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

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

27th June 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 the first rigorous mathematical proof ever 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 Methodology

The double-slit experiment has been qualified by Feynman [1,2] as the only mystery of quantum mechanics (QM). This mystery resides in an apparent paradox between the QM result and what we expect on the basis of our 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 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. It is 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 these axiomatics, 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 appealing 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 [3]). 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 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). 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 [4].[1]

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 the reasonings, 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 the change of perspective, 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 and in a slight *abus de language*) a "meta" analysis which adopts changes of perspective as evoked here "metamathematical". Adopting the term "metaphysical" was out of the question because it means something entirely different. From now on we will drop the quotes when we use the word metamathematical.[2]

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. We have proposed a reconstruction of QM in [5,6,7]. 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 [7,8]. This is pure mathematics and it can be easily checked whether it is right or

[1] 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 [4]. 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 we will have put him into a 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.

[2] I cannot insist enough that the contents of [5] are a prerequisite for reading the present paper, because it is based on [5] in certain very important points, and it is unfortunately just impossible to copy into the present paper the large parts from [5] that would be needed to make it self-contained. Doing this would expose the paper to easy opprobrium of its prohibitive length and its ethics (self-plagiarism). Some of the contents of [5] are truly beyond guessing and there is no royal short-cut to it such that if the reader skips reading [5] he will almost certainly feel upset by some "unsettling" statements in the present article, but wrongly so. Let us be brutally honest. If the reader does not want to make the effort of first reading [5] completely, he must stop reading right now. Anything else will be a waste of his time.

wrong. Using the geometrical meaning of spinors we have been able to derive the Dirac equation from scratch [5,7]. 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 exotic 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".

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. This does not leave any space 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 [5] the isomorphism is provided by the geometrical meaning of spinors, which is beyond the reach of physical re-interpretation or discussion.

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 sect. 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 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 [5]. This actually means that we have understood the double-slit experiment. The reader can appreciate that the chain of reasoning described here depends crucially on the truth of the pretended contents of [5], which is why we insisted in footnote 2 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 is (1) more cursory and far less formal, (2) more based 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 searched 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. 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 the rather casual approach we often take to probability calculus could put us on somewhat shaky grounds.

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 to not get lost in the complexity of the ramifications, the main structure of the reasoning is therefore summarized in sect. 10. A very beautiful result is the proof obtained in subsect. 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.

1.2 Limitations and clarifications

1.2.1 Use of spinors

As mentioned, we have used the meaning of spinors in the rotation group SO(3) and the homogeneous Lorentz group SO(3,1) to propose a reconstruction of QM in [5,7]. When we speak about wave functions in this paper, we therefore think primarily about 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 $\psi \in \mathrm{F}(\mathbb{R}^4, \mathbb{C}^2)$:

$$\psi(\mathbf{r},t) = e^{\imath(Et-\mathbf{p}\cdot\mathbf{r})/\hbar} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{by the short-hand:} \quad \eta \in \mathrm{F}(\mathbb{R}^4, \mathbb{C}): \quad \eta(\mathbf{r},t) = e^{\imath(Et-\mathbf{p}\cdot\mathbf{r})/\hbar}. \tag{1}$$

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 actually been written with electrons in mind, whose wave functions are spinors, rather than scalar wave functions. We will therefore very often refer to results that apply for spinors.

1.2.2 Generality

Based on this statement one could argue that our approach is not general. We have e.g. excluded bosons from the considerations. The derivation of the three equations based on [5] is the only one that is rigorous and deductive. Before our derivation of the Dirac equation from scratch in [5] 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. All this implies 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 [5] 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 has not been explained. This facade 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. This 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. In fact, a criticism that the present approach would not be general can only reveal how gatekeepers of the traditional dogma with its guessed equations and all its internal contradictions claim generality based on the pseudo-mathematical generalization procedures we will describe in sub-subsect. 1.2.3 while they deny the right to exist to any attempt of explaining what QM really means based on rigorous mathematics and derived equations. However, we will provide a better 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 the fact that summing or making linear combinations of spinors is *a priori* meaningless in the group representation theory.

[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.

1.2.3 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. [14], which explicitly claims that spinors form a vector space). It may sound heretic, which is one of the reasons why I insisted in footnote 2 to read [5], where it is developed in sect. 2.5 starting from p.12. We have developed this further in [9].

Before screaming blue murder, the reader should perhaps make up his mind about what is to be qualified as more loony: a rigorous clarification of the geometrical meaning of spinors that flies in the face of unsubstantiated beliefs or postulating the existence of parallel worlds, advanced waves, electrons that are both particles and waves, and cats that are half dead and half alive. The doctrine that spinors can be summed and would be just vectors in Hilbert space is deeply entrenched, not in the least, perhaps, because there exists even a book written by Dirac, entitled "Spinors in Hilbert space" [10]. But this notion is a historical error based on ignorance ([9] and 5). 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. It is therefore admittedly a real watershed to discover so many years later that one cannot sum spinors as vectors as physicists routinely do. We invite the reader to get over it. As he will see it is not as challenging a total disaster he might fear it is.

We will further discuss this issue (which is completely settled in [9]) in sub-subsect. 3.4.2. 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 [11] about them?

As they might experience these mathematical facts as a source of angst, physicists may prefer to stay within their comfort zone by considering them as a nuisance and belittling them as mathematical faultfinding, just like they often ignore the embarrassing fact that the so-called "Dirac delta function" cannot possibly exist. A nice example of how some physicists on their high horse may justify such an irrational attitude of blunt denial is given by the following quote from [12] about this "Dirac delta function":

"This idea caused great distress to mathematicians, some of whom even declared that Dirac was wrong despite the fact that he kept getting consistent and useful results. Wisely, the physicists rejected these extreme criticisms and followed their intuition."

Note the use of the word "wisely". The truth is perhaps more mundane. In fact, the theory of distributions developed by Laurent Schwartz [13], to justify Dirac's cavalier methods, was such a trivial and trite minor feat that it owed him the Fields medal. Another example are the contorsions they use to refuse to admit that $\sum_{n=1}^{\infty} = -\frac{1}{12}$ is just wrong. But whistleblowing always disturbs and in the present context, nothing is more easy than palming off high-handedly the fact that making linear combinations of spinors is not a defined operation. To turn the tables, it suffices to postulate *ad hoc* that it is the notion of a vector in Hilbert space which is the true definition of a spinor. Everything that is wrong becomes then right by definition, *and vice versa*. After all, why should one question (critically) what one has learned (uncritically). In other words, proofs are not good enough. Well, nobody has a license to pull out of a magic hat an "alternative truth" that would allow one to second-guess the definition of spinors and overrule the mathematics. Such attitudes must be stamped out. [4] Apart from this fact that nobody is entitled to fiddle the mathematics at whim, a good reason for being a bit less self-righteous, is that this way the whole reconstruction of QM based on the derivation of the Dirac equation from scratch given in [5] is washed overboard and with it, the whole geometrical interpretation of the formalism of QM naturally provided by the group theory, while there is not the slightest alternative approach available to make really sense of QM. It is perhaps also instrumental to remind that solving a paradox is all about the aspiration of "finding faults", whereby the most subtle mathematical details may come into play and count, and that the solution has to be mathematical, rather than physical as pointed out above. By refusing the mathematical solution by disqualifying it with cheap hearsay and polemics, one just slams the door to the only logical loophole of escape from the paradox of the double-slit experiment. It reflects a total unawareness of the fact that it is possible to reach a perfect dialogue between the algebra and the geometry or between the mathematics and the physics. The reader should realize that he cannot have his cake and eat it too. One cannot feel entitled to be spoon-fed with an explanation of QM that is in perfect alignment with a utopian personal agenda of pet preconceptions, while the internal contradictions built into these preconceptions are the very cause of our conceptual problems with QM. And one cannot just dismiss the mathematical truths rigorously proved in [5] as part of such an agenda.

I can reassure all physicists in a way that is similar to the way they have been reassured by the theory of distributions of Laurent Schwartz [13] about their use of the "delta function". Indeed, part of the present article aims exactly at repairing for the historical error, by showing on a case example that physicists can "wisely" continue to carry out their algebra on the wave functions as they have done up to now and by acting as though spinors really are vectors in

[4] The *ad hoc* claim [15] 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 vector that is being used in SU(2), and which corresponds to the first column of an SU(2) matrix as explained on p. 7 of [5].

Hilbert space. However, this applies only to the calculations. When it comes down to trying to understand 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.

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. In fact, as explained during the derivation of the Dirac equation in subsect. 4.1 of [5], 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.[5]

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 $\psi_j : t \rightarrow \psi_j(t)$ 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 are e.g. representations of SO(2) or U(1) and that these 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 figure out 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 about what we called 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 like robots 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 [9]).

It does not mean anything. It is just mindless, impenetrable and inkhorn, abracadabra calculus, 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 sect. 6.

The results for spinors show that taking advantage of the fact that summing wave functions in a scalar context apparently makes mathematical sense 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.

[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.

1.3 Two possible sources of errors in our intuition

1.3.1 "Non-locality"

In general relativity, which uses Riemannian geometry, we must work with local (and instantaneous) Lorentz frames. The definition of 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 almost feel ensnared by the Orwellian technique of cornering people by depriving them of their language tools.

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 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".

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. [16,17] to test the Bell inequalities. The angles $\varphi^{(A)}$ and $\varphi^{(B)}$ correspond to the orientations of two polarizers A and B which can be as far apart as we like [18,19]. They are thus not defined at the microscopic level of a single interaction point 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".

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 [5, 6]. 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 [5]. 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. That issue is totally out of order and only based on a lack of insight, which is something gatekeepers could keep in mind when they furiously deny or banalize the mathematical, factual truth that spinors cannot be summed (see sub-subsects. 1.2.3 and 3.4.2). To cite Hannah Arendt: *"What confuses them is that my arguments and my approach are different from what they are used to. In other words, the trouble is that I am independent."* It is perhaps high time for them to search their own hearts, because, like it or not, there is unfortunately a cringeworthy laundry list of further, equally shocking errors in standard QM, and what I pointed out in [5,6] is far from exhaustive.[6]

From now on, we will drop the quotes when we use the word non-local.

1.3.2 Contextuality: Bohr's caveat (Conditional probabilities)

Our intuition about probabilities is also 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

[6] E.g. $c\boldsymbol{\alpha} = (c\alpha_x, c\alpha_y, c\alpha_y)$, where $\alpha_x, \alpha_y, \alpha_y, \beta$ are the matrices occuring in the Dirac equation, is believed to be the operator for the electron velocity $\mathbf{v} \in \mathbb{R}^3$. This leads then to the idea of a so-called Zitterbewegung. This is complete nonsense. In reality $(\alpha_x, \alpha_y, \alpha_y)$ just represents the triad of basis vectors $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z) \in \mathbb{R}^9$, and there is absolutely no elbow room for tampering with this factual truth.

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 $\frac{1}{2}\cos^2(\varphi_j^{(A)} - \varphi_k^{(B)})$ defined by four different experimental set-ups, *viz.* the four combinations $(\varphi_j^{(A)}, \varphi_k^{(B)})$ of the settings for the polarizers A and B, within the Bell inequality used by Aspect et al. [16,17] in his photon correlation experiments (see [20,21] and the many references therein).

As will be discussed in sect. 2 and sect. 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 with all these counterintuitive ingredients one can brew a potent potion of a truly magical quantum paradox.

After the derivation of the Dirac equation from scratch in [5] 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. As we will see, these boundary conditions can affect the conditional probabilities in very unexpected ways. This is an observation which points again 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 also confirms the idea that the paradox is a probability paradox. There is thus some self-consistency in our reconstruction of QM.

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 [1,2], 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 [22,23] and of the quantum eraser experiment [24], which can also be understood based on our discussion. Feynman's lecture in the video [2] 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).[7]

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 below).

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

3.1 The ideal logical approach

Let us now take a break until sect. 4, leaving our intuition for what it is, and turn to QM to analyse the last step of our metamathematical analysis described in subsect. 1.1. 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:

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

Here S_j refer to the slits. (We have written Schrödinger equations in eq. 2 and 4 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 [5,8,6]. 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 subsect. 1.1 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 we *assume without proof at this stage* that the exact solution of the equation for ψ_3 yields the interference pattern. We will provide a rigorous proof for it in sub-subsection 3.6.2.

[7] 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 [25].

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:

$$\psi = \sum_{j=1}^{n} c_j \psi_j \quad \Rightarrow \quad p = \begin{cases} \sum_{j=1}^{n} |c_j|^2 \, |\psi_j|^2 & (\textit{incoherent summing}) \\ \left| \sum_{j=1}^{n} c_j \psi_j \right|^2 & (\textit{coherent summing}) \end{cases} \tag{3}$$

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.[8]

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. It is a mystery and we must accept it in an act of faith. There is a particle and a wave, but there is only one electron. That is a quantum mystery which we must accept in an act of faith. Good luck! That this kind of antinomy is still hanging around after a century is very disturbing and makes the solution of the paradox even more difficult. A conceptual clean-up in the form of a deconstruction of the Copenhagen doctrine would be 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:

$$-\frac{\hbar^2}{2m} \Delta\psi + V(\mathbf{r})\,\psi = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi. \tag{4}$$

when all wave functions χ_j are solutions of the *same* Schrödinger equation eq. 4. 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.

[8] We might be taken here further for a ride by the paradigm of wave packets and argue that the electron must be a wave packet that can travel through both slits at the same time and even interfere with itself. 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 [5]. But by explaining the meaning of the superluminal phase velocity c^2/v in sub-subsect. 1.3.1, 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 sub-subsect. 1.3.1. Nevertheless, textbooks tell us that an electron is both a particle and a wave, which is a *contradictio in terminis*.

It is just infringing the rule that we should not mix up conditional probabilities defined by different experimental set-ups. 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. But these boundary conditions for ψ'_j are just those for the solutions ψ_j of the equations for the two single-slit experiments. By imposing the boundary conditions formulated for ψ'_j you just 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 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. You 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 it just resumes to overriding Born's caveat by writing $\psi'_1 = \psi_1$, $\psi'_2 = \psi_2$, and $\psi_3 = \psi_1 + \psi_2$. And most importantly, 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 sect. 8). We will come back on this problem in sub-subsect. 9.2.1. 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 [5,7,8]. 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 [9], see also e.g. eq. 1 in [5]).

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(3,1) 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 sect. 2.5 of [5]. This is further worked out in [9]. 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 sub-subsect. 1.2.3. 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 [5,8] reveals that the purely formal expression $\sum_{j=1}^{n} g_j$ is in reality a definition of the set $\mathcal{S} = \{g_1, g_2, \cdots g_j, \cdots g_n\}$. 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 $h \circ \mathcal{S} = \mathcal{S} \circ h$. 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 $\mathcal{S}$. The probabilities must then be summed *incoherently* as explained in [5,8]. 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: $\mathcal{V}_1 = \{\psi_1(\mathbf{r})\} \neq \emptyset$ & $\mathcal{V}_2 = \{\psi_2(\mathbf{r})\} \neq \emptyset$ & $\mathcal{V}_1 \cup \mathcal{V}_2 = \{\psi_1(\mathbf{r}), \psi_2(\mathbf{r})\} = \emptyset$. 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 [26,27]. 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 $\mathcal{V}_1 = \{\chi_1\} \neq \emptyset$ & $\mathcal{V}_2 = \{\chi_2\} \neq \emptyset$ & $\mathcal{V}_1 \cup \mathcal{V}_2 = \{\chi_1, \chi_2\} = \emptyset$, as explained above.

The sum is thus not only meaningless in physics as pointed out in sub-subsect. 1.2.3, 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 sect. 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 you 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 sect. 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 sub-subsect. 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 sub-subsect. 9.2.1. 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 sub-subsect. 9.2.1 the use of the solution obtained by the Huygens' principle for spinor fields:

$$\begin{array}{rcccl|l} p_3 & = & |\psi_3|^2 & = & |\psi_1 \boxplus \psi_2|^2 & \text{Huygens' principle} \\ & = & |\psi'_1|^2 + |\psi'_2|^2 & = & p'_1 + p'_2 & \text{Boolean algebra} \\ & \neq & |\psi_1|^2 + |\psi_2|^2 & = & p_1 + p_2 & \text{Bohr's caveat, fake superposition.} \end{array} \tag{5}$$

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

We have found an alternative approach to justify the procedure of coherent summing, whereby we do not need the Huygens' principle and an appeal on intuition. It is based on symmetry and as it is much clearer we will describe it now. We were surprised that this idea works. It also confirms that everything we have explained above is correct.

3.6.1 The idea

It often happens that we must take tensor products of identical vector quantities. An example is the tensor product of two identical spinors in SU(2). Let us note the spinor as:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix}. \tag{6}$$

Then

$$\boldsymbol{\xi} \otimes \boldsymbol{\xi} = \begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix} \otimes \begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix} = \begin{bmatrix} \xi_0^2 \\ \xi_0 \xi_1 \\ \xi_1 \xi_0 \\ \xi_1^2 \end{bmatrix}. \tag{7}$$

When we build the spinor starting from an isotropic vector $(x, y, z) = \mathbf{e}_x + \imath\mathbf{e}_y$ we have (see [5], p. 21, equation 31):

$$x = \xi_0^2 - \xi_1^2, \quad y = \imath(\xi_0^2 + \xi_1^2), \quad z = -2\xi_0\xi_1. \tag{8}$$

The tensors of even rank 2ℓ correspond to the harmonic polynomials $Y_{\ell m}$. As $\xi_0\xi_1 = \xi_1\xi_0$, using the 4×1 matrix in eq. 7 is redundant. We could take care of this by e.g. bluntly introducing:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{M}_{3\times 4}} \begin{bmatrix} \xi_0^2 \\ \xi_0\xi_1 \\ \xi_1\xi_0 \\ \xi_1^2 \end{bmatrix} = \begin{bmatrix} \xi_0^2 \\ \xi_0\xi_1 \\ \xi_1^2 \end{bmatrix} . \tag{9}$$

The inconvenience of this is that the 3×4 matrix $\mathbf{M}_{3\times 4}$ is not invertible because it is not a square matrix. We encounter e.g. similar problems in solid-state physics, where the symmetry of the 3×3 matrix which represents the rank-2 Cauchy stress tensor is used to write it in a reduced form as a 6×1 six-dimensional vector in the so-called Voigt notation. When the tensor undergoes a transformation, one must be able to express this transformation also in the reduced formalism of the six-dimensional vectors. It is then necessary to be able to switch backwards and forwards between the Voigt and Cauchy notations. We can obtain the required invertible switch formalism by making use of the n^{th} roots of unity in $\mathbb{C}$.

Let us e.g. consider $\boldsymbol{\xi} \otimes \boldsymbol{\xi} \otimes \boldsymbol{\xi}$. Here certain quantities will occur with some multiplicities. E.g. $\xi_0^2\xi_1$ will occur three times, *viz.* as $\xi_0\xi_0\xi_1$, $\xi_0\xi_1\xi_0$ and as $\xi_1\xi_0\xi_0$. To reduce these three terms to a single one in an invertible way we can use a 3×3 matrix $\mathbf{M}_{3\times 3}$:

$$\frac{1}{3}\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{\imath 2\pi/3} & e^{\imath 4\pi/3} \\ 1 & e^{\imath 4\pi/3} & e^{\imath 8\pi/3} \end{bmatrix}}_{\mathbf{M}_{3\times 3}} \begin{bmatrix} \xi_0\xi_0\xi_1 \\ \xi_0\xi_1\xi_0 \\ \xi_1\xi_0\xi_0 \end{bmatrix} = \xi_0^2\xi_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} . \tag{10}$$

With:

$$\mathbf{v}_0 = \begin{bmatrix} \xi_0^3 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} \xi_0\xi_0\xi_1 \\ \xi_0\xi_1\xi_0 \\ \xi_1\xi_0\xi_0 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} \xi_0\xi_1\xi_1 \\ \xi_1\xi_0\xi_1 \\ \xi_1\xi_1\xi_0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} \xi_1^3 \end{bmatrix}, \tag{11}$$

we can then introduce the block-diagonal matrix:

$$\begin{bmatrix} 1 & & & \\ & \mathbf{M}_{3\times 3} & & \\ & & \mathbf{M}_{3\times 3} & \\ & & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} . \tag{12}$$

This invertible block-diagonal matrix will take care of the full size-reduction of the 8-component tensor to a 4-component one. We can reorder in a second stage the components of the reduced 4-component tensor in such a way that the 4 non-zero entries are located in the first lines at the top of the column matrix and the 4 remaining zero entries relegated to the last lines at the bottom of the column matrix. Such a permutation of the entries is also an invertible transformation. The general expression for the $n \times n$ matrices we must use to reduce n equal terms to a single one in an invertible way is given by:

$$[\,\mathbf{M}\,]_{jk} = \frac{1}{\sqrt{n}} e^{\imath\frac{2\pi}{n}(j-1)(k-1)}, \forall(j,k) \in ([1,n] \cap \mathbb{N})^2. \tag{13}$$

Their inverse is given by:

$$[\,\mathbf{M}^{-1}\,]_{km} = \frac{1}{\sqrt{n}} e^{-\imath\frac{2\pi}{n}(k-1)(m-1)}, \forall(k,m) \in ([1,n] \cap \mathbb{N})^2. \tag{14}$$

This is easily proved by $\forall(j,m) \in ([1,n] \cap \mathbb{N})^2$:

$$[\,\mathbf{M}\mathbf{M}^{-1}\,]_{jm} = \frac{1}{n}\sum_{k=1}^{n} e^{\imath\frac{2\pi}{n}(j-1)(k-1)} e^{-\imath\frac{2\pi}{n}(k-1)(m-1)} = \frac{1}{n}\sum_{k=1}^{n} e^{\imath\frac{2\pi}{n}(j-m)(k-1)} = \delta_{jm}. \tag{15}$$

The matrices $\mathbf{M}$ are thus indeed invertible. We have actually $\mathbf{M}^{-1} = \mathbf{M}^{\dagger}$. We have used the normalization factors $\frac{1}{\sqrt{n}}$ rather than $\frac{1}{n}$ in the definition of $\mathbf{M}$ in order to obtain a more symmetrical formalism. We can use this formalism each time we want to group together entries that we want to consider as indistinguishable.

3.6.2 Application to the double-slit experiment - THE proof of the coherent sum rule

We consider the wave functions ψ_1, ψ_2 and ψ_3 defined in eq. 2. If we cannot make a distinction between the paths through the two slits in the set-up because the interactions are coherent, we must reduce the contributions ψ_1' and ψ_2' to ψ_3 to a single quantity. We want to put them into a same basket because due to the coherence they really ought to be treated within a single basket (see sect. 2 and sect. 8). It runs contrary to the physical reality and the logic to consider two separate baskets, because we are in the impossibility to know which way the particle travelled through the apparatus. We need the 2×2 matrices:

$$\mathbf{M} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{M}^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{16}$$

to reduce two quantities reversibly to a single one. Logically, we would have to apply this to the contributions ψ_1' and ψ_2' to ψ_3. In what follows we will not stick rigorously to this logic of contributions to ψ_3. We will do something weird, that will produce a little miracle. We introduce the shorthands:

$$\mathscr{O}_j = -\frac{\hbar^2}{2m}\Delta + V_j(\mathbf{r}), \quad \forall j \in \{1, 2\}. \tag{17}$$

Let us now write simultaneously the two single-slit Schrödinger equations for ψ_1 and ψ_2 as:

$$\begin{bmatrix} \mathscr{O}_1 & \\ & \mathscr{O}_2 \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}. \tag{18}$$

We have explained in sub-subsect. 3.4.1 that we are unable to write Schrödinger equations with appropriate boundary conditions for ψ_1' and ψ_2'. Let us now apply the 2×2 matrix $\mathbf{M}$ in eq. 16 to both sides of eq. 18:

$$\mathbf{M} \begin{bmatrix} \mathscr{O}_1(\mathbf{r}) & \\ & \mathscr{O}_2(\mathbf{r}) \end{bmatrix} \mathbf{M}^{-1} \cdot \mathbf{M} \begin{bmatrix} \psi_1(\mathbf{r}) \\ \psi_2(\mathbf{r}) \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\mathbf{M} \begin{bmatrix} \psi_1(\mathbf{r}) \\ \psi_2(\mathbf{r}) \end{bmatrix}. \tag{19}$$

Note that this is different of what we had in mind, because we are doing the operations on ψ_1 and ψ_2 rather than on the contributions ψ_1' and ψ_2' to ψ_3. One could qualify what we do as conceptually scatterbrained, because we are so to say making linear combinations of experimental set-ups. The calculation yields:

$$\frac{1}{2} \begin{bmatrix} \mathscr{O}_1(\mathbf{r}) + \mathscr{O}_2(\mathbf{r}) & \mathscr{O}_1(\mathbf{r}) - \mathscr{O}_2(\mathbf{r}) \\ \mathscr{O}_1(\mathbf{r}) - \mathscr{O}_2(\mathbf{r}) & \mathscr{O}_1(\mathbf{r}) + \mathscr{O}_2(\mathbf{r}) \end{bmatrix} \begin{bmatrix} \frac{\psi_1(\mathbf{r})+\psi_2(\mathbf{r})}{\sqrt{2}} \\ \frac{\psi_1(\mathbf{r})-\psi_2(\mathbf{r})}{\sqrt{2}} \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t} \begin{bmatrix} \frac{\psi_1(\mathbf{r})+\psi_2(\mathbf{r})}{\sqrt{2}} \\ \frac{\psi_1(\mathbf{r})-\psi_2(\mathbf{r})}{\sqrt{2}} \end{bmatrix}. \tag{20}$$

We introduce the notations:

$$\begin{bmatrix} \frac{\psi_1(\mathbf{r})+\psi_2(\mathbf{r})}{\sqrt{2}} \\ \frac{\psi_1(\mathbf{r})-\psi_2(\mathbf{r})}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \psi(\mathbf{r}) \\ \chi(\mathbf{r}) \end{bmatrix}. \tag{21}$$

Eq. 20 must be examined on slit S_1, on slit S_2, on the solid part $A\backslash(S_1 \cup S_2)$ of the double-slit apparatus A where there is no slit, and on the detector screen D, which we consider to be far behind the slits such that each point of it is exposed to both ψ_1 and ψ_2 (and thus situated in the interference region). Now we have:

$\forall \mathbf{r} \in S_1$	$V_1(\mathbf{r}) = 0$	$\psi_1 \neq 0$
$\forall \mathbf{r} \in A\backslash(S_1 \cup S_2)$	$V_1(\mathbf{r}) = V$	$\psi_1 = 0$
$\forall \mathbf{r} \in S_2$	$V_1(\mathbf{r}) = V$	$\psi_1 = 0$
$\forall \mathbf{r} \in S_1$	$V_2(\mathbf{r}) = V$	$\psi_2 = 0$
$\forall \mathbf{r} \in A\backslash(S_1 \cup S_2)$	$V_2(\mathbf{r}) = V$	$\psi_2 = 0$
$\forall \mathbf{r} \in S_2$	$V_2(\mathbf{r}) = 0$	$\psi_2 \neq 0$
$\forall \mathbf{r} \in D$	$V_1(\mathbf{r}) = 0$	$\psi_1 \neq 0$
$\forall \mathbf{r} \in D$	$V_2(\mathbf{r}) = 0$	$\psi_2 \neq 0$

(22)

For ψ_1 and ψ_2 really to become zero on $A\backslash(S_1 \cup S_2)$ the value of V must actually be ∞. An extension $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ of $\mathbb{R}$ can be mathematically defined. We will then have to remember that $V - V = 0$. We can avoid making use of this by making V monstrously large, e.g. the term t_a or even the term t_{t_a} in the series:

$$t_1 = 1, \quad t_2 = 2^2, \quad t_3 = 3^{3^3}, \quad t_4 = 4^{4^{4^4}}, \cdots \tag{23}$$

Where $a = t_{10}$. The zero values for ψ_1 and ψ_2 will then not be exact but approximations of a mind-boggling precision. On S_1 we have thus:

$$\frac{1}{2}\begin{bmatrix} 2\Delta+V & -V \\ -V & 2\Delta+V \end{bmatrix}\begin{bmatrix} \psi_1(\mathbf{r}) \\ \psi_1(\mathbf{r}) \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\begin{bmatrix} \psi_1(\mathbf{r}) \\ \psi_1(\mathbf{r}) \end{bmatrix}. \tag{24}$$

This resumes to:

$$\forall \mathbf{r} \in S_1 : \mathscr{O}_1\psi_1(\mathbf{r}) = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi_1(\mathbf{r}), \tag{25}$$

which is just the Schrödinger equation for ψ_1 on S_1. Furthermore, as $\forall \mathbf{r} \in S_1$ we have $\psi_1(\mathbf{r}) - \psi_2(\mathbf{r}) = \psi_1(\mathbf{r}) + \psi_2(\mathbf{r})$, we can also write both equations as:

$$\forall \mathbf{r} \in S_1 : \mathscr{Q}\psi(\mathbf{r}) = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi(\mathbf{r}), \tag{26}$$

where $\psi = \psi_1 + \psi_2$, $\mathscr{Q} = -\frac{\hbar^2}{2m}\Delta + V_3(\mathbf{r})$ and V_3 is the potential of the double-slit experiment. Here $V_3(\mathbf{r}) = 0$ on S_1. On S_2 we have now:

$$\frac{1}{2}\begin{bmatrix} 2\Delta+V & V \\ V & 2\Delta+V \end{bmatrix}\begin{bmatrix} \psi_2(\mathbf{r}) \\ -\psi_2(\mathbf{r}) \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\begin{bmatrix} \psi_2(\mathbf{r}) \\ -\psi_2(\mathbf{r}) \end{bmatrix}. \tag{27}$$

This resumes to:

$$\forall \mathbf{r} \in S_2 : \mathscr{O}_2\psi_2 = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi_2, \tag{28}$$

which is just the Schrödinger equation for ψ_2 on S_2. As $\forall \mathbf{r} \in S_2$ we have $\psi_1(\mathbf{r}) - \psi_2(\mathbf{r}) = -(\psi_1(\mathbf{r}) + \psi_2(\mathbf{r}))$, we can write both equations as:

$$\forall \mathbf{r} \in S_2 : \mathscr{Q}\psi = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi. \tag{29}$$

On $A\backslash(S_1 \cup S_2)$ we have:

$$\frac{1}{2}\begin{bmatrix} 2\Delta+2V & 0 \\ 0 & 2\Delta+2V \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{30}$$

This resumes to:

$$\forall \mathbf{r} \in A\backslash(S_1 \cup S_2) : \mathscr{Q}\psi = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi, \tag{31}$$

with a second independent but similar equation for χ. On the detector screen D we have now:

$$\frac{1}{2}\begin{bmatrix} 2\Delta & 0 \\ 0 & 2\Delta \end{bmatrix}\begin{bmatrix} \psi(\mathbf{r}) \\ \chi(\mathbf{r}) \end{bmatrix} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\begin{bmatrix} \psi(\mathbf{r}) \\ \chi(\mathbf{r}) \end{bmatrix}. \tag{32}$$

This leads to:

$$\forall \mathbf{r} \in D : \mathscr{Q}\psi = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\psi. \tag{33}$$

and a second independent equation for χ:

$$\forall \mathbf{r} \in D : \mathscr{Q}\chi = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\chi. \tag{34}$$

Recollecting eqs. 26, 29, 31, 33, we see that we have found a solution ψ of the Schrödinger equation for ψ_3 that satisfies the boundary conditions for the double-slit experiment, just by combining solutions of the single-slit experiments. This is perhaps not something we could have had in mind as a visionary goal beforehand, given the smoke screen of calculations one has to plow through in order to check all the possible separate cases for the boundary conditions before one can reach the final conclusion. I hope the reader will realize what a gift from heaven this is!

The conditions on A correspond to a near-field solution. The conditions on D have been selected in order to correspond to a far-field solution, by assuming that the screen is located far behind the slits. The solution for the intermediate region is not obtained, but this is also not necessary because we do not measure what happens there. We only measure what happens on the detector screen D. If we do not put the detector screen in the far-field region, it will be difficult in some points $\mathbf{r}$ behind the slits to know if $\psi_1(\mathbf{r}) = 0$ and/or $\psi_2(\mathbf{r}) = 0$ for the given experimental

set-up. It is only in the far-field solution that we can be sure of that both $\psi_1(\mathbf{r}) \neq 0$ and $\psi_2(\mathbf{r}) \neq 0$. The solution does not rely on the superposition principle or on the use of sets but on a symmetry induced by the coherence. It is the first rigorous proof ever that $\psi_3 = \frac{\psi_1+\psi_2}{\sqrt{2}}$ is that solution. Note that $e^{\imath\varphi_1}\psi_1$ and $e^{\imath\varphi_2}\psi_2$ are also valid solutions of the single-slit experiments, such that we could *a priori* end up with $e^{\imath\varphi_1}\psi_1 + e^{\imath\varphi_2}\psi_2$. To make sure that we will really deal with $\psi_1 + \psi_2$ in our calculations we must make the phases of ψ_1 and ψ_2 identical at the source.

With hindsight we can get a feeling for the result obtained, by considering $\mathbf{M}$ as taking care of a change of basis, wherein the functions ψ and χ are the new basis vectors. It takes thereby simultaneously care of creating a set of equations for the new basis vectors ψ and χ. What is strange with this is that the basis vectors we are combining linearly are not taken from the same experiment, i.e. not the same (probability) space, such that we have proceeded in a completely quirky way. We have treated the two separate experiments with their separate probability spaces simultaneously within a higher-dimensional space. To put it grandiloquently, we have introduced a Hilbert space which is the direct sum of two Hilbert spaces. But this formulation is not very thoughtful, because it is obvious that $e^{\imath\varphi_1}\psi_1 + e^{\imath\varphi_2}\psi_2$ which also belongs to the direct sum corresponds to experiments that cannot be carried out. It is something for the beaming adepts of the "shut and calculate!" leitmotiv. In taking this intellectual quantum leap we have been able to stop reasoning on one single equation and actually treat simultaneously three different equations corresponding to three different experiments (if we consider the experiment that would lead to the result $\psi_1 - \psi_2$ as not feasible). We have found a technique to circumvent Bohr's caveat. Retrospectively we can then appreciate that the equation for ψ could have been anticipated to be the Schrödinger equation for the double-slit experiment, because ψ has always been used as the solution of that equation in QM, and this solution has been empirically shown to be in agreement with experiment. But I did not expect that mathematics could do this for me. We have taken apart equations and reassembled their parts to make new equations as though they were pieces of a Meccano game. Of course, this relies on the fact that also the operators $\mathscr{O}_j$ are building a vector space.

Note that this an algebraic solution of the Schrödinger equation for ψ_3, whereby we have ignored the fact that summing spinors and even representations of group elements of SO(2) has no geometrical meaning. But the Schrödinger equation for ψ_3 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! What historically has been introduced based on an intuition based on Copenhagen concepts like the particle-wave duality which do not have any rigorous justification, has now been proved rigorously. And as anticipated in the Introduction, there is absolutely no quantum magic involved in this solution. Just like we promised it is entirely accessible to classical common-sense intuition as it relies only on pure logic and pure mathematics. We have found a rigorous justification for the coherent sum rule. It would be fun to figure out if we can generalize these methods to the case of a periodic array of n slits of equal width. In principle we have the tools to do it and in principle it should work. But let us not forget: We must still give the purely algebraic solution a geometrical meaning in terms of group representations, in conformity with what we pointed out in the discussion of Huygens' principle.

4 Local interactions, non-local probabilities

The answer to Feynman's question quoted at the end of subsect. 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 sub-subsect. 1.3.1. 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 QM probabilities are defined with respect to the global geometry of the set-up. This global geometry is fundamentally non-local in the sense that the local interactions of an electron cannot be affected by all aspects of the geometry. Due to this fact the ensuing probability distribution is also non-locally defined. 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. 35). 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!) [15,28].[9]

[9] Fig. 3 below shows interference patterns in two double-slit experiments that differ only in the distance d between the two slits. In observing the marked difference between the two interference patterns we could also ask the question: How does a particle

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 [29,30] (and especially the video in [31]) 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 sub-subsect. 1.3.1 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 subsect. 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 [26,27].

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:

---

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, 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 be just $\psi_q(x) = e^{\imath q x}$, with $q \in \mathbb{R}$. We see thus that a macroscopic length L operates a selection on the continuum of eigenfunctions ψ_q, despite the fact that the interactions might be just defined between first neighbours. The selection is based on the boundary conditions.

$$\frac{d\sigma}{d\Omega} = \frac{m_0}{4\pi^2} \, |\mathscr{F}(V_s)(\mathbf{q})|^2, \tag{35}$$

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 subsect. 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.[10]

The Born approximation provides us with a second 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 [15,28]. 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. 35 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. 35. People who try to synthesize new materials do indeed use characterization by X-ray diffraction to study what they have manufactured by using eq. 35, this time not unwittingly but really on purpose. That is how quasicrystals were discovered.

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 subsect. 4.3 of [5] 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.

[10] We can consider the set-up to be of zero thickness in the z-direction and to be confined to the Oxy-plane. We can then define a model potential $V_3(x,y)$ on the Oxy-plane as follows (see subsect. 3.6): $\forall x \in S = [-b,-a] \cup [a,b] : V_3(x,y) = 0$, $\forall x \in \mathbb{R}\backslash S : V_3(x,y) = V$. Here $\mathcal{S} \times \mathbb{R} = S_1 \cup S_2$ 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 $\mathbf{r}_0$ 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.

6 Highly simplified descriptions still catch the essence

We can now address the problem of generality discussed in sub-subsect. 1.2.2 on which we promised to come back towards the end of sub-subsect. 1.2.3. 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 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 (see fig. 3).

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 [5] and the Appendix of [6]. 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?[11]

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. 5 in subsect. 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 sub-subsect. 3.4.1 and to some extent in what we explained in subsect. 2.3 of [5], 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 [1], to show that there is a one-to-one correspondence between undecidability and coherence of the interactions with the set-up.

[11] 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.

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 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 sub-subsect. 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. 2 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. 5.

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 subsect. 9.2; 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 subsect. 9.2).

If all this sounds esoteric or not very convincing, the reader will change his mind after reading sub-subsect. 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 TL + 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 TL + 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. In a rigorous formulation, this undecidability becomes a consequence of the fact that the wave function must be a function, because it is the integral transform of the potential, which must represent all the information about the set-up and its built-in undecidability. As the phase of the wave function corresponds to the spin angle of the electron, even changes in this angle are uniquely defined.

9 The correct analysis of the experiment

9.1 An Aharonov-Bohm type argument

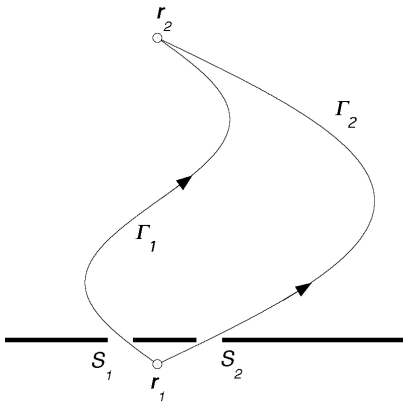

Fig. 1. Two paths Γ_1 and Γ_2 between the points S (with position vector $\mathbf{r}_1$) and P (with position vector $\mathbf{r}_2$) that are threading through the slits of the double slit experiment. The slits have been labeled S_1 and S_2. In the loop integral in eq. 37 the sense of Γ_2 is inverted. This loop cannot be shrunk continuously to a point, because it encircles the matter of the set-up in between the two slits.

This idea is worked out in reference [7], pp. 329-333, and depends critically on the fact that the space traversed by the electrons that end up in the detector is not simply connected. It is based on the simplifying *Ansatz* that the way the electron travels through the set-up from a point S (with position vector $\mathbf{r}_1$) before the slits to a point P (with position vector $\mathbf{r}_2$) behind the slits has no incidence whatsoever on the phase difference of the wave function between $\mathbf{r}_1$ and $\mathbf{r}_2$. The idea is based on an Aharonov-Bohm type of argument: Because the wave function is single-valued, we must on two alternative paths Γ_1 and Γ_2 between the points $\mathbf{r}_1$ and $\mathbf{r}_2$ (see fig. 1) obtain a phase difference:

$$\frac{1}{\hbar}\int_{\Gamma_1}[\,Edt-\mathbf{p}\cdot d\mathbf{r}\,]-\frac{1}{\hbar}\int_{\Gamma_2}[\,Edt-\mathbf{p}\cdot d\mathbf{r}\,]=2\pi n,\quad\text{with:}\quad n\in\mathbb{Z}. \tag{36}$$

The union of the two paths defines a loop, such that we can write eq. 36 under the form:

$$\frac{1}{\hbar}\oint_{\Gamma_1\cup\Gamma_2}[\,Edt-\mathbf{p}\cdot d\mathbf{r}\,]=2\pi n,\quad\text{with:}\quad n\in\mathbb{Z}, \tag{37}$$

whereby the sense of running through Γ_2 has been inverted. In a single-slit experiment this loop can be shrunk continuously to a point which can be used to prove that $n=0$. In fact, the phase cannot make a jump of 2π when the loop is changed by an infinitesimal amount (if we avoid the locations where the amplitude of the wave function becomes zero). In a double-slit experiment the loop cannot be shrunk to a point when Γ_1 and Γ_2 are threading through different slits, such that $n\neq 0$ becomes then possible. Each interference fringe corresponds to one value of $n\in\mathbb{Z}$. A phase difference of $2\pi n$ occurs also in the textbook approach where one argues that to obtain constructive interference the difference in path lengths behind the slits must yield a phase difference $2\pi n$. But this resemblance does not run deep and is superficial. The textbook approach deals with phase differences between ψ_1 and ψ_2 in special points $\mathbf{r}_2$, while our approach deals with different phases built up over paths Γ_1 and Γ_2 between points $\mathbf{r}_1$ and $\mathbf{r}_2$ within an open set $D_n\subset\mathbb{R}^3$, with $n\in\mathbb{Z}$, which is a definition domain for ψ_3, whereby ∂D_n consists of lines $\psi_3(\mathbf{r})=\psi_1(\mathbf{r})\boxplus\psi_2(\mathbf{r})=0$.

9.2 The axiomatic systems of Einstein and Bohr

9.2.1 The solution obtained from the Huygens' principle can be adopted as a meaningful spinor field

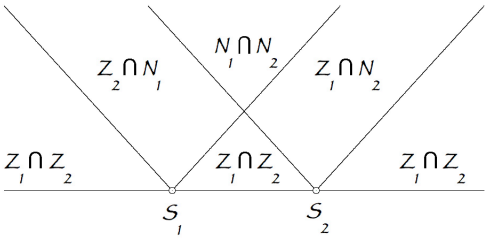

Fig. 2. Schematic diagram illustrating the notations used in the text for the zones of $\mathcal{V}$ considered according to their accessibility to the electrons emerging from the slits S_1 and S_2. The set $\mathcal{N}_1$ is the part of space that can be reached by electrons that traverse S_1. The set $\mathcal{Z}_1$ 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 $\mathcal{N}_1 \cap \mathcal{N}_2$. The shadow zone $\mathcal{Z}_1 \cap \mathcal{Z}_2$, from which the electrons remain absent, consists of three disconnected parts. The diagram is really only schematic because we have represented it for slits with width zero. When the widths are not zero, it is much more difficult to define such a partition for $\mathcal{V}$.

We can approach this somewhat differently, by showing how we can justify that the algebraic result obtained by using the Huygens' principle, the Born approximation, or the reasoning of subsect 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 subsect. 3.5 and 3.6, and we will rely on the analysis laid down in eq. 5.

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 does not know. 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. 5 of subsect. 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]

Let us call the part of $\mathbb{R}^3$ behind the slits $\mathcal{V}$. 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 $\mathcal{V}$ in a region $\mathcal{Z}_1$ where $\psi_1'(\mathbf{r}) = 0$ and a region $\mathcal{N}_1$ where $\psi_1'(\mathbf{r}) \neq 0$ (see fig. 2). We define $\mathcal{Z}_2$ and $\mathcal{N}_2$ similarly. In the region $\mathcal{N}_1 \cap \mathcal{Z}_2$ we can multiply ψ_1' by an arbitrary phase factor $e^{\imath\chi_1}$ without changing $|\psi_1'(\mathbf{r})|^2$. In the region $\mathcal{Z}_1$ 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 $\mathcal{Z}_2$ and $\mathcal{N}_2 \cap \mathcal{Z}_1$.

Let us now address the region $\mathcal{W} = \mathcal{N}_1 \cap \mathcal{N}_2$. 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 $\psi_1'(\mathbf{r}) = |\psi_3(\mathbf{r})| \cos\alpha\, e^{\imath\alpha_1}$, $\psi_2'(\mathbf{r}) = |\psi_3(\mathbf{r})| \sin\alpha\, e^{\imath\alpha_2}$, $\forall \mathbf{r} \in \mathcal{N}_1 \cap \mathcal{N}_2 = \mathcal{W}$. 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 $\mathcal{Z}_2 \cap \mathcal{N}_1$ and for χ_2 in $\mathcal{Z}_1 \cap \mathcal{N}_2$ 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.

Over $\mathcal{N}_1 \cap \mathcal{Z}_2$, we must have $\psi_1'(\mathbf{r}) = \psi_3(\mathbf{r}) = \psi_1(\mathbf{r})$, as $\psi_2(\mathbf{r}) = 0$. Similarly, over $\mathcal{N}_2 \cap \mathcal{Z}_1$, we must have $\psi_2'(\mathbf{r}) = \psi_3(\mathbf{r}) = \psi_2(\mathbf{r})$ as $\psi_1(\mathbf{r}) = 0$. Over $\mathcal{W} = \mathcal{N}_1 \cap \mathcal{N}_2$ integration leads to $\int_{\mathcal{W}} |\psi_1'(\mathbf{r})|^2\, d\mathbf{r} = \cos^2\alpha \int_{\mathcal{W}} |\psi_3(\mathbf{r})|^2\, d\mathbf{r}$ and $\int_{\mathcal{W}} |\psi_2'(\mathbf{r})|^2\, d\mathbf{r} = \sin^2\alpha \int_{\mathcal{W}} |\psi_3(\mathbf{r})|^2\, d\mathbf{r}$. Furthermore, we must have $\int_{\mathcal{W}} |\psi_2'(\mathbf{r})|^2\, d\mathbf{r} = \int_{\mathcal{W}} |\psi_1'(\mathbf{r})|^2\, d\mathbf{r}$, due to the mirror symmetry of the set-up, such that $\alpha = \frac{\pi}{4}$. We see from this that not only $\int_{\mathcal{W}} |\psi_2'(\mathbf{r})|^2\, d\mathbf{r} = \int_{\mathcal{W}} |\psi_1'(\mathbf{r})|^2\, d\mathbf{r} = \frac{1}{2}\int_{\mathcal{W}} |\psi_3(\mathbf{r})|^2\, d\mathbf{r}$, but also $|\psi_2'(\mathbf{r})|^2 = |\psi_1'(\mathbf{r})|^2 = \frac{1}{2}|\psi_3(\mathbf{r})|^2$. In each point $\mathbf{r} \in \mathcal{W}$ 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 have been 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 a solution of a Schrödinger equation with potential V_j, for $j = 1, 2$ cannot be considered as a solution of 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.

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 $\mathcal{W}$:

[12] The argument developed in sub-subsect. 3.4.1 does therefore not hold sway here. In sub-subsect. 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 $\mathcal{N}_1 \cap \mathcal{N}_2$ (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 $\mathcal{N}_1 \cap \mathcal{N}_2$, 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. It must 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 sub-subsect. 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 sub-subsect. 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.

$$\begin{aligned} \psi_1' &= \frac{|\psi_1 \boxplus \psi_2|\, e^{\imath(\chi+\frac{\pi}{4})}}{\sqrt{2}} = \frac{1}{\sqrt{2}}\,(\psi_1 \boxplus \psi_2)\, e^{+\imath\frac{\pi}{4}} \neq \psi_1 \\ \psi_2' &= \frac{|\psi_1 \boxplus \psi_2|\, e^{\imath(\chi-\frac{\pi}{4})}}{\sqrt{2}} = \frac{1}{\sqrt{2}}\,(\psi_1 \boxplus \psi_2)\, e^{-\imath\frac{\pi}{4}} \neq \psi_2 \end{aligned} \tag{38}$$

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 subdivides the region $\mathcal{W} \subset \mathbb{R}^3$ of space behind the slits into open disconnected domains $D_n, n \in \mathbb{Z}$, whereby on each of these domains $|\psi_3| > 0$ and ψ_3 is meaningfully defined. Considering things this way automatically respects the requirement that the loop in fig. 1 should avoid the points where the amplitude of the wave function becomes zero as discussed in subsect. 9.1. Noting the restrictions of ψ_3 to D_n as ζ_n we can consider then $\psi_3 = \sum_{n\in\mathbb{Z}} \zeta_n$ defined on $\cup_{n\in\mathbb{Z}} D_n$ as a sum of wave functions $\zeta_n, n \in \mathbb{Z}$, that can be interpreted in terms of sets. The probabilities must be summed incoherently. These wave functions ζ_n with their limited domains of definition D_n are the true spinor solutions of the Schrödinger equation. Each wave function ζ_n with domain D_n corresponds then to a fringe of the interference pattern, in conformity with the result derived in eq. 37 and subsect. 9.1.

What eq. 38 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 this member from the pool of solutions obtained by the use of the Huygens' principle can be justified as meaningful within the context of spinors. 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 subsect. 3.5 of justifying the use of ψ_3.

9.2.2 Some technicalities

We were able to get an inkling of the possibility of the loophole (that the Huygens' principle yields correct numerical results while the procedure of summing $\psi_3 = \psi_1 \boxplus \psi_2$ has no physically meaningful interpretation) by noticing that $\psi_3 = \psi_1 \boxplus \psi_2$ is not rigorously exact in Feynman's path integral method or in Born's approximation, even if it remains an excellent approximation. In fact, a result that is numerically not rigorously exact cannot correspond to a theoretical principle.

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 $\psi_1 \boxplus \psi_2$.

However, at the boundaries of $\mathcal{N}_1 \cap \mathcal{Z}_2$ and $\mathcal{N}_2 \cap \mathcal{Z}_1$ with $\mathcal{W}$ there are awkward discontinuities. In $\mathcal{N}_1 \cap \mathcal{Z}_2$, we must have $\psi_1'(\mathbf{r}) = \psi_1(\mathbf{r})$, while in $\mathcal{N}_1 \cap \mathcal{N}_2$, we have $\psi_1'(\mathbf{r}) = \frac{1}{\sqrt{2}}\,(\psi_1(\mathbf{r}) \boxplus \psi_2(\mathbf{r}))\, e^{+\imath\frac{\pi}{4}}$. We can consider that we can accept this discontinuity at the boundary, because over $\mathcal{N}_1 \cap \mathcal{Z}_2$, the question through which slit the electron has traveled is decidable, while over $\mathcal{W}$ it is undecidable, such that there is an abrupt change of logical regime at this boundary.

□ Option 1. In reality, the boundary between $\mathcal{W}$ and $\mathcal{N}_1 \cap \mathcal{Z}_2$ could be more diffuse than in the schematic diagram of fig. 2 as a result of an integration over the slits, such that the abruptness is not real. In fact, it is absolutely not straightforward to decide for a point close to slit S_1, whether it belongs to $\mathcal{W}$ or to $\mathcal{N}_1 \cap \mathcal{Z}_2$. The main aim of our calculation is to obtain a qualitative understanding rather than a completely rigorous solution. The same arguments can be repeated at the boundary between $\mathcal{N}_2 \cap \mathcal{Z}_1$ and $\mathcal{W}$. We can thus consider that ψ_1' and ψ_2' will be continuous despite these technical difficulties. Even if they were not continuous, this would only reflect how a change in knowledge changes the definition of the probabilities. If we accept this solution, then $\psi_1'(\mathbf{r})$ will vanish on slit S_2 and $\psi_2'(\mathbf{r})$ will vanish on slit S_1.

□ Option 2. We can also consider these discontinuities as a serious issue. We could then postulate that we must assume that $\mathcal{N}_1 \cap \mathcal{Z}_2 = \emptyset$ & $\mathcal{N}_1 \cap \mathcal{Z}_2 = \emptyset$, in order to avoid the discontinuities. The fact that we have to choose $\mathcal{N}_1 \cap \mathcal{Z}_2 = \emptyset$ & $\mathcal{N}_1 \cap \mathcal{Z}_2 = \emptyset$ would then be a poignant illustration of the possible consequences of undecidability. Contrary to intuition, the value we have to attribute in a point of slit S_1, to the probability that the particle has traveled through slit S_2 is now not zero as we might have expected but $\frac{1}{2}|\psi_3'(\mathbf{r})|^2$. The experimental undecidability biases thus the probabilities such that they are no longer the "divine probabilities" of the axiomatic system BL.

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. 38 do not exist, while Einstein would claim they do exist. Both approaches are logically tenable when the detector screen is completely in the zone $\mathcal{W}$, because the quantities in $\mathcal{N}_1 \cap \mathcal{Z}_2$ and $\mathcal{N}_2 \cap \mathcal{Z}_1$ 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 is a purely numerical recipe that is physically meaningless, because it searches for a correct global solution without caring about the correctness of the partial solutions. It follows the experimental ternary logic of TL and therefore is 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 can 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 spinors 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 $\mathbf{r} \in \mathcal{V}$ 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. 38 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.[13]

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

[13] In Option 2 of the approach with the additional binary constraint based on the axiomatic system TL+D, we even do not reproduce $\psi'_1(\mathbf{r}) = 0$ on slit S_2 and $\psi'_2(\mathbf{r}) = 0$ on slit S_1, because these quantities are not measured if we assume that they are only measured far behind the slits. In the Bohr-like approach, the conditions $\psi'_1(\mathbf{r}) = 0$ on slit S_2 and $\psi'_2(\mathbf{r}) = 0$ on slit S_1 are thus not correct boundary conditions for a measurement far behind the slits, because they violate the *Ansatz* of experimental undecidability.

between the detector and the slits such that they can satisfy the undecidable solution in eq. 38. 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 [5]. 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 a Huygens' principle, which has 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 see also 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. Moreover, imposing the boundary condition that $\psi_1'(\mathbf{r}) = 0$ on slit S_2 and $\psi_2'(\mathbf{r}) = 0$ on slit S_1 remains a matter of choice, depending on which choice one wants to follow (see Option 2 within TL+D evoked above). If we do not clearly point out the axiomatic systems and the choices, then confusion can enter the scene because they bias the definitions of the probabilities.

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. Within the axiomatic system TL+D, the agreement of the numerical result 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. 38 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, which is not a narrative 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 spinors can only be added purely algebraically in a meaningful way if we get their phases right.

9.5 More formal presentation

9.5.1 Criterium for binary logic - orthogonality

Two complex numbers ζ_1 and ζ_2 are orthogonal if:[14]

$$\zeta_1^* \zeta_2 + \zeta_1 \zeta_2^* = 0 \tag{39}$$

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

$$|\zeta_1 + \zeta_2|^2 = |\zeta_1|^2 + |\zeta_2|^2. \tag{40}$$

such that in a point $\mathbf{r} \in \mathcal{V}$ 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:

$$\alpha_1 - \alpha_2 = \pm \frac{\pi}{2} (\text{mod } 2\pi). \tag{41}$$

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 - 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:

$$\zeta(\mathbf{r}) = \zeta_1(\mathbf{r}) \boxplus \zeta_2(\mathbf{r}). \tag{42}$$

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 is a third way to justify the introduction of sums of wave functions $\zeta_1 \boxplus \zeta_2$. 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:

$$|\zeta_1(\mathbf{r})| = |\zeta_2(\mathbf{r})|. \tag{43}$$

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.

9.5.3 Combined ternary and binary logic

When we want to satisfy both binary and ternary logic simultaneously, we must satisfy both eq. 43 and eq. 41, such that:

$$\zeta_1 = \zeta_2 e^{\pm \imath \pi/2}. \tag{44}$$

Furthermore, eq. 40 and 43 imply:

$$|\zeta_1| = |\zeta_2| = |\zeta_1 + \zeta_2|/\sqrt{2}. \tag{45}$$

Combining eqs. 44 and 45 we obtain then eq. 38.

[14] 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: $\zeta_1 \cdot \zeta_2 = \frac{1}{2}[\,\zeta_1^\dagger \zeta_2 + \zeta_2^\dagger \zeta_1\,]$.

9.5.4 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. 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. 42 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. 38 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 $\pi_d \subset \mathcal{W}$. As an afterthought we can see that the solutions of eq. 38 are indeed strongly orthogonal. However, in deriving eq. 38 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.

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.

For these two reasons 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. In QM we take d and the undecidability into account unwittingly. This is the reason why we should take Bohr's caveat seriously and rely on the QM formalism rather than on some "physical intuition".

Eq. 38 was derived from the intuition that ψ_1 and ψ_2 are presumably partial solutions for the double-slit experiment where the electron travels through only one slit. Here we have started from the intuition that ζ_1 and ζ_2 are presumably solutions for the single-slit experiments. And in both cases we obtain the same structure of the solution in TL+D.

We see this way that solutions ζ_1 and ζ_2 exist, because they can even be constructed alternatively as ψ_1' and ψ_2' by starting from ψ_1 and ψ_2 and following the method to construct ψ_1' and ψ_2'. This can be turned into a rigorous

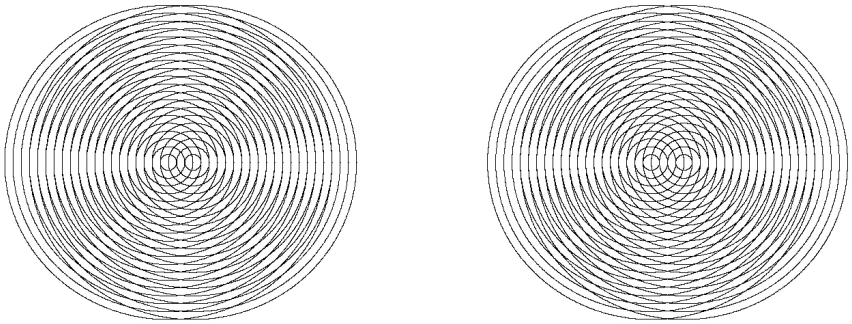

Fig. 3. Two interference patterns for identical wavelengths (visualized by the distances between the crests of the waves) corresponding to two different distances d between the slits in a double-slit experiment. We have represented here the slits as points, rendering the geometrical construction possible. The boundary conditions we impose on the Schrödinger equation define the conditional probabilities which will prevail and will be measured in an experiment. In a double-slit experiment these boundary conditions are the distance d and the undecidability of the question through which the particle will travel. This undecidability is introduced by the left-right symmetry between the two slits in the set-up: look at the figure and ask through which slit the particle will travel. The question is undecided because it is impossible to answer. The interference patterns illustrate the conditional probabilities we then obtain. The interference is the manifestation of the undecidability (see main text). The idea is to look at these two configurations with their boundary conditions from the perspective of a single-slit experiment. The figure visualizes thus two different sets of boundary conditions for a double-slit experiment and their consequences on the conditional probabilities. Both sets of boundary conditions are completely irrelevant for a single-slit experiment, which defines entirely different boundary conditions in its own right. It does not define a distance d and it does not imply undecidability, because there is only one slit. The double-slit boundary conditions are absent from the single-slit universe, for which they are alien and beyond guessing. It is therefore impossible and unphysical to adjust the conditional probabilities p_j of the single-slit experiments to the conditional probabilities p'_j for the double-slit experiment, which depend critically on the interplay between d and the undecidability as the figure illustrates. The single-slit probabilities p_j are therefore totally meaningless for the double-slit experiment. We can draw this conclusion, which is just based on the observation how important the consequences of a change of boundary conditions are, without any further physical brainstorming or digging into mathematical details. This conclusion validates Bohr's caveat, highlighting its all-out importance. We cannot transpose conditional probabilities defined by one set-up to another set-up, because the boundary conditions which define them are different, even if on the basis of our intuition we might be not aware of this or frivolously be convinced that it would not matter. We are trampling then the caveat. The formalism of QM protects us against the intrepidity of following our intuition and lack of vigilance by making us take into account Bohr's caveat unwittingly, as it teaches us to define the boundary conditions for the wave function in the Schrödinger or Dirac equation. The formalism also allows to deposit finer detail about the contextual probabilities into the wave functions.

mathematical proof for the existence of ζ_1 and ζ_2, by justifying the summing with the symmetry argument, which expresses the undecidability. The other two justifications for summing are not rigorous because they rely on theoretical approximations and on the physical evidence that the result $\psi_1 \boxplus \psi_2$ derived from these approximations is in excellent agreement with experiment.

This shows even better that ζ_1 and ζ_2 are not physical solutions for the single-slit experiments because the special solutions $\zeta_1 = \psi'_1$ and $\zeta_2 = \psi'_2$ exhibit interference patterns, which are typical of undecidability. In fact, while we were initially talking about a change of the distance d between the slits that would change the strong orthogonality condition, we can now see that such a change of d together with the condition of undecidability will modify the interference pattern, something which is very obvious (see fig. 3). The functions ζ_1 and ζ_2 are therefore single-slit solutions that could be mathematically correct but that are satisfying completely unreasonable constraints for the equations for ψ_1 and ψ_2. Probably a whole set of functions with unreasonable additional constraints for ψ_1 and ψ_2 are also mathematically correct single-slit solutions. E.g. just imagine something wild for the shape and position of the other slit S_2 and then perform the calculations that yield the additional constraints for ψ_1. Using the Born

approximation you might even be able to design S_2 in such a way that it yields a chosen profile for ψ_1. What makes all these phantom mathematical solutions different from the physical solutions ψ_1 and ψ_2 is that ψ_1 and ψ_2 are obtained with an electron gun that shoots at random.

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. An interference pattern (with its quantization of momentum $\mathbf{p} = h\mathbf{q}$ according to the alternative approach in subsect. 9.1) is therefore the only solution for the wave function that satisfies the axiom of undecidability. Interference is the fingerprint of undecidability.

Of course we can ask now to extend the wave function we found on $\pi_s \cup \pi_d$ to $\mathcal{V}$. That would require imposing strong orthogonality over whole $\mathcal{V}$ and raise the question how we make the transition from decidability on π_s to undecidability on π_d, and especially if that transition can be rendered smooth. That is what led to the discussion about the discontinuities at the boundaries in fig. 2. The new approach avoids this kind of discussion. It just leaves it up to the mathematics to find a function whose domain is $\mathcal{V}$ and which is an extension of the function we found on $\pi_d \cup \pi_s$. Mathematics should find a way, just as physics finds a way. And if there are discontinuities in the physics, then they will be physical. This approach is much more rigorous than using the crude idealized guesses for the partition of $\mathcal{V}$ into domains we introduced in fig. 2. From the physical viewpoint the approach restricted to $\pi_s \cup \pi_d$ to $\mathcal{V}$ also does not introduce more assumptions than needed.

What we say here is valid for purely coherent scattering. In fact, the general solution in $\mathbf{r} \in \mathcal{V}$ 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 $(\forall j \in \{1,3\},\, \lambda_j(\mathbf{r}) \in [0,1]\,)$ & $\lambda_1(\mathbf{r}) + \lambda_2(\mathbf{r}) + \lambda_3(\mathbf{r}) = 1$, 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 $\lambda_2(\mathbf{r})$. Hence the general case is not correctly treated by textbooks, which assume purely coherent scattering. This coherent scattering evolves from a decided situation at the slits to an undecided situation far behind the slits, 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.

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 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.

Finally we may point out that the difference of $\pm\frac{\pi}{2} (\mathrm{mod}\ 2\pi)$ between the phases of ψ_1' and of ψ_2' is not in contradiction with the phase difference $2\pi n$ over the paths SS_1P and SS_2P through ψ_3 we discussed in subsect. 9.1, because the latter phase difference occurs within $\psi_3 = \psi_1' + \psi_2'$, while the former occurs between ψ_1' and ψ_2'.

9.6 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. 35, 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 [5,6, 7]. 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 [5,7]. Eq. 35 hinges 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 [5] 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.

[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. Their phase is a dial 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 $p_3(\mathbf{r}) = p_1(\mathbf{r}) + p_2(\mathbf{r})$ 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 physical nonsense, 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.

[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 a mathematically proved but physically meaningless Huygens' principle. 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 subsect. 4.3 of [5].

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. Even so, it may not be accepted by the whole community. Some physicists may indeed want to dismiss it, feeling chilled by the mathematical subtleties or entrenched in a certain dogmatic way of thinking. They might be shell-shocked by the factual truth that spinors are just not "vectors in Hilbert space", as they have always believed. Nevertheless, this is the only tiny loophole of escape to break away from the contradiction in point [3]. Such a rejection takes then place without offering a reasonable alternative explanation. All one has been able to come up with, are 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 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. Complaining about the impenetrability of QM while refusing at the same time to open up by questioning what you have learned is logically inconsistent. Every physicist can decide for himself if he wants to accept the mathematical construction or otherwise, but we cannot let a few self-righteous censors reject it single-handedly on behalf of the whole scientific community, because they do not represent the consensus of that community. Especially when they ram their decisions through by means of questionable methods [15,28,45,46,47,48].

11 Conclusion

In summary, we have proposed an intelligible solution for the paradox of the double-slit experiment. 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.

Apart from the fact that it can insist on the importance of the context created by a set-up in defining the conditional probabilities, the language of probabilities p_j is too poor to take into account these contexts correctly. To correctly deal with the contextuality we must use spinor analysis. Spinors are not counter-intuitive. If you take the time to study spinors, you will discover that you can perfectly understand them intuitively [5,8]. But it corresponds 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.

This is why the double-slit experiment is just a probability paradox. The two levels of intuition are also using different definitions for the word "understanding". We can state that we understand QM on the advanced level of intuition where we master spinors, as our reconstruction of QM and its derivation of the Dirac equation from scratch show, but we cannot understand it at the zero-level intuition of daily-life conditional probabilities p_j, which cannot account for a surprising type of contextuality with an explosive cocktail of non-locality and ternary logic.

Finally, it must be stressed that justifying coherent summing of wave functions by using the superposition principle is conceptually totally wrong and misleading. The justification of coherent summing must be based on a symmetry argument or a Huygens' principle. The Huygens' principles take into account the non-locality (and also the undecidability). The handshake in Cramer's transactional interpretation of QM [35] catches the essence of what is going on in the principle. 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 [6] constitute convincing evidence for the value of the reconstruction of QM based on the geometrical meaning of spinors in [5,7]. It permits for a perfect dialogue between the mathematics and the physics.

Declarations

Funding Not applicable.

Conflicts of interest/Competing interests Not applicable.

References

1. R.P. Feynman, R. Leighton and M. Sands, The Feynman Lectures on Physics, Vol. 3, (Addison-Wesley, Reading, MA, 1970).
2. R.P. Feynman, "Probability & uncertainty - the quantum mechanical view of nature" https://www.youtube.com/watch?v=2mIk3wBJDgE (video).
3. K.T. McDonald, https://physics.princeton.edu//mcdonald/examples/clock.pdf.
4. G. Coddens, Eur. Phys. J. Plus 135, 152 (2020).
5. G. Coddens, Symmetry, 13, 659 (2021).
6. G. Coddens, Symmetry 13,134 (2021).
7. G. Coddens; From Spinors to Quantum Mechanics, (Imperial College Press, 2015).
8. G. Coddens, https://hal.archives-ouvertes.fr/cea-01572342v1.
9. G. Coddens, https://hal.archives-ouvertes.fr/hal-03289828.
10. P.A.M. Dirac, Spinors in Hilbert Space, (Springer, Heidelberg/New York, 2012).
11. E. Cartan, The Theory of Spinors, (Dover, New York, NY, USA, 1981).
12. B. Schutz, Geometrical Methods of Mathematical Physics, (Cambridge University Press, Cambridge, UK, 1999).
13. L. Schwartz, Théorie des Distributions, (Hermann, Paris, 1966).
14. J. Hladik, J.M. Cole, Spinors in Physics, (Springer, New York, NY, USA, 2012).
15. Anonymous referee, Symmetry, (2021).
16. A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. 49, 1804 (1982).
17. A. Aspect, P. Grangier, G. Roger, Phys. Rev. Lett. 49, 91 (1982).
18. Juan Yin *et al.*, Science 356, 1140 (2017).
19. T. Herbst, T. Scheidl, M. Fink, J. Handsteiner, B. Wittmann, R. Ursin, and A. Zeilinger, PNAS, 112, 14202 (2015).
20. M. Kupczynski, Frontiers in Physics, 8, 273 (2020);
21. M. Kupczynski, Entropy 20 (2018), https://doi.org/10.3390/e20110877.
22. J.A. Wheeler, in Mathematical Foundations of Quantum Theory, A.R. Marlow editor, (Academic Press, 1978), pp. 9-48.
23. J.A. Wheeler, and W.H. Zurek, Quantum Theory and Measurement (Princeton Series in Physics), (Princeton University Press, Princeton, NJ, 1983).
24. Y.-H. Kim, R. Yu, S.P. Kulik, Y. Shih, and M.O. Scully, Phys. Rev. Lett. 84, 1 (2000).
25. E. Gerdau, R. Rüffer, H. Winkler, W. Tolksdorf, C.P. Klages, J.P. Hannon, Phys. Rev. Lett., 54, 835 (1985).
26. J.J. Duistermaat, in Huygens' Principle 1690-1990: Theory and Applications, H. Blok, H. Ferwerds, and H.K. Kuiken eds., (Elsevier, 1992), p. 273.
27. J. Hadamard, Le problème de Cauchy et les Equations aux Dérivées partielles linéaires hyperboliques, reprint of the 1923 publication, (Dover, New York, 1952).

28. Website Effectiviology, https://effectiviology.com/straw-man-arguments-recognize-counter-use/.
29. A. Tonomura, J. Endo, T. Matsuda, and T. Kawasaki, Am. J. Phys. 57, 117 (1989).
30. Silverman, M.P.; More than one Mystery, Explorations in Quantum Interference, (Springer, 1995), p. 3.
31. A. Tonomura, J. Endo, T. Matsuda, and T. Kawasaki, (video) https://www.youtube.com/watch?v=1LVkQfCptEs.
32. R.P. Feynman, Feynman's thesis, edited by L.M. Brown, (World Scientific, Singapore, 2005).
33. R.S. Longhurst, Geometrical and Physical Optics, 2nd edition, (Longman Group, London, 1967).
34. R.P. Feynman, QED: The Strange Theory of Light and Matter, (Princeton University Press,1985).
35. J. Cramer, Rev. Mod. Phys. 58, 795 (2009).
36. P.A.M. Dirac, Physikalishe Zeitschrift der Sowjetunion, Band 3, Heft 1 (1933).
37. G. Coddens, https://hal.archives-ouvertes.fr/hal-02636464v3.
38. W. Rueckner, and J. Peidle, Am. J. Phys. 81, 951 (2013).
39. L.E. Ballentine, Rev. Mod. Phys. 42, 358 (1970).
40. L.E. Ballentine, Quantum Mechanics, A Modern Development, 2nd edition, (World Scientific, Singapore, 1998).
41. K. Gödel, On Formally Undecidable Propositions of Principia Mathematica and Related Systems, (Dover, New York, 1992).
42. J. Dieudonné, Pour l'honneur de l'esprit humain - les mathématiques aujourd'hui, (Hachette, Paris,1987).
43. H.S.M. Coxeter, Introduction to Geometry, Second Edition, Wiley Classics Library, (John Wiley & Sons, New York, 1989).
44. N. Harrigan, and R.W. Spekkens, Found. Phys. 40, 125 (2010).
45. S. Carvahlo, Sc. Am., 323 (16th September 2020).
46. R. Bullock, https://www.universityaffairs.ca/opinion/in-my-opinion/the-trolls-have-infested-academic-peer-review/.
47. R.L. Glass, J. Systems & Software, 54, 1 (2000).
48. P. Knoepfler, https://ipscell.com/2012/05/dirty-dozen-easy-steps-to-killing-a-paper-during-review-elephant-in-the-lab-series/.