

INFORMATION THEORETICAL RECONSTRUCTIONS OF QUANTUM THEORY

BACHELOR THESIS
for the degree of
BACHELOR OF SCIENCE
at the
UNIVERSITY OF VIENNA



submitted by
Ferdinand Horvath

guided by
Ao. Univ. Prof. Dr. Reinhold A. Bertlmann

Vienna, September 2013

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht sind und dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt wurde.

Wien, den November 15, 2013

Contents

1	Introduction	1
1.1	The Idea of Reconstructing Quantum Theory	1
1.2	An outlook on the work in hand	2
2	Hardy’s ”Five Reasonable Axioms”	3
2.1	Classical Probability Theory (CPT)	4
2.2	Quantum Theory (QT)	5
2.3	Deriving Quantum Theory	8
3	Zeilinger’s ”Fundamental Principle”	12
3.1	Limited Information	12
3.2	Time Evolution	13
3.3	Entanglement in the light of the ”fundamental principle”	15
4	Brukner and Dakič: Quantum Theory from three Axioms	17
4.1	Composite Systems and Locality	18
4.2	Ruling Out Higher-Order Theories	20
5	A Bayesian Approach to the Measurement Problem	21
5.1	Quantum Collapse as Bayes’ Rule	21
5.2	Gleason’s Theorem, Born’s Rule and Entanglement	25
6	Interpreting the Information Theoretic Approach to Quantum Theory	27

1 Introduction

1.1 The Idea of Reconstructing Quantum Theory

Quantum theory is a very complex and abstract thing that can't even properly be explained to non-physicists because its core ideas are all about Hilbert spaces, eigenvectors and Hermitean operators. The idea of reconstructing quantum theory is to base the theory on a couple of intuitive axioms from which the abstract formalism can then be derived.

An often cited dictum is a comment Christopher Fuchs made in [1] on the purpose of reconstructing quantum theory. Fuchs compares the present state of quantum theory with that of electro-magnetism before Einstein's invention of special relativity theory. As is well known, Lorentz transformations, the heart of relativity theory, were known long before relativity theory itself – they were published as early as 1895. At that time, much of the empirical predictions of relativity could already be produced but couldn't possibly be understood. What was needed was an understanding for the origin of the mathematical structure of Lorentz transformations. That is, what was asked for were the physical concepts that give rise to the theory's unintuitive mathematics. Only when $x' = \frac{x-vt}{\sqrt{1-v^2/c^2}}$ and $t' = \frac{t-vx/c^2}{\sqrt{1-v^2/c^2}}$ were replaced by the simple statements

1. The speed of light is constant.
2. The laws of physics are the same in all inertial frames.

could the phenomena really be comprehended. Fuchs claims that it is this phase we're currently stuck in as far as quantum theory is concerned: we know all the equations but nothing of their conceptual origin. Therefore, the task that Fuchs and many others try to take on is to find simple principles, expressed in common language, from which quantum theory can be derived, and thus to provide the mathematical axioms of quantum theory with a real meaning.

It should be noted that this approach is quite different from *interpreting* quantum theory. Interpretations like Bohm's mechanic [2] or Everett's many worlds [3] merely take the existing formalism and endow it with a different ontology. The intention of the reconstructions discussed here is different from such interpretations in that they try to justify the formalism by finding a conceptual foundation for it, not by building a conceptual framework around the existing mathematics. As Grindbaum puts it in [4], "as the physics of today is inseparable from mathematics, a meaning cannot be physical and thus satisfactory if it is merely heaped over and above the mathematical formalism of quantum mechanics, instead of coming all the way along with the formalism as it rises in a derivation of the theory."

Specifically, the idea of reconstructing quantum theory is that all the mathematical axioms about Hilbert spaces and Hermitean operators have to be based on how information about a system is acquired and knowledge is updated. Therefore, it is (quantum) information theory that Fuchs and many others chose as the venue of their quest for simple principles.

The idea to reconstruct quantum theory from first principles is not new, by any means.¹ Already in 1927, Hilbert, von Neumann and Nordheim, establishing this axiomatic approach, published a paper in which "[the theory's] analytical apparatus, and the arithmetic quantities occurring in it, receives on the basis of the physical postulates a physical interpretation. Here, the aim is to formulate the physical requirements so completely that the analytical apparatus is just uniquely determined. Thus the route is of axiomatization."² Subsequently, a lot of work was done on the axiomatic foundations of quantum theory. However, until recently most of this work utilized very abstract mathematical notions as their axioms for the theory – not simple physical principles. For this reason, those attempts do not fall under the notion of reconstruction that is meant here.

1.2 An outlook on the work in hand

Probably the best example of a reconstruction of the theory, and the most influential, is Hardy's work [6], in which he manages to derive the full formalism of quantum theory from five plausible axioms. Much of the current research is in some way based on this article, which is why it will be discussed here at some length. Since its appearance, some progress has been made in making the choice of axioms more elaborate and their list shorter. One example from this line of work is Brukner's and Dakič's [7], which makes do with only three axioms. This paper is also based on an idea of Zeilinger and Brukner, discussed in Section 3, the so-called fundamental principle, which postulates that the information carrying capacity of any quantum system is finite. With this assumption, Zeilinger and Brukner manage to make sense of quantum entanglement and randomness.

The fundamental principle is an interesting example of how the information theoretic view opens new doors on the path to a bigger picture of quantum theory. Another example of such a prospect is the Bayesian approach Fuchs discusses in [1]. This program, while not a reconstruction of the theory, still grants a natural way of dealing with one of the most troubling aspect of the quantum world – the so-called measurement problem or wavefunction collapse. The Bayesian solution to this problem is presented in Section 5.

Although this work is focused on the reconstructions presented by Hardy and

¹Compare [4].

²Hilbert, von Neumann, Nordheim, Pergamon Press 1961, 104-133.

Brukner and Dakič (Sections 2 and 4, respectively), it also tries to generally give an overview of the gains that the information theoretic view provides us with – that is, the new perspectives on entanglement, quantum randomness and the measurement process.

2 Hardy’s ”Five Reasonable Axioms”

Since the majority of papers on reconstruction of quantum theory (QT) is based on Lucien Hardy’s paper ”Quantum theory from five reasonable axioms” [6], Hardy’s work will be reproduced here first as an introduction to the topic. Hardy shows that it is possible to derive quantum theory from the following five axioms. Two numbers play a special role in his work, which we will now define. We introduce K as the number of degrees of freedom of a system, i.e. the minimum number of probability measurements required to determine the state of the system. Here, we implicitly assume that there is enough structure in our theory so that we need only a finite amount of information to determine the state. Furthermore, we call N the dimension of the system, i.e. the maximum number of states that can be distinguished in a single measurement.

Axiom 1 Probabilities. Relative frequencies tend to the same value for any case where a given measurement is performed on an ensemble of n systems prepared by some given preparation in the limit as n becomes infinite.

Axiom 2 Simplicity. The theory exhibits a certain simplicity such that $K = K(N)$ and K takes the minimum value consistent with the other axioms.

Axiom 3 Subspace. A system whose state is constrained to belong to an M dimensional subspace of a larger state behaves like a system of dimension M .

Axiom 4 Composite systems. For a system consisting of the subsystems A and B , $N = N_A N_B$ and $K = K_A K_B$.

Axiom 5 Continuity. There is a continuous reversible transformation between any two pure states of a system.

Hardy also shows that it is only Axiom 5 that distinguishes quantum theory from classical probability theory (CPT) while Axioms 1 - 4 are also obeyed by CPT. To arrive at this result, Hardy shows how QT can be formulated in a manner similar to CPT. To this end, he first describes CPT.

2.1 Classical Probability Theory (CPT)

A classical system can be in N distinguishable states, each occurring with a probability p_n . The state of the system can thus be described by a (non-normalized) vector

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix}.$$

Here, a state is determined through N probabilities, therefore for CPT $K = N$. Furthermore, we define the extremal states

$$\mathbf{p}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{etc.}$$

and the null state

$$\mathbf{p}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

which symbolizes the system not being present. Now we define pure states to be all extremal states except the null state, so a pure state corresponds to a system being in one definite state.

We call the (non-unique) minimal set of K probability measurements required to determine the state of a system the *fiducial* measurements. Associated with these are the vectors $\mathbf{r}_1 = (1, 0, \dots, 0)^T$, $\mathbf{r}_2 = (0, 1, \dots, 0)^T$ etc. The k th fiducial measurement gives the k th component of \mathbf{p} , i.e. the probability that the system is in the state with label k . If we measure whether a system \mathbf{p} is in a pure state, denoted by \mathbf{r}_n , then the probability measured is given by

$$p_{meas} = \mathbf{r}_n \cdot \mathbf{p}. \tag{1}$$

However, it is still unclear if this holds for any general measurement. Let's assume that the probability is determined by some general function $f(\mathbf{p})$ and consider the situation of Alice randomly preparing a system in state \mathbf{p}_1 with probability λ and in state \mathbf{p}_2 with probability $1 - \lambda$ before sending it to Bob. So the state prepared by Alice is

$$\mathbf{p} = \lambda \mathbf{p}_1 + (1 - \lambda) \mathbf{p}_2. \tag{2}$$

The probability that Bob then measures is

$$f(\mathbf{p}) = \lambda f(\mathbf{p}_1) + (1 - \lambda)f(\mathbf{p}_2). \quad (3)$$

Combining these equations yields

$$f(\lambda\mathbf{p}_1 + (1 - \lambda)\mathbf{p}_2) = \lambda f(\mathbf{p}_1) + (1 - \lambda)f(\mathbf{p}_2), \quad (4)$$

thus f is linear and we can generally write³

$$p_{meas} = \mathbf{r} \cdot \mathbf{p}. \quad (5)$$

We define the identity measurement

$$\mathbf{r}^{\mathbb{1}} = \sum_{l=1}^L \mathbf{r}_l,$$

which represents measuring whether the system is present. \mathbf{r}_l are the measurement vectors corresponding to the L different outputs our measurement device can display for a given setting.

CPT satisfies Axioms 1 to 4, but not Axiom 5, which demands the existence of a reversible continuous transformation between pure states. When a reversible transformation is applied to a pure state, it is always mapped onto a pure state. This can easily be shown by reductio: consider the mixture $Z\mathbf{p} = \lambda\mathbf{p}_A + (1 - \lambda)\mathbf{p}_B$, where \mathbf{p} is pure and Z^{-1} is an allowed transformation and Z is linear. Then

$$\mathbf{p} = \lambda Z^{-1}\mathbf{p}_A + (1 - \lambda)Z^{-1}\mathbf{p}_B$$

is a mixture for distinct $\mathbf{p}_{A,B}$, which is a contradiction. Therefore, a reversible continuous transformation would take pure states along a continuous trajectory through the space of pure states. This can't work for CPT, because it exhibits only a finite number of N pure states for a system of dimension N . For this reason, it is Axiom 5 that distinguishes CPT from QT, which satisfies all five axioms.

2.2 Quantum Theory (QT)

We will now follow Hardy's description of QT in a formulation similar to the one used to describe CPT. First, note that Hermitean operators (such as the operator $\hat{\rho}$ used to describe a state) acting on a N dimensional complex Hilbert space can be spanned by N^2 linearly independent projection operators \hat{P}_K for $K = 1$ to $K = N^2$. This is because, in its matrix representation, a Hermitean operator

³See Appendix 1 in [6] for further details.

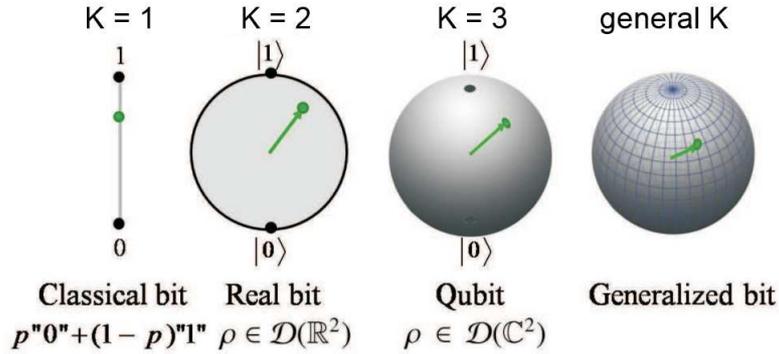


Figure 1: State spaces of two dimensional systems for different $K(N)$. Left to right: A classical bit with the weight p of the mixture as sole parameter, a real bit with two real parameters, a qubit with three real parameters and $\hat{\rho}$ represented by a 2×2 complex density matrix, and a generalised bit for which n real parameters are needed to specify the state. Note that when one moves continuously from one pure state (a spoint on the surface) to another, only in CPT one has to traverse set of mixed states. Whether probability theories with higher-order K can meet our demands will be investigated in Section 4.2.

is characterized by N real numbers along its diagonale and $\frac{1}{2}N(N-1)$ complex number above its diagonale, which add up to N^2 real paramters. We thus note that, unlike CPT where $K=N$, QT represents the case where $K(N)=N^2$. We collect the above projectors into

$$\hat{\mathbf{P}} = \begin{pmatrix} \hat{P}_1 \\ \hat{P}_2 \\ \vdots \\ \hat{P}_{K=N^2} \end{pmatrix}. \quad (6)$$

Since any Hermitean operator \hat{A} is a linear combination of these projectors, we can write $\hat{A} = \mathbf{a} \cdot \hat{\mathbf{P}}$, for real \mathbf{a} . We define the state probability vector

$$\mathbf{p}_S = \text{tr}(\hat{\mathbf{P}}\hat{\rho}) \quad (7)$$

such that its k th component is the probability when \hat{P}_k is measured on $\hat{\rho}$. \mathbf{p}_S thus contains all the information that the density matrix $\hat{\rho}$ contains and can alternatively be used to describe a system. We define the measurement vector \mathbf{r}_M corresponding to a QT-measurement operator \hat{A} as

$$\hat{A} = \mathbf{r}_M \cdot \hat{\mathbf{P}}. \quad (8)$$

Like above, \mathbf{r}_M can also be used to represent a measurement as \hat{A} usually does. Since for QT the Born-rule

$$p_{meas} = \text{tr}(\hat{A}\hat{\rho}), \quad (9)$$

holds, we now have

$$p_{meas} = \mathbf{r}_M \cdot \mathbf{p}_S, \quad (10)$$

using equations (7) and (8). We can also represent the state by an \mathbf{r} -type vector defined as

$$\hat{\rho} = \hat{\mathbf{P}} \cdot \mathbf{r}_S \quad (11)$$

and the measurement by a \mathbf{p} -type vector defined as

$$\mathbf{p}_M = \text{tr}(\hat{A}\hat{\mathbf{P}}). \quad (12)$$

We insert into the QT-trace formula:

$$p_{meas} = \text{tr}(\hat{A}\hat{\rho}) = \text{tr}(\mathbf{r}_M^T \hat{\mathbf{P}} \hat{\mathbf{P}}^T \mathbf{r}_S) \quad (13)$$

$$= \mathbf{r}_M^T D \mathbf{r}_S, \quad (14)$$

where

$$D := \text{tr}(\hat{\mathbf{P}} \hat{\mathbf{P}}^T). \quad (15)$$

Comparing (10) and (14), we obtain the connection between the \mathbf{p} and \mathbf{r} representation of the state

$$\mathbf{p}_S = D \mathbf{r}_S \quad (16)$$

and inserting the \mathbf{r} -type definition of the measurement vector (8) into the definition of its \mathbf{p} -type representation (12) yields

$$\mathbf{p}_M = D^T \mathbf{r}_M. \quad (17)$$

In quantum mechanics, the evolution of a system is most generally described by a superoperator \mathcal{S} : $\hat{\rho} \rightarrow \mathcal{S}(\hat{\rho})$, where \mathcal{S} is linear, completely positive and doesn't increase the trace (see [8]). In our formalism, \mathcal{S} acts as

$$\mathbf{p}_S = \text{tr}(\hat{\mathbf{P}}\hat{\rho}) \rightarrow \text{tr}(\hat{\mathbf{P}}\mathcal{S}(\hat{\rho})) \quad (18)$$

$$= \text{tr}(\hat{\mathbf{P}}\mathcal{S}(\hat{\mathbf{P}}^T D^{-1} \mathbf{p}_S)). \quad (19)$$

Defining

$$Z := \text{tr}(\hat{\mathbf{P}}\mathcal{S}(\hat{\mathbf{P}})^T) D^{-1}, \quad (20)$$

we find that the evolution of a system represented by a \mathbf{p} -vector can simply be described by a linear transformation

$$\mathbf{p} \rightarrow Z\mathbf{p}. \quad (21)$$

2.3 Deriving Quantum Theory

Summarizing the above, Hardy shows that states and measurements can be represented by probability vectors \mathbf{p} and \mathbf{r} , respectively, each with N^2 components. Measured probabilities are given by $p_{meas} = \mathbf{r} \cdot \mathbf{p}$ and evolution of a system is given by $\mathbf{p} \rightarrow Z\mathbf{p}$. We will now reproduce Hardy's proof that this formulation of QT can be derived from his five axioms. First, these axioms are revised and explained some more.

Axiom 1 *Probabilities.* Relative frequencies tend to the same value for any case where a given measurement is performed on an ensemble of n systems prepared by some given preparation in the limit as n becomes infinite.

The first axiom ensures that the notion of probability is well defined. It states that, when a particular outcome is observed in n_+ cases for a given measurement on n systems with same preparation, the probability

$$\text{prob}_+ = \lim_{n \rightarrow \infty} \frac{n_+}{n}$$

takes the same value for any ensemble. This enables us to identify a system with a preparation, because only the preparation determines the probabilities measured.⁴

Axiom 2 *Simplicity.* The number of degrees of freedom K is determined by a function of the dimension of a system N where $N = 1, 2, \dots$ and takes the minimum value consistent with the axioms.

This axiom is motivated by the wish to obtain the simplest theory consistent with our other axioms. We will see in section 4 that it is possible to derive QT without this axiom.

Axiom 3 *Subspace* A system whose state is constrained to belong to an M dimensional subspace of a larger state behaves like a system of dimension M .

Hardy says about this axiom

"In logical terms, we can think of distinguishable states as corresponding to a propositions [sic]. We expect a probability theory pertaining to m propositions to be independent of whether these propositions are a subset [of] some larger set or not."⁵

Axiom 4 *Composite systems.* For a system consisting of the subsystems A and B , $N = N_A N_B$ and $K = K_A K_B$.

⁴Axiom 1 was actually shown to be redundant for Hardy's purpose (see [9]).

⁵L. Hardy, 2008, 14.

In principle, it could be possible that the whole system AB exhibits more than $N_A N_B$ distinguishable states, but this would be rather unintuitive and is thus disregarded. Hardy shows that, under some assumptions, $K = K_A K_B$ follows from $N = N_A N_B$. It should also be noted that Axiom 4 is equivalent to the notion of locality: $K = K_A K_B$ means that the state of the entire system can be determined by performing only local measurements on the two subsystems A and B .

Axiom 5 *Continuity*. There is a continuous reversible transformation between any two pure states of a system.

Hardy motivates this axiom with the observation that there are generally no discontinuities in physics.

To prove that QT derives from these Axioms, Hardy first shows that $K(N) = N^r$ for $r = 1, 2, \dots$. To show this, we first note that $K(N)$ has to be a strictly increasing function. Consider an $N + 1$ dimensional system composed of an N dimensional subsystem W and its 1 dimensional complement \overline{W} . By Axiom 3, subsystem W behaves like a genuinely N dimensional system and therefore has $K(N)$ degrees of freedom. \overline{W} has to have at least one degree of freedom. If the state of the whole system is a mixture of W and \overline{W} , weighted with λ and $1 - \lambda$, respectively, this system has to have at least $K(N) + 1$ degrees of freedom. We obtain that $K(N + 1) \geq K(N) + 1$. By Axiom 4, $K(N_A N_B) = K_A K_B$, which means that K is a so-called completely multiplicative function. Hardy shows that any such function has to be of the form $K = N^\alpha$ (see Appendix 2 in [6]). Since K has to be integer, we arrive at $K = N^r$ where $r = 1, 2, \dots$.

Now we want to rule out the case of $K = N$, i.e. $r = 1$. For $K = N$ there exist N fiducial measurement vectors which we can set to be the basis vectors. For pure states, we have

$$p_{meas} = \mathbf{r}^T D \mathbf{r} = 1 \quad (22)$$

(compare equation (14)). Since the fiducial vectors are the basis vectors, D (defined in equation (15) is the identity:

$$D = \mathbf{1}. \quad (23)$$

This means that

$$\sum_{k=1}^N (p^k)^2 = 1, \quad (24)$$

where p^k is the k th component of \mathbf{p} and $0 \leq p^k \leq 1$. However, normalization requires that

$$\sum_{k=1}^N p^k = 1. \quad (25)$$

Taken together, these requirements can only be met by \mathbf{p} having one p^k equal to one and all others equal to zero. This means, only the basis vectors can represent pure states, so the pure states would form a discrete set. As we have seen above, this is in conflict with Axiom 5, which requires a continuous set of pure states. Therefore, the case $K = N$ can be ruled out.

This leaves us with the case of $K = N^2$. By the simplicity axiom we don't need to consider cases with higher powers of N . The simplest non-trivial case here is $N = 2$ and $K = 4$. Normalized states of this kind are contained in a $K - 1 = 3$ dimensional set with a two-dimensional surface corresponding to the set of pure states. By proving that D takes just the form required by QT, Hardy shows that with an appropriate choice of fiducial states, this surface is exactly the Bloch sphere.⁶ With the help of the subspace axiom, the case $N = 2$ can easily be generalized to higher N . Now, from D we obtain $\hat{\mathbf{P}}$ and subsequently states $\hat{\rho}$ and measurements \hat{A} by equations (11) and (8). Representing the basis state \mathbf{r}_n by $|n\rangle\langle n|$ ($n = 1$ to N) and successively transforming the basis states, any state can be generated as

$$\hat{\rho} = |\psi\rangle\langle\psi|, \quad (26)$$

where

$$|\psi\rangle = \sum_{n=1}^N c_n |n\rangle \quad (27)$$

and $\sum |c_n|^2 = 1$. Since for any such state all measurements must be non-negative, i.e. $\text{tr}(\hat{A}|\psi\rangle\langle\psi|) \geq 0 \quad \forall |\psi\rangle$, we obtain the positivity condition for operators \hat{A} . Similarly, if we associate with the measurement the state it identifies, we can write $\hat{A} = |\psi\rangle\langle\psi|$ and thus get

$$\text{tr}(|\psi\rangle\langle\psi|\hat{\rho}) \geq 0 \quad \forall |\psi\rangle, \quad (28)$$

which is the positivity condition for states. Now, a set of measurements \mathbf{r}_l that can be performed for a given measurement setting has to satisfy $\sum_l^L \mathbf{r}_l = \mathbf{r}^1$. Using the relations $\mathbb{1} = \mathbf{r}^1 \cdot \hat{\mathbf{P}}$ and $\hat{A} = \mathbf{r} \cdot \hat{\mathbf{P}}$, we obtain

$$\sum_{l=1}^L \hat{A}_l = \mathbb{1}, \quad (29)$$

as required in QT.

We will now show that the constraints on the transformation of Z on \mathbf{p} , which was

⁶See section 8.6 in [6].

shown to be equivalent to $\mathcal{S}(\hat{\rho})$, are exactly those imposed in QT. As was shown above (see equation (20)), Z takes the form

$$Z = \text{tr} \left(\hat{\mathbf{P}} \mathcal{S}(\hat{\mathbf{P}})^T \right) D^{-1}.$$

We will now consider a composite system consisting of the subsystems A and B . According to axiom 4, there are $K_A K_B$ fiducial measurements for the entire system spanned by $\hat{P}_i^A \otimes \hat{P}_j^B$ ($i = 1$ to K_A , $j = 1$ to K_B), where $\hat{P}^{A,B}$ are the fiducial sets for A and B , respectively. The projector $\hat{P}_i^A \otimes \hat{P}_j^B$ can be taken to correspond either to a preparation – the i th fiducial state at A and the j th fiducial state at B – when the operator is to represent a *state*, or to correspond to the joint probability p_{ij} for a positive outcome at A and B when performing the i th fiducial measurement on A and the j th on B , when the operator is regarded as representing a *measurement*.⁷ The joint probabilities p_{ij} will be collected in a $K_A \times K_B$ matrix, \tilde{p}_{AB} , with ij -th entry p_{ij} . In this matrix, the state of A is represented by the columns, the state of B by the rows.

We now picture an experimental setup where a preparation device sends system A (B) through the transformation device Z_A (Z_B) towards measurement A (B). By the above considerations, the columns of \tilde{p}_{AB} have to transform under Z_A and the rows under Z_B . The effect of both transformation devices acts as

$$\tilde{p}_{AB} \longrightarrow Z_A \tilde{p}_{AB} Z_B^T. \quad (30)$$

If we want to employ \mathbf{r} type vectors for the two subsystems, we can also write

$$\tilde{p}_{AB} = D_A \tilde{r}_{AB} D_B^T. \quad (31)$$

The transformations then act as

$$\tilde{r}_{AB} \longrightarrow X_A \tilde{r}_{AB} X_B^T, \quad (32)$$

where

$$X_{A,B} = D_{A,B}^{-1} Z_{A,B} D_{A,B}. \quad (33)$$

Inserting these definitions and that of Z , we get

$$\tilde{p}_{AB} \longrightarrow Z_A D_A \tilde{r}_{AB} D_B Z_B^T \quad (34)$$

$$= \text{tr} \left(\hat{\mathbf{P}} \mathcal{S}(\hat{\mathbf{P}})^T \right) \tilde{r}_{AB} \left(\text{tr} \left(\hat{\mathbf{P}} \mathcal{S}(\hat{\mathbf{P}})^T \right) \right)^T. \quad (35)$$

⁷Compare Appendix 4 in [6].

Now, we use the trace property $\text{tr}(X)\text{tr}(Y) = \text{tr}(X \otimes Y)$ and the above mentioned representation of the state as $\hat{P}_i^A \otimes \hat{P}_j^B$ to get the transformation formula of QT,

$$\begin{aligned}
p_{AB}^{ij} &\longrightarrow \sum_{kl} \text{tr} \left(\hat{P}_i \mathcal{S}_A(\hat{P}_k) \right) r_{AB}^{kl} \text{tr} \left(\hat{P}_j \mathcal{S}_B(\hat{P}_{kl}) \right) \\
&= \sum_{kl} \text{tr} \left(\hat{P}_i \otimes \hat{P}_j \mathcal{S}_A(\hat{P}_k) \otimes \mathcal{S}_B(\hat{P}_l) \right) r^{kl} \\
&= \text{tr} \left(\hat{P}_i \otimes \hat{P}_j \mathcal{S}_A \otimes \mathcal{S}_B \left(\sum_{kl} \hat{P}_k \otimes \hat{P}_l r^{kl} \right) \right) \\
&= \text{tr} \left(\hat{P}_i \otimes \hat{P}_j \mathcal{S}_A \otimes \mathcal{S}_B(\hat{\rho}) \right). \tag{36}
\end{aligned}$$

Therefore, we find equation (30) to be responsible for the tensor product structure of composite systems. The transformation Z is found to be subject to the following constraints. It preserves Hermitivity since the transformation matrix Z is real; it must not increase the trace in order for probabilities always to be bounded by 0 and 1; furthermore it is linear and completely positive. Thus, we conclude that the most general transformation consistent with Hardy's axioms is the most general transformation of quantum theory.

3 Zeilinger's "Fundamental Principle"

3.1 Limited Information

The approach of Anton Zeilinger presents a further step of presenting QT as a theory about information. Zeilinger regards physical properties as, basically, nothing but propositions (see [10]). Even more than in the work of Hardy, who depicts QT as merely a special kind of probability theory, this approach emphasizes the role information plays in QT, i.e. the question of what can be known about a system. Zeilinger and Brukner aptly cite Heisenberg on this matter:

"The laws of nature which we formulate mathematically in quantum theory deal no longer with the particles themselves but with our knowledge on the elementary particles. [...] The conception of objective reality [...] evaporated into the [...] mathematics that represent no longer the behaviour of elementary particles but rather *our knowledge of this behaviour*." ⁸

Quantum systems have the intrinsic feature of exhibiting mutually exclusive properties like those corresponding to non-commuting observables. Propositions

⁸Heisenberg, W: 1958, Daedalus 87, 95.

like "The velocity of this particle is v " and "The position of the particle is x " are cases of quantum complementarity: such propositions cannot be assigned to a system simultaneously. The key idea Zeilinger and Brukner present in [10] is to consider this as a consequence of a limit of the amount of information such a system can contain. This is what they call their "fundamental principle", which will also play a role in Section 4. In accordance with this principle, Zeilinger and Brukner make the assumption that *the information content of a quantum system is finite*. For this reason, a quantum system cannot give definite answers to all the questions we might ask it experimentally. Zeilinger sees the distinct irreducible quantum randomness as a consequence of this peculiarity – if a system gave more definite "answers" about, say, its location when the momentum is fixed, it would contain more information than the fundamental principle allows, so it has to produce random results upon measurement.

Regarding systems as carriers of propositions, the most elementary system can intuitively be identified as one that carries only one bit of information, i.e. a single proposition of "yes" or "no". For example, in the case of a Stern-Gerlach measurement on a spin- $\frac{1}{2}$ particle, "yes" and "no" could correspond to the two detector plates the particle can hit. Consequently, N elementary systems can be assumed to carry N bits of information. If these N bits are contained by the N systems jointly rather than by the individuals, then this is the case of what we have come to know as quantum entanglement. Similarly to Hardy's probability vectors, Zeilinger and Brukner introduce the quantity

$$i = p_1 - p_2,$$

where $p_{1,2}$ are the probabilities that a "yes" or "no", respectively, is obtained upon measurement. For an elementary system of dimension 2 (i.e. a spin- $\frac{1}{2}$ particle), the vector

$$\mathbf{i} = \begin{pmatrix} i_1 \\ i_2 \\ i_3 \end{pmatrix} = \begin{pmatrix} p_x^+ - p_x^- \\ p_y^+ - p_y^- \\ p_z^+ - p_z^- \end{pmatrix} \quad (37)$$

can be used as a complete description of the system insofar as it determines the probabilities of all possible measurement outcomes. Here, p_x^+ corresponds to the probability of finding the particle's spin being "up" along the x -axis, etc.

3.2 Time Evolution

Surprisingly, Zeilinger and Brukner manage to produce the evolution law of QT using the vector \mathbf{i} as their starting point. Assuming that, during time evolution, no information should be exchanged with the environment of a system, the total information of the system must be conserved. As an ansatz for the total information,

we use $I_{total} = \sum i_n^2$. Therefore,

$$I_{total}(t) = \sum_{n=1}^3 i_n^2(t) = \sum_{n=1}^3 i_n^2(t_0) = I_{total}(t_0). \quad (38)$$

This means, time evolution can be represented only through rotations of the information vector \mathbf{i} ,

$$\mathbf{i}(t) = \hat{R}\mathbf{i}(t_0). \quad (39)$$

Taking the time derivative of this equation, we get

$$\frac{d\mathbf{i}}{dt} = \frac{d\hat{R}}{dt}\mathbf{i}(t_0) = \hat{K}\mathbf{i}(t), \quad (40)$$

where $\hat{K} = \frac{d\hat{R}}{dt}\hat{R}^T$ and $\hat{R}^T = \hat{R}^{-1}$. \hat{K} can easily be shown to be antisymmetric, so when applied on a vector \mathbf{y} it acts as

$$\hat{K}\mathbf{y} = \mathbf{u} \times \mathbf{y} \quad (41)$$

for all \mathbf{y} with a unique "rotation vector" \mathbf{u} . Now, consider the decomposition of a density matrix in terms of the Pauli matrices $\hat{\sigma}_j$,

$$\hat{\rho} = \frac{1}{2} \left(\mathbb{1} + \sum_{j=1}^3 i_j(t) \hat{\sigma}_j \right). \quad (42)$$

The spin $i_j(t)$ along j is given by

$$i_j(t) = \text{tr}(\hat{\rho}\hat{\sigma}_j). \quad (43)$$

Taking the time derivative of equation (42), we get

$$i\hbar \frac{d\hat{\rho}}{dt} = \frac{1}{2} \sum_{j=1}^3 \frac{di_j(t)}{dt} \hat{\sigma}_j. \quad (44)$$

Now we insert equations (40) and (41),

$$i\hbar \frac{d\hat{\rho}}{dt} = \frac{i}{2} \sum_{i,j,k=1}^3 \epsilon_{ijk} u_i(t) i_j \hat{\sigma}_k. \quad (45)$$

The Pauli matrices satisfy

$$[\hat{\sigma}_i, \hat{\sigma}_j] = 2i \sum_{k=1}^3 \epsilon_{ijk} \hat{\sigma}_k, \quad (46)$$

therefore

$$i\hbar \frac{d\hat{\rho}}{dt} = \frac{1}{4} \sum_{i,j=1}^3 u_i(t) i_j [\hat{\sigma}_i, \hat{\sigma}_j]. \quad (47)$$

By defining

$$u_i(t) := \frac{1}{\hbar} \text{tr}(\hat{H} \hat{\sigma}_i), \quad (48)$$

we obtain the von Neumann equation

$$i\hbar \frac{d\hat{\rho}}{dt} = [\hat{H}, \hat{\rho}]. \quad (49)$$

Naturally, this derivation employs a lot assumptions that remain unjustified, but it still presents an interesting result.

3.3 Entanglement in the light of the "fundamental principle"

Another interesting perspective the work of Zeilinger and Brukner grants is on the meaning on quantum entanglement. Again, this phenomenon can be understood in the light of the "fundamental principle" of limited information capacity. For a pure product state, one bit of information is required to determine the correlation of measurement results. E.g., for a state $|\Psi\rangle = |+x\rangle|-x\rangle$, the correlation can be expressed as

$$I_{xx} = 1. \quad (50)$$

Again we use the measure of information defined in Section 3.1. For a binary experiment, it is defined as

$$I(p_1, p_2) = (p_1 - p_2)^2. \quad (51)$$

Accordingly, I_{xx} is defined as the squared difference of the probabilities of two particles having spin up along x and spin down along x , $I_{xx} = (p_{xx}^+ - p_{xx}^-)^2$.

Now, consider the maximally entangled Bell state $|\Psi^-\rangle$, of which we know that it takes the same form in all bases,

$$\begin{aligned} |\Psi^-\rangle &= \frac{1}{\sqrt{2}} (|+x\rangle|-x\rangle - |-x\rangle|+x\rangle) \\ &= \frac{1}{\sqrt{2}} (|+y\rangle|-y\rangle - |-y\rangle|+y\rangle). \end{aligned}$$

For this state, two bits of information are required to determine the correlation of the spins along x and y . That is, $|\Psi^-\rangle$ represents the double negation of the

two proposition "The two spins are equal along x " and "The two spins are equal along y ", i.e. $I_{xx} = I_{yy} = 1$, $I_{xy} = I_{yx} = 0$. In particular, the corresponding statements about the two spins along z are completely determined by the two bits of information about x and y . This means that the two bits available to the two-particle system are already exhausted in defining joint properties expressed by their spin correlations. Thus, there is no further possibility to also encode information about the individual particles with the information capacity available to the system. This characterization of quantum entanglement fits the words Erwin Schrödinger used to coin the term quite well:

"When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described [...] by endowing each of them with a representative of its own. [...] Another way of expressing the peculiar situation is: the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts, even though they may be entirely separate and therefore virtually capable of [...] possessing, each of them, a representative of its own. The lack of knowledge is by no means due to the interaction being insufficiently known — at least not in the way that it could possibly be known more completely — it is due to the interaction itself."⁹

Schrödinger wrote these words as a clarification on the so-called EPR-paradox (a topic too large to discuss it here in greater detail), which is mainly about whether the collapse of the wave function upon measurement entails a kind of spooky "action-at-a-distance" in the case of states like $|\Psi^-\rangle$. The information theoretic approach provides a new perspective on that matter, too. As was implied before, Zeilinger and Brukner emphasize that the wave function usually used to describe the state of a system is really a mathematical description of our knowledge about that system. As far as measurements are concerned, this means that the abrupt collapse of the wave function merely represents our sudden change of knowledge about a system. When keeping in mind that the wavefunction is not a physical object itself, it becomes obvious that there is nothing strange about it changing so suddenly. In particular, this process does in no way need to involve any sort of spooky interaction. Zeilinger and Brukner write,

"When a measurement is performed, our knowledge of the system changes, and therefore its representation, the quantum state, also changes. In agreement with the new knowledge, it instantaneously changes all

⁹Schrödinger, 1935; p. 555.

its components, even those which describe our knowledge in the regions of space quite distant from the site of measurement.”¹⁰

Again, this goes back to Schrödinger himself, who wrote,

”The abrupt change by measurement [...] is the most interesting part of the entire theory. It is exactly that point which requires breaking with naive realism. For this reason, the ψ -function cannot take the place of the model or of something real. Not because we can’t expect a real object or a model to change abruptly and unexpectedly, but because from a realist point of view, observation is a natural process like any other which can’t cause a disruption of the course of nature.”¹¹

While this view on the collapse of the wave function is certainly very natural, it should be noted that it omits contemplating just how the quantum state probabilistically controls the occurrence of actual events.¹² We will return to the topic of measurements in the information theoretic approach in Section 5.

4 Brukner and Dakič: Quantum Theory from three Axioms

In this section the work of Brukner and Dakič [7] will be examined, which represents an extension of Hardy’s work [6] along the lines of Zeilinger’s fundamental principle. Brukner and Dakič manage to reconstruct QT from three axioms, partly inspired by [6]:

Axiom 1 Information capacity. An elementary system has the information capacity of at most one bit. All systems of the same information capacity are equivalent.

Axiom 2 Locality. The state of a composite system is completely determined by local measurements on its subsystems and their correlations.

Axiom 3 Reversibility. Between any two pure states there existst a reversible transformation.

¹⁰Brukner and Zeilinger, 2008, 19.

¹¹Schrödinger 1935 b, 323.

¹²See Shimony, Abner, ”Bell’s Theorem”, The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/win2012/entries/bell-theorem/>

Axiom 1 could alternatively be formulated as, "Any state of a two dimensional system can be prepared by mixing at most two basis states, i.e. it can be represented as a mixture of two classical bits." The second part of Axiom 1, which we already know from Hardy's work as the subspace axiom, states that there should be no difference between systems of the same information carrying capacity. According to Axiom 1, two basis states correspond to two binary propositions like "The measurement outcome of A is $+1$." and "The measurement outcome of A is -1 ." or, alternatively, two propositions about the joint system, such as, "The measurement outcomes of A on systems one and two are correlated" or "The measurement outcomes of A on systems one and two are anticorrelated". These two choices of for the pair of propositions expressed by two basis states correspond to two choices of basis states which span the full state space of a so-called generalized bit (compare figure ??).

As before, Axiom 3 can be strengthened such that it requires,

Axiom 3' *Reversibility plus Continuity*. Between any two pure states there exists a continuous reversible transformation.

In this form, this axiom suffices to separate QT from CPT, as we've seen in section 2.1. Brukner and Dakič manage to reconstruct QT from these axioms. In particular, they explore the possibility of probability theories with higher orders of $K(N)$ than the ones examined by Hardy. Recall that, by Hardy's Simplicity axiom, he had no need to consider theories other than $K(N) = N^2$. Yet it is an interesting question to investigate whether such higher order theories can be in agreement with the other axioms.

Brukner and Dakič use Hardy's representation of a state by a probability vector $\mathbf{p} = (p_1, \dots, p_N)^T$ to define the state's Bloch vector representation

$$x_i = 2p_i - 1. \tag{52}$$

With this representation, measured probabilities are given by

$$p_{meas} = \frac{1}{2}(1 + \mathbf{r} \cdot \mathbf{x}), \tag{53}$$

which is in accord with equation (1). Again, \mathbf{r} is a vector representation of the measurement.

4.1 Composite Systems and Locality

Axiom 2 states that a composite system is completely determined by local measurements on its subsystems and the correlations of the measurement results. We consider a composite system consisting of two generalized bits on each of which N

fiducial measurements are performed to determine the state. We thus obtain $2N$ parameters plus N^2 parameters that define the correlations between measurement results, making $2N + N^2 = (N + 1)^2 - 1$ parameters in all. These define the local Bloch vectors of the two subsystems and the correlation tensor T :

$$\begin{aligned}x_i &= p^i(A = 1) - p^i(A = -1), \\y_i &= p^i(B = 1) - p^i(B = -1), \\T_{ij} &:= p^{ij}(AB = 1) - p^{ij}(AB = -1).\end{aligned}$$

Measurement A is performed on the subsystem described by \mathbf{x} and B on the one described by \mathbf{y} . Hence, $p^i(A = 1)$ is the probability of obtaining the result $A = 1$ upon performing the i th measurement on the first subsystem, and $p^{ij}(AB = 1)$ denotes the joint probability of obtaining $A = B = 1$ or $A = B = -1$ from the i th and j th measurement on the two systems, respectively.

Note that Axiom 2 implies no-signalling since, according to it, local measurements suffice to determine a state. That is, the measurement setting j for the second subsystem does not influence x_i of the first one – the experimenter at one end of the setup cannot influence the outcome of the measurement at the other end with her measurement setting j .

Now, we represent the entire state of such a system as a triple consisting of the Bloch vectors and the correlation tensor,

$$\psi = (x, y, T).$$

For a product state, the tensor is of the form $T = \mathbf{x}\mathbf{y}^T$. A state for which the correlation tensor is not of this form is called entangled. With this representation, the measured probability can be written as

$$p_{meas} = \frac{1}{4}(1 + (r, \psi)), \quad (54)$$

where $r = (\mathbf{r}_1, \mathbf{r}_2, K)$ is a composite measurement triple made of two local measurement Bloch vectors and their correlations, and (\dots, \dots) is a scalar product defined as

$$(r, \psi) := \mathbf{r}_1 \cdot \mathbf{x} + \mathbf{r}_2 \cdot \mathbf{y} + \text{tr}(K^T T). \quad (55)$$

The Bloch vectors satisfy $\|\mathbf{x}\|, \|\mathbf{y}\| \leq 1$, which implies $\|T\| \geq 1$ for all pure states, where $\|T\|^2 = \text{tr}(T^T T)$. It can be shown that the lower bound $\|T\| = 1$ is saturated if and only if the state is a product state, i.e. $T = \mathbf{x}\mathbf{y}^T$ (see appendix in [7] for proof).

When a measurement corresponding to a vector $r = \psi$ is performed on the state ψ which identifies the measurement, the probability has to satisfy $p_{meas}(\psi, \psi) = 1$, therefore

$$1 = \frac{1}{4}(1 + \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + \|T\|^2) \Leftrightarrow \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + \|T\|^2 = 3, \quad (56)$$

which already hints at the solution $K = 3$ we're going to obtain.

4.2 Ruling Out Higher-Order Theories

In order to finally arrive at $K = N^2 - 1$ from our axioms, first we are going to rule out the case of even K .¹³ Consider the transformation of total inversion $E\mathbf{x} = -\mathbf{x}$ and apply it to the two subsystems separately. Generally, we can describe the action of two local transformation $R_{1,2}$ as

$$\psi \rightarrow (R_1, R_2)\psi = (R_1\mathbf{x}, R_2\mathbf{y}, R_1TR_2^T).$$

Subsequently, totally inverting the first subsystem comes down to

$$\psi \rightarrow \psi' = (E, \mathbb{1})(\mathbf{x}, \mathbf{y}, T) = (-\mathbf{x}, \mathbf{y}, -T). \quad (57)$$

Now, using equation (56), the probability

$$\begin{aligned} p_{meas}(\psi, \psi') &= \frac{1}{4}(1 - \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|T\|^2) \\ &= \frac{1}{2}(\|\mathbf{y}\|^2 - 1) \end{aligned}$$

can't be negative, therefore $\|\mathbf{y}\| = 1$. By applying E to the second subsystem, by the same argument we obtain $\|\mathbf{x}\| = 1$ and subsequently $\|T\| = 1$. Accordingly, ψ has to be a product state. Moreover, we define a basis set of product states

$$\begin{aligned} \psi_1 &= (\mathbf{e}_1, \mathbf{e}_1, T_0 = \mathbf{e}_1\mathbf{e}_1^T), \\ \psi_2 &= (-\mathbf{e}_1, -\mathbf{e}_1, T_0), \\ \psi_3 &= (-\mathbf{e}_1, -\mathbf{e}_1, -T_0), \\ \psi_4 &= (\mathbf{e}_1, -\mathbf{e}_1, -T_0). \end{aligned}$$

Let the subspace spanned by ψ_1 and ψ_2 be S_{12} and the subspace (orthogonal to S_{12}) spanned by $\psi_{3,4}$ be S_{34} . It can now be shown that the only product states belonging to S_{12} are $\psi_{1,2}$ (see appendix in [7]). This means that there also have to be entangled states in S_{12} . This contradicts what we have found above, i.e. that the general state ψ has to be a product state. We thus rule out the case of even K .

Now we will rule out the case of $K > 3$. This can be done in a manner similar to the one used for even K , by constructing two transformations acting on $\psi \in S_{12}$. Let R_i be an operator which flips the first and the i th component of the local

¹³Remember that in Hardy's case, the result of $K = N^2$ corresponded to non-normalized probability vectors. When normalized, one degree of freedom has to be subtracted.

Bloch vectors and R_{jkl} be an operator which flips the first and j th, k th and l th coordinate where $j \neq k \neq l \neq 1$. With the requirement that $\psi_i = (R_i, \mathbb{1})\psi$ and $\psi_{jkl} = (R_{jkl}, \mathbb{1})\psi$ belong to S_{34} , it can be shown that the correlation tensor has to satisfy $\|T\| = 1$, which would again imply that ψ has to be a product state and thus produced the desired contradiction. In this way, it is possible to get rid of Hardy's Simplicity axiom and restrict our theory to $K = 3$ or $K = N^2 - 1$ for $N = 2$, respectively, solely by the use of the other axioms. The further reconstruction can then be done as in [6].

Brukner and Dakić conclude their paper by discussing the interpretational role of their first axiom about the finite information carrying capacity of quantum systems, which corresponds to Zeilinger's foundational principle. It is astonishing to see that this principle, although vital to the above reconstruction, is in thorough disagreement to the position of realism. Realism holds that outcomes of measurements correspond to actualities objectively existing prior to, and being revealed by, a measurement. That is, realism asserts that, say, the location of an electron is an objective fact that could, in principle, be determined even when in an interferometer experiment the path the electron took doesn't seem to be a well-defined notion. Any realistic theory therefore demands an infinite information carrying capacity. Yet, if the information which a quantum system can hold is finite, then only a limited amount of measurements yield definite values. What a fundamental limit of the information in a system means ontologically is obviously a question not duly answered so far.

5 A Bayesian Approach to the Measurement Problem

5.1 Quantum Collapse as Bayes' Rule

In [1] and [14], Christopher Fuchs aims to strengthen the position that wavefunctions really are nothing more than a representation of our knowledge. He shows that it is possible to integrate QT into a theory that is all about knowledge in the light of different data: the Bayesian probability theory. Fuchs does this by showing that the effects of the measurement process is perfectly equivalent to the application of Bayes' rule. In doing so, Fuchs provides a solution to the so-called measurement problem, which investigates how exactly the wavefunction collapse occurs upon measurement. This collapse appears incomprehensible, because the reversible unitary evolution of the wave function described by Schrödinger's equation is, at the moment of measurement, suddenly replaced by an utterly different behaviour – the irreversible collapse. By showing that this seemingly problematic demeanor is just a version of Bayes' rule, Fuchs means to integrate quantum

measurement into the well-known process of information acquisition.

In QT, measurements can be mathematically treated with so-called POVMs (Positive Operator-Valued Measures)¹⁴. The well-known Born rule states that when a measurement described by a measurement operator M_m is performed, the probability of outcome m is given by $p(m) = \langle \psi | M_m^\dagger M_m | \psi \rangle$. Now, it is convenient to define the POVM element E_m , a positive operator, associated with this measurement as

$$E_m := M_m^\dagger M_m, \quad (58)$$

where $\sum_m E_m = \mathbb{1}$, so that we simply have $p(m) = \langle \psi | E_m | \psi \rangle$ or, alternatively, $p(m) = \text{tr}(\hat{\rho} E_m)$. In quantum information theory, it is customary to use POVMs instead of (mathematically equivalent) projective measurements, because they are simpler to handle and lack the restrictions of projection operators. The complete set $\{E_m\}$ is the POVM corresponding to this measurement, which acts upon measurement on a system of state $\hat{\rho}$ as

$$\hat{\rho} \longrightarrow \hat{\rho}_m = \frac{1}{\text{tr}(\hat{\rho} E_m)} \sum_i M_{m,i} \hat{\rho} M_{m,i}^\dagger, \quad (59)$$

where $\sum_i M_{m,i}^\dagger M_{m,i} = E_m$. Conversely, we will consider any set $\{E_m\}$ of positive operators satisfying $\sum_m E_m = \mathbb{1}$ as the mathematical representation of a measurement. This characterization of measurements contains the collapse postulate as a special case when $\{E_m\}$ and $\{M_{m,i} = E_m\}$ are taken to be orthogonal projectors. It is Fuchs' aim to show that this notion of measurement is very much alike to how measurements are depicted in Bayesian probability theory.

In Bayes' theory, an initial state of an agent's knowledge is captured by a probability distribution $p(h)$ for a certain hypothesis h . New data d is taken into account according to Bayes' rule:

$$p(h) = \sum_d p(h|d) = \sum_d p(d)p(h|d) \quad (60)$$

$$\xrightarrow{d} p(h|d) = \frac{p(h, d)}{p(d)}. \quad (61)$$

$p(h)$ is called prior and represents a belief in h , $p(h|d)$ is called posterior and represents the belief in h in the light of d . Generally, $p(a|b)$ is the conditional probability of a given that b has occurred, being defined as the quotient of the joint probability of a and b , and the probability of b , $p(a|b) = p(a, b)/p(b)$.

Unlike the abrupt manner in which measurements change a state $\hat{\rho}$ in QT, here the new state of knowledge was already initially averaged in with other potential

¹⁴Compare Chuang and Nielsen, Cambridge UP 2010, Section 2.2.6.

states of knowledge. Fuchs wants to show that the collapse postulate is quite alike to Bayes' rule and appears unintuitive only because it's an "artifact of a problematic representation". Note that for the Bayesian, probabilities do not denote the relative frequencies measured in experiments but rather a subjective belief in a hypothesis. Fuchs writes,

"According to the quantum-Bayesian approach to quantum foundations [...], the probabilities $p(i)$ represent an agent's Bayesian degrees of belief, or personalist probabilities [...]. They are numbers expressing the agent's uncertainty about which measurement outcome will occur and require an operational meaning through decision theory [...]. Quantum-Bayesian state assignments are personalist in the sense that they are functions of the agent alone, not functions of the world external to the agent [...]"¹⁵

This view on the quantum state goes even further than the ones already discussed, which too emphasize that the state only represent our knowledge. Here, the de-ontologization of the ψ function is even more advanced – in no way is it regarded as something real like a guiding field, but only as a representation of a personal belief.

We will now demonstrate the Bayesian nature of the quantum collapse. First, for simplification we consider the case where the index i in (59) takes only one value, yielding

$$\hat{\rho}_m = \frac{1}{\text{tr}(\hat{\rho}E_m)} M_m \hat{\rho} M_m^\dagger \quad \text{and} \quad E_m = M_m^\dagger M_m. \quad (62)$$

Unfortunately, quantum states do not in general take a form corresponding to equation (60), i.e. in general

$$\hat{\rho} \neq \sum_m P(m) \hat{\rho}_m.$$

However, we can write

$$\begin{aligned} \hat{\rho} &= \hat{\rho}^{1/2} \mathbb{1} \hat{\rho}^{1/2} \\ &= \sum_m \hat{\rho}^{1/2} E_m \hat{\rho}^{1/2} \\ &:= \sum_m P(m) \hat{\sigma}_m, \end{aligned} \quad (63)$$

¹⁵Christopher A. Fuchs and Rüdiger Schack, A Quantum-Bayesian Route to Quantum-State Space, arXiv:0912.4252v1.

where

$$\hat{\sigma}_m = \frac{1}{P(m)} \hat{\rho}^{1/2} E_m \hat{\rho}^{1/2} = \frac{1}{P(m)} \hat{\rho}^{1/2} M_m^\dagger M_m \hat{\rho}^{1/2}. \quad (64)$$

Now, the key observation is that $\hat{\sigma}_m$ and $\hat{\rho}_m$ from equation (59) have the same eigenvalues. This is because $X^\dagger X$ and XX^\dagger have the same eigenvalue for any operator X . For our case, we simply choose $X = M_m \hat{\rho}^{1/2}$. Therefore, we can outline the measurement process as the mental adjustment of an observer's knowledge, who selects one term from the distribution representing her initial state of knowledge:

$$\hat{\rho} = \sum_m P(m) \hat{\sigma}_m \longrightarrow \hat{\sigma}_m. \quad (65)$$

This already very much resembles Bayes' rule (60) \longrightarrow (61). Finally, we make a "mental adjustment" of this state, which takes into account the impact the measurement had on the system, and obtain $\hat{\rho}_m$ again via a unitary operation V_m :

$$\hat{\sigma}_m \longrightarrow \hat{\rho}_m = V_m \hat{\sigma}_m V_m^\dagger. \quad (66)$$

Fuchs justifies this adjustment through the breakdown that is sometimes made of the usual measurement process. This breakdown utilizes the polar decomposition theorem, according to which any square complex matrix A can be written as $A = UP$, where U is a unitary matrix and $P = \sqrt{A^\dagger A}$ is a positive-semidefinite Hermitian matrix. By this theorem, we can rewrite (62) as

$$\hat{\rho}_m = \frac{1}{P(m)} U_m E_m^{1/2} \hat{\rho} E_m^{1/2} U_m^\dagger, \quad (67)$$

where U_m is unitary. Thus, we can imagine the measurement process to be partitioned as

$$\hat{\rho} \longrightarrow \hat{\rho}' = \frac{1}{P(m)} E_m^{1/2} \hat{\rho} E_m^{1/2} \longrightarrow \hat{\rho}_m = U_m \hat{\rho}' U_m^\dagger. \quad (68)$$

Fuchs' breakdown into $\hat{\sigma}_m \longrightarrow \hat{\rho}_m$ is equivalent to this usual picture of measurement.

In this manner, we obtain exactly Bayes' rule: our agent starts with an initial judgement about some hypothesis h (expressed in terms of a measurement $\{E_h\}$)

$$P(h) = \text{tr}(\hat{\rho} E_h) \quad (69)$$

and measures some other observables $\{E_d\}$ where

$$\hat{\rho} = \sum_d P(d) \hat{\sigma}_d \implies P(h) = \sum_d P(d) P(h|d), \quad (70)$$

with

$$P(h|d) = \text{tr}(\hat{\sigma}_d E_h). \quad (71)$$

By showing that the quantum collapse is really just a variant of Bayes' rule, and thus part of pure probability theory, we deprive the so-called measurement problem of all foundation. In this picture, there really is no violent rupture in the natural, unitary evolution of the wave function. All that happens is the acquisition of new information. As Fuchs and Peres put it, "Collapse is something that happens in our description of the system, not to the system itself."¹⁶

5.2 Gleason's Theorem, Born's Rule and Entanglement

While in the previous subsection, we made sense of the quantum collapse in Bayesian terms, we now want to justify Born's rule for measurement itself. To this end, Fuchs refers to a theorem by Andrew Gleason [15]. Gleason's theorem states, that the Born rule is the only rule that satisfies a noncontextuality requirement for measurement outcomes, and derives the state-space structure for states in QT. Said noncontextuality condition requires that when one measures two distinct observables $\{E_a\}$ and $\{\tilde{E}_b\}$, which share one POVM element E_c , the outcome probability does not depend on whether E_c is associated with the POVM $\{E_a\}$ or $\{\tilde{E}_b\}$. Formally, this means there is a function

$$f : \mathcal{E}_d \longrightarrow [0, 1], \quad (72)$$

where

$$\mathcal{E}_d = \{E : 0 \leq \langle \psi | E | \psi \rangle \leq 1, \forall |\psi\rangle \in \mathcal{H}_d\}, \quad (73)$$

such that whenever $\{E_a\}$ forms a POVM,

$$\sum_m f(E_m) = 1. \quad (74)$$

Now, Gleason's theorem in its POVM formulation¹⁷ states that there must exist a density operator $\hat{\rho}$ such that

$$f(E) = \text{tr}(\hat{\rho}E). \quad (75)$$

That is, the theorem reproduces exactly Born's rule (9). The proof for this theorem works as follows. We start off by working only with the field of (complex) rational numbers. Consider a three-element POVM $\{E_1, E_2, E_3\}$. By equation (74), $f(E_1) + f(E_2) + f(E_3) = 1$, but we can also combine the first two elements into a single POVM element to obtain $f(E_1 + E_2) + f(E_3) = 1$. We thus obtain additivity for f ,

$$f(E_1 + E_2) = f(E_1) + f(E_2). \quad (76)$$

¹⁶Christopher A. Fuchs and Asher Peres. Quantum Theory Needs No 'Interpretation'.

¹⁷See [1], [17] and [18].

Subsequently, for any integers n and m we have $f(E) = mf(\frac{1}{m}E) = f(E) = nf(\frac{1}{n}E)$. With $\frac{n}{m} \leq 1$ and $E = nG$, we obtain

$$f\left(\frac{n}{m}G\right) = \frac{n}{m}f(G). \quad (77)$$

In this way, we have shown a limited linearity of f on \mathcal{E}_d . We now want to extend f to a function that is fully linear on the d^2 dimensional vector space \mathcal{O}_d of Hermitian operators on \mathcal{H}_d . For any positive semi-definite operator E we can always find a positive rational number g such that $E = gG$, where $G \in \mathcal{E}_d$. We thus define our extension as $f(E) \equiv gf(G)$. It can easily be shown that this extension exhibits full linearity. Now, we can write any operator $E \in \mathcal{E}_d$ as

$$E = \sum_{i=1}^{d^2} \alpha_i E_i. \quad (78)$$

By linearity of f ,

$$f(E) = \sum_{i=1}^{d^2} \alpha_i f(E_i). \quad (79)$$

We can now define $\hat{\rho}$ to solve

$$\text{tr}(\hat{\rho}E_i) = f(E_i) \quad (80)$$

and obtain

$$f(E) = \sum_{i=1} \alpha_i \text{tr}(\hat{\rho}E_i) = \text{tr}\left(\hat{\rho} \sum_i \alpha_i E_i\right) = \text{tr}(\hat{\rho}E). \quad (81)$$

It is not difficult to extend this theorem from rational numbers to the full continuum of (complex) rational numbers.

As astonishing as it is to see that the Born rule can be presented as simply a consequence of the noncontextuality of measurement, it is even more astounding that the theorem has to be adapted only slightly to also produce the tensor product structure of QT – and thus to tell us something about entanglement.

To this effect, consider two quantum systems A and B on which POVM measurements are performed. The measurement $\{E_j^i\}$ on A may be conditioned on the outcome i of the measurement $\{F_i\}$ on B , and vice versa. We denote the ordered pair of operators (E_i, F_j^i) or (E_i^j, F_j) by S_{ij} and call $\{S_{ij}\}$ a locally-measurable POVM tree. In analogy to Gleason's theorem, we demand that the joint probability for outcomes of such measurements should not depend upon which tree S_{ij} the corresponding POVMs are embedded in. This assumption is very similar to the one above, but applied to a composite system. Formally, it states that for a function

$$f : \mathcal{E}_{d_A} \times \mathcal{E}_{d_B} \longrightarrow [0, 1] \quad (82)$$

with the requirement of

$$\sum_{ij} f(S_{ij}) = 1 \quad (83)$$

for S_{ij} consisting of valid POVM elements, there is a density operator $\hat{\rho}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ such that

$$f(E, F) = \text{tr}(\hat{\rho}(E \otimes F)). \quad (84)$$

If we define, for a fixed $E \in \mathcal{E}_{d_A}$,

$$g_E(F) := f(E, F) \quad (85)$$

we can use the same methods as in the original POVM-Gleason theorem to extend $g_E(F)$ to a linear functional on the Hermitean operators on \mathcal{H}_A , and vice versa for

$$h_F(E) := f(E, F) \quad (86)$$

for fixed $F \in \mathcal{E}_{d_B}$. Combined, this gives full bilinearity for the function $f(E, F)$. Now, we let $\{E_i\}_1^{d_A^2}$ and $\{F_j\}_1^{d_B^2}$ be complete bases for the Hermitean operators on \mathcal{H}_A and \mathcal{H}_B . $E = \sum \alpha_i E_i$ and $F = \sum_j \beta_j F_j$ yields

$$f(E, F) = \sum_{ij} \alpha_i \beta_j f(E_i F_j). \quad (87)$$

As above, we let $\hat{\rho}$ be a linear operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ satisfying

$$\text{tr}(\hat{\rho}(E_i \otimes E_j)) = f(E_i, F_j) \quad (88)$$

to obtain

$$\begin{aligned} f(E, F) &= \sum_{ij} \alpha_i \beta_j \text{tr}(\hat{\rho}(E_i \otimes E_j)) \\ &= \text{tr}(\hat{\rho}(E \otimes F)). \end{aligned} \quad (89)$$

The point of this theorem is that it constructs the probability rule for composite systems solely from Gleason-like noncontextuality and the concept of local measurement. With these assumptions alone, we also arrive at the tensor product structure of composite systems which is the mathematical origin of quantum entanglement.

6 Interpreting the Information Theoretic Approach to Quantum Theory

As mentioned at the beginning of this work, the task of reconstructing QT is quite different from that of finding an interpretation of the theory. In the previous sections, we have seen multiple methods of finding information theoretic foundations

for various features of QT that are typically the core of what most interpretations try to make sense of. Examples are quantum randomness, which was explained as a consequence of the "fundamental principle" (Section 3). Entanglement and the tensor product structure of composite systems, on which the fundamental principle also sheds some light (see Section 3.3) can be explained with Hardy's axioms (Section 2.3) and Fuchs' variant of Gleason's theorem (Section 5.2). Moreover, the wave collapse and Born's rule can naturally be clarified from a Bayesian view on QT (section 5). Finally, in Sections 2 and 4 we saw that the whole of quantum mechanics can be shown to arise from just three modest axioms which can certainly be called rather intuitive, or at least aren't as exotic as the quantum formalism itself.

All these endeavors deal quite pragmatically with phenomena that are traditionally considered to be exceptional or even "spooky", but still they manage to make sense of them without employing any metaphysical pretensions. Consequently, they raise the question of whether we really want to find underlying mechanisms or processes that give rise to, say, entanglement, or whether we can simply subsume it under a number of principles or axioms that are not further justified. C. Fuchs and A. Peres adress this question in their paper "Quantum Theory Needs No Interpretation" [19]. They claim that what the results discussed so far should tell us is that all the metaphysic superstructure made up by our interpretations are redundant, because QT really is solely about mathematically representing and updating our state of knowledge about some system. That is, the theory really is not as much about the system but rather about what we *know* about it. At its core, QT is just a tool to make predictions for future measurements. It is not a theory about the processes internal to the systems measured which give rise to the measurement outcome – or at least that's what the information theoretic construction strongly suggests.

While, for example, in the Zeilinger papers discussed in this work, the central notion of information is still considered something internal to the quantum system (information intrinsically carried by the system), especially in Fuchs' Bayesian approach we observe an extensive de-ontologization of the quantum system and its state. Here, the information the theory governs solely denotes an agent's personal belief and not something physical with an independent reality. This viewpoint is central to all the work presented here: quantum mechanics is depicted as something unexotic and free of underlying metaphysical grounding that sounds quite like science fiction. Instead, the theory is turned into something rather instrumental. As Fuchs and Peres put it,

"The thread common to all the nonstandard 'interpretations' is the desire to create a new theory with features that correspond to some reality independent of our potential experiments. But, trying to fulfill

a classical worldview by encumbering quantum mechanics with hidden variables, multiple worlds, consistency rules, or spontaneous collapse, without any improvement in its predictive power, only gives the illusion of a better understanding. Contrary to those desires, quantum theory does not describe physical reality. What it does is provide an algorithm for computing probabilities for the macroscopic events ('detector clicks') that are the consequences of our experimental interventions. This strict definition of the scope of quantum theory is the only interpretation ever needed, whether by experimenters or theorists."¹⁸

In any case, although interpretations like Bohm's nonlocal realistic mechanics [2] are empirically equivalent to QT and can thus never be ruled out, the mere fact that all phenomena of the quantum world can be reproduced from a small number of "reasonable axioms" is enough to call for Occam's razor. In this case, the famous metaphoric blade tells us to abandon all metaphysic notions associated with quantum mechanics and just consider it a kind of probability theory – a theory about information.

¹⁸C. A. Fuchs, A. Peres: Quantum Theory needs no "Interpretation". In: Physics Today. Nr. 53(3), 2000, Seite 70.

References

- [1] C. A. Fuchs, "Quantum Foundations in the Light of Quantum Information," in *Decoherence and its Implications in Quantum Computation and Information Transfer: Proceedings of the NATO Advanced Research Workshop, Mykonos Greece, June 25–30, 2000*, edited by A. Gonis and P. E. A. Turchi (IOS Press, Amsterdam, 2001), pp. 38–82. [quant-ph/0106166](#)
- [2] Bohm, David (1952). "A Suggested Interpretation of the Quantum Theory in Terms of "Hidden Variables" I". *Physical Review* 85: 166–179.
- [3] Hugh Everett, III: "Relative State" Formulation of Quantum Mechanics. In: *Rev. Mod. Phys.* 29, 1957, S. 454–462.
- [4] Alexei Grinbaum, On the notion of reconstruction of quantum theory, [arXiv:quant-ph/0509104v2](#), 2005.
- [5] D. Hilbert, J. von Neumann, and L. Nordheim. *Über die Grundlagen der Quantenmechanik*. *Math. Ann.*, 98:1-30, 1927. (Reprinted in J. von Neumann *Collected Works* Pergamon Press, Oxford, 1961, Vol. I, pp. 104-133).
- [6] L. Hardy. Quantum theory from five reasonable axioms. [arXiv:quant-ph/0101012](#) (2001). <http://arxiv.org/abs/quant-ph/0101012>.
- [7] B. Dakic and Č. Brukner, *Quantum Theory and Beyond: Is Entanglement Special?*, Contribution to "Deep beauty", Editor Hans Halvorson (Cambridge Press, 2011)
- [8] K. Kraus, *States, effects, and operations: Fundamental notions of quantum theory* (Springer-Verlag, Berlin, 1983); B. Schumacher, [quant-ph/9604023](#) (appendix A); J. Preskill *Lecture notes for physics 229: quantum information and computation*, (see chapter 3).
- [9] Ruediger Schack. Quantum theory from four of Hardy's axioms. [arXiv:quant-ph/0210017](#)
- [10] Brukner, Č.; Zeilinger, A. Information and fundamental elements of the structure of quantum theory. In *Time, Quantum, Information*; Castell, L., Ischebeck, O., Eds.; Springer: Berlin, Heidelberg, Germany, 2003; pp. 323–354.
- [11] Schrödinger, E. (1935) "Discussion of Probability Relations Between Separated Systems,," *Proceedings of the Cambridge Philosophical Society*, 31: 555–563; 32 (1936): 446–451.

- [12] Nielsen, Michael A., and Isaac L. Chuang. Quantum Computation and Quantum Information. Cambridge: Cambridge UP, 2010.
- [13] Christopher A. Fuchs and Rüdiger Schack, A Quantum-Bayesian Route to Quantum-State Space, [arXiv:0912.4252v1](#).
- [14] Christopher A. Fuchs, Quantum Mechanics as Quantum Information (and only a little more), [arXiv:quant-ph/0205039](#).
- [15] A. M. Gleason, "Measures on the closed subspaces of a Hilbert space," J. Math. Mech. 6, 885 (1957).
- [16] Carlton M. Caves, Christopher A. Fuchs, Rüdiger Schack (2001). "Quantum probabilities as Bayesian probabilities"
- [17] P. Busch, "Quantum States as Effect Valuations: Giving Operational Content to von Neumann's No-Hidden-Variables Theorem," [quant-ph/9909073](#).
- [18] C. M. Caves, C. A. Fuchs, K. Manne, and J. Renes, "Gleason's Theorem with POVMs," unpublished, 2000.
- [19] C. A. Fuchs, A. Peres: Quantum Theory needs no "Interpretation". In: Physics Today. Nr. 53(3), 2000, Seite 70.