Radboud University Nijmegen

Faculty of Science

The Frauchiger-Renner paradox

Bachelor Thesis Mathematics

Supervisor:
prof. dr. Klaas Landsman

Author:
Ian Koot

Second reader:
dr. Michael Müger

July 2019

Contents

Introduction

> I think I can safely say that nobody understands quantum mechanics.

- Richard Feynman, The character of Physical Law, p. 129, 1995

Few things are as counterintuitive as quantum mechanics. The probabilistic structure, the scale at which it takes place, wave-particle duality, and many more peculiarities make the quantum world one of the most difficult things to truly understand (if that is even possible). It is also very intriguing, this world where seemingly everything is possible (just look at the fact that the word "quantum" appeared 30 times in this summer's blockbuster *Avengers: Endgame*).

However, as strange as quantum mechanics might appear, it at least has to be *consistently strange*. One of the definitions of 'paradox' given by the Merriam-Webster dictionary is "a statement that is seemingly contradictory or opposed to common sense and yet is perhaps true". In this sense there have been quite a few paradoxes in quantum mechanics, most famously Schrödingers cat, but the Frauchiger-Renner paradox is different. In [Frauchiger and Renner, 2018], Daniela Frauchiger and Renato Renner claim to have found a situation so strange, that when quantum theory is applied to it, it *contradicts* itself.

This thesis is certainly not the first to address this paradox. It seems, however, that the discussion about this subject has been plagued by long explanations and subtle analyses. The goal of this thesis is to try to take a mathematical perspective to this paradox, in order to uncover what makes it 'tick', and if possible, debunk it.

In order to achieve this, first, in chapter 1, some time is spent looking at the physical side of this problem. Some basic quantum mechanics is introduced, and some historical influences are treated to give context to the Frauchiger-Renner paper. Then we critically evaluate the Frauchiger-Renner Paper, first in chapter 2 by introducing notation and taking a look at the thought experiment that lies at the heart of the paradox. Lastly, in chapter 3, a detailed analysis is put forward, along with possible weak spots in the proof of the contradiction given by Frauchiger and Renner, after which we will take a look at some other reactions to the paper.

Chapter 1

Quantum mechanical concepts

The section that follows serves a dual purpose: on the one hand, for anyone not familiar with either the physics or mathematics of quantum mechanics, it is meant as a need-to-know summary of concepts in order to read this thesis. This is by no means a sufficient summary to understand even elementary quantum physics, but it should provide the directions in which one could look for further reading on the topic.

On the other hand, since much of the literature concerned with the subject of this thesis is hindered by a lack of precision or agreement on what "the rules" are (to use terminology used in [Tausk, 2018]), this section aims to provide a solid basis on which to conduct the discussion of the Frauchiger-Renner paradox.[1].

1.1 Mathematics of quantum states

The mathematical structure with which we try to describe reality in quantum mechanics , is a **complex Hilbert space**: a vector space $\mathcal{H}$ over $\mathbb{C}$ endowed with a inner product $\langle\cdot,\cdot\rangle : \mathcal{H}\times\mathcal{H} \longrightarrow \mathbb{C}$ which gives rise to a norm $\|\psi\| = |\langle\psi,\psi\rangle|^{\frac{1}{2}}$ and a metric $d(\psi,\phi) = \|\psi-\phi\|$, which must make $\mathcal{H}$ a complete metric space. When we say that a physical system is in a certain pure state, we mean that the physical system corresponds to a unit vector ψ in $\mathcal{H}$, and we often write it as $|\psi\rangle$.[2] This notation, which might be somewhat unusual for mathematicians, is called *Dirac notation*, and perhaps it becomes more clear if we look at the inner product of two vectors $|\psi\rangle$ and $|\phi\rangle$. The function $\langle\psi,\cdot\rangle : \mathcal{H} \longrightarrow \mathbb{C}$ defined by $\phi \mapsto \langle\psi,\phi\rangle$ is a linear functional,[3] which we

[1]There are many introductory books on quantum mechanics, for example [Griffiths, 2005], which was used for the following section, together with [Landsman, 2017].

[2]It turns out that states that only differ by a constant phase actually behave the same: we will come back to this at the end of this section.

[3]See any textbook on Functional Analysis, for example [MacCluer, 2009], p. 15

denote by $\langle\psi| := \langle\psi, \cdot\rangle$, so that $\langle\psi, \phi\rangle = \langle\psi|\phi\rangle$. In this context, the vectors are often called "kets" and the corresponding linear functionals are called "bras".

We then relate this mathematical structure to physical systems via **observables**. Observables in the physical sense are just that: things we can observe or measure. In our mathematical model, they correspond to **self-adjoint linear operators** on $\mathcal{H}$. So there's a position operator, a spin operator, an angular momentum operator, etc.[4]

We would now like something in our mathematical description that describes those properties of our physical system which we are interested in, for example, because we want to measure them to test our mathematical description. But this is where our "classical" ideas of properties start to break down. We connect our mathematical model in terms of the outcomes of measurements, which for a given observable lie in its set of eigenvalues. This brings us to the following:

Postulate 1.1 (Born rule)**.** *Let $\mathcal{H}$ be the Hilbert space describing a physical system, and let a be a self-adjoint operator corresponding to an observable with a discrete spectrum. Let λ be an eigenvalue of a, and e_λ be the projection operator of $\mathcal{H}$ onto the eigenspace belonging to λ. Furthermore, let $|\psi\rangle \in \mathcal{H}$ be a unit vector, representing the state of the system. Then the probability that a measurement of that observable will result in λ is*

$$P_\psi(\lambda) = \|e_\lambda |\psi\rangle\|^2.$$

I have labeled this a postulate since this is essentially where the physics come in. We assume that reality (in the form of measurements) is described by the mathematical structure that is presented. This is of course very difficult to prove, if at all possible, and therefore the label 'postulate' seems appropriate.

As we will see in a moment, this, and especially what happens to a state *after* it has been measured, doesn't correspond to our physical intuition and gives rise to the many paradoxical situations with which this thesis is concerned (and to even more with which it is not). It is important to note that this means that most states don't have definite values for observables. Only eigenstates of a certain observable will give the same answer upon repeated measurements when we ask what the "value" of that observable is.

Alternatively, and more generally, we could say that quantum mechanics is described by **C*-algebras**, which are basically associative complex algebras with a compatible norm and involution map $x \mapsto x^*$. The most important C*-algebra for quantum mechanics is $B(\mathcal{H})$ of bounded operators on Hilbert space. As before, self-adjoint operators are observables, so it is very natural to consider linear functions $\omega : B(H) \to \mathbb{C}$ such that $\omega(a^*a) \geq 0$ (positivity) and

[4] These operators, in general, are not the same for different systems, even though they are sometimes called by the same name. A spin-$\frac{1}{2}$ particle is described by a two-dimensional Hilbert space, while for a spin-1 particle it is three-dimensional, so they can't be the same operator.

$\omega(1_{\mathcal{H}}) = 1$, which are called "states". It turns out that those states for a finite dimensional Hilbert space take the form

$$\omega(a) = \mathrm{Tr}(\rho a)$$

for some $\rho \in B(\mathcal{H})$ with $\mathrm{Tr}(\rho) = 1$ and $\rho = b^*b$ for some $b \in B(\mathcal{H})$.[5] Such a ρ is called a **density operator**, which is a generalization of the unit vectors in a Hilbert space mentioned earlier, since the simplest example of a density matrix is a one-dimensional **projection**. We can now see that vectors in a Hilbert space differing by a phase produce the same behaviour; the density operator corresponding to such a pure state is a projection on a one-dimensional subspace, and since the Hilbert space is a complex space, a vector lies in the same one-dimensional subspace as that same vector multiplied by a complex phase. So their density operators would be the same, since they project on the same subspace.

1.1.1 Time evolution

Physics is very much concerned with the way physical systems evolve over time. It is here that we need to make a specification. The Frauchiger-Renner paradox is mostly discussed within the foundations of quantum theory, and in order not to obscure the underlying concepts, we will primarily examine *non-relativistic* quantum theory, as opposed to *relativistic* quantum theory, which is the attempt at unifying quantum theory and special/general relativity.

In non-relativistic quantum mechanics, a system can evolve in two ways. Most of the time, it will evolve in a *unitary* way. This can be realized in two ways. The first, called the *Schrödinger picture*, realizes this time evolution by making the state vector of a system dependent on time:

$$|\psi(t_2)\rangle = U_{t_1}^{t_2} |\psi(t_1)\rangle ,$$

where $U_{t_1}^{t_2}$ is some unitary operator describing the time evolution. In non-relativistic quantum mechanics, this unitary time evolution is given by the *Schrödinger equation*:

$$i\hbar \frac{\partial |\psi(t)\rangle}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 |\psi(t)\rangle + V(x) |\psi(t)\rangle$$

The second way in which this unitary evolution can be realized is called the *Heisenberg picture*. Here, the operators are time-dependent, instead of the state vectors. Say we have an operator a, then

$$a(t_2) |\psi\rangle = {U_{t_1}^{t_2}}^* a(t_1) U_{t_1}^{t_2} |\psi\rangle .$$

One can show using the Born rule that the expectation value of a measurement of an observable a given that the system is in state $|\psi\rangle$ is $\langle\psi| a |\psi\rangle$.

[5] For infinite-dimensional Hilbert spaces, this can be done if the states are "normal": see [Landsman, 2017], section 2.2.

Using this expression, we can see that the two pictures give the same expectation value for all observables at all times (also using that the linear functional corresponding to a vector $b\,|\psi\rangle$ for an operator b is $b^*\,\langle\psi|$).

Things are different when we perform a measurement of an observable, and therefore have to apply (1.1). If the state was $|\psi\rangle$ before the measurement, it "jumps" to $e_\lambda\,|\psi\rangle\,/\|e_\lambda\,|\psi\rangle\|$, where λ is the value of the observable we measured (and $e_\lambda\,|\psi\rangle\,/\|e_\lambda\,|\psi\rangle\|$ is therefore the projection of the state onto the eigenspace of λ, scaled to be a unit vector again). This evolution is in general *not* unitary.

When we measure a system that is not in an eigenstate of our observable, multiple different outcomes of the measurement have a non-zero probability of happening. However, since the state jumps to an eigenvector of the observable after the measurement, if we do a measurement again immediately after the first measurement (so that there is no significant time evolution of the state) we are certain to find the same value. The system behaves *differently* after the first measurement, and so we must have changed the system by measuring it for the first time.

1.2 Relevant quantum effects

Now that we have set up the general framework of quantum mechanics, it is time to consider some of the implications that are relevant to this thesis.

Superposition

Suppose we have an observable a with eigenvectors $|\psi_i\rangle$ corresponding to eigenvalues λ_i. According to postulate 1.1, when the state of a system corresponds to an eigenstate of the observable, say $|\psi_1\rangle$, then if we measure a, there is only one value for which we have a non-zero probability to measure it, namely λ_1. We could therefore say that the value of a is well-defined.

However, there is also the possibility that a state corresponds to a *linear combination* of eigenvectors, for example $|\psi_1\rangle + |\psi_2\rangle$ (where $\lambda_1 \neq \lambda_2$). Now postulate 1.1 predicts that both λ_1 *and* λ_2 have a non-zero probability of occurring. The system is now said to be in a *superposition* of states ψ_1 and ψ_2 with respect to the observable a.

Expanding on this idea, we can see that the Schrödinger equation, which as stated above dictates the time-evolution of a quantum system, is a *linear* partial differential equation. This means that a linear combination of solutions also satisfies the equation. Given an observable a, we might be able to find solutions to the Schrödinger equation that are at any time t eigenvectors of a. But a linear combination of those solutions is again a solution, albeit without a well-defined value for a. A system in such a linear combination of states over time behaves like it is in a *superposition* of multiple well-defined values for a existing simultaneously, until we measure the system with respect to a, at which point we *collapse* the superposition and force the system to take on a well-defined value for a.

Composed systems and Entanglement

Most of the difficulties that arise in this thesis have to do with systems that consist of multiple subsystems. We describe these kinds of systems by the *tensor product*: given vector spaces U and V, the tensor product is a vector space $U \otimes V$ with a bilinear map $\otimes : U \times V \to U \otimes V$ (written as $\otimes(u, v) = u \otimes v$), such that for every bilinear map to a third vector space $b : U \times V \to W$, there is a linear map $B : U \otimes V \to W$ with $b(u, v) = B(u \otimes v)$. However, we will not use this definition explicitly. Instead, for the purposes of this thesis, it can best be thought of as follows: take bases $(u_i) \in U$ and $(v_i) \in V$, then $U \otimes V$ has as a basis $(u_i \otimes v_j) \in U \otimes V$.

Let's look at an example. Suppose we have two non-interacting, totally independent spin-$\frac{1}{2}$ systems. The first system is in a state corresponding to $a\left|\uparrow\right\rangle_1 + b\left|\downarrow\right\rangle_1 \in \mathcal{H}_1$ with respect to some z-axis, and the second is in a state corresponding to $\alpha\left|\uparrow\right\rangle_2 + \beta\left|\uparrow\right\rangle_2 \in \mathcal{H}_2$. Then the whole system corresponds to the state

$$
\begin{aligned}
&(a\left|\uparrow\right\rangle_1 + b\left|\downarrow\right\rangle_1) \otimes (\alpha\left|\uparrow\right\rangle_2 + \beta\left|\downarrow\right\rangle_2) \\
&\qquad = a\alpha\left|\uparrow\right\rangle_1 \otimes \left|\uparrow\right\rangle_2 + a\beta\left|\uparrow\right\rangle_1 \otimes \left|\downarrow\right\rangle_2 + b\alpha\left|\downarrow\right\rangle_1 \otimes \left|\uparrow\right\rangle_2 + b\beta\left|\downarrow\right\rangle_1 \otimes \left|\downarrow\right\rangle_2 \\
&\qquad = a\alpha\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2 + a\beta\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2 + b\alpha\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2 + b\beta\left|\downarrow\right\rangle_1\left|\downarrow\right\rangle_2
\end{aligned}
$$

where the last line is an often used abbreviation. Sometimes, in order to improve legibility, we will denote $\left|\psi\right\rangle\left|\phi\right\rangle$ as $\left|\psi, \phi\right\rangle$, especially when describing the bra of a composed ket, which in this case would be $\left\langle\psi, \phi\right|$.

Note that not all vectors in $U \otimes V$ can be written as the product of a vector in U with a vector in V: for example, an often encountered state of a system consisting of two spin-$\frac{1}{2}$ subsystems is

$$
\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2 - \left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2, \tag{1.1}
$$

which cannot be written as a single product of linear superpositions of ups and downs. Such a state is said to be *entangled*. Though it is maybe not immediately clear from the mathematical structure why this would be remarkable, if we know a system is in such a state, then whenever we measure one particle, we *immediately* know the state of the other; wherever that other particle might be. Considering that in relativity, signals cannot travel faster than the speed of light, this seems strange. Properties of a theory that state that objects that are sufficiently far away from each other cannot influence each other are in general referred to as *locality* properties, and there are a few subtle variations on this concept. I will discuss this further in section 1.4.2.

1.3 Measurements

Until now, we have considered "measurements" to be somewhat of a mysterious building block of our description of reality. It is very unclear what exactly happens when a measurement is performed. To illustrate this, let's look at an example: we have some kind of measuring device D, and a spin-$\frac{1}{2}$ particle S that is measured by D. Assume that S is in the state $\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) \in \mathcal{H}_S$ (using the subscript on $\mathcal{H}_S$ to denote which system it is the Hilbert space of). When we turn on the device, it will display either $\uparrow$ or $\downarrow$: if we perform the same experiment many times, we expect to see both outcomes about half of the time. But what exactly is it that makes S decide to jump? Even more peculiarly, when we look at $\mathcal{H}_S \otimes \mathcal{H}_D$, we see that the time evolution is given by

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_S + |\downarrow\rangle_S) \otimes |w\rangle_D \;\mapsto\; \frac{1}{\sqrt{2}}(|\uparrow\rangle_S \otimes |\uparrow\rangle_D + |\downarrow\rangle_S \otimes |\downarrow\rangle_D), \tag{1.2}$$

where $|w\rangle_D$ is some kind of waiting state the device has before the measurement goes through, and $|\uparrow\rangle_D$ and $|\downarrow\rangle_D$ are states which the device takes on after it has measured a certain value. So if we look at only S, the system goes from one state to two possible different states, and the evolution is therefore *not* invertible, but if we look at S and D together, suddenly the evolution *is* invertible. It appears that the entanglement is spreading out like a bubble, entangling with more and more of its surroundings as all the systems interact. Since the beginning of quantum theory, many ways of looking at this so-called *measurement problem* have arisen.[6]

1.4 Historical perspective of the Frauchiger-Renner Paradox

As is to be expected, the Frauchiger-Renner paradox did not just come from nowhere: the paradox is a combination of some other peculiar phenomena in quantum mechanics. To understand what exactly makes the Frauchiger-Renner paradox work, we need to look at what it adds to previous concepts. In this section, I introduce some of those concepts in order to discuss which new aspects the Frauchiger-Renner paper introduced. Later, in section 3.2, we will compare this historical context with the paper itself.

1.4.1 Wigner's Friend Paradox

The Frauchiger-Renner Paradox is primarily based on another thought experiment called *Wigner's Friend Paradox*, first published by Eugene Wigner in [Wigner, 1961]. One of the main questions Wigner poses in the article is whether our mind can influence physical reality:

> Does, conversely, the consciousness influence the Physico-chemical conditions? In other words, does the human body deviate from the laws of physics, as gleaned from the study of inanimate nature? The traditional answer to this question is, "No": the body influences the mind but the mind does not influence the body. Yet at least two reasons can be given to support the opposite thesis, [...]

[6]For an example of an overview, see: [Landsman, 2017], ch. 11

One of the reasons which Wigner gives for believing that the human body deviates from the laws of physics, which I will only mention briefly due to the mere tangential relation to the subject of this thesis, is that Wigner observes that "we do not know of any phenomenon in which one subject is influenced by another without exerting an influence thereupon". Basically, he claims that since we have not found any such phenomenon (which is already quite a significant claim, one I find to be unsupported in the article), quantum mechanics will also likely not be such a phenomenon.

The other reason is the most important for our purposes. Wigner proposes a thought experiment: there is a quantum system in a quantum state, say $\alpha |\psi_A\rangle + \beta |\psi_B\rangle$. Now, instead of observing the system ourselves, we let a friend observe the system. This means we assign to the joint system of the quantum system and our friend the state $\alpha |\psi_A\rangle \otimes |\chi_A\rangle + \beta |\psi_B\rangle \otimes |\chi_B\rangle$. Here $|\chi_A\rangle$ and $|\chi_B\rangle$ are vectors in the Hilbert space of our friend (which is very complex, but if quantum mechanics is universally applicable, it should exist). When we then ask our friend which state he observed, he will answer either "A" or "B". But when we ask what he thought that he had measured before we asked him about it, he will respond the same: even though for us, the friend and quantum system were in a joint quantum superposition before we asked him. Wigner says that this "appears absurd because it implies that my friend was in a state of suspended animation before he answered my question". This, according to Wigner, implies that the consciousness of our friend was likely the reason that the quantum state collapsed.

To my knowledge, this view of consciousness as the cause of wavefunction collapse is not widely shared in the field of quantum mechanics, and certainly not by authors discussed in this thesis. The focus on consciousness is therefore not relevant for the discussion of the Frauchiger-Renner paper and all the literature that it has motivated. However, the experiment itself, viewed separately from Wigner's interpretation, still seems paradoxical, and it is an important building block for the Frauchiger-Renner paradox.

For example, an alternative analysis can be found in [Healey, 2017], which goes roughly as follows. We have Wigner, observing a lab containing a friend, whom we will call F,[7] and a quantum system, which we will call S. This quantum system is in the state $\sum_i c_i |\phi_i\rangle \in \mathcal{H}_S$, with $\mathcal{H}_S$ the Hilbert space corresponding to S. At the start of the experiment, just before the lab becomes totally isolated, both Wigner and his friend agree that the state of the lab corresponds to some state $|\psi_0\rangle \in \mathcal{H}_L$, where $\mathcal{H}_L$ is the Hilbert space corresponding to the lab, including the friend and S (so $\mathcal{H}_L = \mathcal{H}_S \otimes \mathcal{H}'$, where $\mathcal{H}'$ describes the friend, the measuring apparatus, and the rest of the lab). Then, the lab is sealed so that it becomes an isolated system. We will call the time at which this happens t_i. When everything is sealed, the friend in the lab performs a measurement with respect to the observable $\sum_i i |\phi_i\rangle \langle\phi_i|$ at time t_e, and we will fix a time $t_f > t_e$.

[7]In [Healey, 2017], Healey uses the names Eugene and John, for Eugene Wigner and John von Neumann. However, for the sake of consistency with notation used later, W and F will be used here.

Let the state of the lab at time t be given by $|\psi(t)\rangle$. Both Wigner and the friend will agree that $|\psi(t_i)\rangle = |\psi_0\rangle$. But for Wigner, there is only unitary evolution in the lab, since he doesn't actually measure the lab, and the state $|\psi(t_f)\rangle$ will be $U_{t_i}^{t_f} |\psi_0\rangle$, where $U_{t_i}^{t_f}$ is the unitary evolution predicted by quantum mechanics. The friend, on the contrary, *does* measure a quantum system, and will correspondingly collapse the state of the lab. Suppose he measures j, then $|\psi(t_f)\rangle = U_{t_i}^{t_m} \circ e_j \circ U_{t_m}^{t_f} |\psi_0\rangle$, which will generally not be the same state. However, for each measurement that the friend can make, he would assign a different state. This means that our friend would be in a superposition of assigning states. This seems strange, but it does not imply any contradictions.

1.4.2 Bell-type experiments

Another type of experiment that illustrates the peculiarity of quantum mechanics is one proposed by John Stewart Bell in [Bell, 1964], based on a thought experiment proposed by Einstein, , and Rosen (collectively known as EPR) in [Einstein et al., 1935], and its refinement by Bohm in [Bohm, 1960], p. 614.[8] EPR claimed that quantum mechanics was not complete and that certain hidden variables should be considered to complete the theory. In his paper, Bell shows that if such a hidden variable theory is *local*, meaning the state of a system can only be influenced by things sufficiently close to that system (for example, within its past light cone), it cannot explain the outcome quantum mechanics predicts for the thought experiment.

Specifically, Bell considers spin measurements of two entangled spin particles in a "singlet state", meaning they have zero total spin (like the state in (1.1)), by two different observers, A and B, which are assumed to be very far away. The directions in which they measure the spin are $\vec{a}$ and $\vec{b}$, respectively. It is then assumed that the outcome of the measurement of A only depends on $\vec{a}$ and a hidden variable λ, *not on* $\vec{b}$ (and *vice versa*). This condition is called *signal-locality*, meaning that the probability with which one of the observers find a certain outcome given the settings of both observers, does not depend on the other observer's settings.

Mathematically, this is achieved by assuming that the outcome of the measurement of A is given by the function $A(\lambda, \vec{a})$ which assigns 1 or -1 to every hidden variable and every vector $\vec{a}$, and the same for $B(\lambda, \vec{b})$. He then defines $P(\vec{a}, \vec{b})$ as the expectation value of $A(\lambda, \vec{a}) \cdot B(\lambda, \vec{b})$; this means that $P(\vec{a}, \vec{a}) = -1$ since with the same setting both observers will always observe the spin to be in the opposite direction, and $P(\vec{a}, \vec{b}) = 0$ if $\vec{a}$ and $\vec{b}$ are orthogonal. In general, for the quantum-mechanical singlet state, we have $P(\vec{a}, \vec{b}) = -\vec{a} \cdot \vec{b}$.

Bell now deduces the following, which has come to be known as *Bell's inequality*:

$$|P(\vec{a}, \vec{b}) - P(\vec{a}, \vec{c})| - P(\vec{b}, \vec{c}) \leq 1.$$

[8]This section will only briefly mention the contradiction between locality and various kinds of hidden variable theory that underlie the Bell, Kochen-Specker and GHZ articles. For extensive mathematical analysis, see [Landsman, 2017], ch. 6

This inequality holds for *all* theories in which the outcome for each observer is determined solely by their own setting and the hidden variable, or in other words, all signal-local hidden variable theories. However, taking

$$\vec{a} = (-1/\sqrt{2}, 1/\sqrt{2}), \vec{b} = (0,1), \vec{c} = (1,0),$$

we have

$$|\vec{a}\cdot\vec{c} - \vec{a}\cdot\vec{b}| - \vec{b}\cdot\vec{c} = \sqrt{2} \geq 1.$$

This means, that there is no signal-local hidden variable theory which reproduces the results from quantum mechanics.

In their 1969 paper [Clauser et al., 1969], Clauser, Horne, Shimony, and Holt (collectively known as CHSH) aim to make Bell's experiment and inequality easier to test. To do this, they dropped the assumption that if the two observers were to choose the *same* orientation, they would measure exact correlation, since this is practically unattainable. They then arrive at the *CHSH inequality*, of which Bell's inequality is a special case.

Bell's thought experiment, however, relied on observations of averages, which could be considered non-ideal. Greenberger, Horne and Zeilinger (collectively GHZ) in [Greenberger et al., 1989] devised a similar experiment, but this time, it was for four particles, and later, in [Greenberger et al., 1990] (together with Shimony), they designed one for three particles, in the latter also supplying a way to make the experiment easier to do in real life. These experiments only relied on events happening with absolute certainty, so they would be easier to realize.

1.4.3 Hardy's paradox

While the experiments mentioned above show the peculiarities of quantum entanglement, and show how quantum entanglement defies our traditional understanding of reality, there is another thought experiment that shows the dangers of trying to reason too hastily within a quantum mechanical world. Consider the following paradox, called *Hardy's paradox*, after Lucien Hardy, who published it in [Hardy, 1992]. Like the previous articles Hardy's article is concerned with the possibility of local hidden variable theories as an explanation for quantum theory. However, I will follow an analysis in [Aharonov et al., 2002] which I think better shows the kinds of reasoning errors one can make in quantum mechanics, and which is, therefore, more applicable to this case.[9]

The experiment goes as follows (see Figure 1.1): An electron and a positron are both fired at their own beamsplitter, following paths s^- and s^+ towards the beamsplitter, respectively. The electron will either take path v^- or u^- to its second beamsplitter, and the positron will either take path v^+ or u^+ to its own second beamsplitter. However, u^+ and u^- overlap, so that when the positron takes the path u^+ and the electron takes path u^-, the particles annihilate and produce photons.

[9] I would also like to mention the short but very interesting video by MinutePhysics on YouTube about Hardy's paradox: `https://www.youtube.com/watch?v=Ph3d-ByEA7Q`

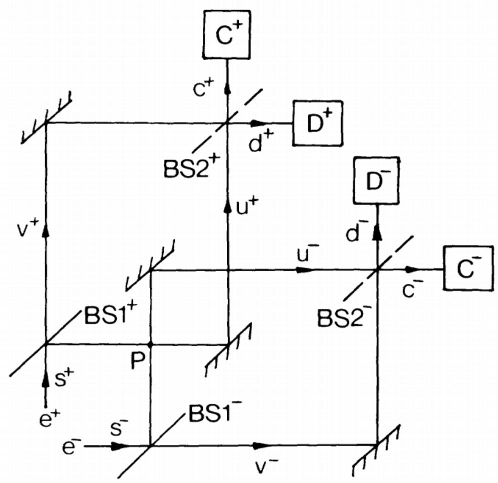

Figure 1.1: The thought experiment used in [Hardy, 1992]. BS1$^\pm$ and BS2$^\pm$ are beamsplitters for postitrons and electrons. The electron comes to BS1−, and "chooses" between v$^-$ and u^-. However, when the positron "chooses" path u^+, the electron and positron would annihilate. Image source: [Hardy, 1992]

The beamsplitters are tuned so that the time evolution is described by:

$$\left|s^\pm\right\rangle \longrightarrow \frac{1}{\sqrt{2}}\left(i\left|u^\pm\right\rangle + \left|v^\pm\right\rangle\right) \tag{1.3}$$

Then, at beamsplitter two, we have

$$\left|u^\pm\right\rangle \longrightarrow \frac{1}{\sqrt{2}}\left(\left|c^\pm\right\rangle + i\left|d^\pm\right\rangle\right) \tag{1.4}$$

$$\left|v^\pm\right\rangle \longrightarrow \frac{1}{\sqrt{2}}\left(i\left|c^\pm\right\rangle + \left|d^\pm\right\rangle\right), \tag{1.5}$$

which means that we have a total evolution of

$$\left|s^+\right\rangle\left|s^-\right\rangle \longrightarrow \frac{1}{4}\left(-2\left|\gamma\right\rangle - 3\left|c^+\right\rangle\left|c^-\right\rangle + i\left|d^+\right\rangle\left|c^-\right\rangle - \left|d^+\right\rangle\left|d^-\right\rangle\right), \tag{1.6}$$

where γ is the state where the particles have annihilated and produced a photon. Now consider the probability that both the d detectors fire: this corresponds to the state $|d^+, d^-\rangle$. By postulate 1.1, this has a probability of

$$\left\|\left|d^+, d^-\right\rangle\left\langle d^+, d^-\middle|s^+, s^-\right\rangle\right\|^2 = \left\|\frac{1}{4}\left|d^+, d^-\right\rangle\right\| = |-\frac{1}{4}|^2 = \frac{1}{16}$$

of happening (in the sense that if we repeat the pattern sufficiently many times, both d detectors will simultaneously fire on average 1 in 16 times).

However, reasoning from the 'point of view' of the electron and the positron, it seems that the d detectors *never* fire simultaneously![10] Consider what happens when we remove the positron from the setup: the time evolution of the electron is now given by composing equation (1.3) with equations (1.4) and (1.5) to form

$$\left|s^-\right\rangle \to \frac{1}{2}\left(i\left(\left|c^-\right\rangle + i\left|d^-\right\rangle\right) + \left(i\left|c^-\right\rangle + \left|d^-\right\rangle\right)\right),$$

or, writing out the right-hand side,

$$\left|s^-\right\rangle \longrightarrow i\left|c^-\right\rangle.$$

This means that when the positron is not in the setup, the d^- detector will never fire, and the c^- detector always fires. The same holds for the positron if we remove the electron: it that case the d^+ detector will never fire, and the c^+ detector will always fire.

The argument continues by noting that if the positron would take the outer route v^+, it would be like the electron was fired without the positron, since they cannot annihilate. This means, that if the positron would take the v^+ route, the c^- detector will *always* fires. We can reason the same for the electron: if it takes the v^- path, the positron proceeds like there would be no electron, and the c^+ detector *always* fires.

Finally, consider the case that the d^+ detector fires. By the above reasoning, this means that the electron did not take the v^- path, because if it did, the c^+ detector would fire and not the d^+ detector. But if we now consider the case that the d^- detector fires, we can conclude that the positron did not take the v^+ path for the same reason. But this means that d^+ and d^- can *never* fire simultaneously, because if neither the c^+ nor the c^- detector fire, both the electron and the positron will have taken the u path, and that means that they would have annihilated, leaving no electron and positron for the d detectors to detect!

Errors and connection to Wigner's friend paradox

What goes wrong here is that in reasoning from the 'point of view' of the positron, we concluded that it must either take path v^+ or path u^+, and similarly that the electron must either take path v^- or path u^-. This does not take into account that the different options might interfere with each other: if both the d detectors fire, then after the experiment the particles are *still* in a superposition of which path they took. By reasoning from these different cases, we assume that those cases are well-defined, which they are not. We are essentially assuming that one particle measures the other to force it into a path instead of a superposition of paths, and as noted in section 1.1.1, this changes the system, something which will not happen if we actually perform the experiment, since there is no reason for the particles to measure each other.

[10] This is very similar to the analysis of the Frauchiger-Renner paradox that will be made in section 3.1; we will compare the reasoning in Hardy's paradox and the Frauchiger-Renner paradox in section 3.3.

Hardy's paradox shows that one must be very careful with which states are well-defined and which are not. There is also a very strong link with Wigner's friend paradox. In Wigner's friend paradox, the friend in the lab measures a quantum system, and the question is which state the friend is in. The above reasoning shows that the state is really undefined until we measure the state of the friend, because if we would reason like the friend was in an actual state (even if we would not know which one), we get contradictions like the one above. With that in mind, we are ready to take a look at the Frauchiger-Renner paradox, to see if the reasoning of Frauchiger and Renner is significantly different or has the same problems.

Chapter 2

The Frauchiger-Renner Paper

Now that we have acquainted ourselves with some appropriate concepts, it is time to discuss the main subject of this thesis: the paper [Frauchiger and Renner, 2018]. In this chapter, first, some notation will be introduced in order to facilitate our discussion of the contents of the Frauchiger-Renner paper. After this, a description of the thought experiment central to the paper is given. The analysis will be given in the next chapter.

2.1 Definitions and assumptions

The Frauchiger-Renner paper is for the most part about 'agents' reasoning using quantum mechanics, and we will be making statements about the validity of their statements. We would like to make a definition similar to the following:

'Definition' 2.1. *A Quantum Theory is a theory in the language with constants for every agent and physical system, together with the language of Hilbert spaces, and a 5-place relation m and for every agent A a binary relation $\sim_A$.*

Here $m(S, t, T, A, \lambda)$ is T if A measures system S at time t to be λ with respect to the observable T, and $S(t) \sim_A \psi$ is true if agent A must assign state ψ to system S at time t.

However, this definitions presents many subtleties. For example, it is not clear from this definition that there even exists a model for such a theory, and if there is a model, there might be many. Additionally, it is not clear what exactly a 'physical system' or an 'agent' is. Considering this, we can still use the ideas of mathematical logic to closely examine the Frauchiger-Renner paradox in a more systematic way. This is not a true axiomatization, just the basis of a more mathematical view.

Having established this, we assume that all agents in the paradox reason with the same quantum theory, and write $\phi \vdash \psi$ if every agent can logically conclude ψ if they now ϕ. The difference in knowledge only arises when someone has actually done a measurement, and therefore knows that some statement $m(S,t,T,A,\lambda)$ is true, while others do not. If $m(S,t,T,A,\lambda) \vdash \psi$, then all agents know that ψ follows form $m(S,t,T,A,\lambda)$, but only agents who know that $m(S,t,T,A,\lambda)$ is true for them, will be able to say that ψ is true for them. If only a certain agent A is able to conclude ψ, we denote this as $\vdash_A \psi$.

So what do these agents reason about? They make statements about the assignment of states to physical (quantum) systems, and then predict the outcome of measurements. Beginning with the former, we should account for a situation where person A, due to having more information than person B, might reason that person B would have to assign a different state to a physical system than they do themselves. Take, for example, Wigner's Friend paradox (section 1.4.1). Here Wigner would assign $U_{t_i}^{t_f} |\psi_0\rangle$ to the lab, but the friend, who is *inside* the superposition, would assign $U_{t_i}^{t_e} e_{c_i} U_{t_e}^{t_f} |\psi_0\rangle$.

So, as briefly described above, we introduce:

Definition 2.2. *If an agent A should assign a state $|\psi\rangle$ to a physical system X at time t, we say that*

$$X(t) \sim_A |\psi\rangle .$$

As mentioned in 'definition' 2.1, we need a precise way to describe measurements. One would be tempted to consider a measurement as some random variable, whose distribution is determined by postulate 1.1. In that case, a definition such as the following might be in order:

'Definition' 2.3. *We write $m(X(t),T,A)$ to denote the outcome of a measurement of operator T of a system X at time t by agent A. Specifically, $m(X(t),T,A)$ is in the spectrum of T.*

However, writing $m(X(t),T,A) = \lambda$ does then not answer the question whether that measurement was even made at all. That is why the 5-place relation in 'definition' 2.1 was introduced, so we could say the following:

Definition 2.4. *Let T be an observable of the Hilbert Space corresponding to the system X at time t. If agent A will measure λ with certainty, where λ is some eigenvalue of T, we say $m(X(t),T,A) = \lambda$.*

We thus have $m(X,t,T,A,\lambda) \leftrightarrow m(X(t),T,A) = \lambda$. From time to time it will be clearer to number experiments, so $m(X(t),T,A) = \lambda$ would become $m(1) = \lambda$ if the measurement of observable T on system X at time t by A was measurement 1. We will write $m(X(t),T,A) \neq \lambda$ for $\neg(m(X(t),T,A) = \lambda)$ (so a measurement will definitely *not* result in λ), as another notational convenience (but keep in mind, we don't suppose $m(X(t),T,A)$ to mean anything on its own).

Note that "measurement 1 will certainly not result in λ" is not the same as "measurement 1 will not certainly result in λ". If an agent A can conclude the former, we write

$$\vdash_A m(1) \neq \lambda,$$

or equivalently

$$\vdash_A \neg(m(1) = \lambda),$$

but if they can only conclude the latter, we write

$$\nvdash m(1) = \lambda.$$

As mentioned above, the notation used throughout this thesis is based upon mathematical logic. The fact that the above two statements are not equivalent, reveals the fact that it is based upon *intuitionistic logic*. In intuitionistic logic, the law of the excluded middle is not true in general (while it is in classical logic), which corresponds to quantum mechanics: when we show that we cannot prove that a certain result of a measurement will definitely *not* occur, it could be in a superposition, so that we can also not guarantee that it definitely will.

We will now look at the three explicit assumptions made in the Frauchiger-Renner paper and try to systematically describe what they say.

2.1.1 Q: partial Born rule

As the authors comment in the article, the first assumption is basically Postulate 1.1, but only considering events that have unit probability of happening, and, as noted above, where everything is observer-dependent. In the article, Frauchiger and Renner use a so-called "family of Heisenberg operators" to formulate this assumption, which I find to be somewhat confusing. I think that what is meant, is the measurement of some self-adjoint operator a with eigenvalues $\sigma(a)$, which is measured at time t. We could then reformulate the assumption as follows:

Postulate 2.5 (Assumption Q)**.** *Let $X(t)$ be a physical system at time t, $|\psi\rangle \in \mathcal{H}$ where $\mathcal{H}$ is the Hilbert space corresponding to $X(t)$, let a be a self-adjoint operator on $\mathcal{H}$, and $a\,|\psi\rangle = \lambda\,|\psi\rangle$. Furthermore, let A be an agent. Then*

$$X(t) \sim_A |\psi\rangle \;\;\vdash\;\; m(X(t), a, A) = \lambda.$$

However, in the article, Frauchiger and Renner often use the inverse statement: if a state is *orthogonal* to a certain eigenstate, than that eigenvalue will definitely *not* occur as the outcome of the measurement.

This does not follow from Postulate 2.5, but we can prove it by including another statement that Frauchiger and Renner include in their assumptions, namely lemma 2.10, which we will discuss later. Formulated more precisely:

Lemma 2.6. *Let $X(t)$ be a physical system at time t, $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ where $\mathcal{H}$ is the Hilbert space corresponding to $X(t)$, let a be a self-adjoint operator on $\mathcal{H}$ with two orthogonal eigenvectors $|\phi\rangle$ and $|\psi\rangle$, and let A be an agent. Additionally, let λ be the eigenvalue corresponding to $|\psi\rangle$, and suppose a measurement of a was made at (or just before) time t by A. Then:*

$$X(t) \sim_A |\phi\rangle \;\;\vdash\;\; m(X(t), a, A) \neq \lambda.$$

Proof. Let $a\,|\phi\rangle = \mu\,|\phi\rangle$. Then by postulate 2.5,

$$X(t) \sim_A |\phi\rangle \;\vdash\; m(X(t), a, A) = \mu.$$

Since a measurement has just been made at time t, we can use lemma 2.10:

$$m(X(t), a, A) = \mu \;\vdash\; m(X(t), a, A) \neq \lambda.$$

□

2.1.2 C: Self-consistency

Frauchiger and Renner call the next assumption “Self-Consistency”:

> if A knows that $x = \xi$, and B knows that A knows that $x = \xi$, than B knows that $x = \xi$.

This is where the difference in perspective comes in. Translating this to our notation we get:

Postulate 2.7 (Assumption C)**.** *Let A and B be agents, $X(t)$ a physical system at time t, a an observable on X, and $\lambda \in \sigma(a)$. Then*

$$(\vdash_A m(X(t), a, B) = \lambda) \vdash m(X(t), a, B) = \lambda.$$

However, as noted in section 2.1, since all agents use the same logic, the only differences in what they can prove as true are statements that follow from a measurement they made. If someone can conclude $\vdash_A m(X(t), a, B) = \lambda$, meaning some observer A can conclude $m(X(t), a, B) = \lambda$, but others cannot conclude this, then $m(X(t), a, B) = \lambda$ follows from some other assumptions $m(X_i(t_i), a_i, A) = \lambda_i$, which only A knows to be true. This can only happen if A actually made those measurements of a_i on X_i at times t_i and got λ_i as a result. But if we know that A can conclude $m(X(t), a, B) = \lambda$, we know that they must have been able to conclude $m(S_i(t_i), a_i, A) = \lambda_i$ for all i, meaning that they must have measured those values. But this means that we essentially have measured those systems ourselves! Formulating this more precisely:

Theorem 2.8. *If all agents reason with the same logic, then*

$$(\vdash_A \phi) \vdash \phi$$

for all agents A and all sentences ϕ.

Proof. All agents reason with the same logic, so the only difference in provability is caused by actual measurements. This means that $\vdash_A \phi$ is equivalent to

$$\{m(X_i(t_i), a_i, A) = \lambda_i \,|\, i \in I\} \vdash \phi$$

together with the fact that A measured λ_i at t_i for all i, where I is some finite index set (making infinitely many measurements seems somewhat unrealistic). But knowing that A measured λ_i at t_i is exactly the statement $m(X_i(t_i), a_i, A) = \lambda_i$, so if we know those to be true, then ϕ is true. □

Theorem 2.8 shows that Postulate 2.7 does not really introduce anything new.

2.1.3 S: Single world

The last explicit assumption is mentioned in an off-hand way at the end of Frauchiger and Renner's analysis:

> We have thus reached a contradiction - unless agent W would accept that w simultaneously admits multiple values. For our discussion below, it will be useful to introduce an explicit assumption, termed Assumption (S), which disallows this;

They formulate this assumption as follows:

> If an agent A is has established that $x = \xi$ at time t, then they must necessarily deny that $x \neq \xi$ at time t.

By x Frauchiger and Renner mean the result of a measurement, so when we make this explicit, we arrive at the following:

Postulate 2.9 (Assumption S)**.** *Let $X(t)$ be a physical system at time t, and let a be a self-adjoint operator on the Hilbert space corresponding to $X(t)$, let $\lambda \in \sigma(a)$, and let A and B be agents. Then*

$$(m(X(t), a, A) = \lambda) \ \wedge \ (m(X(t), a, A) \neq \lambda) \ \vdash \ \bot.$$

The goal of adding this assumption is to exclude many-worlds interpretations of quantum mechanics, which Frauchiger and Renner believe to circumvent the paradox. It is not entirely clear to me how exactly those interpretations would do this. Perhaps it is intended as the assumption that the outcome of a measurement is described by a single variable. However, as the assumption is formulated in the paper, it is trivially satisfied by all forms of logic.

A statement often used in their analysis is the following:

Lemma 2.10. *Let $X(t)$ be physical system at time t, and let a be a self-adjoint operator on the Hilbert space corresponding to $X(t)$. Let A and B be agents. Additionally, suppose that a has 2 eigenvalues λ and λ', and that a measurement is actually made at (a time just before) t. Then*

$$m(X(t), a, A) \neq \lambda \ \vdash \ m(X(t), a, A) = \lambda'.$$

This doesn't follow directly from postulate 2.9: it essentially states one of the fundamental properties of quantum mechanics, namely that directly after a measurement a system *must* be in a well-defined state (and so if it isn't in one eigenstate it must be in the other). When a system is in a superposition, this does not hold, and so we must include the condition that a measurement has just been made in order for this lemma to hold.

2.2 Set-up of the gedankenexperiment

Having covered the assumptions and some notation, let us take a look at the thought experiment that underlies the Frauchiger-Renner paper.

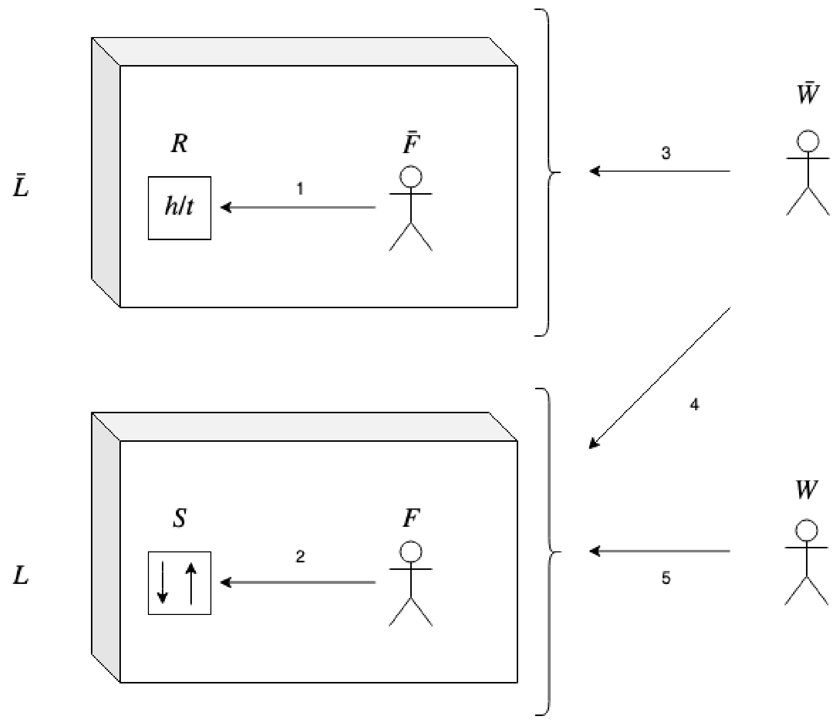

Figure 2.1: An overview of the Frauchiger-Renner thought experiment. There are two labs, $\bar{L}$ and L, each with their own observers inside and outside the lab, $\bar{F}$, $\bar{W}$, F and W, respectively. The arrows 1 through 5 are measurements made by observers, and quantum systems R and S are entangled. Note that measurement 4 does not explicitly appear in the original paper.

An overview of the experiment is provided in Figure 2.1. First I will explain what the basic idea is, then describe the initial state of the experiment, and lastly provide a description of the proceedings of the measurement.[1]

In principle, the experiment considers isolated laboratories containing observers who measure a two-state quantum system, to be two-state quantum-system themselves, in the sense that they are in a superposition of the observer having measured state 1 and the observer having measured state 2. Some person outside of the lab can then in theory measure the entire lab as a two-state system. Additionally, the super-observers outside of the labs in the experiment measure with respect to a different basis of the state-space than the observers inside the labs do. So if an observer inside a lab measures the quantum system and collapses its state, by definition the whole lab is in a superposition for the super-observer, and *vice versa*. The whole experiment is based on this idea.

[1]In the rest of this thesis, the unitary evolution of time will be given by the identity operator, since its effects are not relevant to the discussion.

Let us take a close look at all parts of the experiment. There are two isolated labs, $\bar{L}$ and L, each containing a quantum system and an observer. The quantum systems are called R and S respectively, and the observers inside the labs are called $\bar{F}$ and F. Outside each lab, there is another observer, the one outside $\bar{L}$ being called $\bar{W}$ and the one outside L being called W.

In the paper, the system S is prepared by $\bar{F}$ and then transported to the other lab, all the while keeping both labs isolated. I think this is somewhat confusing and difficult to imagine. We could instead imagine S and R to be entangled in the specific state

$$|\psi_{RS}\rangle = \sqrt{1/3}\,|h\rangle\,|{\downarrow}\rangle + \sqrt{2/3}\,|t\rangle\,(|{\uparrow}\rangle + |{\downarrow}\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2, \tag{2.1}$$

which produces the same effect as $\bar{F}$ passing system S to F in one of the two states, depending on the result of measurement 1 (see Figure 2.1).[2] Here $|h\rangle$, $|t\rangle$, $|{\uparrow}\rangle$ and $|{\downarrow}\rangle$ are the two orthogonal states of R and S respectively.

So we have entangled systems in two different labs, each with their own observer inside, and each lab has a super-observer outside as well. As described above, Frauchiger and Renner model the labs as being two-state systems, so $\bar{L}$ would have the space spanned by $|h\rangle_{\bar{L}}$ and $|t\rangle_{\bar{L}}$ as a state-space, and L would have the space spanned by $|{\uparrow}\rangle_L$ and $|{\downarrow}\rangle_L$ as a state-space. With that in mind, we can define other vectors in their state-space, such as

$$\begin{aligned}
|\overline{\text{ok}}\rangle_{\bar{L}} &= \sqrt{1/2}\,|h\rangle_{\bar{L}} - \sqrt{1/2}\,|t\rangle_{\bar{L}} \\
|\overline{\text{fail}}\rangle_{\bar{L}} &= \sqrt{1/2}\,|h\rangle_{\bar{L}} + \sqrt{1/2}\,|t\rangle_{\bar{L}} \\
|\text{ok}\rangle_L &= \sqrt{1/2}\,|{\downarrow}\rangle_L - \sqrt{1/2}\,|{\uparrow}\rangle_L \\
|\text{fail}\rangle_L &= \sqrt{1/2}\,|{\downarrow}\rangle_L + \sqrt{1/2}\,|{\uparrow}\rangle_L\,.
\end{aligned} \tag{2.2}$$

The whole experiment is performed in four steps:

1. $\bar{F}$ measures R with respect to the observable

   $$|h\rangle\langle h| - |t\rangle\langle t| \in B(\mathbb{C}^2),$$

   getting either 1 (which we will call h for clarity) or -1 (which we will call t); this will be measurement (1), made at time t_1.

2. F measures S with respect to

   $$|{\uparrow}\rangle\langle{\uparrow}| - |{\downarrow}\rangle\langle{\downarrow}| \in B(\mathbb{C}^2),$$

   getting either 1 (↑ for clarity) or -1 (↓ for clarity); this will be measurement (2), made at time t_2.

[2] When we have two entangled systems, we can measure one by measuring the other. This differs from how the systems R and S are treated in the Frauchiger-Renner though experiment in the sense that $\bar{F}$ cannot measure S by measuring R, since $\bar{F}$ physically prepares the state and then moves it to the other lab, while if R and S would be entangled, $\bar{F}$ could measure S by measuring R. However, in this experiment $\bar{F}$ always measures first, so the extra option does not constitute a problem.

3. $\bar{W}$ measures $\bar{L}$ with respect to

$$|\overline{\text{ok}}\rangle\langle\overline{\text{ok}}| - |\overline{\text{fail}}\rangle\langle\overline{\text{fail}}| \in B(\mathbb{C}^2),$$

getting either 1 ('$\overline{\text{ok}}$') or -1 ('$\overline{\text{fail}}$'); this will be measurement (3).
Note: as we will see later, it seems that $\bar{W}$ implicitly makes another measurement here, one of L with respect to

$$|\uparrow\rangle_L \langle\uparrow|_L - |\downarrow\rangle_L \langle\downarrow|_L \in B(\mathbb{C}^2),$$

which we call measurement (4). Both measurements (3) and (4) will be made at time t_3.

4. Finally, W measures L with respect to

$$|\text{ok}\rangle \langle\text{ok}| - |\text{fail}\rangle \langle\text{fail}| \in B(\mathbb{C}^2),$$

getting either 1 ('ok') or -1 ('fail'); this will be measurement (5), made at time t_4.

Sometimes, agents implicitly make 2 measurements at the same time. For example, at a certain point, $\bar{W}$ implicitly measures both L with respect to the basis $(|\uparrow\rangle, |\downarrow\rangle)$ and $\bar{L}$ with respect to the basis $(|\text{ok}\rangle, |\text{fail}\rangle)$. We will denote this as measuring with respect to the observable

$$|\text{ok}, \uparrow\rangle \langle\text{ok}, \uparrow| + 2\,|\text{fail}, \uparrow\rangle \langle\text{fail}, \uparrow| + 3\,|\text{ok}, \downarrow\rangle \langle\text{ok}, \downarrow| + 4\,|\text{fail}, \downarrow\rangle \langle\text{fail}, \downarrow|$$

and denote the outcome 1 as $(\text{ok}, \uparrow)$, 2 as $(\text{fail}, \uparrow)$, etc.

We repeat the experiment until both $\bar{W}$ and W measure 'ok'. As we will see in the next section, according to Frauchiger and Renner, there are two ways in which we can calculate the probability that this happens, which give different results.

Chapter 3

Discussion of the paradox

3.1 Analysis of information and the paradox

We will now analyze the reasoning Frauchiger and Renner use to come to their contradiction.[1] On the one hand, they note that all the agents agree that the initial state of the system $\bar{L}L$ is

$$
\begin{aligned}
|i\rangle_{L\bar{L}} &= \sqrt{\frac{1}{3}}\Big(|t\rangle_{\bar{L}} \,|\uparrow\rangle_L + |t\rangle_{\bar{L}} \,|\downarrow\rangle_L + |h\rangle_{\bar{L}} \,|\downarrow\rangle_L \Big) && (3.1)\\
&= \sqrt{\frac{1}{6}}\Big(|\overline{\text{fail}}\rangle_{\bar{L}} \,|\uparrow\rangle_L + 2|\overline{\text{fail}}\rangle_{\bar{L}} \,|\downarrow\rangle_L - |\overline{\text{ok}}\rangle_{\bar{L}} \,|\uparrow\rangle_L \Big) && (3.2)\\
&= \sqrt{\frac{1}{6}}\Big(2\,|t\rangle_{\bar{L}} \,|\text{fail}\rangle_L + |h\rangle_{\bar{L}} \,|\text{fail}\rangle_L + |h\rangle_{\bar{L}} \,|\text{ok}\rangle_L \Big) && (3.3)\\
&= \sqrt{\frac{1}{12}}\Big(3|\overline{\text{fail}}\rangle_{\bar{L}} \,|\text{fail}\rangle_L + |\overline{\text{fail}}\rangle_{\bar{L}} \,|\text{ok}\rangle_L - |\overline{\text{ok}}\rangle_{\bar{L}} \,|\text{fail}\rangle_L \\
&\quad + |\overline{\text{ok}}\rangle_{\bar{L}} \,|\text{ok}\rangle_L \Big), && (3.4)
\end{aligned}
$$

using notation from (2.2). Frauchiger and Renner claim that if we prepare a system in a state that is *not* orthogonal to a given eigenspace of some eigenvalue λ of an observable a, and then measure the system with respect to a, and if we then repeat the preparing and the measuring, eventually we will find λ (as opposed to when the initial state is orthogonal to the eigenspace belonging to λ, because then we will never measure λ). They then claim that this follows from postulate 2.5. I find the proof that they give somewhat confusing (starting with the fact that they equate a function to a vector), and I will just assume it, since I don not believe this to be a widely contested claim. When assuming this, we see that

$$\left\langle \overline{\text{ok}}_{\bar{L}}, \text{ok}_L \middle| i_{\bar{L}L} \right\rangle = \frac{1}{12}$$

[1]In the interest of making this analysis as easily comparable to the Frauchiger-Renner paper as possible, I should note that in this analysis I have made some adjustments to the order in which the calculations are made for narrative purposes.

implies that if we perform the experiment sufficiently many times, the outcome

$$m(3) = \overline{\text{ok}} \ \wedge \ m(5) = \text{ok} \tag{3.5}$$

will occur. On the other hand, Frauchiger and Renner claim, quantum mechanics *also* predicts that this never happens. Their reasoning goes as follows. First, using (3.2), we see that

$$\left\langle \overline{\text{ok}}_{\bar{L}}, \downarrow_L \middle| i \right\rangle_{L\bar{L}} = 0$$

so we apply Lemma 2.6 to agent $\overline{W}$ and system $\bar{L}L$ at time t_3:

$$\bar{L}L(t_3) \sim_{\bar{W}} |i\rangle \ \vdash \ m(\bar{L}L, r_{\bar{L}} \otimes s_L, Z) \neq \overline{\text{ok}}, \downarrow \tag{3.6}$$

Frauchiger and Renner now argue that if we assume $\overline{\text{ok}}$ is found, that means that $\bar{W}$ knows that S cannot be in state $\downarrow$. There aren't many details about this reasoning in the paper, and so I have tried to fill in with what I think was meant. This involves two parts; the 'breaking up' of the fact that $\bar{W}$ can't measure $\overline{\text{ok}}$ and $\downarrow$ at the same time, and the way $\bar{W}$ measuring the lab L doesn't result in $\downarrow$ so that the system S *inside* L would then be in state $|\uparrow\rangle$.

The first statement we need to make is about the relationship between physical systems: how can we translate a measurement of a larger system to measurements of subsystems? As noted in section 1.2, composite systems in quantum mechanics are usually described by the tensor product. So we need an axiom that describes that a measurement will 'behave well' under some kind of composition operation between physical systems, such as

$$\left(\left(m(\bar{L}L(t_3), r_{\bar{L}} \otimes s_L, \bar{W}) \neq (\overline{\text{ok}}, \downarrow) \right) \ \wedge \ \left(m(3) = \overline{\text{ok}} \right) \right) \vdash \ m(4) \neq \downarrow \tag{3.7}$$

In section 3.2, we will discuss whether this assumption is a good reflection of the current understanding of the measuring of multi-body systems.

Frauchiger and Renner claim that this means that F must have found $\uparrow$ earlier, since a measurement of the whole lab at time t_3 cannot result in $\downarrow$. But this looks a lot like Wigner's Friend paradox: what are the claims that we can make about a measurement made in a superposition *before* the time that a measurement collapses that superposition? For now, we will denote this as

$$\left(m(4) = \ \uparrow \right) \ \vdash \ \left(m(2) = \ \uparrow \right). \tag{3.8}$$

We now apply lemma 2.10 to get

$$\left(m(2) \neq \downarrow \right) \ \vdash \ \left(m(2) = \uparrow \right). \tag{3.9}$$

Putting what we have until now together for the experiment as described above,

$$\left(m(3) = \overline{\text{ok}} \right) \ \vdash \ \left(m(2) = \uparrow \right), \tag{3.10}$$

where we have dropped the assumption $\bar{L}L(t_3) \sim_{\bar{W}} U_{t_1}^{t_3} |i\rangle$ for legibility.

Now, we will be reasoning from the perspective of F. The argument in the paper is the same as used for (3.7), only this time, different observers measure both subsystems:

$$((m(RS(t_1), r \otimes s, F) \neq (h, \uparrow)) \;\;\wedge\;\; (m(2) = \uparrow)) \;\;\vdash\;\; (m(1) \neq \;\; h)\,. \tag{3.11}$$

Note also that measurement (2) is made later than t_1, and so here we must take into account that F might re-collapse the state. We will discuss this together with the similar equation (3.7) in section 3.2

Now Frauchiger and Renner use this to imply that when $\bar{W}$ measures $\overline{\text{ok}}$, he knows that $\bar{F}$ measured t. In our formulas, this would be the statement

$$\begin{aligned} & m(3) = \overline{\text{ok}} \;\;\vdash\;\; m(2) = \uparrow \\ \wedge\;\; & m(2) = \uparrow \;\vdash\;\; m(1) = t \\ \Longrightarrow\; & m(3) = \overline{\text{ok}} \;\;\vdash\;\; m(1) = t. \end{aligned} \tag{3.12}$$

Frauchiger and Renner seem to invoke assumption 2.7, transferring the inference from one agent to another. Whether these two statements actually agree will be discussed in section 3.2.

Now starting from $m(1) = t$, we again use the entanglement of R and S to say

$$\begin{aligned} & RS(t_1) \sim_{\bar{F}} \sqrt{1/3}\,(|h\rangle\,|{\downarrow}\rangle + |t\rangle\,|{\uparrow}\rangle + |t\rangle\,|{\downarrow}\rangle) \;\;\wedge\;\; m(1) = t \\ & \vdash S(t_1) \sim_{\bar{F}} \sqrt{1/2}(|{\uparrow}\rangle + |{\downarrow}\rangle), \end{aligned} \tag{3.13}$$

which again is not an assumption made explicitly, and we will discuss the validity later. Finally, Frauchiger and Renner conclude that

$$S(t_1) \sim_{\bar{F}} \sqrt{1/2}(|{\uparrow}\rangle + |{\downarrow}\rangle) \vdash m(5) \neq \text{ok}; \tag{3.14}$$

Putting everything together, we get

$$m(3) = \overline{\text{ok}} \;\;\vdash\;\; m(5) \neq \text{ok}. \tag{3.15}$$

This, according to Frauchiger and Renner, forms a contradiction with (3.5), which has to be true at some point.

3.2 Explicit and Implicit assumptions

We have now gone through the entire proof that Frauchiger and Renner give. This means that we can try to 'fill in the gaps': which statements can be considered logical assumptions, and which cannot?

Assumption Q is, as noted above, a special case of quantum mechanics, which we use all the time. This does not seem controversial to me. **Assumption S**, as it is formulated in Postulate 2.9, never explicitly used, and as stated in section 2.1.3, it is just a result of the logic system that seems appropriate here. Lemma 2.10 is what I think they really meant, since this is actually used in the paper. Again, the fact that after a measurement, a quantum system is in a well-defined state, is well-established.

Frauchiger and Renner claim to use **assumption C** in (3.12). As stated in section 2.1, the only difference in information different people have is that they made measurements and so have a formula of the form $m(n) = \lambda$ which they know to be true. But an agent cannot reason what another agent will measure unless some outcome has a probability 1 of happening, otherwise it would not be a random measurement. So assumption C does not add anything. The way the reasoning is worded in the paper, like (3.12), also does not introduce anything new: it is essentially

$$((a \to b) \wedge (b \to c)) \to (a \to c),$$

and if such a formula would not be true in general, proving statements would become very difficult.

This leaves us with the assumptions that were not stated beforehand. Firstly, eq. (3.7) describes a measurement of a composite system and how it relates to the subsystems, as does eq. (3.13). Everything happens at the same time by the same agent, so this does not seem very controversial to me.

Eq. (3.11) does something similar, but I disagree with this assumption: while it *is* true that a measurement done by F of R at time t_2 wouldn't have resulted in h, the formula at the right hand side of eq. (3.11) is about $\bar{F}$ at t_1! At that point, from the perspective of F, $\bar{F}$ was in a superposition, so it may well be that $m(1) = h$ was not valid, but that does not mean that $\neg m(1) = h$ was. This assumption is therefore questionable at best.

This is the same in (3.8): we cannot definitively say that the measurement at time t_2 would have resulted in $\uparrow$, even though at time t_4 it has a well-defined value. So, this is also a questionable assumption.

These assumptions have a lot in common with Wigner's friend paradox. When someone is measuring in a lab, and the lab is in a superposition of all the possible outcomes of the measurement, even though when we look inside the lab later we can conclude what was measured earlier, we cannot conclude that the outcome was welln defined at an earlier time.

Finally, (3.14). This would have been true if measurement 5 was also made at time t_1. However, now it is only true if the state of S isn't altered between t_1 and t_4. But the lab L is implicitly measured by $\bar{W}$. This is the same situation as a Bell measurement, and there we have seen, through Bell's Theorem, that measurements of entangled systems indeed *can* influence each other. So this assumption is also invalid.

3.3 Comparison to historical perspective: Hardy's paradox

Just like in Hardy's paradox, the Frauchiger-Renner paradox calculates a certain probability in 2 ways: one by considering the system as a whole (the straightforward way), and one by considering subsystems (reasoning from a certain 'point of view'). Like mentioned in section 1.4.3, in a superposition the states are not well-defined and trying to reason by looking at all the different possibilities, even if it is not known which possibility occurred, will result in contradictions. Specifically, the systems S and R correspond to the electron and the proton in Hardy's paradox, and $\bar{F}$ and F correspond to the first beam splitters and the annihilation, in the sense that they create the superposition and entangle the particles. The second beam splitters then correspond to $\bar{W}$ and W. The only difference is exactly how the intial state is entangled, which gives slightly different probabilities. By then changing the basis with which all the agents measure, Hardy's paradox can be exactly replicated. Perhaps the only significant difference is the fact that in the Frauchiger-Renner paradox, there are conscious beings in the superpositions. However, to me, there seems to be no evidence of consciousness playing a special role in quantum mechanics, and the two paradoxes are therefore equivalent.

3.4 Reactions to the paper

As is to be expected, the Frauchiger-Renner paper spawned many reactions, ranging from dismissal of the proof to claiming that the assumptions are not satisfied in a certain interpretation of quantum mechanics and developing new interpretations that circumvent the theorem. There are, however, too many papers to discuss here, many of which contain reasoning which I am unable to follow. I would, however, like to mention some papers with ideas similar to this paper.

For example, in [Sudbery, 2019] it is argued that there are hidden assumptions which make the theorem invalid. Notably, one of the 5 (!) hidden assumptions is locality, which through Bell's theorem is in contradiction with quantum mechanics, and which is also what we concluded in section 3.2.

On the other hand, [Tausk, 2018] takes issue with the reasoning that a measurement of tails by $\bar{F}$ leads to a measurement of W resulting in "fail". They argue that since $\bar{F}$ is in a superposition with respect to W, this means that the "heads" possibility interferes with the state, and therefore the direct inference is invalid. This is very similar to the reasoning we used to disqualify (3.11) and (3.8). [Healey, 2018] also finds an additional assumption, which he calls *Intervention insensitivity*:

> The truth-value of an outcome-counterfactual is insensitive to the occurrence of a physically isolated intervening event.

An outcome-counterfactual is then defined as a statement of the form 'If the outcome of a quantum measurement a had been x, then the outcome of another measurement b would have been y'. Healey also suggests that this assumption violates Bell's theorem, and should be rejected, saying that the Frauchiger-Renner paradox "raises no *new* worry about non-locality".

3.5 Conclusion

In conclusion, it seems that the Frauchiger-Renner paradox does not offer a restriction on the interpretations of quantum mechanics which we can use, instead offering another example of why careful analysis is often needed in quantum mechanics. Careful analysis of the reasoning used in [Frauchiger and Renner, 2018] suggests that assumptions were used that do not reflect quantum mechanics and instead imply a contradiction through Bell's Theorem, and a closer inspection of the thought experiment suggests that the thought experiment is almost identical to the one used in Hardy's paradox.

The logical formalism used in this thesis is not a true axiomatization of quantum logic, and although it is based upon concepts in mathematical logic like formal theories and proofs, a closer examination of the possibilities an problems of such a logical system in order to avoid future confusion might be desirable.

Bibliography

[Aharonov et al., 2002] Aharonov, Y., Botero, A., Popescu, S., Reznik, B., and Tollaksen, J. (2002). Revisiting Hardy's paradox: counterfactual statements, real measurements, entanglement and weak values. *Physics Letters A*, 301(3-4):130–138.

[Bell, 1964] Bell, J. S. (1964). On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195.

[Bohm, 1960] Bohm, D. (1960). *Quantum Theory*. Prentice-Hall.

[Clauser et al., 1969] Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1969). Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880.

[Einstein et al., 1935] Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777.

[Frauchiger and Renner, 2018] Frauchiger, D. and Renner, R. (2018). Quantum theory cannot consistently describe the use of itself. *Nature communications*, 9(1):3711.

[Greenberger et al., 1990] Greenberger, D. M., Horne, M. A., Shimony, A., and Zeilinger, A. (1990). Bells theorem without inequalities. *American Journal of Physics*, 58(12):1131–1143.

[Greenberger et al., 1989] Greenberger, D. M., Horne, M. A., and Zeilinger, A. (1989). Going beyond Bells theorem. In *Bells theorem, quantum theory and conceptions of the universe*, pages 69–72. Springer.

[Griffiths, 2005] Griffiths, D. (2005). *Introduction to Quantum Mechanics*. Pearson international edition. Pearson Prentice Hall.

[Hardy, 1992] Hardy, L. (1992). Quantum mechanics, local realistic theories, and lorentz-invariant realistic theories. *Physical Review Letters*, 68(20):2981.

[Healey, 2017] Healey, R. (2017). *The quantum revolution in philosophy*. Oxford University Press.

[Healey, 2018] Healey, R. (2018). Quantum theory and the limits of objectivity. *Foundations of Physics*, 48(11):1568–1589.

[Landsman, 2017] Landsman, K. (2017). *Foundations of Quantum Theory*, volume 188 of *Fundamental Theories of Physics*. Springer.

[MacCluer, 2009] MacCluer, B. D. (2009). *Elementary Functional Analysis*. Springer.

[Sudbery, 2019] Sudbery, A. (2019). The hidden assumptions of Frauchiger and Renner. *arXiv preprint arXiv:1905.13248*.

[Tausk, 2018] Tausk, D. V. (2018). A brief introduction to the foundations of quantum theory and an analysis of the Frauchiger-Renner paradox. *arXiv preprint arXiv:1812.11140*.

[Wigner, 1961] Wigner, E. (1961). Remarks on the mind-body problem. *The Scientist Speculates*. Heineman, London.