must square the amplitude of the wave function to obtain an *exact* probability distribution, based on the Born rule. This justifies then coherent summing, based on the argument that the solution obtained using the Huygens' principle is an exact pure-state solution rather than a superposition.

The double-slit paradox is so difficult that it has the same destabilizing effect as gaslighting. One starts doubting about one's own mental capabilities. But the very last thing we can do in face of such adversity is to capitulate and think that we are not able to think straight. We will thus categorically refuse to yield to such defeatism. If we believe in logic, the rule $p_3 = p'_1 + p'_2$, where p'_1 and p'_2 are the mutually exclusive probabilities to traverse the slits in the double-slit experiment, *must* still be exact, even if in principle we cannot measure these probabilities (see Section 8) without modifying the set-up (and therefore the conditional probabilities). We are using here the accents to distinguish the conditional probabilities p'_1 and p'_2 which occur in the double-slit experiment from the conditional probabilities p_1 and p_2 which occur in the single-slit experiments, thereby acknowledging Bohr's caveat. This principle of adding mutually exclusive probabilities p'_1 and p'_2 can never be questioned by claiming that it would no longer be true on the quantum level and that it would have to be replaced there by a rule of summing probability amplitudes. Because that

would destroy Boolean logic and also entail the possibility that the union of two non-empty sets would be an empty set (as evoked in Subsection 3.4.1).

In summary, we must have $\psi_3 = \psi'_1 + \psi'_2$, where ψ'_1 and ψ'_2 are the contributions to ψ_3 in the double-slit experiment. These contributions cannot be obtained by imposing boundary conditions, but must be obtained by applying mathematical surgery on ψ_3 after its calculation, as will be explained in Subsection 9.1. These contributions are mutually exclusive such that $p_3 = p'_1 + p'_2 = |\psi'_1|^2 + |\psi'_2|^2$ (according to Boolean algebra), but $p_3 \neq p_1 + p_2$ (according to Bohr's caveat). The inequality $p_3 \neq p_1 + p_2$ justifies also our rejection of the fake superposition principle. We must also have $\psi_3 = \psi_1 \boxplus \psi_2$ (Huygens' principle), whereby ψ_1 and ψ_2 are the solutions of the single-slit experiment, because it expresses that this sum is an accurate numerical solution. This leads to $p_3 = |\psi_1 \boxplus \psi_2|^2$ according to the Born rule. We will have then the following identities and inequalities we will use to justify in Subsection 9.1 the use of the solution obtained by the Huygens' principle for spinor fields:

$$\begin{array}{lllll} p_3 & = & |\psi_3|^2 & = & |\psi_1 \boxplus \psi_2|^2 \\ & = & |\psi'_1|^2 + |\psi'_2|^2 & = & p'_1 + p'_2 \\ & \neq & |\psi_1|^2 + |\psi_2|^2 & = & p_1 + p_2 \end{array} \quad \left| \begin{array}{l} \text{Huygens' principle} \\ \text{Boolean algebra} \\ \text{Bohr's caveat, fake superposition.} \end{array} \right. \tag{6}$$

We are then compelled to conclude that in QM the probability p'_1 for traversing slit S_1 when slit S_2 is open is manifestly different from the probability p_1 for traversing slit S_1 when slit S_2 is closed. We can then ask with Feynman how the particle can "know" if the other slit is open or otherwise if its interactions are local.

3.6 Coherence-induced symmetry: Rigorous proof of the coherent-sum rule

3.6.1 We are conceptually stuck

As it stands after the preceding survey, the coherent-sum rule has been introduced as a mere rule that has been empirically verified, but for which we have not found a transparent justification. The Huygens' principle could constitute a justification for it but the calculations it entails to render it rigorous are leading us into impenetrable algebra rather than some true physical insight. We need something better.

3.6.2 Geometry

We will therefore now give a proof for the coherent-sum rule. We will immediately give a general proof for an n-slit experiment. We consider a plate with n slits S_k. We consider the thickness of this plate to be zero. We call the plane of the plate π_s. The plane π_d of the detector screen is considered to be parallel to π_s. We introduce a reference frame with its z-axis perpendicular to $\pi_d \parallel \pi_s$, and its origin $O \in \pi_s$. We will call $\mathscr{V}$ the part of $\mathbb{R}^3$ between π_d and π_s (including the boundaries).

Let us discuss a number of geometrical assumptions. Let us assume that the source is situated in a plane parallel to $\pi_d \parallel \pi_s$ and emitting electrons whereby the electrons can only travel on straight lines. When they hit the plate in the plane π_s, we assume that the electrons are absorbed. Else they just pass through a slit S_k and travel further to the detector. If the particles were neutrons, they could e.g. be absorbed by neutron capture. This implies that we have a binary situation. We only can have transmission or absorption. Behind the slits we will therefore only be interested in the electrons which traverse the slits.

If the source where infinite and emitting plane waves propagating parallel to the plane π_s, the straight lines followed by the electrons would have to be parallel to the z-axis. The projection parallel to the z-axis of the slits S_k on the plane π_d of the detector screen would then define n disjoint sets on π_d. We can call these sets the illumination spots. Under the given assumptions, we cannot obtain interference because the illumination spots do not overlap.

If the source were a point source S and emitting spherical waves, then the electrons would give rise to illumination spots corresponding to n disjoint sets $\mathscr{P}_k$ on the detector screen, where $\mathscr{P}_k$ is the projection with centre S of S_k on π_d. Again the illumination spots would not overlap. Therefore, in order to obtain n sets $\mathscr{P}_k$ such that $\exists \mathscr{B} \subset \cap_{k=1}^{n} \mathscr{P}_k \neq \emptyset$ whereby the set $\mathscr{B}$ has some extension with a non-zero surface, the source $\mathscr{Q}$ must be extended. If it is sufficiently extended, we will obtain illumination spots which all overlap mutually and can give rise to interference.

We could also have assumed that the electrons bounce back as billiard balls from the plate, but that would lead to considering a complicated mix of trajectories before the slits, with incoming and reflected paths. It could then also be questioned why the interaction has to be a billiard ball collision. That is not necessarily true. On the atomic level other scattering laws than rigid-sphere collisions may prevail. For all these reasons, assuming absorption is simpler.

All these purely geometrical assumptions are crude. They would not explain the existence of diffraction fringes in single-slit experiments. But their discussion illustrates already how difficult it would be to formulate and solve rigorously single-slit Schrödinger equations that could be combined in order to give rise to an interference region

on the detector screen. In what follows we will therefore just assume that we have formulated and solved single-slit equations whose solutions have the required properties to give rise to a common illumination zone on the detector, that gives rise to an interference pattern when the solutions are combined in coherent summing, without specifying any details about the formulation and the solution of these single-slit equations.

Let us note finally that a trajectory can be identified with a Lorentz transformation, the Lorentz transformation that describes the propagation of the electron on its trajectory. This assumption has been used also in the derivation of the Dirac equation in [1]. We have explained that it does not makes sense to add up representations of two different Lorentz transformations. But such sums do occur in the algebra of coherent summing. We must therefore construct the wave function by first doing the algebra of brute-force coherent summing, as though this were permitted for group representations. Afterwards one must then find an explanation for this result in terms of an incoherent sum of group representations which can be treated in terms of sets. That will require rewriting the algebraic coherent sum as an incoherent sum of group representations. That is what we will do in Subsection 9.1.

3.6.3 Behind the slits

After all these preparations to set the stage, we can now assume that we have n single-slit equations corresponding to the case that slit S_k is open and all other slits are closed, such that $\forall k \in [1,n] \cap \mathbb{N}$:

$$[\,-\frac{\hbar^2}{2m_0}\Delta + V_k(\mathbf{r})\,]\,\psi_k(\mathbf{r}) = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi_k(\mathbf{r}), \tag{7}$$

with:

$$(\forall \mathbf{r} \in \mathrm{S}_k)\,(V_k(\mathbf{r}) = 0),\ (\forall \mathbf{r} \in \pi_s \backslash \mathrm{S}_k)\,(V_k(\mathbf{r}) = \infty),\ (\forall \mathscr{V} \backslash \pi_s)(V_k(\mathbf{r}) = 0). \tag{8}$$

Summing the n equations yields:

$$-\frac{\hbar^2}{2m_0}\Delta\,[\,\boxplus_{k=1}^{n}\psi_k(\mathbf{r})\,] + \boxplus_{k=1}^{n} V_k(\mathbf{r})\psi_k(\mathbf{r}) = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}[\,\boxplus_{k=1}^{n}\psi_k(\mathbf{r})\,]. \tag{9}$$

We have used here the convention introduced in Subsection 3.5 to use the symbol $\boxplus$ for the sums to remind that results obtained by purely algebraically summing of group representations requires further geometrical justification. The second term on the lefthand side impedes us to identify Eq. 9 with the solution:

$$[\,-\frac{\hbar^2}{2m_0}\Delta + W(\mathbf{r})\,]\,[\,\boxplus_{k=1}^{n}\psi_k(\mathbf{r})\,] = \frac{\hbar}{\imath}\frac{\partial}{\partial t}[\,\boxplus_{k=1}^{n}\psi_k(\mathbf{r})\,], \tag{10}$$

we would like to find for the multi-slit equation corresponding to the case that all slits are open:

$$[\,-\frac{\hbar^2}{2m_0}\Delta + W(\mathbf{r})\,]\,\psi = \frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi, \tag{11}$$

with:

$$\begin{aligned}&(\forall \mathbf{r} \in \overset{n}{\underset{k=1}{\cup}}\mathrm{S}_k)(W(\mathbf{r}) = 0), \quad (\forall \mathbf{r} \in \pi_s \backslash (\overset{n}{\underset{k=1}{\cup}}\mathrm{S}_k)\,(W(\mathbf{r}) = \infty),\\ &(\forall \mathscr{V} \backslash \pi_s)(W(\mathbf{r}) = 0),\end{aligned} \tag{12}$$

if the coherent-sum rule is correct, because it stipulates that the solution should be:

$$\psi(\mathbf{r}) = \boxplus_{k=1}^{n}\psi_k(\mathbf{r}). \tag{13}$$

At the time we discussed Eq. 3 in Subsection 3.1 we already pointed out that it is in general not possible to identify eqs. 9 and 10. But we will render it possible by adopting some special boundary conditions. We introduce the following assumption:

$$\forall k \in [1,n] \cap \mathbb{N}, \forall \mathbf{r} \in \pi_s \backslash \mathrm{S}_k : (V_k(\mathbf{r}) = \infty) \Rightarrow (V_k(\mathbf{r})\psi_k(\mathbf{r}) = 0). \tag{14}$$

Equivalently, instead of fighting with products of the type $V_k(\mathbf{r})\psi_k(\mathbf{r}) \to \infty \times 0$ we can just assume that the plate material does not transmit. This leads then to:

$$(\forall \mathbf{r} \in \mathrm{S}_k)\,(V_k(\mathbf{r})\psi_k(\mathbf{r}) = 0), \quad (\forall \mathbf{r} \in \pi_s \backslash \mathrm{S}_k)\,(V_k(\mathbf{r})\psi_k(\mathbf{r}) = 0),$$
$$(\forall \mathbf{r} \in \mathscr{V} \backslash \pi_s)\,(V_k(\mathbf{r})\psi_k(\mathbf{r}) = 0). \tag{15}$$

This way Eq. 9 yields:

$$\forall \mathbf{r} \in \mathscr{V} : -\frac{\hbar^2}{2m_0} \Delta [\, \boxplus_{k=1}^{n} \psi_k(\mathbf{r})\,] = -\frac{\hbar}{\imath} \frac{\partial}{\partial t} [\, \boxplus_{k=1}^{n} \psi_k(\mathbf{r})\,]. \tag{16}$$

We assume further:

$$\forall \mathbf{r} \in \pi_s \backslash (\bigcup_{k=1}^{n} \mathrm{S}_k) : (W(\mathbf{r}) = \infty) \Rightarrow (W(\mathbf{r})\psi_k(\mathbf{r}) = 0). \tag{17}$$

This follows from the assumptions we made in Eq. 14 for the single-slit equations. This leads then to:

$$(\forall \mathbf{r} \in \bigcup_{k=1}^{n} \mathrm{S}_k)\,(W(\mathbf{r})\psi_k(\mathbf{r}) = 0), \quad (\forall \mathbf{r} \in \pi_s \backslash (\bigcup_{k=1}^{n} \mathrm{S}_k))\,(W(\mathbf{r})\psi_k(\mathbf{r}) = 0),$$
$$(\forall \mathbf{r} \in \mathscr{V} \backslash \pi_s)\,(W(\mathbf{r})\psi_k(\mathbf{r}) = 0). \tag{18}$$

Hence Eq. 10 yields $\forall \mathbf{r} \in \mathscr{V}$:

$$-\frac{\hbar^2}{2m_0} \Delta [\, \boxplus_{k=1}^{n} \psi_k(\mathbf{r})\,] = \frac{\hbar}{\imath} \frac{\partial}{\partial t} [\, \boxplus_{k=1}^{n} \psi_k(\mathbf{r})\,]. \tag{19}$$

Now Eqs. 16 and 19 are identical on $\mathscr{V}$. Therefore the boundary conditions have rendered Eqs. 9 and 10 equivalent on $\mathscr{V}$. Note that the proof is critically contingent on the modified assumptions $\forall k \in [1,n] \cap \mathbb{N}, \forall \mathbf{r} \in \pi_s \backslash \mathrm{S}_k : V_k(\mathbf{r})\psi_k(\mathbf{r}) = 0$ in Eq. 14, and $\forall \mathbf{r} \in \pi_s \backslash (\cup_{k=1}^{n} \mathrm{S}_k) : W(\mathbf{r})\psi_k(\mathbf{r}) = 0$ in Eq. 17, in other words on an idealized situation whereby the plate material does not transmit whatsoever. A presence of tunneling would destroy the proof. We have obtained thus for all potentials V, that $V(\mathbf{r})\psi(\mathbf{r}) = 0 \Leftrightarrow V(\mathbf{r}) = 0 \vee \psi(\mathbf{r}) = 0$ by the condition $V(\mathbf{r}) \neq 0 \Rightarrow \psi(\mathbf{r}) = 0$.

In other words, as pointed out with respect to the three lines in Eq. 3 of Subsection 3.1 it is absurd to expect that summing the solutions of the two different equations on lines 1 and 2 will yield the solution of yet another different equation on line 3. However, it has now been rigorously proved that the special set of boundary conditions given by Eqs. 14 and 17 renders this procedure viable. It takes some reasonable approximations to justify the application of these boundary conditions to the true physical situation. A true physical situation would also require a finite thickness for the plate, and this would already be enough to spoil the simplicity of the mathematics.

Nevertheless, the algebra looks weird because it gives the impression that the waves are traveling all the time in free space. But in reality, the wave functions have boundary conditions: $(\forall \mathbf{r} \in \pi_s \backslash \mathrm{S}_k)(\psi_k(\mathbf{r}) = 0)$, and $(\forall \mathbf{r} \in \pi_s \backslash (\cup_{k=1}^{n} \mathrm{S}_k))(\psi_k(\mathbf{r}) = 0)$. As we will discuss below these boundary conditions, together with those for the partial derivatives of ψ_k, introduce shadow regions behind the plate. We have lost these conditions on the way by considering only the boundary conditions for the products $V_k(\mathbf{r})\psi_k(\mathbf{r})$ and $W(\mathbf{r})\psi(\mathbf{r})$, such that we lost the equivalence with the initial conditions which were for $V_k(\mathbf{r})$, $\psi_k(\mathbf{r})$, $W(\mathbf{r})$ and $\psi(\mathbf{r})$. To maintain equivalence we should consider the boundary conditions for $V_k(\mathbf{r})\psi_k(\mathbf{r})$, $\psi_k(\mathbf{r})$, $W(\mathbf{r})\psi(\mathbf{r})$ and $\psi(\mathbf{r})$. With this reformulation of the boundary conditions we can also avoid discussing infinite values for the potential.

In the shadow regions, the equations correspond indeed also to free-space equations. Apart from these boundary conditions the waves are therefore indeed really travelling all the time in free space. The point is however that the regions that are not in the shadow in the single-slit experiments add up to the regions that are not in the shadow in the multi-slit experiment. In fact, the boundary conditions for the derivatives of the wave function at the back of the plate, which are necessary to impose these shadow regions remain unchanged. They are the same on $\pi_s \backslash (\cup_{k=1}^{n} \mathrm{S}_k)$ for ψ and for $\psi_k, \forall k \in [1,n] \cap \mathbb{N}$, and of the type $[\, \frac{\partial \psi}{\partial z}\,]_{+0} = 0$, when the z-axis is perpendicular to π_s and oriented such that $\forall (x,y,z) \in \pi_d : z > 0$. .

These supplementary boundary conditions are needed to make sure that it is not just accidentally that the wave function becomes zero on the plate, such that it could propagate beyond it, but that there are really shadow zones behind the plate. This argument deals then with the boundary conditions for ψ_k and ψ. This really corresponds to the intuition we obtain from a water tank. Therefore the proof is conceptually close to Huygens' principle. However, we cannot say that it rigorously expresses Huygens' principle, because Huygens' principle allows also for non-physical wave propagation, as e.g. in the all-histories approach of Feynman's path integral method (see Section 5).

It must be very clear now that the solution does not rely on the superposition principle or on the use of sets but on a symmetry preserved by coherence (and broken by incoherence). It is perhaps the first proof for the coherent-sum rule ever that is both transparent and rigorous. The rôle played by the coherence is essential. In fact, $e^{\imath\varphi_j}\psi_j$ are also

valid solutions of the single-slit experiments, such that we could *a priori* end up with $\boxplus_{j=1}^{n} e^{\imath\varphi_j}\psi_j$. To make sure that we will really deal with $\boxplus_{j=1}^{n}\psi_j$ in our calculations we must make the phases of all wave functions ψ_j identical at the source. This is indeed the only experiment that is physically feasible. We cannot want to impose physically different phases for different paths through the slits, when these paths cannot be distinguished due to the coherence (as will be further discussed in Section 8). The coherence denies us the possibility to make distinctions between the paths, because distinguishing between them implies a knowledge we can never acquire. Coherence implies that there is only one wave function. If we could really select the phases independently one would have to sum the contributions $e^{\imath\varphi_j}\psi_j$ incoherently. That is, when we loose coherence, we recover the classical tennis ball regime.

To construct a mathematical theory of this we might propose that we are not dealing with sets of group elements, but with sets of sets of group elements. Incoherent scattering would correspond to a set of n singletons $\{\{g_1\},\{g_2\},\cdots\{g_n\}\}$, while coherent scattering would correspond to the singleton $\{\{g_1,g_2,\cdots g_n\}\}$ which has the set $\mathscr{S}=\{g_1,g_2,\cdots g_n\}$ of n elements g_j as its unique element. If we identified the brackets $\{\ \}$ with a wall surrounding a set, then access to the distinction between group elements would be denied by hiding it behind secondary walls. We could call the sets $\{g_j\}$ incoherent and compare them with atoms and the set $\mathscr{S}$ coherent and compare it with a a molecule.

This construction would lay down the mathematical foundations for the Hilbert space formalism and the rules physicists use to calculate probabilities for histories and paths. The two rules could be formulated then as a single rule, whereby for all sets belonging to a set of sets the probabilities must be added incoherently, while for all the elements that belong to these member sets the probabilities must be added coherently. There is nothing self-evident about all this, because the proof of the coherent-sum rule found for the specific situation of a multi-slit experiment described here does not warrant generalization of the rule to other experiments. In view of the large differences that may exist between various experimental set-ups, a general proof might just not be feasible. Different experiments may require entirely different proofs. Therefore, the construction based on sets of sets evoked here is not a proof of the coherent-sum rule but a way to implement the rules and justify the use of Hilbert spaces once the coherent-sum rule has been proved.

Note that this is an algebraic solution ψ of the Schrödinger equation for the multi-slit experiment, whereby we have ignored the fact that summing representations of group elements (i.e. spinors of SO(1,3), but also scalar wave functions of SO(1,1), SO(2) or U(1)) has no geometrical meaning. But the Schrödinger equation for ψ has been formulated for group representations. What we have found is thus a solution based on vector calculus of an equation for group representations for which in principle such vector calculus is not defined!

In summary: It is meaningless to add equations and meaningless to sum group representations, but with all this nonsense we get an algebraic result that is a member of the pool of exact algebraic solutions. Let us therefore not forget: We must still give the purely algebraic solution a geometrical meaning in terms of group representations, in conformity with the approach we outlined in the discussion of Huygens' principle in Subsection 3.5 and summarized in Eq. 6. This is why we have used the symbol $\boxplus$ for the sums. This justification will be given in Subsection 9.1 and especially in Eq. 21. What historically has been introduced based on a flawed intuition based on Copenhagen concepts like the particle-wave duality which do not have any rigorous justification, will then be proved rigorously. And as anticipated in the Introduction, there is absolutely no quantum magic involved in this solution. Just like we promised it relies only on pure logic and pure mathematics.

3.6.4 Before the slits

We should actually also prove the solution for the region before the slits. Let us call the region before the plate $\mathscr{U}$. Physically, we could consider that before the slits, all pre-slit wave functions are the same, because all waves emanate the same way from a same source. Hence $\forall\mathbf{r}\in\mathscr{U},\forall k\in[1,n]\cap\mathbb{N}:\psi_k(\mathbf{r})=\psi(\mathbf{r})$. But in fact, also before the slits we must write $\psi=\boxplus_{k=1}^{n}\psi_k$. Otherwise we do not have a logical continuity with the equation $\psi=\boxplus_{k=1}^{n}\psi_k$ behind the slits. If we are adding up all n single-slit wave equations then we will have $\forall\mathbf{r}\in\mathscr{U},\forall k\in[1,n]\cap\mathbb{N}:\psi(\mathbf{r})=n\psi_k(\mathbf{r})$. There is here a normalization problem, that we while treat in a while, but there is also another problem. We must remove all trajectories that lead to absorption from the single-slit wave functions ψ_k. Then the wave function ψ that enters slit k will be identical to ψ_k rather than to $n\psi_k$. The difference $(n-1)\psi_k$ comes from the trajectories in the $n-1$ wave functions ψ_j with $j\neq k$ that are absorbed at slit S_k which is closed when $j\neq k$.

The most elegant way to treat the wave function before the slits is to inverse the time. This is no physical travel in time but an imaginary mathematical travel within the definition domain $\mathbb{R}^4$ of the wave function. In the mathematics, the fact that time flows does not enter the considerations. We just study the wave function on its definition domain $\mathbb{R}^4$. This way the interference region, i.e. the set $\mathscr{B}$ will now be the source of the trajectories travelling backwards in time, and the extended source $\mathscr{Q}$ will be a region where we add up the advanced single-slit wave functions. But here there will be no interference because we have stated before-hand that we must give all wave functions ψ_k the same phase on the source. We must thus take $\forall\mathbf{r}\in\mathscr{U}:\psi_k(\mathbf{r})=\frac{1}{\sqrt{n}}\psi(\mathbf{r})$, such that $\forall\mathbf{r}\in\mathscr{U}:|\psi_k(\mathbf{r})|^2=\frac{1}{n}|\psi(\mathbf{r})|^2$ and $\forall\mathbf{r}\in\mathscr{U}:\sum_{k=1}^{n}|\psi_k(\mathbf{r})|^2=|\psi(\mathbf{r})|^2$. This settles the normalization problem.

This way what comes before the slits must be treated just the same way as what comes after the slits, by time inversion symmetry, except for the fact that there will be no interference. The waves are not travelling truly backwards in time, but if we remove the trajectories from the incoming wave function that will be absorbed and therefore do not contribute to the wave function at the detector, then indeed the wave function will have this time symmetry. The removal of the trajectories "foreshadows" the fact that they will not make their way to the detector. We can restrict the incoming wave functions ψ_k by this procedure of removing the histories of the electrons that are absorbed by the plate. This looks like a preselection based on advanced knowledge, but it is a preselection *a posteriori* within the mathematical wave function, such that it does not violate causality. It avoids discussing the absorption or reflection of waves. After doing this, there will be now also "foreshadows" before the slits, as a consequence of the preselection. They act as shadow regions for the electrons that would travel backwards in time from the detector to the plane π_s where they hit the plate. There are no such electrons linking the detector to the plate in the wave function. They can be considered to be removed from the wave function by what is a preselection procedure in backward logic. We do not have these electrons in the backward wave function because they are in the shadow region in forward logic.

For one choice of the arrow of time, a region will be a shadow, for the opposite choice, the same region will have been removed by preselection and be a "foreshadow". There is no true back-action in time in this procedure. It is just that we render the wave function compatible with all the boundary conditions by traveling mathematically through the wave function in all directions to check if the wave functions complies with the macroscopic boundaries. This is mathematically feasible but does not have physical meaning, because the direction of the propagation we take through the wave function can be non-physical. It is a kind of Huygens' principle without physical meaning.

In Subsection 9.1 the redefinition of the coherent sum $\psi_1 \boxplus \psi_2$ as an incoherent sum $\psi_1' + \psi_2'$ in Eq. 21 leads to expressions whereby both single-slit beams ψ_1' and ψ_2' are containing $\psi_1 \boxplus \psi_2$. Actually, this analysis must be extended all the way backwards to the source such that we must also make a decomposition $\psi_1' + \psi_2'$ of the wave function before the slits. On the source and at the slits there will be no interference, because $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ do not lead to interference in these places. On the slits this is due to the fact that $\psi_2(\mathbf{r}) = 0$, $\forall \mathbf{r} \in S_1$ and $\psi_1(\mathbf{r}) = 0$, $\forall \mathbf{r} \in S_2$. On the source we will not encounter the problem that e.g. $\psi_j \boxplus \psi_k = 0$ while $\psi_j = -\psi_k \neq 0$, leading to a collision with set theory, because we have explained that all single-slit wave functions must be given the same phase on the source, i.e. $\forall \mathbf{r} \in \mathscr{Q} : \psi_2(\mathbf{r}) = \psi_1(\mathbf{r})$. It is this condition that breaks the symmetry under time reversal, such that it is no longer globally valid.

4 Local interactions, non-local probabilities

The answer to Feynman's question quoted at the end of Subsection 3.5 is that *the interactions* of the electron with the device *are locally defined* while the probabilities defined by the wave function are not. *The probabilities are non-locally, globally defined* in the "deviant" sense we defined for domestic use only in Subsection 1.4.1. When we follow our intuition, the electron interacts with the device in one of the slits. The corresponding probabilities are local interaction probabilities. We may take this point into consideration. Following our intuition we may then think that after doing so we are done.

But in QM the story does not end here! The probabilities are globally defined and we must solve the wave equation with the global boundary conditions. We may find locally a solution to the wave equation based on the consideration of the local interactions, but that is not good enough. The wave equation must also satisfy boundary conditions that are far away from the place where the electron is interacting. The probabilities are not only defined by the local interactions but also by the global geometry of the set-up. This global geometry is fundamentally non-local. This non-locality leads to the unexpected result that the ensuing probability distribution is also non-locally defined, such that the argument that the interactions of an electron are local misses the point that the probability pattern is global. We need the local interactions of many electrons to account for all aspects of the global probability pattern. This claim may look startling. To make sense of this global aspect we propose the following slogan, which we will explain below: *"We are not studying electrons with the measuring device, we are studying the measuring device with electrons"*. This slogan introduces a paradigm shift that will grow to a leading principle as we go along. We can call it the holographic principle (see below, in the discussion based Eq. 20). The slogan should not be lifted out of its context which still has to be explained below. My argument can thus not be distorted to the straw man *"that you build a setup and do experiments to understand what you have built"* (sic!) [14].[8]

[8] One can compare interference patterns from two double-slit experiments that differ only in the distance d between the two slits. We can observe then a marked difference between the two interference patterns. We could then also ask the question: How does a particle passing through a given slit "know" at which distance the other slit is positioned? This clearly illustrates that the probabilities are non-locally defined. Something analogous is encountered with the phonons of a linear chain. If the chain contains e.g. n identical atoms linked by identical springs, the eigenfunctions will be $\psi_k(j) = e^{\imath \frac{2\pi}{n}(k-1)(j-1)}$, where $k \in [1,n] \cap \mathbb{N}$ labels the modes and $j \in [1,n] \cap \mathbb{N}$ labels the atoms. One introduces then $L = na$, where a is the interatomic distance, $x = (j-1)a$ and $q_k = (k-1)\frac{2\pi}{L}, k \in [1,n] \cap \mathbb{N}$. We obtain then $\psi_{q_k}(x) = e^{\imath q_k x}$. For an infinite chain, the eigenfunctions would

In fact, we cannot measure the interference pattern in the double-slit experiment with one electron impact on a detector screen. We must make statistics of many electron impacts. We must thus use many electrons and measure a probability distribution for them. The probabilities must be defined in a globally self-consistent way. The definitions of the probabilities that prevail at one slit may therefore be subject to compatibility constraints imposed by the definitions that prevail at the other slit. We are thus measuring the probability distribution of an ensemble of electrons in interaction with the whole device. While a single electron cannot "know" if the other slit is open or otherwise, the ensemble of electrons will "know" it, because all parts of the measuring device will eventually be explored by the ensemble of electrons if this ensemble is large enough, *i.e.* if our statistics are good enough. When this is the case, the interference pattern will appear. References [30] (especially the corresponding video in [31] from which some snap shots are reproduced on p. 3 of [7]) give actually a nice illustration of how the interference pattern builds up with time.

The geometry of the measuring device is non-local in the sense that a single electron cannot explore all aspects of the set-up through its local interactions. But the wave function which represents an infinite statistical ensemble of electrons can probe all aspects of the set-up. There is no contradiction with relativity in the fact that the probabilities for these local interactions must fit into a global probability scheme, such that they are dictated by parts of the set-up a single electron cannot probe simultaneously. We must thus realize how Euclidean geometry contains information that in essence is non-local, because it cannot all be probed by a single particle, but that this is not in contradiction with the theory of relativity. We have pointed this out in Subsection 1.4.1 and it does not constitute a violation of the theory of relativity if we observe a number of safeguards. E.g. it is not true that for an observer in a distant galaxy traveling at a relativistic speed with respect to the Earth the future on Earth already exists. That is an artefact of using a Lorentz frame with clocks synchronized up to infinite distances and such a frame just does not exist. Such a frame has actually been adopted from Newtonian mechanics, whereby the true change resides in replacing the Galilei transformation by the Lorentz transformation. Unfortunately this folk lore about space-time as a completed infinite mathematical structure $\mathbb{R}^4$ rather than a work in progress is promoted in several texts about relativity.

5 A classical analogy

We can render these ideas clear by an analogy. Imagine a country that sends out spies to an enemy country. The electrons behave as this army of spies. The double-slit set-up is the enemy country. The country that sends out the spies is the physicist. Each spy is sent to a different part of the enemy's country, chosen by a random generator. They will all take photographs of the part of the enemy country they end up in. The spies may have an action radius of only a kilometre: their interactions are local. Some of the photographs of different spies will overlap. These photographs correspond to the spots left by the electrons on your detector. If the army of spies you send out is large enough, then in the end the army will have made enough photographs to assemble a very detailed complete map of the country. That map corresponds to the interference pattern. In assembling the global map from the small local patches presented by the photographs we must make sure that the errors do not accumulate such that everything fits together self-consistently. This is somewhat analogous to the boundary conditions of the wave function that must be satisfied globally, whereby we can construct the global wave function also by assembling patches of local solutions.

The tool one can use to ensure this global consistency is a Huygens' principle. An example of such a Huygens' principle is Feynman's path integral method [32] or Kirchhoff's method in optics [33]. The principle is non-local and is therefore responsible for the fact that we must carry out calculations that are purely mathematical but have no real physical meaning. They may look incomprehensible if we take them literally, because they may involve e.g. backward propagation in space and even in time [34,35], not to mention photons traveling faster than light [34]. Interpretations of QM that rely on advanced potentials or signalling backwards in time just correspond to this Huygens' principle, without realizing that it is only a mathematical expedient without any physical meaning. In fact, each point of the wave front is the source of new spherical waves that can propagate in any direction, which very obviously flies in the face of physical reality. This is the physical basis announced in Subsection 3.5 for disqualifying the Huygens' principle as physically meaningless. As already mentioned, the Huygens' principle has been shown to be valid for elliptic partial differential equations [28,29].

The interference pattern presents this way the information about the whole experimental set-up. It does not present this information directly but in an equivalent way, by an integral transform. This can be seen from Born's treatment of the scattering of particles of mass m_0 by a potential V_s, which leads to the differential cross section:

$$\frac{d\sigma}{d\Omega} = \frac{m_0}{4\pi^2} |\mathscr{F}(V_s)(\mathbf{q})|^2, \tag{20}$$

be just $\psi_q(x) = e^{\imath qx}$, with $q \in \mathbb{R}$. We see thus that a global quantity L can intervene in the definition of the eigenfunctions ψ_{q_k}, due to the boundary conditions, while the interactions can be entirely locally defined, e.g. as mere first-neighbour interactions. Another example is the following: If we want to solve a differential equation on a curved manifold, we need to make the local solutions on the members of the atlas fit together to construct the global solution.

where $\mathbf{p} = \hbar\mathbf{q}$ is the momentum transfer. The integral transform is here the Fourier transform $\mathscr{F}$, which is even a one-to-one mapping. This result is derived within the Born approximation and is therefore an approximate result. In a more rigorous setting, the integral transform could be e.g. the one proposed by Dirac [36], which Feynman was able to use to derive the Schrödinger equation [32]. The Huygens' principles used by Feynman and Kirchhoff are derived from integral transforms to which they correspond. (In Feynman's path integral there will be paths that thread through both slits, which interjects the possibility that $\psi_3 = \psi_1 \boxplus \psi_2$ might not be rigorously exact. We have shown that it is exact in Subsection 3.6). In the Born approximation, which is not exact, the rule $\psi_3 = \psi_1 \boxplus \psi_2$ can be derived using the linearity of the Fourier transform.[9]

The Born approximation provides us with a supplementary mathematical method to justify the result $\psi_3 = \psi_1 \boxplus \psi_2$. Combined with a reference beam $\mathscr{F}(V_s)(\mathbf{q})$ would yield the hologram of the set-up. Of course, it is very obvious that this observation in retrospective that in measuring the interference pattern you have unwittingly collected detailed information about the set-up cannot be ridiculed by distorting it to a claim that the purpose of an experiment would be to build a set-up and then use particles to study what you have built [14]. No initial purpose can be added to this pure post-factum observation, particularly for a trivial set-up with just two slits. The observation is that the interference pattern is by Eq. 20 related to the (square of the) Fourier transform of the set-up and contains therefore rather detailed information about it, just like the X-ray diffraction pattern of a crystal contains detailed information about the atomic structure of that crystal, based on the very same Eq. 20. People who try to synthesize new materials do indeed use characterization by X-ray diffraction to study what they have manufactured by using Eq. 20, this time not unwittingly but really on purpose.

The spies in our analogy are not correlated and not interacting, but the information about the country is correlated: It is the information we put on a map. The map will e.g. show correlations in the form of long straight lines, roads that stretch out for thousands of miles, but none of the spies will have seen these correlations and the global picture. They just have seen the local picture of the things that were situated within their action radius. The global picture, the global information about the enemy country is non-local, and contains correlations, but it can nevertheless be obtained if one sends out enough spies to explore the whole country, and it will show on the map assembled. That is what we are aiming at by invoking the non-locality of the experimental set-up and the non-locality of the wave function.

We may note that the wave functions used in QM are coherent which creates the impression that the electrons are correlated. But this is certainly not true. You can send an electron through the set-up every quarter of an hour and if you wait long enough the interference pattern will nevertheless build up. Such electrons are not correlated. This could lead to the objection that this shows that an electron must be a wave after all. We have discussed this in Subsection 4.3 of [1] and the Appendix of [37] where we have explained why we can use a coherent wave even when the electrons remain point particles and the source which emits the electrons is incoherent.

The global information gathered by many electrons contains the information how many slits are open. It is that kind of global information about the set-up that is contained in the wave function. One needs many single electrons to collect that global information. A single electron gives us just one impact on the detector screen. That is almost no information. Such an impact is a Dirac delta measure, derived from the Fourier transform of a flat distribution. It contains hardly any information about the set-up because it does not provide any contrast. This global geometry contains thus more information than any single electron can measure through its local interactions. And it is here that the paradox creeps in. The probabilities are not defined locally, but globally. The interactions are local and in following our intuition, we infer from this that the definitions of the probabilities will be local as well, but they are not. The space wherein the electrons travel in the double-slit experiment is not simply connected, which is, as we will see, a piece of global, topological information apt to profoundly upset the way we must define probabilities.

6 Highly simplified descriptions still catch the essence

We can now address the problem of generality discussed in Subsection 1.3.2 on which we promised to come back towards the end of Subsection 1.3.3. The description of the experimental set-up we use to calculate a wave function is conventionally highly idealized and simplified. Writing an equation that would make it possible to take into account

[9] We can consider the set-up to be of zero thickness in the z-direction and to be confined to the Oxy-plane π_s. We can then define a model potential $V_3(x,y)$ on the Oxy-plane as follows (see Subsection 3.6): $\forall x \in \mathscr{A} = [-b,-a] \cup [a,b] : V_3(x,y) = 0$, $\forall x \in \mathbb{R}\backslash S : V_3(x,y) = V$. Here $\mathscr{A} \times \mathbb{R} = \mathrm{S}_1 \cup \mathrm{S}_2$ defines the two slits. The model function V_3 just expresses the geometry of the set-up. The Fourier transform can be calculated by considering $V - V_3(x,y)$. In the double-slit experiment, the potential V_3 just reflects the geometry of the set up by translating the presence or absence of matter in a point by the numbers V and 0. The origin of the Fourier transform can then intuitively be understood by considering contributions $e^{-\imath(Et-\mathbf{p}\cdot\mathbf{r})/\hbar}$ originating from each point $\mathbf{r}_0 \in \pi_s$ that has been attributed the value $V_3(\mathbf{r}_0)$. These contributions are produced by the spinning motion of the electron, which travels from such a point of the set-up to the detector, whereby the spinning motion is expressed by a spinor containing $e^{-\imath\omega\tau/2} = e^{-\imath(Et-\mathbf{p}\cdot\mathbf{r})/\hbar}$ in its rest frame.

all atoms of the macroscopic device in the experimental set-up is a hopeless task. Moreover, the total number of atoms in "identical" experimental set-ups is only approximately identical. In such a description there is no thought for the question if the local interactions of the spin of the neutron are coherent or otherwise. Despite its crudeness, such a purely geometrical description is apt to seize a crucial ingredient of any experiment whereby interference occurs. It introduces the phase of the wave function which represents the periodicity of the internal dynamics of the particles, without addressing its specific details. This is very likely the explanation for what we called a fluke, *viz.* that the Schrödinger equation is apt to describe so many different kinds of particles successfully. It also introduces unwittingly the essentials about the boundary conditions which our intuition does not pay intention to in inspecting the geometry. It thereby respects the fact that the probabilities are both non-local and conditional, something that escapes our vigilance.

It is able to account for the difference between set-ups with one and two slits, as in solving the wave equation we automatically avoid the pitfall of ignoring the difference between globally and locally defined probabilities, rendering the solution adopted tacitly global. We are also avoiding the pitfall of forgetting that the probabilities are conditional. In this sense the probability paradox we are confronted with is akin to Bertrand's paradox in probability calculus. It is not sufficient to calculate the probabilities some way. We must further specify how we will use these probabilities later on in the procedure to fit them into a global picture. The probabilities will be only unambiguously defined if we define simultaneously the whole protocol we will use to calculate with them. Here the whole protocol is dictated by the global, non-local set-up. These are the elements which produce what we have called the fluke that all particles seem to comply with a universal formalism with a limited set of equations. And it is for the reasons evoked here that the present paper is general despite the fact that its development is based uniquely on an understanding of the wave function of the electron, i.e. only one special type of particle.

7 Winnowing out the over-interpretations

7.1 No particle-wave duality

It is now high time to get rid of the particle-wave duality. Electrons are always particles, never waves. As pointed out by Feynman, electrons are always particles because a detector detects always a full electron at a time, never a fraction of an electron. Electrons never travel like a wave through both slits simultaneously. But in a sense, their wave function does, behaving like a dense liquid of possible particle positions. It is the probability amplitude distribution of many electrons which displays wave behaviour and acts like a flowing liquid in a water tank, not the individual electrons themselves. And it is the behaviour of this fluid that reminds us of a Huygens' principle. These observations also apply to photons: e.g. in reference [38] it is clearly mentioned that one wants to make sure that a single photon excites only one pixel of the detector at the time, such that photons are also detected as particles. The authors also discuss how to avoid photon bursts.

This postulate only reflects literally what QM says, viz. that the wave function is a probability amplitude, and that it behaves like a wave because it is obtained as the solution of a wave equation. It also transpires in the details of our procedure to define the wave function mathematically in [1] and the Appendix of [2]. Measuring the probabilities requires measuring many electrons, such that the probability amplitude is a probability amplitude defined by considering an ensemble of electrons [39,40] with an ensemble of possible histories.

Although this sharp dichotomy is very clearly present in the rules, we seem to loose sight of it when we are reasoning intuitively. This is due to a tendency towards *"Hineininterpretierung"* in terms of Broglie's initial idea that the particles themselves, not their probability distributions, would be waves. These heuristics have historically been useful but are reading more into the issue than there really is. Their addition blurs again the very accurate sharp pictures provided by QM. With hindsight, we must therefore dispense with the particle-wave duality. The rules of QM are clear enough in their own right: *In claro non interpretatur!* Wave functions also very obviously do not collapse. They serve to describe a statistical ensemble of possible events, not outcomes of single events.

7.2 No superposition principle

It is also time to kill the traditional reading of $\psi_3 = \psi_1 \boxplus \psi_2$ in terms of a "superposition principle", based on the wave picture. It is is only a convenient numerical recipe, a Huygens' principle without true physical meaning. We can make the experiment in such a way that only one electron is emitted by the source every quarter of an hour. Still the interference pattern will build up with time if we wait long enough. But if ψ_1 and ψ_2 were to describe the correct probabilities from slit S_1 and slit S_2, we would never be able to explain destructive interference. How could a second

electron that travels through slit S_2 erase the impact made on the detector screen of an electron that traveled through slit S_1 hours earlier?[10]

We must thus conclude that $\psi_3 = \psi_1 \boxplus \psi_2$ is a numerical expression for the true wave function ψ_3, whereby the physically meaningful identity reads $\psi_3 = \psi'_1 + \psi'_2$ in terms of other wave functions ψ'_1 and ψ'_2 *with incoherent summing* (see Eq. 6 in Subsection 3.5) The wave functions ψ'_1 and ψ'_2 must now both be zero, $\psi'_1(\mathbf{r}) = \psi'_2(\mathbf{r}) = 0$, in all places $\mathbf{r}$ where we have "destructive interference", because $p_3 = p_1 + p_2$ must still be valid. In other words $\psi_1 \neq \psi'_1$ and $\psi_2 \neq \psi'_2$.

8 Undecidability

8.1 Incomplete axiomatic systems

We can further improve our intuition for this by another approach that addresses more the way we study electrons with the set-up and is based on undecidability. The concept of undecidability has been formalized in mathematics, which provides many examples of undecidable questions. Examples occur e.g. in Gödel's theorem [41]. The existence of such undecidable questions may look arcane to common sense but this does not need to be. In fact, the reason for the existence of such undecidable questions is that the set of axioms of the theory is incomplete. We can complete then the theory by adding an axiom telling the answer to the question is "yes", or by adding an axiom telling the answer to the question is "no". The two alternatives permit to stay within a system based on binary logic (*"tertium non datur"*) and lead to two different axiomatic systems and thus to two different theories.

An example of this are Euclidean and hyperbolic geometry [42]. In Euclidean geometry one has added on the fifth parallels postulate to the first four postulates of Euclid, while in hyperbolic geometry one has added on an alternative postulate that is at variance with the parallels postulate. We are actually not forced to make a choice: We can decide to study so-called absolute geometry [43], wherein the question remains undecidable. The axiom one has to add can be considered as information that was lacking in the initial set of four axioms. Without adding it one cannot address the yes-or-no question. This reveals that the axiomatic system without the parallels postulate added is incomplete. We can compare the situation with a joke whereby a person enumerates a long list of commercial items that have been stowed into a large ship and at the end asks you for the age of the captain. Of course the information to answer that question was not provided in his account, such that the answer to the question cannot be given. In a more formal mathematical context, such a question that cannot be answered is undecided. By analogy, we will adopt the same terminology here.

As Gödel has shown, we will almost always run eventually into such a problem of incompleteness. On the basis of Poincaré's mapping between hyperbolic and Euclidean geometry [42], we can appreciate which information was lacking in the first four postulates. The information was not enough to identify the straight lines as really straight, as we could still interpret the straight lines in terms of half circles in a half plane. The straight lines of Euclidean geometry were physically straight in our heads but not in what we laid down about them in the first four postulates. The situation is analogous to what we explained in Subsection 3.4.1 and to some extent in what we explained in Subsection 2.3 of [1], viz. that in the abstract group theory of SU(2), spinors must be group elements.

When the interactions are coherent in the double-slit experiment, the question through which one of the two slits the electron has traveled is very obviously also experimentally undecidable. Just like in mathematics, this is due to lack of information. We just do not have the information that could permit us telling which way the electron has gone, because absolutely no information about that issue has been created by the interactions. The coherent interaction has withheld the information. This is exactly what Feynman has pointed out so carefully. In his lecture he considers three possibilities for our observation of the history of an electron: "slit S_1", "slit S_2", and "not seen". The third option corresponds exactly to this concept of undecidability. He works this out with many examples in reference [5], to show that there is a one-to-one correspondence between undecidability and coherence of the interactions with the set-up.

Coherence can already occur in a single-slit experiment, where it is at the origin of the diffraction fringes. But in the double-slit experiment the lack of knowledge becomes all at once amplified to an objective undecidability of

[10] We may speculate that the electron "feels" whether the other slit is open or otherwise. E.g. the electron might polarize the charge distribution inside the measuring device and the presence of the other slit might influence this induced charge distribution. This would be an influence at a distance that is not incompatible with the theory of relativity. But this scenario is not very likely. As pointed out by Feynman interference is a universal phenomenon. It exists also for photons, neutrons, ^{4}He atoms, etc... which are neutral particles. We already capture the essence of this universal phenomenon in a simple, crude geometrical description of the macroscopic set-up of the experiment. While this could be a matter of pure luck according to the principle that fortune favours fools, it is not likely that one could translate the scenario evoked for electrons to an equivalent scenario in all these different situations. E.g. how could the fact that another slit is open (in a nm-sized double-slit experiment) affect the process at the fm scale of the interaction of the spin of a nucleus with the spin of a neutron? The generality of the scenario based on an influence at a distance is thus not very likely.

the question through which slit the electron has traveled, which does not exist in the single-slit experiment. What happens here in the required change of the definition of the probabilities has nothing to do with a change in local physical interactions. It even has nothing to do with some interactions that would become incoherent. It has only to do with the question how we define a probability with respect to a body of available information. The probabilities are conditional because they depend on the information available. As the lack of information is different in the double-slit experiment, the body of information available changes, such that the probabilities must be defined in a completely different way. This is Bertrand's paradox. Information biases probabilities, rendering them conditional, which is why insurance companies ask their clients to fill forms requesting information about them.

8.2 Incompatibility of the axiomatic systems

We have tried-and-proved methods to deal with such bias. According to common-sense intuition whereby we reason only on the local interactions, opening or closing the other slit would not affect the probabilities or only affect them slightly, but this is wrong (See Subsection 9.5.1). We may also think that the undecidability is just experimental such that it would not matter for performing our probability calculus. We may reckon that in reality, the electron must have gone through one of the two slits anyway. We argue then that we can just assume that half of the electrons went one way, and the other half of the electrons the other way, and that we can then use statistical averaging to simulate the reality, just like we do in classical statistical physics to remove bias. We can verify this argument by detailed QM calculations. We can calculate the solutions of the three wave equations in Eq. 3 and compare $|\psi_3|^2$ with the result of our averaging procedure based on $|\psi_1|^2$ and $|\psi_2|^2$. This will reproduce the disagreement between the experimental data and our classical intuition, confirming QM is right and that we have failed to respect Bohr's caveat. We have failed to discern that the probabilities are conditional, whereby the conditions are non-local. In fact, the correct identity is not $p_3 = p_1 + p_2$ but $p_3 = p'_1 + p'_2$, as already summarized in Eq. 6.

To make sense of this we may argue that we are not used to logic that allows for undecidability. Decided histories with labels S_1 or S_2 occur in a theory based on a system of axioms BL (binary logic), while the undecided histories occur in a theory based on an all together different system of axioms TL (ternary logic). And within the axiomatic system TL it is not possible to define an averaging procedure, because the averaging is based on binary logic. In reality, it can be somewhat more complicated because one can consider that the electron does indeed travel either through S_1 or through S_2 following binary logic, even if it is physically impossible to know which path it has taken due to the coherence of the interactions. Then the axiomatic system is not TL but TL+D, which combines the binary and ternary aspects of the logic of the set-up in a non-contradictory way (see Subsection 9.1.1; we want to point out that we are using here the + sign in the notation TL+D in a completely informal way, not in the way it is being used in the notation ZF+AC for Zermelo-Fraenkel set theory enriched with the Axiom of Choice within specialized literature. We just want to express that we allow simultaneously for two different logical points of view).

In this hybrid axiomatic system TL+D there is nothing logically wrong about the intuitive idea of adopting an averaging procedure for the double-slit experiment. However, the conditional probabilities we must add then are not those from the single-slit experiments because the information about the electron's path does not follow the binary logic according to the axiomatic system BL, but the binary logic according to the axiomatic system TL+D, which respects also the undecidability. But this is then a purely mental construction beyond testing because these conditional probabilities are not experimentally knowable (see Subsection 9.1.1).

If all this sounds esoteric or not very convincing, the reader will change his mind after reading Subsection 9.5.1 where we point out a number of subliminal errors we do not suspect. Due to the information bias the probabilities $|\psi'_1|^2$ and $|\psi'_2|^2$ to be used in $\mathsf{TL} + \mathsf{D}$ are very different from the probabilities $|\psi_1|^2$ and $|\psi_2|^2$ to be used in BL. The paradox results thus from the fact that we just did not imagine that such a difference could exist. We have underestimated the importance of the boundary conditions which are intervening in the definition of conditional probabilities and neglected Bohr's caveat. Assuming $\psi'_j = \psi_j$, for $j = 1, 2$ amounts to neglecting the bias imposed by the axiomatic system $\mathsf{TL} + \mathsf{D}$ on the information contained in our data and reflects the fact that we are not aware of the global character of the definition of the probabilities. We have thus probabilities that are *non-local*, *conditional (i.e. not absolute)* and *undecided*. No wonder the double-slit paradox exhales an exotic flagrance of mystery! The axiomatic system TL+D imposes a global constraint that has a spectacular impact on the definition of the probabilities.

Einstein is perfectly right that the Moon is still out there when we are not watching. But we cannot find out that the Moon is there if we do not register any of its interactions with its environment, even if it is there. If we do not register any information about the existence of the Moon, then the information contained in our experimental results must be biased in such a way that everything looks as though the Moon were not there (This tallies with ideas developed on a toy model, see [44]). Therefore, in QM the undecidability must affect the definition of the probabilities and bias them, such that $p'_j \neq p_j$, for $j = 1, 2$. The experimental probabilities must reflect the experimental undecidability, else reality would contradict itself.

9 The correct analysis of the experiment

9.1 The solution $\psi_3 = \psi_1 \boxplus \psi_2$ can be adopted as a meaningful spinor field

.

9.1.1 Adopting the combination of binary and ternary logic

[11]

We will now show how we can justify that the algebraic result $\psi_3 = \psi_1 \boxplus \psi_2$ (obtained by using the Huygens' principle, the Born approximation, or the reasoning of Subsection 3.6), and which has no geometrical meaning, can be adopted as a meaningful spinor wave function. We stick thereby to the methodology of fulfilling the logical obligation to validate a solution selected from the pool of purely algebraic solutions of the spinor equation, outlined in Subsections 3.5 and 3.6, and we will rely on the analysis laid down in Eq. 6.

It is the fact that our binary logic tells us that the electron can only have gone through slit S_1 or S_2 (whereby these options are mutually exclusive), which creates the conceptual tension because it clashes intuitively with the factual reality that the coherent interactions have rendered the answer to the "which way" question undecided, by withholding the information. As human beings we want to know better and refuse to bother about what nature knows or otherwise. Furthermore the rule $p_3 = p'_1 + p'_2$ we expect seems to clash with the rule $\psi_3 = \psi_1 \boxplus \psi_2$ provided by the traditional theory.

We will show that the textbook QM prescription $\psi_3 = \psi_1 \boxplus \psi_2$ belongs to absolute geometry in the analogy we discussed above. We accept that the question through which slit the electron has traveled is undecidable and accept ternary logic with its axiomatic system TL. We should then play the game and not attempt in any instance to reason about the question which way the electron has traveled, because this information is not available. The averaging procedure used in the axiomatic system BL can therefore not be used and the values ψ_1 and ψ_2 taken separately cannot be given any physical meaning in the double-slit experiment.

But we can also add a new axiom, the axiom of the existence of a divine perspective, rendering the "which-way" question decidable for a divine observer who can also observe the information withheld by the set-up, without needing to rely on any physical interaction. This way we will have God on our side! We call this the axiomatic system TL+D. This axiomatic system is as close as we can get in reconciling our intuition that a particle has only two mutually exclusive options for traversing the slits of the set-up and the ternary logic imposed by the coherent interactions of the particle with the set-up. The axioms of the axiomatic system TL+D are not contradictory. They are only divine. We must then also play the game and accept the fact that the probabilities we will discuss can no longer be measured, such that the conclusions we draw can no longer be checked by experimental evidence. The probabilities are only accessible mentally through the logic imposed by the addition of the binary axiom of divine knowledge.

In the axiomatic system TL+D we can then take the exact solution ψ_3 of the double-slit experiment and try to determine the parts ψ'_1 and ψ'_2 of it that stem from slits S_1 and S_2. We can mentally imagine such a partition without making a logical error because each electron must go through one of the slits, even if we will never know which one. We obtain then the logical constraints given in Eq. 6 of Subsection 3.5. As already said the probabilities p'_1 and p'_2 are not accessible to experimental measurement, they are just logical consequences of the addition of the binary axiom of the divine observer. We will carry out a post-calculation dissection of ψ_3, calculated with the boundary conditions for the double-slit experiment.[12]

With this dissection procedure we will achieve two goals. We will justify the algebraic solution as geometrically meaningful and we will preserve the traditional rules of probability calculus for mutually exclusive possibilities.

[11] In reference [10], pp. 329-333 and in [3,4], we proposed an analysis based on following the "phase of the wave function" on a loop. This analysis is wrong because the wave function is not everywhere of the type $e^{-\frac{i}{\hbar}(Et-\mathbf{p}\cdot\mathbf{r})}$. When we add ψ_1 and ψ_2 in an interference region, cosine functions can come into play. The wave function can never become zero if we stick to an expression $e^{-\frac{i}{\hbar}(Et-\mathbf{p}\cdot\mathbf{r})}$. The wave function must also become zero on certain boundaries in the proof of the coherent-sum rule given in Subsection 3.6, where this extinction is considered to be the consequence of reflection or absorption.

[12] The argument developed in Subsection 3.4.1 does therefore not hold sway here. In Subsection 3.4.1 we anticipated constructing wave functions ψ'_1 and ψ'_2 from two equations for the double-slit experiment. Equation j for ψ'_j was defined by the boundary condition $\psi'_j(\mathbf{r}) = 0$, $\forall \mathbf{r} \in S_{1+|j-2|}$ and some supplementary conditions. We pointed out that these boundary conditions are in reality closing the slits $S_{1+|j-2|}$, such that they rather define wave function solutions ψ_1 and ψ_2 for single-slit experiments. As these boundary conditions were laid down before we started the calculations, they automatically imposed experimental binary logic, removing experimental undecidability from the equations once and for ever. In the present approach we will try to make a post-calculation dissection of the solution ψ_3 of a single double-slit equation, whose boundary condition allows for undecidability. This boundary condition results in the existence of a zone $\mathscr{N}_1 \cap \mathscr{N}_2$ (see below) where experimental undecidability rules. If we put the detector in this zone the path taken by the particle will be really undecidable. But if we put the detector too close to the slits, outside $\mathscr{N}_1 \cap \mathscr{N}_2$, the undecidability will go away even if both slits are open. Undecidability requires thus more than coherent interactions when both slits are open. The position of the detector plays also a rôle. The position of the detector must

9.2 The proof

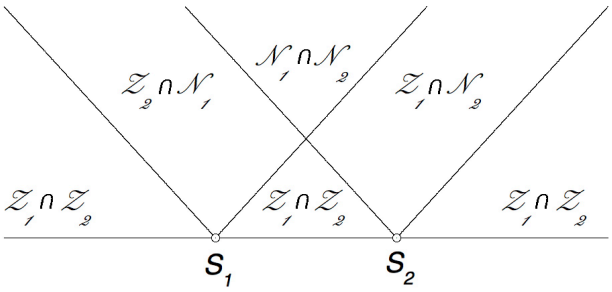

Fig. 1. Schematic diagram illustrating the notations used in the text for the zones of $\mathscr{V}$ considered according to their accessibility to the electrons emerging from the slits S_1 and S_2. The set $\mathscr{N}_1$ is the part of space that can be reached by electrons that traverse S_1. The set $\mathscr{Z}_1$ is the part of space that cannot be reached by these electrons. The conventions for the slit S_2 are analogous. The interference pattern occurs in the zone $\mathscr{N}_1 \cap \mathscr{N}_2$. The shadow zone $\mathscr{Z}_1 \cap \mathscr{Z}_2$, from which the electrons remain absent, consists of three disconnected parts. The drawing is schematic as the slits are represented with zero widths. In the real situation, the widths of the slits are not zero. Note that the regions $\mathscr{N}_j$ are assumed to include the penumbrae where the intensity of the beam emanating from slit S_j varies smoothly from a maximum on a plateau region to zero on the border between $\mathscr{N}_j$ and $\mathscr{Z}_j$.

9.2.1 Conventions and preparations

Let us recall the definitions of $\mathscr{V}$, π_s and π_d from Subsection 3.6. Following the idea that $\psi_1'(\mathbf{r})$ would have to vanish on slit S_2 and $\psi_2'(\mathbf{r})$ on slit S_1, we subdivide $\mathscr{V}$ in a region $\mathscr{Z}_1$ where $\psi_1'(\mathbf{r}) = 0$ and a region $\mathscr{N}_1$ where $\psi_1'(\mathbf{r}) \neq 0$. We define $\mathscr{Z}_2$ and $\mathscr{N}_2$ similarly. These conventions are represented in Fig. 1, which is only schematic. A more detailed partition of $\mathscr{V}$ would allow for penumbral zones $\mathscr{T}_j$ whereover ψ_j are varying smoothly from a maximum value to zero. In the diagram we have simplified the representation by incorporating $\mathscr{T}_j \subset \mathscr{N}_j$ by definition. Within the set of assumptions made in Subsection 3.6 the penumbral zones can be attributed to the finite sizes of the source and the slits. The idea is that in principle all these details have already been settled at the time we have formulated and solved the single-slit Schrödinger equations, thereby imposing that the solutions ψ_j are smooth. It is also understood that these solutions have then been seamlessly combined in the solution of the double-slit Schrödinger equation according to Subsection 3.6. The schematic diagram in Fig. 1 distorts the logic of Subsection 3.6, in that it represents the slits with zero rather than finite widths.

In the region $\mathscr{N}_1 \cap \mathscr{Z}_2$ we can multiply ψ_1' by an arbitrary phase factor $e^{\imath\chi_1}$ without changing $|\psi_1'(\mathbf{r})|^2$. In the region $\mathscr{Z}_1$ this is true as well as $|\psi_1'(\mathbf{r})|^2 = 0$. Similarly, $\psi_2'(\mathbf{r})$ can be multiplied by an arbitrary phase factor $e^{\imath\chi_2}$ in the regions $\mathscr{Z}_2$ and $\mathscr{N}_2 \cap \mathscr{Z}_1$.

Over $\mathscr{N}_1 \cap \mathscr{Z}_2$, we must have $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r})$, as $\psi_2(\mathbf{r}) = 0$. Let us put here $\psi_1'(\mathbf{r}) = C_1\psi_1(\mathbf{r})$. This takes into account that there is also a normalization problem as discussed in Subsection 3.6.3. We will determine the value of C_1

be selected, but we can do this after the calculation by cutting through the wave function with the plane that we want to choose as the detector plane. This procedure, which adds the constraints related to the detector at the end instead of imposing them at the beginning of the calculations is a flexible way to deal with experimental undecidability. It infringes in a sense Bohr's caveat, but in a harmless way. Adding constraints after the calculation is also the only way to enable the intended dissection of ψ_3 by imposing the required unphysical "divine" constraints. We have to proceed this way as the formalism of QM with its initial boundary conditions has of course been designed for physical constraints, not for unphysical "divine" constraints. The results ψ_1' and ψ_2' obtained by the dissection procedure will be the wave functions anticipated in Subsection 3.4.1. The procedure of dissecting ψ_3 after its calculation is the only way to obtain functions ψ_1' and ψ_2' on which we could apply an alleged superposition principle as anticipated in Subsection 3.4.1. As it turns out that $\psi_j' \neq \psi_j$, the superposition principle cannot possibly be used to justify the calculation $\psi_1 \boxplus \psi_2$ used in textbooks.

in Subsection 9.2.3. Similarly, over $\mathscr{N}_2 \cap \mathscr{Z}_1$, we must have $\psi_3(\mathbf{r}) = \psi_2(\mathbf{r})$ as $\psi_1(\mathbf{r}) = 0$, and we put $\psi_2'(\mathbf{r}) = C_2\psi_2(\mathbf{r})$. We will determine the value of C_2 also in Subsection 9.2.3.

Let us now address the region $\mathscr{W} = \mathscr{N}_1 \cap \mathscr{N}_2$. We must certainly have $|\psi_1'(\mathbf{r})|^2 + |\psi_2'(\mathbf{r})|^2 = |\psi_3'(\mathbf{r})|^2$, because the probabilities for going through slit S_1 and for going through slit S_2 are mutually exclusive and must add up to the total probability of transmission. We might have started to doubt about the correctness of this idea, due to the way textbooks present the problem, but we should never have doubted. Let us thus put $\psi_1'(\mathbf{r}) = |\psi_3(\mathbf{r})| \cos\alpha\, e^{\imath\alpha_1}$, $\psi_2'(\mathbf{r}) = |\psi_3(\mathbf{r})| \sin\alpha\, e^{\imath\alpha_2}$, $\forall \mathbf{r} \in \mathscr{N}_1 \cap \mathscr{N}_2 = \mathscr{W}$. In fact, if ψ_j' is a partial solution for the slit S_j, $\psi_j' e^{\imath\alpha_j}$ will also be a partial solution for the slit S_j. We must take here α, α_1 and α_2 as constants. If we took a solution whereby α, α_1 and α_2 were varying functions of $\mathbf{r}$, the result obtained would no longer be a solution of the Schrödinger equation in free space, due to the terms containing the spatial derivatives of α, α_1 and α_2 which are then no longer zero.

In first instance this argument shows also that we must take $\chi_1 = \alpha_1$ and $\chi_2 = \alpha_2$ in order to keep α_1 and α_2 constant everywhere. The original idea that is does not matter what value we pick for χ_1 in $\mathscr{Z}_2 \cap \mathscr{N}_1$ and for χ_2 in $\mathscr{Z}_1 \cap \mathscr{N}_2$ must thus be revised. This is because we do not have to care about the phases in the single slit experiments but this changes in the double-slit experiment because we must make things work out globally.

We are thus forced to adjust our choices for χ_1 and χ_2 to those for $\alpha_1 \neq 0$, $\alpha_2 \neq 0$. Therefore, the choices $\chi_1 \neq 0$, $\chi_2 \neq 0$ we have to impose on the phases, embody the idea that not every solution of a Schrödinger equation with potential V_j, for $j = 1, 2$ can be considered as an obvious contribution to a Schrödinger equation with potential V_3. The conditions we have to impose on α_1 and α_2 are thus a kind of disguised boundary conditions. They are not true boundary boundaries, but a supplementary condition (a logical constraint) for the surgery we want to carry out on ψ_3 to obtain the functions ψ_1' and ψ_2' which must obey "divine" binary logic.

9.2.2 The interference region

Over $\mathscr{W} = \mathscr{N}_1 \cap \mathscr{N}_2$ integration leads to $\int_{\mathscr{W}} |\psi_1'(\mathbf{r})|^2\, d\mathbf{r} = \cos^2\alpha \int_{\mathscr{W}} |\psi_3(\mathbf{r})|^2\, d\mathbf{r}$ and $\int_{\mathscr{W}} |\psi_2'(\mathbf{r})|^2\, d\mathbf{r} = \sin^2\alpha \int_{\mathscr{W}} |\psi_3(\mathbf{r})|^2\, d\mathbf{r}$. Furthermore, we must have $\int_{\mathscr{W}} |\psi_2'(\mathbf{r})|^2\, d\mathbf{r} = \int_{\mathscr{W}} |\psi_1'(\mathbf{r})|^2\, d\mathbf{r}$, due to the mirror symmetry of the set-up, such that $\alpha = \frac{\pi}{4}$. We see from this that not only $\int_{\mathscr{W}} |\psi_2'(\mathbf{r})|^2\, d\mathbf{r} = \int_{\mathscr{W}} |\psi_1'(\mathbf{r})|^2\, d\mathbf{r} = \frac{1}{2}\int_{\mathscr{W}} |\psi_3(\mathbf{r})|^2\, d\mathbf{r}$, but also $|\psi_2'(\mathbf{r})|^2 = |\psi_1'(\mathbf{r})|^2 = \frac{1}{2}|\psi_3(\mathbf{r})|^2$. In each point $\mathbf{r} \in \mathscr{W}$ the probability that the electron has traveled through a slit to get to $\mathbf{r}$ is equal to the probability that it has traveled through the other slit. This is due to the undecidability.

We can summarize these results as $\psi_1' = \frac{1}{\sqrt{2}}|\psi_1 \boxplus \psi_2|\, e^{\imath\alpha_1}$ and $\psi_2' = \frac{1}{\sqrt{2}}|\psi_1 \boxplus \psi_2|\, e^{\imath\alpha_2}$. Let us write $\psi_1 \boxplus \psi_2 = |\psi_1 \boxplus \psi_2|\, e^{\imath\chi}$. We can now calculate α_1 and α_2 by identification. This yields on $\mathscr{W}$:

$$\begin{aligned} \psi_1' &= \frac{|\psi_1 \boxplus \psi_2|\, e^{\imath(\chi + \frac{\pi}{4})}}{\sqrt{2}} = \frac{1}{\sqrt{2}}(\psi_1 \boxplus \psi_2)\, e^{+\imath\frac{\pi}{4}} \neq \psi_1 \\ \psi_2' &= \frac{|\psi_1 \boxplus \psi_2|\, e^{\imath(\chi - \frac{\pi}{4})}}{\sqrt{2}} = \frac{1}{\sqrt{2}}(\psi_1 \boxplus \psi_2)\, e^{-\imath\frac{\pi}{4}} \neq \psi_2 \end{aligned} \tag{21}$$

9.2.3 The other regions

We must still calculate C_1 and C_2. We will base this calculation on the postulate that the solution should be devoid of any discontinuities at the boundaries of $\mathscr{N}_1 \cap \mathscr{Z}_2$ and $\mathscr{N}_2 \cap \mathscr{Z}_1$ with $\mathscr{W}$. In $\mathscr{N}_1 \cap \mathscr{Z}_2$, we have $\psi_1'(\mathbf{r}) = C_1\psi_1(\mathbf{r})$, while in $\mathscr{N}_1 \cap \mathscr{N}_2$, we have $\psi_1'(\mathbf{r}) = \frac{1}{\sqrt{2}}(\psi_1(\mathbf{r}) \boxplus \psi_2(\mathbf{r}))\, e^{+\imath\frac{\pi}{4}}$. On the boundary between $\mathscr{N}_1 \cap \mathscr{Z}_2$ and $\mathscr{N}_1 \cap \mathscr{N}_2$ we must have $\psi_2(\mathbf{r}) = 0$ if ψ_2 is continuous, such that then $\psi_1'(\mathbf{r}) = \frac{1}{\sqrt{2}}\psi_1(\mathbf{r})\, e^{+\imath\frac{\pi}{4}}$. This shows that we must take $C_1 = \frac{1}{\sqrt{2}}\, e^{+\imath\frac{\pi}{4}}$. In other words, in $\mathscr{N}_1 \cap \mathscr{Z}_2$ we obtain the expression for $\psi_1'(\mathbf{r})$ by just extending $\psi_1'(\mathbf{r}) = \frac{1}{\sqrt{2}}(\psi_1(\mathbf{r}) \boxplus \psi_2(\mathbf{r}))\, e^{+\imath\frac{\pi}{4}}$ from $\mathscr{N}_1 \cap \mathscr{N}_2$ to $\mathscr{N}_1 \cap \mathscr{Z}_2$ and expressing the fact that $\psi_2(\mathbf{r}) = 0$. It is thus just a special case of the expression $\psi_1'(\mathbf{r}) = \frac{1}{\sqrt{2}}(\psi_1(\mathbf{r}) \boxplus \psi_2(\mathbf{r}))\, e^{+\imath\frac{\pi}{4}}$ that becomes now generally valid on $\mathscr{N}_1$. By considering the boundary between $\mathscr{N}_2 \cap \mathscr{Z}_1$ and $\mathscr{N}_1 \cap \mathscr{N}_2$, we can derive in a completely analogous way that $C_2 = \frac{1}{\sqrt{2}}\, e^{-\imath\frac{\pi}{4}}$. The expression $\psi_2'(\mathbf{r}) = \frac{1}{\sqrt{2}}(\psi_1(\mathbf{r}) \boxplus \psi_2(\mathbf{r}))\, e^{-\imath\frac{\pi}{4}}$ becomes this way generally valid on $\mathscr{N}_2$. The wave functions ψ_j' and ψ_3 are continuous when ψ_j are continuous. In summary, the dissection operator $X : (\psi_1, \psi_2) \to X(\psi_1, \psi_2) = (\psi_1', \psi_2')$ is continuous and it has the same general algebraic expression given by Eq. 21 on all subsets of the partition of $\mathscr{V}$.

9.2.4 Epilogue

The regions behind the slits where $\psi_3(\mathbf{r}) = 0$ must be excluded from the definition domain of ψ_3 because $\psi_3(\mathbf{r}) = 0$ can never represent a spinor. This defines on the region $\mathscr{W} \subset \mathbb{R}^3$ of space behind the slits open domains $D_n, n \in \mathbb{Z}$,

whereby on each of these domains $|\psi_3| > 0$ and ψ_3 is meaningfully defined. The restrictions of ψ_3 to the domains D_n are the true solutions of the wave equation and they must be summed incoherently.

What Eq. 21 shows is that the rule $\psi_3 = \psi_1 \boxplus \psi_2$ is physically meaningless, because it violates Bohr's caveat (by using ψ_1 and ψ_2) and because the correct expression in the only axiomatic system TL+D where a sum can be written is $\psi_3 = \psi_1' + \psi_2'$. This shows how the member $\psi_1 \boxplus \psi_2$ from the pool of solutions obtained by coherent summing can be justified as meaningful within the context of spinors or other group representations. Moreover, it soothes our intuition by allowing for the binary logic of divine knowledge combined with the ternary logic imposed by the absence of information left behind by coherent interactions. This way we have accomplished the task defined in Subsection 3.5 of justifying the use of ψ_3.

The differences between ψ_j' and ψ_j, for $j = 1, 2$, are not negligible. The phases of ψ_1' and ψ_2' always differ by $\frac{\pi}{2}$ such that they are fully correlated. The difference between the phases of ψ_1 and ψ_2 varies, whereby these phases can be opposite (destructive interference) or identical (constructive interference). In contrast to ψ_1 and ψ_2, ψ_1' and ψ_2' reproduce the oscillations of the interference pattern, reflecting the lack of information. In this respect, the fact that α_1 and α_2 are different by a fixed amount is crucial. It permits to make up for the normalization factor $\frac{1}{\sqrt{2}}$ and end up with the correct numerical result of the flawed calculation $\psi_1 \boxplus \psi_2$. The phases of ψ_1' and ψ_2' conspire to render $\psi_1' + \psi_2'$ equal to $\psi_1 \boxplus \psi_2$. Note that we already obtained the normalization factor $\frac{1}{\sqrt{2}}$ in Subsection 3.6.

9.3 Analyzing the experiment according to Bohr and Einstein

The two approaches TL+D and TL correspond to Einstein-like and Bohr-like viewpoints respectively. Perhaps we are cheating somewhat in laying down this claim and awarding too much credit to Einstein, because Einstein may just have been thinking within the axiomatic system BL, even if he was close friends with Gödel in Princeton. The axiomatic system TL+D can then be seen as a third way that improves both Einstein's and Bohr's axiomatic systems by cherrypicking from them those axioms we consider as pertinent. Anyway, we will refer in the following to the axiomatic system TL+D by calling it Einstein's viewpoint. Bohr would just claim that the probabilities calculated in Eq. 21 do not exist, while Einstein would claim they do exist. Both approaches are logically tenable when the detector screen is completely in the zone $\mathscr{W}$, because the quantities in $\mathscr{N}_1 \cap \mathscr{Z}_2$ and $\mathscr{N}_2 \cap \mathscr{Z}_1$ are then not measured quantities. Of course we could try to measure them by putting the detector screen closer to the slits, but in Bohr's view this would be a different experiment and violate his caveat. Of course it is harmless, because it does not change the probabilities between the slits and the position chosen for the detector, as we have pointed out.

In analyzing the results from Feynman's path integral method, one should recover in principle the same results. However, the pitfall is here that one might too quickly conclude that $\psi_j' = \psi_j$, which leads us straight into the paradox. We see thus that the Huygens' principle and coherent summing in general are purely numerical recipes that are physically meaningless, because they search for a correct global solution without caring about the correctness of the partial solutions. They follow the experimental ternary logic of TL and therefore are allowed to mistreat the phase difference that exists between the partial solutions ψ_1' and ψ_2' obtained in TL+D, by just bluntly using ψ_1 and ψ_2 without bothering to give any meaning to these values, because any form of splitting up ψ_3 must be considered as meaningless. The rule $\psi_3 = \psi_1 \boxplus \psi_2$, whereby the phase difference between ψ_1 and ψ_2 can vary, is thus perfectly acceptable in ternary logic. This corresponds to Bohr's viewpoint, who uses the axiomatic set TL. It is tenable because it will not be contradicted by experiment. It is not the theory based on the axiomatic set TL which is incomplete, it is the information contained in the data it must describe that is "incomplete" due to the coherent processes.

This changes if one wants to impose also binary logic on the wave function and use Einstein's set of axioms TL+D, arguing that we know that the particle can only go through slit S_1 or through slit S_2 and that these options are mutually exclusive. Therefore the question through which the slit has traveled is decidable from the perspective of a divine observer who could see what happens without interaction. We do need then a correct decomposition $\psi_3 = \psi_1' + \psi_2'$. We then find out that $\psi_j' \neq \psi_j$ and we can attribute this change between the single-slit and the double-slit probabilities to the difference between the ways we must define probabilities in both types of logic. The partial solutions ψ_1' and ψ_2' have then always the same phase difference, which prevents them from interfering destructively, because interference is a meaningless concept for group representations in general. If we were able by divine knowledge to assign to each electron impact on the detector the corresponding number of the slit through which the electron has traveled, we would recover the experimental frequencies $|\psi_j'|^2$. This is Einstein's viewpoint, who uses TL+D.

Having made this difference clear, everybody is free to decide for himself if he prefers to study the analogue of geometry in TL+D or the analogue of absolute geometry in TL. But refusing Einstein's axiomatic system TL+D based on the argument that ψ_1' and ψ_2' cannot be measured appears to us a stronger and more frustrating *Ansatz* than accepting the introduction of ψ_1' and ψ_2', despite the fact that they cannot be measured. This is because refusing Einstein's axiomatic system TL+D in favour of Bohr's axiomatic system TL comes down to denying that the particle has only two mutually exclusive options: traveling through S_1 or through S_2. That is hard to accept, because it just does not agree with our macroscopic intuition and - much more seriously - with the way we can give a meaning to sums of spinors in terms of sets.

The refusal is of course in direct line with Heisenberg's initial program of removing from the theory all quantities that cannot be measured. However, Heisenberg's program is violated by the very wave function $\psi_3(\mathbf{r})$, which is claimed to describe probabilities for all $\mathbf{r} \in \mathscr{V}$ and therefore provides also probabilities for observing the particles in the space between the slits and the detectors. These probabilities are never measured in the set-up chosen. Of course we could put the detector screen closer to the slits and this does not change the probabilities between the slit and the new position of the detector, but *sensu stricto* we are this way taking exception with Bohr's caveat.

It is Heisenberg's minimalism which preserves the experimental undecidability and ternary logic within the theory. We can afford being less strict by adding a supplementary logical constraint which preserves our binary logic about the "which way" question, by adopting the axiomatic system TL+D. As Eq. 21 shows, the difference between the fake partial ternary solutions ψ_j of TL and the correct partial ternary solutions ψ'_j (where $j = 1, 2$) of TL+D is much larger than we ever might have expected on the basis of the logical loophole that $\psi_3 = \psi_1 \boxplus \psi_2$ is not rigorously exact.

The bias in the experimentally measured probabilities due to the undecidability cannot be removed by the divine knowledge about the history. If we want a set-up with the boundary conditions that correspond to the unbiased case whereby $\psi'_1(\mathbf{r}) = 0$ on slit S_2 and $\psi'_2(\mathbf{r}) = 0$ on slit S_1, we must assume that the detector screen is put immediately behind the slits, and the interference pattern can then not be measured, while everything becomes experimentally decidable. We are then in the axiomatic system BL. Otherwise, we must assume that $\psi'_1(\mathbf{r})$ and $\psi'_2(\mathbf{r})$ are not measured between the detector and the slits such that they can satisfy the undecidable solution in Eq. 21. The partial probabilities given by $\psi'_1(\mathbf{r})$ and $\psi'_2(\mathbf{r})$ are thus extrapolated quantities. But in the pure Heisenberg approach whereby one postulates that we are not allowed to ask through which slit the electron has traveled, the wave function contains also extrapolated quantities that are not measured, despite the original agenda of that approach.

9.4 The discussions of Einstein and Bohr - who is right and who is wrong?

Bohr's and Einstein's viewpoints correspond to different axiomatic systems and both agree with experiment. Both systems are internally consistent and contradiction-free. It is therefore *logically flawed* to promote one axiomatic system to a touchstone of absolute truth and to attack the other axiomatic system from such a self-appointed stronghold. That would be like attacking hyperbolic geometry by pointing out that it is in contradiction with Euclidean geometry or, even more accurately, like attacking both Euclidean and hyperbolic geometry because your pet axiomatic system is that of absolute geometry, wherein the question whether the parallels postulate is true or otherwise is kept undecided. This attack would then promote the axiomatic system of absolute geometry to an absolute touchstone. It is thus logically flawed to attack Einstein's viewpoint by arguing that it cannot be tested, because that is *arbitrarily* promoting Bohr's axiomatic system, which refuses to consider quantities that cannot be measured, to such a touchstone. By discussing the arguments of Einstein and Bohr in terms of axiomatic systems in a metamathematical approach we can avoid polemics, such that the feathers can settle.

The extremism of Bohr's postulate could be rebutted by pointing out that it is violated by its very own formalism, because it considers also the quantity $|\psi(\mathbf{r}, t)|^2$ for values of $(\mathbf{r}, t)$ where it cannot be measured, such that the approach is wrong by its very own standards. Bohr could save his viewpoint by rejecting the use of the wave function over the part of $\mathbb{R}^3$ outside the detector screen. After all, the wave function is only a mathematical tool, not a physical wave, as we have shown in [1]. However, Bohr might have believed that ψ_3 is truly a matter wave. But it is not a matter wave and there is no justification for summing ψ_1 and ψ_2 other than given in Subsections 3.5 and 3.6 which have no physical meaning. The axiomatic system TL shows thus little cracks, because Bohr was forced to reason on equations that were guessed rather than derived.

We also see that there is *always* an additional logical constraint that must be added in TL+D in order to account for the fact that the options of traveling through the slits S_1 or S_2 are mutually exclusive. This has been systematically overlooked, with the consequence that one obtains the result $p \neq p_1 + p_2$, which is impossible to make sense of. It is certainly not justified to use $p \neq p_1 + p_2$ as a starting basis for raising philosophical issues. As we explained, the solution of the paradoxes of QM is not a matter of epistemology, but a matter of pure logic and mathematical rigour.

In summary, $\psi_3 = \psi_1 \boxplus \psi_2$ is wrong if we cheat by wanting to satisfy also binary logic in the analysis of an experiment that follows the axiomatic system TL by attributing meaning to ψ_1 and ψ_2. But it yields the correct numerical result for the total wave function if we play the game and respect the empirical undecidability by not asking which way the particle has traveled. We can thus only uphold that the textbook rule $\psi_3 = \psi_1 \boxplus \psi_2$ is correct if we accept that the double-slit experiment experimentally follows the axiomatic system TL. From the viewpoint of the axiomatic system TL+D, the agreement of the numerical result $\psi_1 \boxplus \psi_2$ with the experimental data is misleading, as such an agreement does not provide a watertight proof for the correctness of a theory. If a theory contains logical and mathematical flaws, then it must be wrong despite its agreement with experimental data.

In the axiomatic system TL+D, Eq. 21 shows that interference does not exist, because the phase factors $e^{\imath\frac{\pi}{4}}$ and $e^{-\imath\frac{\pi}{4}}$ of ψ'_1 and ψ'_2 always add up to $\sqrt{2}$. We may note in this respect that the phase χ itself is only determined up to an arbitrary constant within the experiment. The wave function can thus not become zero due to phase differences between ψ'_1 and ψ'_2 like happens with ψ_1 and ψ_2 in $\psi_1 \boxplus \psi_2$. When ψ_3 is zero, both ψ'_1 and ψ'_2 are zero. Interference

thus only exists within the purely numerical, virtual reality of the Huygens' principle or coherent summing in general, which are not narratives of the real world. We must thus not only dispose of the particle-wave duality, but we must also be very wary of the wave pictures we build based on the intuition we gain from experiments in water tanks. These pictures are apt to conjure up a very misleading imagery that leads to fake conceptual problems and stirs a lot of confusion. The phase of the wave function has physical meaning, and group representations can only be added purely algebraically in a meaningful way if we get their phases right.

We think that Einstein's axiomatic system TL+D is a better axiomatic system for normal human beings than Bohr's axiomatic system TL because it respects our binary intuition and this way does not lead to a perceived paradox. Its construction of the wave function respects the meaning of spinors and is clear about what has physical meaning or otherwise. It permits to analyze the situation according to the logic of human beings. The fast-lane, brute-force rule $\psi_3 = \psi_1 \boxplus \psi_2$ of the traditional approach is logically of an unfazed, brazen mathematical clumsiness, to the point that it becomes completely unintelligible to human beings, even if its algebra agrees with the experimental results. It has even led to the firm conviction that the double-slit experiment is full of quantum magic that nobody can understand.

9.5 More formal presentation

9.5.1 Criterium for binary logic - orthogonality

Two complex numbers ζ_1 and ζ_2 are orthogonal if:[13]

$$\zeta_1^* \zeta_2 + \zeta_1 \zeta_2^* = 0 \tag{22}$$

The orthogonality condition for two complex numbers ζ_1 and ζ_2 implies that:

$$|\zeta_1 + \zeta_2|^2 = |\zeta_1|^2 + |\zeta_2|^2. \tag{23}$$

such that in a point $\mathbf{r} \in \mathscr{V}$ the orthogonality condition for $\psi_1'(\mathbf{r})$ and $\psi_2'(\mathbf{r})$ expresses that they are describing mutually exclusive probabilities, according to binary logic: $p_3 = p_1' + p_2'$.

Let us define for two complex numbers ζ_1 and ζ_2, $\zeta_1 = |\zeta_1| e^{\imath \alpha_1}$, $\zeta_2 = |\zeta_2| e^{\imath \alpha_2}$. The orthogonality implies then that:

$$\alpha_1 - \alpha_2 = \pm \frac{\pi}{2} (\text{mod } 2\pi). \tag{24}$$

Of course when $\zeta_1 = 0$ then there is no constraint on $\alpha_1 - \alpha_2$ because ζ_1 and ζ_2 are then automatically orthogonal. The same is true *mutatis mutandis* when $\zeta_2 = 0$.

9.5.2 Condition for undecidability - Coherence-induced symmetry

For a wave function ζ constructed from two spinor wave functions ζ_1 and ζ_2 to be undecidable in some point $\mathbf{r}$ we must use a completely symmetrical expression:

$$\zeta(\mathbf{r}) = \zeta_1(\mathbf{r}) \boxplus \zeta_2(\mathbf{r}). \tag{25}$$

This does not use the Huygens' principle. It is a pure symmetry argument (The set-up is completely symmetrical. Incoherent interactions break this symmetry, coherent interactions preserve this symmetry). This justifies the introduction of sums of wave functions $\zeta_1 \boxplus \zeta_2$ based on Subsection 3.6. It is didactically superior because it provides an explicit direct link with undecidability, while in the two other ways this undecidability is introduced tacitly and implicitly. It is also superior because it does not rely on physical approximations. We must nevertheless use the symbol $\boxplus$ when we use spinors, because the sum of two spinors is *a priori* not defined. Furthermore we must have:

$$|\zeta_1(\mathbf{r})| = |\zeta_2(\mathbf{r})|. \tag{26}$$

In fact, an information of the type $|\zeta_1(\mathbf{r})| > |\zeta_2(\mathbf{r})|$ would make it more likely that the particle in $\mathbf{r}$ has gone through slit S_1. This would violate the absolute undecidability induced by the coherent interactions.

[13] The definition of the Hermitian scalar product $\zeta_1 \cdot \zeta_2 = \frac{1}{2}[\,\zeta_1^* \zeta_2 + \zeta_1 \zeta_2^*\,]$ for two complex numbers ζ_1 and ζ_2 is completely analogous to the definition of the Euclidean scalar product for two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. One can just use the analogue of $(\mathbf{r}_1 + \mathbf{r}_2)^2 - \mathbf{r}_1^2 - \mathbf{r}_2^2 = 2\,(\mathbf{r}_1 \cdot \mathbf{r}_2)$ as a definition for $\mathbf{r}_1 \cdot \mathbf{r}_2$. Orthogonality is then defined by $\zeta_1 \cdot \zeta_2 = 0$ in complete analogy with $\mathbf{r}_1 \cdot \mathbf{r}_2 = 0$. For spinors, we can use the same methodology. As the norm of a spinor ζ is now given by $\zeta^\dagger \zeta$, the scalar product would just be: $\zeta_1 \cdot \zeta_2 = \frac{1}{2}[\,\zeta_1^\dagger \zeta_2 + \zeta_2^\dagger \zeta_1\,]$.

9.5.3 Combined ternary and binary logic

When we want to satisfy both binary and ternary logic simultaneously, we must satisfy both Eqs. 26 and 24, such that:

$$\zeta_1 = \zeta_2 e^{\pm \imath\pi/2}. \tag{27}$$

Furthermore, Eqs. 23 and 26 imply:

$$|\zeta_1| = |\zeta_2| = |\zeta_1 + \zeta_2|/\sqrt{2}. \tag{28}$$

Combining Eqs. 27 and 28 we obtain then Eq. 21.

9.5.4 Construction of a wave function

We are now going to construct by surgery two partial solutions ζ_1 and ζ_2 of the Schrödinger equation for the double-slit experiment. We do not assume interference. We will impose conditions on ζ_1 and ζ_2, hoping that solutions ζ_1 and ζ_2 which meet these criteria will exist.

We consider two planes, the plane of the set-up π_s and the plane π_d of the detector screen, as defined in Subsection 3.6. When the strong orthogonality condition is fulfilled by ζ_1 and ζ_2 for the points $\mathbf{r} \in \pi_d \cup \pi_s$ then the restriction of the sum $\zeta_1 + \zeta_2$ to $\pi_d \cup \pi_s$ will be a meaningful sum of spinor functions, which can be interpreted in terms of sets, because ζ_1 and ζ_2 are then representing mutually exclusive probabilities.

Normally, we impose an orthogonality condition $\int \zeta_1^*(\mathbf{r})\zeta_2(\mathbf{r})\, d\mathbf{r} = 0$ (implying automatically $\int [\, \zeta_1^*(\mathbf{r})\zeta_2(\mathbf{r}) + \zeta_1(\mathbf{r})\psi_2'^*(\mathbf{r})\,]\, d\mathbf{r} = 0$), which is a much weaker criterion for considering linear combinations $c_1\zeta_1 + c_2\zeta_2$ as acceptable mixed states. This is why we can call the orthogonality condition which implies binary logic on a whole set now a strong orthogonality condition. On function spaces we can have strong and weak convergence. Here we discover strong and weak orthogonality.

On π_s we satisfy the strong orthogonality condition by just requiring that ζ_1 is zero on slit S_2 and ζ_2 is zero on slit S_1. On the plane π_s of the slits, the question through which slit the particle travels is then decidable for both ζ_1 and ζ_2, such that the axiom of the divine observer of the axiomatic system TL+D is satisfied. If such solutions exist, they represent partial solutions (obtained by surgery) for the double-slit Schrödinger equation whereby the particle is only allowed to travel through a single slit.

We only add the condition of undecidability on the plane π_d of the detector screen because it is only at the detector that we want the "which way" question to be completely undecidable. This takes then into account that the interactions with the set-up have been coherent, such that we should satisfy ternary logic.

This way all axioms of TL+D will be satisfied by this combination of the conditions on $\pi_d \cup \pi_s$. We may note that we can then write Eq. 25 with the normal symbol + rather than with the symbol $\boxplus$, because we have established that $\zeta_1 + \zeta_2$ is now a meaningful sum of spinors. We obtain then again Eq. 21 where we can replace the symbol $\boxplus$ by the symbol +. The difference is just due to the different order in which we have imposed the binary and ternary conditions and the solution is this time only established on $\pi_d \subset \mathscr{W}$. As an afterthought we can see that the solutions of Eq. 21 are indeed strongly orthogonal. However, in deriving Eq. 21 we have this time not used the unphysical Huygens' principle and the sum which occurs here is physically meaningful. It corresponds to a mere juxtaposition of strongly orthogonal spinor functions which defines a set, whereby the spinor functions must be summed incoherently. It does not correspond to the meaningless operation $\boxplus$ of the Huygens' principle because we have respected the fact that ζ_1 and ζ_2 are spinors.

9.6 Mixtures of coherent and incoherent scattering

This formal approach just expresses the axioms of TL+D, nothing more and nothing less. It directly links interference with undecidability, and binary logic with strong orthogonality. In a sense we have defined a set $\{\psi_1', \psi_2'\}$ that contains two elements but that has to be considered as a set with a single element $\{\psi_3\}$ because we cannot distinguish the two elements one from another, an object with some traits of the Holy Trinity after all (but without mystery).

Of course we can ask now to extend the wave function we found on $\pi_s \cup \pi_d$ to $\mathscr{V}$. That would require imposing strong orthogonality over whole $\mathscr{V}$ and raise the question how we make the transition from decidability on π_s to undecidability on π_d. In Subsections 9.1 and 9.2.4 we have solved that problem for a fully coherent wave function whereby the solution $\psi_3 = \psi_1 + \psi_2$ can be considered as proved in Subsection 3.6. Despite the coherence, there are zones $\mathscr{Z}_1$ and $\mathscr{Z}_2$ where we can still answer the question through which slit the particle has traveled. But, we have shown that there are general expressions for ψ_j' which are equally valid on the zones $\mathscr{Z}_j$ where we can answer that question and the zones $\mathscr{N}_j$ where we cannot answer it.

For purely coherent scattering, what we have done is therefore universally valid over whole $\mathscr{V}$. However, the general solution in $\mathbf{r} \in \mathscr{V}$ would be of the type $\lambda_1(\mathbf{r})p_1(\mathbf{r}) + \lambda_2(\mathbf{r})p_2(\mathbf{r}) + \lambda_3(\mathbf{r})p_3$, whereby $(\forall j \in \{1,3\}, \lambda_j(\mathbf{r}) \in [0,1])$ & $\lambda_1(\mathbf{r}) + \lambda_2(\mathbf{r}) + \lambda_3(\mathbf{r}) = 1$, in order to allow for a simultaneous occurrence of coherent scattering taken into account by $\lambda_3(\mathbf{r})$ and incoherent scattering taken into account by $\lambda_1(\mathbf{r})$ and $\lambda_2(\mathbf{r})$. This general case should in principle be able to deal with all more elaborate versions of the double-slit experiment that fiddle with the knowledge we can acquire about the answer to the "which-way" question. But deriving exactly from the experimental details of a set-up of such a more complex type how the coefficients $\lambda_j(\mathbf{r})$ will depend on $\mathbf{r}$ for that set-up could be very difficult if not impossible. The general case is not treated in textbooks, which just consider purely coherent scattering. This purely coherent scattering evolves from a decided situation at the slits (the near-field regime) to an undecided situation far behind the slits (the far-field regime), while for incoherent scattering the situation is always decided. The general theory of neutron scattering deals with such mixtures of ternary and binary logic by using coherent and incoherent scattering lengths.

9.7 The boundary conditions are the only means to correctly take into account the conditional probabilities

The double-slit solutions ζ_j are presumably solutions for the single-slit Schrödinger equations for ψ_j, but they are not physical because ζ_1 satisfies a strong orthogonality condition with respect to ζ_2 which is not imposed by the boundary conditions of the single-slit experiment described by ψ_1. The same is true *mutatis mutandis* for ζ_2. If we changed the distance d between the slits a bit, the strong orthogonality condition $\zeta_1 \cdot \zeta_2 = 0$ would be different. The strong orthogonality that will have to be met in the double-slit experiment with its specific choice for d can thus not be anticipated in the single-experiments. Furthermore, ζ_1 and ζ_2 satisfy a condition of undecidability which is also unphysical for the single-slit Schrödinger equations for ψ_j. Undecidability is not imposed by the boundary conditions of the single-slit experiments described by ψ_1 and ψ_2.

The other way around there is also incompatibility. The single-slit conditional probabilities p_1 and p_2 are different from $p_1' = |\zeta_1|^2$ and $p_2' = |\zeta_2|^2$, because the latter satisfy supplementary constraints that are completely unphysical for the single-slit experiments. The single-slit conditional probabilities p_1 and p_2 are therefore deemed to be useless for the double-slit experiment. It is just impossible to prepare ψ_1 and ψ_2 in the single-slit experiments for the requirements to be imposed on the conditional probabilities by the double-slit experiment. These conditions are irrelevant for the single-slit experiments and in a single-slit experiment we just cannot know that somebody will consider the experiment within the framework of a discussion about a double-slit experiment.

We really see here Bohr's caveat at work. One cannot transpose conditional probabilities from one set-up to another set-up. Ignore his warning based on your intuition and you are bound to make subliminal errors, even if you think that you are way too clever to make mistakes! The distance d and the undecidability are supplementary boundary conditions in the double-slit experiment, which are not present in the single-slit experiments. The way we learn to solve Schrödinger equations makes us take into account all the subtleties in these boundary conditions automatically. Without Feynman's analysis we might even have not been aware of the existence of the issue of undecidability. But even if we overlook some unsuspected features which intervene in the definition of the conditional probabilities, they will be introduced unwittingly by following the algebraic procedures, such that we get the definitions right anyway, in spite of ourselves. This is the reason why we should take Bohr's caveat seriously and rely on the QM formalism with its boundary conditions rather than on some casual "physical intuition". It is a safeguard that protects us against making subliminal errors. But if we want to learn to make sense of this algebra we must sharpen our awareness of the unsuspected exotic aspects that may intervene in the definition of the conditional probabilities. This is why we have started our journey by drawing up an inventory of such aspects in Subsection 1.4.

9.8 Final remarks

The true reason why we can calculate $\psi_3 = \psi_1' + \psi_2'$ as $\psi_3 = \psi_1 \boxplus \psi_2$ can within the Born approximation also be explained by the linearity of the Fourier transform used in Eq. 20, which is a better argument than wrongly invoking the linearity of the wave equation. The path integral is just a more refined approach, based on a more refined integral transform.

The reason for the presence of the Fourier transform in the formalism is the fact that the electron spins [1,2, 10]. One can derive the whole wave formalism purely classically, just from the assumption that the electron spins as we have shown by deriving the Dirac equation from scratch in [1,10]. Eq. 20 relies also crucially on the Born rule $p = |\psi|^2$. There is no universal proof for this rule. We can only give proofs on a case-by-case basis, although the rule is just taken for granted in the Hilbert space formulation of QM. For photons we can use the total energy of a monochromatic photon beam and divide the result by the energies $h\nu$ of the individual photons. For electrons we can use the argument given in [1] that each electron carries a spinor ψ for which $\psi^\dagger\psi = 1$. The undecidability is completely due to the properties of the potential which defines both the local interactions and the global symmetry.

10 Synopsis: why is it compelling?

It is perhaps instrumental to provide a general overview of the mathematical construction to show how it is actually obtained by deductive reasoning from some strong anchoring points and how the various pieces of the puzzle fall into place within the globally consistent scheme.[14]

[1] As single electrons are detected as particles, they must be always particles and therefore also traverse the slits as particles. The interactions of the particles with the measuring device must be local. To understand the probability distributions observed one must therefore conclude that they are globally defined, despite the fact that the interaction probabilities are local. They must be conditioned by the global set-up, expressed by the boundary conditions. Particles cannot be waves. They have individual phases which are clocks for their periodic internal dynamics. What behaves then as a wave flowing through the two slits is an imaginary infinitely dense liquid of virtual particles represented by the wave function. A large but finite sample of particles taken from this liquid models the physical reality of the many particles which build the interference pattern in an experiment. It is due to the fact that the interference pattern is created by many particles, which interact with different parts of the set-up, that the probabilities can be global rather than local. We can have the particles travel through the set-up one by one such that there is absolutely no interaction between the real particles or the virtual particles of the infinitely dense liquid.

[2] For reasonable paths of real single particles, going through slit S_1 and going through slit S_2 are mutually exclusive possibilities, even when we cannot possibly know through which slit the particle has travelled due to the coherence of its interactions. Therefore we intuitively expect that $p_3(\mathbf{r}) = p_1(\mathbf{r}) + p_2(\mathbf{r})$ must be true.

[3] But the interference pattern is explained theoretically by the calculation $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r})$. In the case of destructive interference this leads to a contradiction because $\psi_1(\mathbf{r}) = -\psi_2(\mathbf{r}) \neq 0$ simultaneously implies $p_3(\mathbf{r}) = p_1(\mathbf{r}) + p_2(\mathbf{r}) > 0$ and $p_3(\mathbf{r}) = 0$ (as a consequence of $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) = 0$). This contradiction can be made very poignant by allowing for only particle at the time within the set-up, with large time intervals in between, e.g. one quarter of an hour.

[4] The only way out is to forbid the calculations $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) = 0$ (which leads to $\psi_1(\mathbf{r}) = -\psi_2(\mathbf{r}) \neq 0$) as algebra without physical meaning, and replace it by $\psi_3(\mathbf{r}) = \psi'_1(\mathbf{r}) + \psi'_2(\mathbf{r}) = 0$. The case of destructive interference can then be explained by $\psi'_1(\mathbf{r}) = -\psi'_2(\mathbf{r}) = 0$ leading to $p_3(\mathbf{r}) = p'_1(\mathbf{r}) + p'_2(\mathbf{r}) = 0$ and $p'_1(\mathbf{r}) = p'_2(\mathbf{r}) = 0$.

[5] To underpin this theoretically one can use the fact that $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r})$ is indeed meaningless for spinors, especially in the case $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) = 0$. The argument can be generalized to particles whose wave functions are not spinor fields by pointing out that it is in general meaningless to sum expressions for the internal dynamics of particles, and also meaningless to sum scalar group representations (e.g. of SO(1,1), SO(2) or U(1)).

[6] Nevertheless we need the calculation $\psi_3(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r})$ in the theory, to reproduce the interference. The only way out is therefore to explain the calculation as mathematically proved but physically meaningless. This can be justified by ruling out the superposition principle, by showing that it is always boils down to the fake superposition principle.

[7] We then still have the problem that the solution of the equations is done using coherent wave functions while the source will often not be coherent, e.g. if we only allow for one particle at the time to traverse the set-up every quarter of an hour. The use of the coherent wave functions when the source is incoherent can be justified by a reasoning based on the coherence of the interactions, proposed in Subsection 4.3 of [1].

Mathematically there is absolutely no problem with the whole construction, which is derived *deductively* from the observations and in agreement with our reconstruction of QM.

11 Conclusion

In summary, we have proposed an intelligible analysis of the paradox of the double-slit experiment. It avoids not testable, doubt-casting assumptions like travelling backwards in time, advanced waves or the outright *contradictio in terminis* of the particle-wave duality. Our proposal is free of such eerie assumptions, such that these can be weeded

[14] The scheme represents the only possibility to avoid a logical contradiction. Based on the derivation of the Dirac equation from scratch in [1] and the proof in Subsection 3.6 we have devised a completely logical and mathematically rigorous proof for the way we must calculate the interference pattern of the double-slit experiment. This is the result we anticipated in Subsection 1.2. It is like a miraculous theorem of advanced mathematics. There was no rigorous algebraic proof beforehand. We have also explained why the second intuitive approach completely fails, by pinpointing all the pitfalls we must avoid and the many instances where our intuition goes wrong. Despite the rigour of the proof, we might still feel that the correct reasoning takes a huge detour to justify the coherent-sum rule. We would have preferred something more direct and synthetic, that does not rely on wave equations and group representations, such that we could end up with two reasonings, a long formal one and a short intuitive one which now agree. But we do not have direct synthetic methods to deal with conditional probabilities within a context of ternary logic and long-distance correlations. It is hard to find justifications for the Born rule without using spinors and for making coherent sums of group representations.

out using the principle of Occam's razor. We have also made the effort to reassure less open-minded physicists that they can continue to "shut up and calculate" with "vectors in Hilbert space" as usual.

What we must learn from it is that there is no way one can use probabilities obtained from one experiment in the analysis of another experiment. The probabilities are conditional and context-bound. Combining probabilities conditioned by different contexts in a same calculation is logically flawed and should therefore be considered as taboo. This is Bohr's caveat at work. It stresses the importance of contextuality.

The traditional language we use to deal with probabilities p_j is too poor to take into account these contexts with long-distance correlations and ternary logic correctly. To correctly deal with the contextuality we must use group representation theory. The group representations are not counter-intuitive (see [1,11]), but they correspond to an unexpected advanced level of sharpened intuition that is not part of the congenital zero-level intuition we use to deal with daily-life probabilities p_j. The two levels of intuition are using therefore different definitions of probability.

Justifying coherent summing of wave functions by using the superposition principle is conceptually wrong and misleading. The justification of coherent summing must be based on Subsections 3.6 and 9.1. There could exist a proof based on the Huygens principle that is as rigorous as the proof in 3.6 if the calculations can be carried out in full detail.

Our approach takes into account the "non-locality" and the undecidability. The handshake in Cramer's transactional interpretation of QM [35] catches the essence of what is going on in the algebra. It permits to fine-tune the wave function such that it can satisfy simultaneously all boundary conditions from mutually distant places. The handshake is a two-way process for adjusting the wave function to two such mutually remote boundary conditions. But the advanced waves used in Cramer's transaction do not have physical reality.

We think that the results of the present paper and those about the Stern-Gerlach experiment in [2] constitute convincing evidence for the value of the reconstruction of QM based on the geometrical meaning of spinors in [1,10]. It permits for a perfect dialogue between the mathematics and the physics. Of course, as we admitted in Subsection 1.2, there is more than one mystery in QM. The list of mysteries is long but what we have achieved should give us confidence in the reconstruction.

Appendix A. Why spinors do not form a vector space

11.1 Introduction

A matrix of SU(2) is of the form:

$$\mathbf{R} = \begin{bmatrix} u & -v^* \\ v & u^* \end{bmatrix}, \quad \text{with} \quad (u,v) \in \mathbb{C}^2, \quad \det(\mathbf{R}) = uu^* + vv^* = 1, \tag{29}$$

and represents a rotation. It suffices to know the first column of this matrix to know the whole matrix, because the second column can be derived unambiguously from the first column. This leads to the idea of using this first column as a shorthand notation for the full matrix. Such a first column is a spinor of SU(2). We have already explained this earlier, e.g. based on equation 4 on p. 7 of [1]. Therefore a spinor also represents a rotation around the origin of $\mathbb{R}^3$. The group SO(3) of rotations around the origin of $\mathbb{R}^3$ forms a three-dimensional manifold, because a rotation is defined by three independent real parameters, e.g. by the three Euler angles (α,β,γ) or by $(\mathbf{s},\varphi)$, where φ is the rotation angle around the rotation axis ℓ defined by the unit vector $\mathbf{s} \parallel \ell$. The 4×1 column matrices we use in the Dirac equation are also a stenographic notation for the 4×4 matrices which represent the group elements of the Lorentz group.

A linear combination of two (or more) rotations is not defined by the group axioms, only the products of group elements are defined. Therefore, also the linear combination of two SU(2) matrices or of two spinors is geometrically not defined. But these spinors belong to $\mathbb{C}^2$ which is a two-dimensional complex vector space, equivalent to the four-dimensional real vector space $\mathbb{R}^4$. This leads to the speculation that by definition every element of $\mathbb{C}^2$ could be called a spinor and that $\forall(c_1,c_2) \in \mathbb{C}^2$, $c_1\psi_1 + c_2\psi_2$, would be defined in the vector space $\mathbb{C}^2$ (see e.g. [12]).

We will show that this widespread belief is shortsighted and not correct. In fact, while these operations are undeniably algebraically defined, this is not the end of the story. The problem resides in finding a geometrical meaning for the purely algebraic result. When a curved manifold M is embedded in a vector space V, the geometrical meaning one would have to attribute to the points of the set $V \backslash M$ may not make sense. This is a well-known mathematical fact (see below in Subsection 11.2), which is overlooked in such speculations.

In all physics textbooks I was able to consult, this misconception is allowed to linger on, if not overtly promoted. Most physics textbooks define a representation of a group $(G,\circ)$ with group product $\circ$, as an isomorphism between the group $(G,\circ)$ and a matrix group $(G',\bullet)$, where $\bullet$ is the usual product of matrices. (See e.g. p. 68 of [45], p. 37 of [46], p.10 of [47]). We use the extravagant notations $\circ$ and $\bullet$ to render the composition laws explicit because they are in general implied by a pure juxtaposition of group elements. It is very obviously an illicit over-interpretation to extend this definition from the matrix *group* to the matrix *ring* (the so-called group ring, see e.g. [48]), obtained

by allowing also for making linear combinations of matrices. This group ring would correspond then by an extended isomorphism to a so-called group module, which is only purely formally defined and *a priori* has no geometrical meaning. The matrices of the group are bijective linear transformations (i.e. automorphisms) of vector spaces. Two caveats are lacking in the textbooks. First of all, none of their authors mentions the fact that it might be necessary to restrict the definition domain of the matrices to a subset of a vector space, e.g. a manifold M embedded in that vector space V (see Subsection 11.2). Furthermore, the textbooks do not mention explicitly that linear combinations of such matrices are *a priori* not considered to be part of the definitions. It would therefore be instrumental if physics textbooks mentioned explicitly those two caveats at the time they introduce these definitions. We would then need also the discussion in section 2.5.2 of [1] to justify certain calculations of group representation theory where linear combinations are nevertheless used, e.g. in the calculation of Casimir operators. (Such linear combinations are also routinely used in quantum mechanics (QM)).

In fact, on p. 7 of his monograph [49], Sagan states clearly that the linear combinations are *purely formal*. He calls the set of these purely formal linear combinations which are isomorphic to the group ring a G-module, and the corresponding calculus the group algebra. This shows that at least some mathematicians are aware of the subtleties in the definitions whereas physicists are blissfully unaware of them. That no geometrical meaning is being provided for this G-module and its group algebra is something a mathematician should not be satisfied with, and even less a physicist because he applies the formalism to the real world.

The mathematical fact that spinors do not form a vector space flies in the face of what one can read in physics textbooks (see e.g. [12], which explicitly claims that spinors form a vector space) and of the very existence of a monograph written by Dirac with the title "Spinors in Hilbert space". This should not refrain us from stoically pointing out the truth. Traditional QM is teeming with such misinterpretations of the mathematics due to the way physicists consistently use the algebra as a blackbox without caring what it means. Nothing summarizes this head-over-heels approach better than the expeditious slogan: *"Shut up and calculate!"* There is actually no pride to be drawn from this fast-and-furious philosophy, because in the end we find ourselves confronted with a medal that has two sides. The bright side is that the calculations carried out in blindfold mode are reproducing the experimental data with stunning precision. The dark side is that such a blitzkrieg approach outplays any attempt at understanding what the calculations may mean.

It is natural to object that the statement that spinors do not form a vector space must be very obviously wrong because the calculations we are carrying out in QM yield correct and useful results. But considering it as self-evident that one can treat spinors as vectors in the calculations, just because it is algebraically feasible and leads to satisfactory results takes a blunt short-cut to the correct argument, which comprises two steps. The first step consists in showing that spinors truly do not form a vector space. The second step consists in introducing a mathematical construction that provides the group module with a geometrical meaning in terms of sets, whereby the algebra of the group ring becomes a calculus of probabilities or probability amplitudes, as explained in section 2.5.2 of [1]. In its most general form this becomes a theory of sets of pure states (i.e. group elements) with a complex probability amplitude defined on them (used to define mixed states).

11.2 The points of the vector space V which do not belong to the embedded manifold M

Column matrices representing group elements are not proverbial rare birds. We encounter them in every regular representation of a finite group. Let us consider e.g. the permutation group S_n. We can label arbitrarily each of the $n!$ permutations $p_j \in S_n$ with a number $j \in [1, n!] \cap \mathbb{N}$. The order the group elements acquire this way has no importance. Any order will do. The regular representation is then given by $n! \times n!$ representation matrices and each group element p_j is also represented by a $n! \times 1$ matrix $\mathbf{p}^{(j)}$, whose entries are $[\,\mathbf{p}^{(j)}\,]_k = \delta_{kj}$. That is, all entries take the value 0 except the one on line j, which takes the value 1. The square matrix representing a group element p_j just represents p_j by the group automorphism: $T_{p_j} : q \in S_n \to T_{p_j}(q) = p_j \circ q \in S_n$. We could call the $n! \times 1$ column matrices $\mathbf{p}^{(j)}$ "column vectors" and they would span a "vector space" $(V, \mathbb{K})$ over some number field $\mathbb{K}$ that could be $\mathbb{R}$ or $\mathbb{C}$. The group S_n would then be a finite discrete subset (of $n!$ points) of the vector space $(V, \mathbb{K}) = (\mathbb{K}^{n!}, \mathbb{K})$. The $n!$ column vectors constitute an orthonormal basis for $(V, \mathbb{K})$. But this is just shallow nonsense, because in the representation the sum of two such $n! \times 1$ column matrices $\mathbf{p}$ and $\mathbf{q}$ would by isomorphism correspond to:

$$\begin{pmatrix} 1 & 2 & \cdots & k & \cdots & n \\ p(1) & p(2) & \cdots & p(k) & \cdots & p(n) \end{pmatrix} + \begin{pmatrix} 1 & 2 & \cdots & k & \cdots & n \\ q(1) & q(2) & \cdots & q(k) & \cdots & q(n) \end{pmatrix}.$$

It is very obvious that such a sum of permutations p and q is just not defined, and the same applies for any other linear combination $\sum_{j=1}^{n!} c_j \mathbf{p}^{(j)}$, with $c_j \in \mathbb{K}, \forall j \in [1, n!] \cap \mathbb{N}$, that does not belong to S_n. All points of $\mathbb{K}^{n!} \backslash S_n$ are *a priori* meaningless. This example illustrates the pitfalls of heedlessly carrying out algebraic calculations without bothering what they mean, as implied by the leitmotiv to *"Shut up and calculate!"*

It is for the same reason that in general relativity the curved space-time manifold should not be considered as embedded in a vector space, but described *intrinsically*. The points one would have to add to obtain an extension in the form of a vector space wherein the curved space-time manifold could be embedded do not exist physically. That vector space would have to be $\mathbb{R}^5$, just like the two-dimensional surface of a sphere is embedded in $\mathbb{R}^3$. The points of the extension to $\mathbb{R}^5$ that do not belong to space-time would be physically meaningless. It is in order to avoid such utter nonsense and to describe accordingly space-time *intrinsically* that we need a whole artillery of concepts from differential geometry like manifolds, Riemann and Ricci tensors, curvilinear coordinates, covariant derivatives and parallel transport.

Within the representation $\mathrm{SO}(3) \subset \mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$ it is possible to attribute a meaning to the sum of two rotations $R_1 + R_2$ because we can figure out the result of the action of $R_1 + R_2$ on a general vector $\mathbf{r} \in \mathbb{R}^3$. In fact $R_1(\mathbf{r}) + R_2(\mathbf{r})$ is a sum of vectors, and we know what this means, such that it can be used to define $R_1 + R_2$. But the resulting definition does not introduce a very enlightening geometrical concept. It does not correspond to a Gestalt. A similar approach for the sum of two SU(2) matrices $\mathbf{R}_1$, $\mathbf{R}_2$ and a general spinor ψ yields $\mathbf{R}_1\psi + \mathbf{R}_2\psi$ which is a sum of spinors. This makes it impossible to figure out the geometrical meaning of $\mathbf{R}_1 + \mathbf{R}_2$ for $\mathbf{R}_j \in \mathrm{SU}(2)$ without knowing the geometrical meaning of sums of spinors.[15]

But these spinors ψ_j are the first columns of the matrices $\mathbf{R}_j$, such that to figure out their meaning we must know the meaning of $\mathbf{R}_1 + \mathbf{R}_2$. If we wanted to figure out what the sum $\psi_1 + \psi_2$ of two spinors means, we could also try to calculate $\mathbf{R}\psi_1 + \mathbf{R}\psi_2$, but this is again a sum of two spinors. Whatever we try to make sense of such sums, we end up running in circles. In summary, we cannot figure out what the sum of two spinors means because we cannot figure out what the sum of two rotation matrices means, and we cannot figure out what the sum of two rotation matrices means because we cannot figure out what the sum of two spinors means. What we did for SO(3) can this way not be transposed to SU(2). We cannot solve the problem in SU(2) with the method we used in SO(3), because the column vectors the rotation matrices are operating on do not correspond to vectors but to group elements, which form a manifold rather than a vector space. To make sense of sums of rotation matrices or spinors in SU(2) we must therefore try again to figure out how they operate on vectors $\mathbf{r} \in \mathbb{R}^3$, but the problem is that the calculations to be performed are now no longer linear but "quadratic", in the sense to be developed below, starting from Eq. 35.

Consider now the complex vector space $\mathbb{C}^2$ and a general point $(\zeta_1, \zeta_2) \in \mathbb{C}^2$ of it. Until something is done about it, such a point has no obvious geometrical meaning. A first step consists in pointing out that the spinors $\psi = [\,\zeta_1, \zeta_2\,]^\top$ of SU(2), which (as explained above and on p. 7 of [1]) can be given the meaning of rotations around the origin in $\mathbb{R}^3$, belong to the set $\mathscr{G} = \{(\zeta_1, \zeta_2) \parallel \zeta_1^*\zeta_1 + \zeta_2^*\zeta_2 = 1\} \subset \mathbb{C}^2$, such that $\mathrm{SU}(2) \subset \mathscr{G}$. These spinors are isometries, i.e. special elements of the vector space $\mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$, which conserve the metric. They are geometrical operators. Each rotation corresponds to two spinors which are identical up to a factor ± 1. Each rotation defines therefore two points of the curved manifold $\mathscr{G}$, but the converse is also true: each such pair of points of $\mathscr{G}$ corresponds to a rotation, i.e. a pair of spinors of SU(2), which is a double covering of SO(3) due to the very existence of these two possible factors ± 1. The simplest way to prove this is to identify (ζ_1, ζ_2) with the expression for a spinor that corresponds to the rotation $R(\alpha, \beta, \gamma)$ where (α, β, γ) are its Euler angles, as e.g. given by the first column of:

$$\mathbf{R}(\alpha, \beta, \gamma) = \begin{bmatrix} e^{-\imath(\alpha+\gamma)/2} \cos\frac{\beta}{2} & -\imath e^{-\imath(\alpha-\gamma)/2} \sin\frac{\beta}{2} \\ -\imath e^{\imath(\alpha-\gamma)/2} \sin\frac{\beta}{2} & e^{+\imath(\alpha+\gamma)/2} \cos\frac{\beta}{2} \end{bmatrix}, \tag{30}$$

in equation 1.2.29 of [12]. The definition of the Euler angles used here is given in Fig. 1.5 of [12]. Therefore we have also $\mathscr{G} \subset \mathrm{SU}(2)$, such that $\mathscr{G} \equiv \mathrm{SU}(2)$. We can consider the manifold $\mathscr{G}$ as embedded in $\mathbb{C}^2$. In terms of real numbers, the rotation group, represented by $\mathscr{G}$, is then a three-dimensional manifold embedded in the four-dimensional vector space $\mathbb{C}^2 \equiv \mathbb{R}^4$. This is analogous to four-dimensional space-time embedded in $\mathbb{R}^5$ as explained above, such that analogous caveats must prevail.

It is of course algebraically feasible to calculate linear combinations $c_1\psi_1 + c_2\psi_2$, where $(c_1, c_2) \in \mathbb{C}^2$, or to consider elements of $\mathbb{C}^2 \backslash \mathscr{G}$ but this is purely formal and *a priori* devoid of any geometrical meaning in terms of some element of $\mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$, which is the natural embedding for the rotation group $\mathrm{SO}(3) \subset \mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$ [1]. What other kind of embedding of SO(3) could we else imagine to give $(\zeta_1, \zeta_2) \in \mathbb{C}^2 \backslash \mathscr{G}$ meaning? This argument is similar to the one for S_n above. However, this time the group is a continuous Lie group and therefore no longer a discrete finite set but a differentiable manifold.

11.3 Attempt to attribute a meaning to the sum of two spinors of SU(2) within $\mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$

We can further illustrate that the meaning we would have to attribute to a sum of two spinors of SU(2) in $\mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$ is spurious, by making the following calculation on vectors anticipated above. Let us consider two rotations $R_j(\mathbf{s}_j, \varphi_j)$,

[15] In reality $\mathbf{R}_1$, $\mathbf{R}_2$ must here be considered as elements of F(SU(2),SU(2)), more precisely of the group of group automorphisms, which is isomorphic to SU(2).

$j \in \{1,2\}$, and a vector $\mathbf{r} \in \mathbb{R}^3$. Here φ_j are the rotation angles around the rotation axes ℓ_j defined by the unit vectors $\mathbf{s}_j \parallel \ell_j$. The SU(2) representation matrices of $R_j(\mathbf{s}_j, \varphi_j)$ are given by $\mathbf{R}_j = \cos(\varphi/2)\mathbb{1} - \imath \sin(\varphi/2)[\,\mathbf{s}_j \cdot \boldsymbol{\sigma}\,]$, which is called the Rodrigues formula. The corresponding spinors ψ_j are obtained by taking the first columns of $\mathbf{R}_j$. The vector $\mathbf{r}$ is represented by $[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]$.

We know that for a matrix $\mathbf{R}$ of SU(2) its inverse is given by: $\mathbf{R}^{-1} = \mathbf{R}^\dagger$. In fact, the inverse of a general 2×2 matrix:

$$\mathbf{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{with:} \quad D = \det(\mathbf{M}) = ad - bc, \tag{31}$$

exists when $D \neq 0$ and is then given by:

$$\mathbf{M}^{-1} = \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{32}$$

Applying this to a matrix of the form:

$$\mathbf{S} = \begin{bmatrix} u & -v^* \\ v & u^* \end{bmatrix}, \quad \text{with} \quad (u,v) \in \mathbb{C}^2, \quad \det(\mathbf{S}) = uu^* + vv^* \in \mathbb{R}, \tag{33}$$

one obtains:

$$\mathbf{S}^{-1} = \frac{1}{uu^* + vv^*} \begin{bmatrix} u^* & v^* \\ -v & u \end{bmatrix} = \frac{1}{uu^* + vv^*} \begin{bmatrix} u & -v^* \\ v & u^* \end{bmatrix}^\dagger. \tag{34}$$

For matrices $\mathbf{R} \in \mathrm{SU}(2)$, we have $uu^* + vv^* = 1$, such that then $\mathbf{R}^{-1} = \mathbf{R}^\dagger$. For a sum $\mathbf{R}_1 + \mathbf{R}_2$ of two SU(2) matrices, $\det(\mathbf{R}_1 + \mathbf{R}_2) \in \mathbb{R}$ will not be equal to 1. But to calculate the inverse of $\mathbf{R}_1 + \mathbf{R}_2$ it suffices to calculate $[\,\mathbf{R}_1 + \mathbf{R}_2\,]^\dagger$ and to divide the result by $\det[\,\mathbf{R}_1 + \mathbf{R}_2\,]$, provided $D = \det[\,\mathbf{R}_1 + \mathbf{R}_2\,] \neq 0$. Otherwise the inverse will not be defined, which will happen if and only if $u = v = 0$, i.e. $\mathbf{R}_2 = -\mathbf{R}_1$. This is equivalent to $\mathbf{s}_2 = \mathbf{s}_1$ & $\varphi_2 \equiv \varphi_1 + 2\pi \pmod{4\pi}$.

Let us calculate the effect of the sum $\mathbf{R}_1 + \mathbf{R}_2$ on $[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]$. The original idea behind the speculation is that all elements of $\mathbb{C}^2$ should be meaningful spinors. They therefore should be the first columns of 2×2 matrices which work on vectors $\mathbf{r} \in \mathbb{R}^3$ in the same way as the rotation matrices of SU(2). Hence $\psi_1 + \psi_2$ corresponds to $\mathbf{R}_1 + \mathbf{R}_2$ which works on $[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]$ according to:

$$[\,\mathbf{R}_1 + \mathbf{R}_2\,]\,[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\,[\,\mathbf{R}_1 + \mathbf{R}_2\,]^{-1} = \frac{1}{D}[\,\mathbf{R}_1 + \mathbf{R}_2\,]\,[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\,[\,\mathbf{R}_1 + \mathbf{R}_2\,]^\dagger. \tag{35}$$

From section 2.6 of [1], where we introduced the parallel formalism for vectors, it must be clear that for $\mathbf{R} \in \mathrm{SU}(2)$, it is the rule $[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,] \to \mathbf{R}\,[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\,\mathbf{R}^{-1}$ which is primal, while the rule $[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,] \to \mathbf{R}\,[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\,\mathbf{R}^\dagger$ is derived and relies on the specificity $\mathbf{R}^{-1} = \mathbf{R}^\dagger$ which is no longer valid for $\mathbf{R}_1 + \mathbf{R}_2$. Thus, the result must really be calculated according to Eq. 35, not according to Eq. 36 below. In view of this, when $\mathbf{R}_1 + \mathbf{R}_2 = 0$, the operation of $\mathbf{R}_1 + \mathbf{R}_2$ on a vector $\mathbf{r} \in \mathbb{R}^3$ is not defined. Let us now first calculate the result of the transformation:

$$\begin{aligned} &[\,\mathbf{R}_1 + \mathbf{R}_2\,]\,[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\,[\,\mathbf{R}_1 + \mathbf{R}_2\,]^\dagger = \\ &\underbrace{\mathbf{R}_1[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\mathbf{R}_1^\dagger}_{R_1(\mathbf{r})} + \underbrace{\mathbf{R}_2[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\mathbf{R}_2^\dagger}_{R_2(\mathbf{r})} + \underbrace{\mathbf{R}_1[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\mathbf{R}_2^\dagger}_{T_3} + \underbrace{\mathbf{R}_2[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\mathbf{R}_1^\dagger}_{T_4}. \end{aligned} \tag{36}$$

As already mentioned, this action is no longer linear but "quadratic" or rank-2.

The term T_4 can be obtained from the term T_3 by carrying out the substitution $(1,2)|(2,1)$ on all indices. It suffices therefore to calculate T_3. Using the algebraic identity $[\,\mathbf{a} \cdot \boldsymbol{\sigma}\,][\,\mathbf{b} \cdot \boldsymbol{\sigma}\,] = (\mathbf{a} \cdot \mathbf{b})\mathbb{1} + \imath[\,(\mathbf{a} \wedge \mathbf{b}) \cdot \boldsymbol{\sigma}\,]$, this yields:

$$\begin{aligned} T_3 &= \{\cos(\varphi_1/2)\mathbb{1} - \imath \sin(\varphi_1/2)[\,\mathbf{s}_1 \cdot \boldsymbol{\sigma}\,]\}\,[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,]\,\{\cos(\varphi_2/2)\mathbb{1} + \imath \sin(\varphi_2/2)[\,\mathbf{s}_2 \cdot \boldsymbol{\sigma}\,]\} \\ &= \{\cos(\varphi_1/2)[\,\mathbf{r} \cdot \boldsymbol{\sigma}\,] - \imath \sin(\varphi_1/2)(\,\mathbf{s}_1 \cdot \mathbf{r}\,)\,\mathbb{1} + \sin(\varphi_1/2)[\,(\mathbf{s}_1 \wedge \mathbf{r}) \cdot \boldsymbol{\sigma}\,]\} \\ &\quad \times \{\cos(\varphi_2/2)\mathbb{1} + \imath \sin(\varphi_2/2)[\,\mathbf{s}_2 \cdot \boldsymbol{\sigma}\,]\}. \end{aligned} \tag{37}$$

This equation contains a scalar and a vector term. The scalar term of T_3 is:

$$-\imath \sin(\varphi_1/2)\cos(\varphi_2/2)(\,\mathbf{s}_1 \cdot \mathbf{r}\,) + \imath \sin(\varphi_2/2)\cos(\varphi_1/2)(\,\mathbf{s}_2 \cdot \mathbf{r}\,) + \imath \sin(\varphi_1/2)\sin(\varphi_2/2)(\mathbf{s}_2 \cdot (\mathbf{s}_1 \wedge \mathbf{r})). \tag{38}$$

The substitution $(1,2)|(2,1)$ in the indices yields for the corresponding scalar term within T_4:

$$-\imath\sin(\varphi_2/2)\cos(\varphi_1/2)(\,\mathbf{s}_2{\cdot}\mathbf{r}\,)+\imath\sin(\varphi_1/2)\cos(\varphi_2/2)(\,\mathbf{s}_1{\cdot}\mathbf{r}\,)+\imath\sin(\varphi_2/2)\sin(\varphi_1/2)(\mathbf{s}_1{\cdot}(\mathbf{s}_2\wedge\mathbf{r})).$$

The sum of these two scalar terms is zero. In fact, the mixed products can be written under the form of determinants, and these two determinants are obtained one from another by exchanging two lines. Because the scalar terms vanish, T_3+T_4 is a true vector. The vector part of term T_3 is:

$$\begin{aligned}&\cos(\varphi_1/2)\cos(\varphi_2/2)[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]+\sin(\varphi_1/2)\sin(\varphi_2/2)(\,\mathbf{s}_1{\cdot}\mathbf{r}\,)[\,\mathbf{s}_2{\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_1/2)\cos(\varphi_2/2)[\,(\mathbf{s}_1\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]-\cos(\varphi_1/2)\sin(\varphi_2/2)[\,(\mathbf{r}\wedge\mathbf{s}_2){\cdot}\boldsymbol{\sigma}\,]\\&-\sin(\varphi_1/2)\sin(\varphi_2/2)[\,((\,\mathbf{s}_1\wedge\mathbf{r}\,)\wedge\mathbf{s}_2){\cdot}\boldsymbol{\sigma}\,].\end{aligned}\tag{39}$$

This can be rewritten as:

$$\begin{aligned}&\cos(\varphi_1/2)\cos(\varphi_2/2)[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]+\sin(\varphi_1/2)\sin(\varphi_2/2)(\,\mathbf{s}_1{\cdot}\mathbf{r}\,)[\,\mathbf{s}_2{\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_1/2)\cos(\varphi_2/2)[\,(\mathbf{s}_1\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]+\cos(\varphi_1/2)\sin(\varphi_2/2)[\,(\mathbf{s}_2\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_1/2)\sin(\varphi_2/2)[\,(\mathbf{s}_2\wedge(\,\mathbf{s}_1\wedge\mathbf{r}\,)){\cdot}\boldsymbol{\sigma}\,].\end{aligned}\tag{40}$$

Now we can use the identity $\mathbf{a}\wedge(\mathbf{b}\wedge\mathbf{c})=(\mathbf{a}{\cdot}\mathbf{c})\mathbf{b}-(\mathbf{a}{\cdot}\mathbf{b})\mathbf{c}$ to rewrite this as:

$$\begin{aligned}&\cos(\varphi_1/2)\cos(\varphi_2/2)[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]+\sin(\varphi_1/2)\sin(\varphi_2/2)(\,\mathbf{s}_1{\cdot}\mathbf{r}\,)[\,\mathbf{s}_2{\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_1/2)\cos(\varphi_2/2)[\,(\mathbf{s}_1\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]+\cos(\varphi_1/2)\sin(\varphi_2/2)[\,(\mathbf{s}_2\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_1/2)\sin(\varphi_2/2)((\mathbf{s}_2{\cdot}\mathbf{r})[\,\mathbf{s}_1{\cdot}\boldsymbol{\sigma}\,]-(\mathbf{s}_2{\cdot}\mathbf{s}_1)[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]).\end{aligned}\tag{41}$$

With the substitution $(1,2)|(2,1)$ in the indices we obtain the corresponding term in T_4:

$$\begin{aligned}&\cos(\varphi_2/2)\cos(\varphi_1/2)[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]+\sin(\varphi_2/2)\sin(\varphi_1/2)(\,\mathbf{s}_2{\cdot}\mathbf{r}\,)[\,\mathbf{s}_1{\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_2/2)\cos(\varphi_1/2)[\,(\mathbf{s}_2\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]+\cos(\varphi_2/2)\sin(\varphi_1/2)[\,(\mathbf{s}_1\wedge\mathbf{r}){\cdot}\boldsymbol{\sigma}\,]\\&+\sin(\varphi_2/2)\sin(\varphi_1/2)((\mathbf{s}_1{\cdot}\mathbf{r})[\,\mathbf{s}_2{\cdot}\boldsymbol{\sigma}\,]-(\mathbf{s}_1{\cdot}\mathbf{s}_2)[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]).\end{aligned}\tag{42}$$

Summing the vector terms in T_3 and T_4 yields:

$$\begin{aligned}&2\,[\,\cos(\varphi_1/2)\cos(\varphi_2/2)-\mathbf{s}_1{\cdot}\mathbf{s}_2\,]\,\mathbf{r}+2\sin(\varphi_1/2)\sin(\varphi_2/2)[\,(\,\mathbf{s}_2{\cdot}\mathbf{r}\,)\,\mathbf{s}_1+(\,\mathbf{s}_1{\cdot}\mathbf{r}\,)\,\mathbf{s}_2\,]\\&+2\sin(\varphi_1/2)\cos(\varphi_2/2)(\mathbf{s}_1\wedge\mathbf{r})+2\cos(\varphi_1/2)\sin(\varphi_2/2)(\mathbf{s}_2\wedge\mathbf{r}).\end{aligned}\tag{43}$$

Therefore the transformation in Eq. 36 corresponds to the function $\mathbf{h}\in\mathrm{L}(\mathbb{R}^3,\mathbb{R}^3):\mathbf{r}\in\mathbb{R}^3\rightarrow\mathbf{h}(\mathbf{r})$ given by:

$$\begin{aligned}\mathbf{h}(\mathbf{r})=&R_1(\mathbf{r})+R_2(\mathbf{r})+2\,[\,\cos(\varphi_1/2)\cos(\varphi_2/2)-\mathbf{s}_1{\cdot}\mathbf{s}_2\,]\,\mathbf{r}\\&+2\sin(\varphi_1/2)\sin(\varphi_2/2)[\,(\,\mathbf{s}_2{\cdot}\mathbf{r}\,)\,\mathbf{s}_1+(\,\mathbf{s}_1{\cdot}\mathbf{r}\,)\,\mathbf{s}_2\,]\\&+2\sin(\varphi_1/2)\cos(\varphi_2/2)(\mathbf{s}_1\wedge\mathbf{r})+2\cos(\varphi_1/2)\sin(\varphi_2/2)(\mathbf{s}_2\wedge\mathbf{r}).\end{aligned}\tag{44}$$

Let us now treat the division by $D=\det[\,\mathbf{R}_1+\mathbf{R}_2\,]\in\mathbb{R}$, when $D\neq0$. In this case $D>0$, such that $\sqrt{D}\in\mathbb{R}$. The determinant of the matrix $\mathbf{U}=\frac{1}{\sqrt{D}}[\,\mathbf{R}_1+\mathbf{R}_2\,]$ is equal to 1. The matrix $\mathbf{U}$ is therefore unitarian and:

$$[\,\mathbf{R}_1+\mathbf{R}_2\,]\,[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]\,[\,\mathbf{R}_1+\mathbf{R}_2\,]^{-1}=\sqrt{D}\,\mathbf{U}[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]\mathbf{U}^{-1}\,\frac{1}{\sqrt{D}}=\mathbf{U}[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]\mathbf{U}^{-1}.\tag{45}$$

The matrix $\mathbf{U}$ is a rotation which preserves the length of $\mathbf{r}$. Hence, if it is defined, the operation $[\,\mathbf{R}_1+\mathbf{R}_2\,]\,[\,\mathbf{r}{\cdot}\boldsymbol{\sigma}\,]\,[\,\mathbf{R}_1+\mathbf{R}_2\,]^{-1}$ also preserves the length. This means that this operation is nothing else than the rotation R which transforms $\mathbf{r}$ into $|\mathbf{r}|\,\mathbf{h}(\mathbf{r})/|\mathbf{h}(\mathbf{r})|$, provided $\mathbf{R}_1+\mathbf{R}_2\neq\mathbf{0}$ & $\mathbf{h}(\mathbf{r})\neq\mathbf{0}$. We have not investigated if $\mathbf{h}(\mathbf{r})$ can accidentally become $\mathbf{0}$ for some values of $\mathbf{r}$, because this is tedious and not very useful for the point we want to make. The rotation R would then be the transformation that corresponds to $\psi=\psi_1+\psi_2$.

The sum $\psi=\psi_1+\psi_2$ has this way been given a geometrical definition but it is completely abstruse. It is an elaboration of the calculation in section 2.5.1 starting on p. 12 of [1], where we also introduced $(\psi_1+\psi_2)/|\psi_1+\psi_2|$ provided $\psi_1+\psi_2\neq0$. But now we have extended the scope to the action on vectors. We were wondering how we

could justify introducing such a renormalization of $\psi_1+\psi_2$. The present calculation gives the answer. It is a (rather accidental) corollary of the agenda to call all elements of $\mathbb{C}^2$ spinors, based on the assumption that spinors would constitute a vector space, whereby it turns out that the operator that corresponds to the sum $\psi_1+\psi_2$ of two spinors has the same action on a vector as these renormalized sums $(\psi_1+\psi_2)/|\psi_1+\psi_2|$.

The first two terms in $\mathbf{h}(\mathbf{r})$ correspond to $R_1(\mathbf{r})$ and $R_2(\mathbf{r})$. Their sum corresponds to what we expect on the basis of the representation SO(3). But in SU(2) there are extra terms and the sum of these terms is in general not identical to zero, as is easily checked. The result in SU(2) will in general be different from the result in SO(3), be it only because the result in SO(3) does not preserve the length of the vector $\mathbf{r}$ while the result in SU(2) does. But the vectors will in general also not be parallel. This ought to make everybody wake up and smell the coffee. The finding that the value we must attribute to the sum of two group elements would depend on the choice of the representation indicates that the algebraic procedure of summing spinors or group elements does not necessarily define a meaningful geometrical result. We can compare it to a definition that would depend on the choice of a reference frame. The elaboration of $|\psi_1+\psi_2|^2=|\psi_1|^2+|\psi_2|^2+\psi_1\psi_2^\dagger+\psi_2\psi_1^\dagger$ contains also two extra terms $\psi_1\psi_2^\dagger,\psi_2\psi_1^\dagger$ which lead to conceptual problems in the double-slit experiment for electrons, raising also questions about the procedure of summing spinors.

The extravagant result $|\mathbf{r}|\,\mathbf{h}(\mathbf{r})/|\mathbf{h}(\mathbf{r})|$ is admittedly providing $\psi_1+\psi_2$ with a geometrical definition, but one that is just not suited for any use in physics because its meaning is unfathomable. The vector $\mathbf{h}(\mathbf{r})$ is as useful as a point of $\mathbb{R}^5$ in general relativity. It all just looks like inscrutable nonsense, let alone what we would obtain for $\psi=c_1\psi_1+c_2\psi_2$. One could argue that the result is mathematically defined even if it does not look meaningful. But even that argument is invalidated by the case $\mathbf{R}_1+\mathbf{R}_2=0$, which has no inverse, such that $\mathbf{h}$ is then with certainty not defined. When $\mathbf{R}_2=-\mathbf{R}_1$ the result is definitely meaningless, because the function $f\in \mathrm{L}(\mathbb{C}^2,\mathbb{C}^2):(\zeta_1,\zeta_2)\rightarrow f(\zeta_1,\zeta_2)=(0,0),\forall(\zeta_1,\zeta_2)\in\mathbb{C}^2$ is not bijective, i.e., *it is not an automorphism of vector spaces.*

11.4 An approach based on isotropic vectors

A spinor $[\,\xi_0,\xi_1\,]^\top$ in SU(2) has been shown to be the "square root" of an isotropic vector (see e.g. [1], Subsection 2.7.1 and Eq. 29). The isotropic vector $\mathbf{e}'_x+\imath\mathbf{e}'_y=(x,y,z)\in\mathbb{C}^3$ can be used to define a rotated basis $(\mathbf{e}'_x,\mathbf{e}'_y,\mathbf{e}'_z)$, which in turn can be used to define a rotation. E.g.

$$\begin{bmatrix} z & x-\imath y \\ x+\imath y & -z \end{bmatrix}=\sqrt{2}\begin{bmatrix}\xi_0\\ \xi_1\end{bmatrix}\otimes\begin{bmatrix}-\xi_1 & \xi_0\end{bmatrix}\sqrt{2},\quad\text{with:}\quad \xi_0\xi_0^*+\xi_1\xi_1^*=1, \tag{46}$$

corresponds to an isotropic vector $(\mathbf{e}'_x+\imath\mathbf{e}'_y)\cdot\boldsymbol{\sigma}$. This approach based on isotropic vectors is more commonly used to define spinors, with the result that the concept becomes veiled in mystery. We have shown that our approach leads to the same definition of spinors, while solving the enigma of their meaning. In fact, we have shown that both the isotropic vector and the corresponding spinor $[\,\xi_0,\xi_1\,]^\top$ define a rotation. Consider now a second isotropic vector, based on the spinor $[\,\eta_0,\eta_1\,]^\top$ which also defines a rotation:

$$\sqrt{2}\begin{bmatrix}\eta_0\\ \eta_1\end{bmatrix}\otimes\begin{bmatrix}-\eta_1 & \eta_0\end{bmatrix}\sqrt{2}\quad\text{with:}\quad \eta_0\eta_0^*+\eta_1\eta_1^*=1. \tag{47}$$

Then:

$$\sqrt{2}\begin{bmatrix}\xi_0+\eta_0\\ \xi_1+\eta_1\end{bmatrix}\otimes\begin{bmatrix}-\xi_1-\eta_1 & \xi_0+\eta_0\end{bmatrix}\sqrt{2}\quad\text{but:}\quad (\xi_0+\eta_0)(\xi_0^*+\eta_0^*)+(\xi_1+\eta_1)(\xi_1^*+\eta_1^*)\neq 1, \tag{48}$$

defines also an isotropic vector. But this isotropic vector no longer defines a rotation because it is not normalized to 1. Furthermore, in the case $(\eta_0,\eta_1)=-(\xi_0,\xi_1)$ subsequent renormalization to 1 is just out of the question. Otherwise, by renormalizing the result we can indeed define again a rotation, but as we proved in Subsection 11.3, this is rather meaningless. As already stress, the procedure fails completely when $(\eta_0,\eta_1)=-(\xi_0,\xi_1)$. We may finally note that the sum of two isotropic vectors is not necessarily an isotropic vector (unless they are coplanar).

This research has received no funding.

The author declares no conflict of interest.

The latter only reflects my own naive perception of the issues, as illustrated by the way papers proposing a paradigm shift are nowadays stonewalled by nimbies. Here is an anthology:

◦ "We thank you for your kind offer to let us publish your manuscript, but regret to inform you that we have decided not to accept your offer. The paper did not undergo technical review and is not being declined for any technical error." (H. Saller, editor of International of Theoretical Physics).

◦ "We are writing to inform you that we will not be able to process your submission further. Submissions sent for peer-review are selected on the basis of discipline, novelty and general significance, in addition to the usual criteria for publication in scholarly journals. Therefore, our decision is not necessarily a reflection of the quality of your work." (Anonymous editor, Axioms).

◦ "We have examined your submission and unfortunately have to inform you that we feel it to be outside the scope of our journal. (Paolo Facchi, editor of European Physical Journal Plus). This desk rejection was sent to me after three months of peer review. His own acts (sending the paper out for peer review) are contradicting his narrative. I have not been communicated the results of the peer review.

◦ "I regret to inform you that, upon initial screening, I have determined that your manuscript referenced above is not suitable for publication in our journal. The mission of the journal is rapid dissemination of papers reporting new, timely and important advances. Although I make no determination regarding the technical correctness of your work, I think that it does not satisfy the acceptance criteria briefly outlined above". (Matteo Paris, Physics Letters A).

◦ "In addition, I have a serious methodological problem about publishing papers in science. Every scientific paper should be, in my view, self-consistent. I am sorry but I cannot accept the argument that for reading a given paper I will have to read another one as an introduction." (Anonymous referee, Symmetry, editor: Alessandro Sergi). The correct wording would be "self-contained" rather than "self-consistent". This is catch-22 because after yielding to this injunction I would be attacked on the length of the paper and for self-plagiarizing. This punitive enforcement of his own singular personal "views" runs contrary to all editorial guidelines and all currently accepted notions of ethics. This is the same person who weaponized the strawman that "you build a set-up and do experiments to understand what you have built" [14].

◦ "I have spent several hours to read the three papers. I have not found any new idea that makes the questions discussed more understandable. There are more understandable descriptions of spinors than the one suggested by the author." (Anonymous referee, Symmetry, editor: Alessandro Sergi). If this referee truly had read the three papers, he would not have blundered by presenting it as just a matter of "several hours", because such a narrative is completely out of touch. It takes a large number of days, presumably several weeks to assess the three papers. The very first lines of [1] are quoting Fields medallist Michael Atiyah's who has stated that nobody understands spinors. What are then the referee's sources for the more understandable descriptions (*in plural*)? All this renders obvious how incredibly brazen and spiteful these lies are. Not a single technical argument or source is given to underpin the peremptory statements, which ensnare me within an inversion of the charge of proof. I can only wonder if the referee reports Sergi forwarded to me might have been written by his cats.

There does not exist a single valid alternative for the explanations offered by my approach. The gold standard peer review pretends to be is a mere masquerade to disempower and wear out authors into burnout by dishing out relentless adversity using the pretexts of façade quality control, façade expert advice, façade politeness, façade poise, façade objectivity, façade truth, façade due process, façade fairness, façade respect and last but not least *façade empathy* (see [50]). Authors are blackmailed into exposing themselves to this farce of freewheeling vicious maltreatment by the propaganda that a work is worthless if it has not been published in a peer-reviewed journal. Only to charge them an exorbitant open-access fee for this "service", which in reality is just all their own work, including the secretary's job of delivering a camera-ready copy. This kind of mob law is permanent, as evidenced by the following savage desk rejection letter I received for reference [9] in an earlier attempt to have it published:

◦ "We have reviewed your submission, "What is the reason for the asymmetry between the twins in the twin paradox" and determined that, for a number of reasons, it is not appropriate for publication in the American Journal of Physics. One of the reasons is that you seem to be denying that time dilatation is physically real. It has in fact been confirmed by many experiments." (Richard H. Price, editor of the American Journal of Physics). It suffices to read [9] to check that there is no denial whatsoever. It just comes for free and can only be qualified as gaslighting.

There are no denials of QM or relativity in my work, which is meticulous and rigorous. The only denial which is not merely perceived but real and never stops, is the one of my rights.

Eppur si muove.

References

1. Coddens, G.: The geometrical meaning of spinors as a key to make sense of quantum mechanics. Symmetry, **13**, 659 (2021).
2. Coddens, G.: The Exact Theory of the Stern-Gerlach Experiment and Why it Does Not Imply that a Fermion Can Only Have Its Spin Up or Down. Symmetry **13**,134 (2021).
3. Coddens, G.: A solution for the paradox of the double-slit experiment, https://hal.archives-ouvertes.fr/cea-01459890v3.

4. Coddens, G.: A proposal to make sense of the double-slit experiment, https://hal.archives-ouvertes.fr/01383609v5.
5. Feynman, R.P., Leighton, R., Sands, M.: The Feynman Lectures on Physics, Vol. 3. Addison-Wesley, Reading, MA, (1970).
6. Feynman, R.P.: Probability & uncertainty - the quantum mechanical view of nature. https://www.youtube.com/watch?v=2mIk3wBJDgE (video).
7. Silverman, M.P.: More than one Mystery, Explorations in Quantum Interference. Springer, Heidelberg/New York (1995).
8. McDonald, K.T.: The Clock Paradox and Acceleratometers. https://physics.princeton.edu//mcdonald/examples/clock.pdf.
9. Coddens, G.: What is the reason for the asymmetry between the twins in the twin paradox? Eur. Phys. J. Plus **135**, 152 (2020).
10. Coddens, G.: From Spinors to Quantum Mechanics. Imperial College Press, (2015).
11. Coddens, G.: Spinors for everyone. https://hal.archives-ouvertes.fcea-01572342r/v1.
12. Hladik, J., Cole, J.M.: Spinors in Physics. Springer, New York, NY, USA, (2012).
13. Anonymous referee 1 (Symmetry, 2021).
14. Anonymous referee 2 (Symmetry, 2021).
15. Coddens, G.: Why spinors do not form a vector space. https://hal.archives-ouvertes.fr/hal-03289828.
16. Dirac, P.A.M.: Spinors in Hilbert Space. Springer, Heidelberg/New York, (2012).
17. Cartan, E.: The Theory of Spinors. Dover, New York, NY, USA, (1981).
18. Aspect, A., Dalibard, J., Roger, G.: Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. Phys. Rev. Lett. **49**, 1804 (1982).
19. Aspect, A., Grangier, P., Roger, G.: Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. Phys. Rev. Lett. **49**, 91 (1982).
20. Juan Yin *et al.*: Satellite-based entanglement distribution over 1200 kilometers, Science **356**, 1140 (2017).
21. Herbst, T., Scheidl, T., Fink, M., Handsteiner, J., Wittmann, B., Ursin, R., Zeilinger, A.: Teleportation of entanglement over 143 km. PNAS, **112**, 14202 (2015).
22. Kupczynski, M.: Is the Moon there if nobody looks: Bell Inequalities and physical reality. Frontiers in Physics, **8**, 273 (2020);
23. Kupczynski, B.: Closing the Door on Quantum Nonlocality. Entropy **20**, 877 (2018). https://doi.org/10.3390/e20110877.
24. Wheeler, J.A.; Mathematical Foundations of Quantum Theory, Marlow, A.R. (ed), pp. 9-48. Academic Press, (1978).
25. Wheeler, J.A., Zurek, W.H.: Quantum Theory and Measurement (Princeton Series in Physics). Princeton University Press, Princeton, NJ, (1983).
26. Kim, Y.-H., Yu, R., Kulik, S.P., Shih, Y., Scully, M.O.: Delayed "Choice" Quantum Eraser. Phys. Rev. Lett. **84**, 1 (2000).
27. Gerdau, E., Rüffer, R., Winkler, H., Tolksdorf, W., Klages, C.P., Hannon, J.P. : Nuclear Bragg diffraction of synchrotron radiation in Yttrium Iron Garnet. Phys. Rev. Lett. **54**, 835 (1985).
28. J.J. Duistermaat, in Huygens' Principle 1690-1990: Theory and Applications, H. Blok, H. Ferwerds, and H.K. Kuiken eds., (Elsevier, 1992), p. 273.
29. Hadamard, J.: Le problème de Cauchy et les Equations aux Dérivées partielles linéaires hyperboliques (reprint of the 1923 publication). Dover, New York, (1952).
30. Tonomura, A., Endo, J., Matsuda, T., and Kawasaki, T.: Demonstration of single-electron buildup of an interference pattern Am. J. Phys. 57, 117 (1989).
31. Tonomura, A., Endo, J., Matsuda, T., Kawasaki, T.: https://www.youtube.com/watch?v=1L VkQfCptEs (video).
32. Feynman, R.P.: Feynman's thesis, ed. by L.M. Brown. World Scientific, Singapore, (2005).
33. Longhurst, R.S.: Geometrical and Physical Optics, 2nd edition. Longman Group, London, (1967).
34. Feynman, R.P.: QED: The Strange Theory of Light and Matter. Princeton University Press (1985).
35. Cramer, J.: The transactional interpretation of quantum mechanics, Rev. Mod. Phys. **58**, 795 (2009).
36. Dirac, P.A.M.: The Langrangian in Quantum Mechanics. Physikalische Zeitschrift der Sowjetunion, Band 3, Heft 1 (1933).
37. Coddens, G.: A linearly polarized electromagnetic wave as a swarm of photons half of which have spin -1 and half of which have spin +1. https://hal.archives-ouvertes.fr/hal-0 2636464v3.
38. Rueckner, W., Peidle, J.: A lecture demonstration of single photon interference. Am. J. Phys. **81**, 951 (2013).
39. Ballentine, L.E.: The Statistical Interpretation of Quantum Mechanics, Rev. Mod. Phys. **42**, 358 (1970).
40. Ballentine, L.E.: Quantum Mechanics, A Modern Development, 2nd edition. World Scientific, Singapore, (1998).
41. Gödel, K.: On Formally Undecidable Propositions of Principia Mathematica and Related Systems. Dover, New York, (1992).
42. Dieudonné, J.: Pour l'honneur de l'esprit humain - les mathématiques aujourd'hui. Hachette, Paris,(1987).
43. Coxeter, H.S.M.: Introduction to Geometry, Second Edition, Wiley Classics Library. John Wiley & Sons, New York (1989).
44. Harrigan, N., Spekkens, R.W.: Einstein, Incompleteness, and the Epistemic View of Quantum States, Found. Phys. **40**, 125 (2010).
45. Cornwell, J.F.: Group Theory in Physics. Academic Press, London, UK (1984).
46. Jones, H.F.: Groups, Representations and Physics. Adam Hilger, Bristol, UK (1990).
47. Chaichian, M., Hagedorn, R.: Symmetries in Quantum Mechanics, from Angular Momentum to Supersymmetry. IOP, Bristol, UK, (1998).
48. van der Waerden, B.L.: Group Theory and Quantum Mechanics. Springer, Berlin-Heidelberg (1974). Translated from: Die gruppentheoretische Methode in der Quantenmechanik. Springer, Berlin (1932).
49. Sagan, B.E.: The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions; Springer Graduate Texts in Mathematics, Volume **203**, 2nd ed.; Springer, New York, NY, USA (2001).
50. Łobaczewski, A.M.: Political Ponerology, Ed. Laura Knight-Jadcyk, 2nd ed., Red Pill Press, Grande Prairie, Canada (2007).