Concepts surrounding incompatibility in quantum physics

Neil Stevens

Doctor of Philosophy

University of York

Mathematics

September 2014

Abstract

This thesis concerns how attempting to measure certain fundamental properties of a physical theory can lead us to new insights. The idea of general probabilistic models is introduced and developed, in order to provide a way to be able to consider and compare different theories. A measure of incompatibility of two observables is defined, through the amount of smearing needed to make them jointly measureable. This measure is then used to characterise the degree of incompatibility that exists in a given theory. Quantum theory is then shown to be as incompatible as any other theory, but only in a very course grained sense. A related way of measuring the strength of incompatibility of a pair of observables is then shown to put a bound on a measure of the strength of non-local correlations. The notion of steering, or remote ensemble preparation, is then shown to be a sufficient condition for the saturation of that bound. Examples are considered that demonstrate that the given sufficient ntion of steering is not necessary, and it is proposed how the measures can be modified in cases where the link does not hold. The idea of formulating measures of error and disturbance asscociated with a measurement device is discussed. The notion of a direct test of error or disturbance is used to analyse current proposed measures, and show their shortcomings.

Contents

Acknowledgements

I am indebted to numerous people who have provided me with help and support throughout my PhD. Amongst the many I would like to thank are my family for the continuing support they have offered, Leon Loveridge for making the whole process feel less daunting and providing some useful conversations, James Waldron for the sheer range of mathematical discussions we shared, Chris Perry for the support and insights offered over the years, Rosie Napier for being able to make me smile when I most needed to, and most of all to Paul Busch who has been more than I would have expected of a supervisor, and from whom I have learned so much.

Declarations

I declare that the work presented in this thesis, except where otherwise stated, is based on my own research and has not been submitted previously for a degree in this or any other university. Parts of the work reported in this thesis have been published in:

Paul Busch, Teiko Heinosaari, Jussi Schultz, Neil Stevens, "Comparing the degrees of incompatibility inherent in probabilistic physical theories ", EPL, 103 (2013) 10002

Neil Stevens, Paul Busch, "Steering, incompatibility, and Bell inequality violations in a class of probabilistic theories ", Physical Review A 89, 022123 (2014)

Paul Busch, Neil Stevens, "Direct tests of measurement uncertainty relations: what it takes ", arXiv:1407.7752

Neil Stevens

Chapter 1

Introduction

The purpose of the sciences has always been to advance the knowlege of mankind, by seeking to offer a description of the processes that are obwerved to happen in the natural world. Such descriptions are usually arrived at by the process of experimentation and data collection, follwed by analysis in order to form or update hypotheses, which are then subjected to further experimentation. This process leads to refined hypotheses that fit data to a high degree of accuracy. This allows future behaviours of systems to be predicted in a precise way, and is the main reason the scientific method has been so successful throughout history. However such descriptions do not answer one of the most basic and fundamental questions: why?

Many ponderings on the question of why the world behaves the way it does, can end up leading down the road of philosophy, and on to questions about the nature of existance. Such questions tend to stray a long way from the empirical, and so science can rarely offer any guidance into such matters. However many scientists have pondered the question of why, in a more tangible and concrete way. One popular way to do so has been to consider whether the full set of rules usually used to describe systems, can be reduced to a smaller set of more reasonable ones. This usually takes the form of starting with a small set of postulates which it seems reasonable that reality should conform to, and then showing that the theory used is the only, or one of very few theories that conform to such postulates. Another related approach is not to consider a theory as a whole, but rather just certain properties of that theory; often ones that are

considered useful in some way. Then it is possible to ask which, if any, intuitive postulates lead to which properties. Such considerations naturally lead to the need to develop ways of talking about other alternative, potential theories, that do not necessarily describe reality but potentiall could. Otherwise, statements of the form 'these postualtes entail our current model of reality' are merely tautologies when the only model of reality that exists is the current one.

Once the possibility of considering other, alternative models of reality is opened up, a whole new area of consideration is opened up with it. If there are many models of how things could work, it is possible, if not natural, to start to compare and contrast such models. The purpose of this thesis is to explore some of these comparisons that can be drawn. We will look at direct comparisons of quantities that define properties of a theory, which properties a theory must have in order to exhibit certain behaviours, and the exact form of relationships between specific quantities within a theory.

The thesis is concerned with three main topics to do with certain properties of theories, all of which are related, to a greater or lesser extent, to the notion of incompatibility. Incompatibility, or the converse notion joint measurability, pertains to a property of observables being able to be realised at the same time. The first of these topics is to provide a method for determining how far from being jointly measurable a pair of observables is, and in a certain sense assigning a measure of incompatibility to that pair. This measure is then extended to a way to characterise and comapre theories themselves, by considering how 'badly' incompatible pairs of observables can get in that theory.

The second topic seeks to tie together two fundamental, stereotypically non-classical, properties of quantum mechanics, namely incompatibility and non-local correlations. The result that a certain measure of incompatibility gives a limit on the strength of any bipartite correlations, regardless of the theory they are in, is reviewed. It is then shown that under an additional assumption, this limit can always be saturated. This idea is explored by giving examples of where the results do, and do not, hold, as well as possibilities for generalisiations.

The final topic moves away from the perspective of comparing different theories, and looks at the long standing problem of uncertainty relations. Specifically the problem of formulating error-disturbance uncertainty relations pertaining to

measurements is discussed. The idea of a direct test of a relation is formalised, and the suitability and universality of currently proposed measures of error and disturbance is analysed.

Chapter 2 seeks to lay the foundations and provide an introduction to the laguage for the rest of the thesis. Chapters 3, 4 and 5 comprise the contents of three research papers submitted for publication, and are largely self contained. However the material contained in chapter 2 should serve as an introduction ot the topic for a non-expert. First to be addressed is the question of what should, and should not, be considered as a candidate of a physical model worthy of analysis. There is an introduction to the framework that will be used to propose candidate theories, namely the framework of general probabilistic models. The notions of what constitutes physical systems that we wish to model, and the idea of the state of those systems is discussed, along with the restrictions we may place in order to build up an at least potentially viable physical model. This reasoning is then extended to considerations about what knowlege we can gain about a system, leading to the ideas of effects and then observables. This leads the being able to formalise the definition of joint measurability. The concepts of how to model composition of systems and of correlations between the parties are introduced.

The paper that comprises Chapter 3 sets out to provide a way to compare the levels of incompatibility in different theories. First the concept of taking convex combinations of observables to form a further observable that can be considered a mixture of the originals is introduced, as is the notion of a trivial observable as one which does not depend on the state on which the observable measured. Then a smeared, or fuzzy, version of an observable is defined by taking a certain convex combination with a trivial observable, where the convex weightings are determined by the smearing parameter. The concept of a measure of incompatibility is then introduced by considering smeared versions of a pair of observables: if the smearings are taken to be large enough, then the resulting observables will be jointly measurable, even if they were incompatible to start with. The joint measurability region, which represents the smearing values which result in two observables becoming jointly measureable, is intorduced as a way of characterising the incompatibility in a theory. It is then shown that if all observables are considered that quantum theory is, in this sense, as incom-

patible as it can be. However if the more fine grained approach of looking at joint measurability regions for observables with a given number of outcomes is taken, this no longer holds. An example is given, for the case of a pair of two outcome observables, of a probabilistic theory that is more incompatible than quantum theory.

In Chapter 4 the paper analyses the relationship between a certain measure of incompatibility, and the idea of Bell non-locality. The measure of incompatibility of an observable used here is again based off of the idea of how much smearing is needed to make a pair of observables incomaptible. This time only two-outcome observables are considered, and the smearing parameters are taken to be the same on both observables, along with the trivial observables being fixed as being unbiased. This measure is then rewritten as the solution to a conic optimisation problem, where the cone is dependent on the theory being considered. The dual of this problem is then shown to be directly linked to the Bell-CHSH quantity, often taken to be a measure of the strength of non-local correlations. This provides a limit on the strength of such correlations, formulated here as a generalised Tsirelson bound, and also provides a possible constructive method for saturating the limit. The concept of steering is then introduced as the ability to prepare any ensemble that forms a convex decomposition of one half of a bipartite state, by measuring an observable on the other half. A theory possessing this concept of steering is shown to be a sufficient condition for the saturation of the generalised Tsirelson bound. The case of a theory whose state spaces are polygons is then considered. It is shown that the general results hold in the cose of the polygon having an even number of sides, whereas they generally do not for an odd number. The pentagon is examined in more detail, and although the main results of the paper do not hold there, a slightly different measure is proposed that may be useful in such a scenario.

The paper making up chapter 6 lives purely in the domain of quantum mechanics, and delves into the world of uncertainty relations. The notion of uncertainty relations is somethinig that most physicists have a basic understanding of. However when the concept is usually discussed it is with relation to preparation uncertainties. Recently there has been a fair amount of activity attempting to formalise the thoughts of Heisenberg regarding measurement, or error-disturbance relations. There have beeen two approaches to this problem,

one looking at state dependent measures, and one looking at state independant. The paper focuses mainly on the state dependent case, and specifically on the proposal to have garnered the most attention so far. It is suggested that in order to be considered a valid error-disturbance relation, that the quantities involved should admit a direct test, a concept that is formalised here. An experimental suggestion is given to carry out such a direct test of the measures in question, in a given setup. However it is also pointed out that such a test is not always possible to do, which is in line with previous work which has shown the interpretations of the measures as being of error and disturbance as being only confined to a limited scenario.

Chapter 2

General Probabilistic Models

The idea of looking to theories more general than quantum theory is not new, having reared it's head with names such as convex operational theories, convex state approach, and general probabilistic models amongst others. The literature is reasonably extensive, if not so well known in the wider physics community, see for example [1, 2, 3, 4, 5, 6]. The recent rise in popularity of quantum information has brought about something of a change in perspective about the non-classical features of quantum mechanics. Rather than being viewed as something that needs to be explained away, quantum effects are seen as something that can be exploited. This has lead somewhat of a renewed interest in the topic of more generalised models. For a sampling of more recent work in the area see [7, 8, 9, 10], and for an up to date review see [11]. This introductory chapter gives an original take on the topic of general probabilistic models for the purposes of the thesis, but is based on the work in the references given, as well as others.

2.1 Where to start

It is important to consider first what we have in mind, and why we would ever choose to look at general probabilistic models (GPMs). What is the motivation for such considerations? What new insights could they give? How general is general? Why probabilistic? What are we modelling?

2.1.1 What are we trying to model?

The fundamentals of what a GPM aims to achieve is to describe the state of some (possibly hypothetical) physical system, and what information it is possible to extract about the system through a process of experimentation. Of course this immediately begs the question of what counts as a physical system.

In brief, the idea behind the notion of a system is simply anything which can be subjected to a set of one or more experiments. There are many examples of occasions where systems occur, such as balls on a billiard table, a tossed coin or rolled die, an electron shot towards a Stern-Gerlach magnet, or a photon passing through a polarising beam splitter. Here however it is important to note that it is not the objects themselves that comprise the system. Rather the context in which we find the objects is important, since the system depends on which experiments are considered, and it is those which determine the actual system in question.

Instead of imagining individual particles or objects when thinking of a system that we are trying to model, it is more accurate to consider what degrees of freedom that object has. In the example of the tossed coin, one could consider attempting to write down a model that completely described every aspect of the coin. This would include, the exact shape and size, the precise distribution of mass, and of temperature of the material, and potentially many more variables. However, in a situation where a coin has indeed been tossed, it is highly impractical (if not impossible) to give such a precise description, and one is in fact very often not interested in such finery; rather one is merely interested only in the degree of freedom that determines the propensity of the coin to land on one side or another, since that is the only measurement that is likely to be available to any potential observers. Likewise, since we will only seek to observe the deflection of the electron throught the Stern-Gerlach apparatus, the system we would seek to model would only involve its spin degree of freedom.

Any type of model that is to be considered from now on must not be thought of as attempting to describe any given object in the abstract: the context in which the particle resides is also important since it determines which properties of the object we may be able to gain knowlege of, and thus it is only those properties that should be considered to be modelled.

2.1.2 Why a probabilistic model?

The next question that must be adressed is why consider probabilistic models? And what exactly does that mean?

Intuitively one might wonder where probability would ever come into an attempt to model a physical system. Indeed, if there is some set of experiments that may be performed, that define the system in question, then it would be desirable for the model to simply tell us what the outcome of performing any such measurements would be. However this idealised scenario is in practice rather idyllic. As was alluded to in the considerations above, even in a simple scenario of tossing a coin, it may be impossible, even in theory, to be able to ascertain all the information necessary to determine which side the coin will land on for certain. Indeed, whether due to a lack of knowlege, or some more fundamental limitation, any model which is at all realistic could not claim to universally describe with certainty the results of any observations made. Instead the best possible description of the system is one that only gives propensities, for example that an electron is equally likely to be deflected up or down, or that a coin has a small bias and thus will land heads up 55% of the time.

2.1.3 Exactly how general?

Within any given model, there are often scenarios where certain 'features' of that model appear to play a significant role in allowing special tasks to be performed, or behaviour demonstrated, that would differentiate it from another model. However when a model has been established, it can become difficult to determine whether the suspect features that the model possesses are truly responsible for any given behaviour, whether it is some other less obvious feature, or whether it is purely down to the the specifics of the model. In order to make sense of this scenario, it is therefore useful to consider the behavioiur of other models that share certain features, but not necessarily others. This is where the idea of generalised models comes into play.

2.2 States

As described above, the idea of a GPM is to provide a framework for theoretical considerations involving candidate models for (possibly hypothetical) systems. In order to discern what may be considered as a suitable candidate, we must enforce some stipulations; however in the interests of generality it shall be attempted to keep these as unrestrictive as possible.

In order to begin to see the effects of any stipulations, we first must begin to build up a model, and for that we need to start somewhere. Here we choose to take as our starting point the state of a system, which we define as follows:

Definition 1. The *state* of a system is any mathematical object that determines the propensities of outcomes of all measurements defining the system.

The idea here is that the state contains all the information about the system it is possible to know. It may not be possible to access all of that information through any amount of observation, but the state determines any information you can gather.

For any given system, be it a coin's faces, a billiard ball, the spin of an electron etc., we shall denote the class of all possible states Ω, which we shall take to be a set.

2.2.1 Preparation and mixing

To put our first restriction on what will be considered a valid candidate for a physical model of a system, we need to introduce the concept of a *preparation procedure*. Simply put, for a given system (with states in Ω), a preparation procedure for $\omega \in \Omega$ is any process or influence that acts upon the constituent parts of the system, that leave the system in the state ω.

Different processes can be preparation procedures for the same state, indeed there are many, many ways to flip a coin and only two states that the coin will end up in. Therefore the multitude of flips that result in heads are all different preparations procedures for the same state. This insight leads us to an alternative characterisation of states. Since two processes are considered preparation procedures for the same state if and only if the propensities of any

observations on the system are equivalent, the relation of being preparation procedures for the same state is an equivalence relation. Since if it is possible for a system to be found in a state, it must be have been possible to prepare it in that state, there is also clearly a one-to-one mapping between states and the equivalence classes, so we can without loss of generality identify the state with its ascociated equivalence class.

Since we adopted such a broad definition of preparation procedure, we see that a probabilistic mixture of preparation procedures is also a preparation procedure. Suppose we have two preparation procedures for the same system $\mathscr{P}$ and $\mathscr{Q}$, then we can define a process of preparing the system by randomly choosing between the two, with a given weight. Indeed it is true that for any $\lambda \in [0,1]$, a procedure whereby with probability λ we carry out $\mathscr{P}$, and with probability $1-\lambda$ we carry out $\mathscr{Q}$, will be another preparation procedure.

The randomisation procedure used in defining a mixture of preparations could in principle later be read off, and thus the 'actual' procedure used could be determined. We would also expect the system to behave the same whether, in a given mixing of preparations, the readout is taken or not, and that when we do know the result of the randomisation, the propensities of outcomes of observations will be given by the statistical mixing of the propensities of the two original procedures. Thus under these circumstances the details of the preparation procedures and mixing would determine the propensities of outcomes of all measurements, and thus define a state.

Thus we have that for each pair of states $\omega_1, \omega_2 \in \Omega$ and weight $\lambda \in [0,1]$ we get another state, and due to the way the states were defined via probabilistic mixing of propensities, we have that the set of states Ω takes on a convex structure.

2.3 Effects and observations

The first part in our definition of GPM gives us information about the structure of a set of states in the model, namely that of convexity. Whilst the state of a system, by definition, in principle contains all the information that can be accessible, there is no reason to expect in general that it is possible to gain a

full description of the state of any given system through observation. Thus we need to consider exactly what observations it is possible to make on the system in question. For this we will need the concept of effects.

Definition 2. An *effect* on a system with state space Ω is a function that assigns to each state $\omega \in \Omega$ a real number between 0 and 1.

Each observation on the system must have some set of defined outcomes that could be 'observed'. Each effect then corresponds to some outcome in a possible observation, and tells us for any given state the probability of achieving that outcome. For a system with state space Ω, we shall deonte the set of all effects on that system as $\mathscr{E}(\Omega)$.

Each effect $e \in \mathscr{E}(\Omega)$ is then clearly represented by a function $e : \Omega \to [0, 1]$, but we must ask if indeed every such function should be classed as an effect, or if there is some restriction that should be required in order for the probability assignments of the effect to make physical sense. The answer to this comes from the sole restriction that we have so far established on the structure of the state space, that of convexity. We have noted that some states can be viewed as a convex mixture of other states, in the sense that the propensities of outcomes of observations on the mixed state are given as a fixed convex combination of the propensitites of the same outcomes on the states that constitute the mixture. Since each effect gives the probabilities of a given outcome on all states, this condition is exactly that each effect, considered as a map, should preserve the convex structure of the state space. Therefore we have that each effect $e \in \mathscr{E}(\Omega)$ in fact lives in the set of affine maps from Ω to $[0, 1]$, i.e. $\mathscr{E}(\Omega) \subset A(\Omega, [0, 1])$.

In an analogous way to that considered above of defining a new preparation procedure by randomly choosing between old ones, we can consider making an observation on a system by randomly choosing between two observations that share the same outcomes. As we want this random choise to preserve all probabilities, similarly to the situation with states, we would expect the probability of any given outcome to be a well-defined mixture of the probabilities of the two original observations. This means that since effects give the probabilities of the outcomes, that we also wish the set of effects $\mathscr{E}(\Omega)$ to have a convex structure.

2.4 Double dual embedding

We have now established that effects are affine maps taking values in the interval $[0,1] \subset \mathbb{R}$, and therefore the set of effects indeed lives in the linear space of real valued affine functionals on Ω, $A(\Omega, \mathbb{R})$. This useful fact not only allows us to talk about effects as elements of a vector space, but also gives us a way to do likewise about the set of states.

For each state $\omega \in \Omega$ we can define a related function on effects $\hat{\omega}$, that acts by evaluating the effect on ω:

$$\hat{\omega}(e) = e(\omega), \quad \forall e \in \mathscr{E}(\Omega).$$

Since effects act in an affine way, it can easily be seen that the functions $\hat{\omega}$ are also affine:

$$\begin{aligned}\hat{\omega}(\lambda e_1 + (1-\lambda)e_2) &= (\lambda e_1 + (1-\lambda)e_2)(\omega) \\ &= \lambda e_1(\omega) + (1-\lambda)e_2(\omega) \\ &= \lambda\hat{\omega}(e_1) + (1-\lambda)\hat{\omega}(e_2).\end{aligned}$$

Here we observe a useful duality: whilst effects are affine functions on states taking values in $[0,1]$, states can be viewed as affine functions on effects with values in $[0,1]$. Again, since we can consider having $\Omega \subset A(\mathscr{E}(\Omega), [0,1])$, we can also view the set of states as living in the linear space of real valued functionals $A(\mathscr{E}(\Omega), \mathbb{R})$.

Now consider an arbitrary element $s \in A(\mathscr{E}(\Omega), [0,1])$. By definition s will ascribe to each effect $e \in \mathscr{E}(\Omega)$ a number between 0 and 1, $s(e)$. Since each outcome of any observation that can be made on the system corresponds to an effect, which determines the probability of the outcome, and s gives a probability to each effect, it seems that s has much in common with a state. Indeed here we make the simplifying assumption that indeed s is ascociated with a state. When we attempt to theoretically model a possibly hypothetical system, we may well want to consider all theoretically possible 'states', and any $s \in A(\mathscr{E}(\Omega), [0,1])$ fits the criterion of being theoretically possible. Even though the exact characteristics of s may not represent anything physically realisable,

in the sense of there existing a known preparation procedure that prepares a state that matches all the statistics of s, is it not wholly unreasonable to consider such a state as, at least theoretically, possible since we have no way of knowing beforehand which states can be prepared by known means, nor if there could be an unknown procedure that would prepare such a state.

Now we have a one-to-one correspondence between Ω and $A(\mathscr{E}(\Omega), [0,1])$. Viewing the latter as a subset of $A(\mathscr{E}(\Omega), \mathbb{R})$, it will form the base of the positive cone $A(\mathscr{E}(\Omega), \mathbb{R}_+) \equiv \mathbb{R}_+\Omega$. The cone in turn defines a partial ordering on $A(\mathscr{E}(\Omega), \mathbb{R})$, namely that inherited from the pointwise partial order of functions on $\mathscr{E}(\Omega)$.

In order to simplify notation at this point, from here on the vector space $A(\mathscr{E}(\Omega), \mathbb{R})$ will be referred to as V, and Ω will be used to refer to both the state space itself and its identification as a convex subset of V. At this point two more small assumptions will be made. Firstly, since it applies to all cases that will be considered in the following, it will be assumed that the vector space V is finite dimensional. The second assumption is that the convex set $\Omega \subset V$ is closed. This assumption can be justified by considering that, similarly to the assumption that $\Omega \equiv A(\mathscr{E}(\Omega), [0,1])$, any boundary point of Ω can be approximated arbitrarily well by actual states, and so it is reasonable to consider such a 'state' to theoretically be possible.

To complete the structure of the dual nature of states and effects we note that effects are affine functionals on the set Ω, which is now considered as a convex subset of V. This means that we can now extend each $e \in \mathscr{E}(\Omega)$ to a linear functional $\tilde{e}$ on V that agrees with e on Ω. Through such a correspondence we can consider $\mathscr{E}(\Omega)$ as a subset of the dual vector space V^*. Similarly to the case with states, it is not unreasonable to consider any affine funcional on Ω that takes values in $[0,1]$ to be a theoretically possible effect, since it assigns what could be considered a probability value to each state. Because of this from now on we will consider the whole of $A(\Omega, [0,1])$ to constitute the set of effects, and use $\mathscr{E}(\Omega)$ to deonte both the actual set of effects and it's identification with the set of linear functionals agreeing on Ω which live in V^*. Here we can also note that if we have two elements $e, f \in V^*$ which both take values in $[0,1]$ on Ω, the the functional $\lambda a + (1-\lambda)f$ will also take values in $[0,1]$.

This leaves us with effects in a dual position to states, with $\mathscr{E}(\Omega)$ a convex

subset of V^*. Indeed the set $\mathbb{R}_+\mathscr{E}(\Omega)$ will form a positive cone in V^*, which defines a partial ordering that comes from the pointwise partial ordering of functions on Ω. In fact the two cones $V_+ = \mathbb{R}_+\Omega$ and $V_+^* = \mathbb{R}_+\mathscr{E}(\Omega)$ are dual to each other.

2.5 Examples

We have now established what mathematical structures we mean when we talk about a basic GPM for a system, namely

- An ordered linear space V, with positive cone V_+
- A convex subset $\Omega \subset V$ - the *state space* - such that $V_+ = \mathbb{R}_+\Omega$
- The dual space V^*, with positive cone V_+^*
- The convex subset $\mathscr{E}(\Omega)$ - the *effect space* - of functionals taking values in $[0, 1]$ on Ω, with $V_+^* = \mathbb{R}_+\mathscr{E}(\Omega)$

In order to understand how this abstract framework fits into the way we currently model physical systems, and to see how it contains even more possibilities, it is useful to look at some examples. Some of the examples, namely those from quantum theory and classical probability theory, are well established as ways of modelling systems, whereas some other models have no known realisation in nature.

2.5.1 Classical

Probably the simplest examples of GPMs come from classical probability theory, which can be used to model events involving macroscopic objects with a degree of randomness. Such situations may not involve 'true' randomness, indeed classical physics is usually considered to be completely deterministic. However the system we are seeking to model may only involve a certain few of the degrees of freedom of the macroscopic objects, and thus behave in a random way, notably like a flipped coin or tossed die.

In a classical probability model there are a fixed number of points, or outcomes, that are used to define a system (e.g. the sides of a die), and the state of the system is then determined simply by the weighting of how relatively likely each outcome is. So if a given system is determined by n outcomes then the state of that system is determined by a list of numbers $p_1, \ldots, p_n$ taking values in $[0, 1]$, each representing the propensity of the related outcome.

Now fitting this in to the framework of GPMs we have

- $V = \mathbb{R}^n$
- $\Omega = \{v \in \mathbb{R}^n \mid 0 \leq v_i,\ \sum_i v_i = 1\}$

where v_i is the ith component of the vector $v = (v_1, \ldots, v_n)$.

There is also another characterisation of the set Ω, that is often useful, coming from the fact that it is, by definition, a closed and bounded convex subset of a real vector space. Indeed such a set is always equal to the convex hull of its extreme points, i.e. in this case we can also write

$$\Omega = \text{conv}\,\{(1, \ldots, 0, \ldots, 0), \ldots, (0, \ldots, 1, \ldots, 0), \ldots, (0, \ldots, 0, \ldots, 1)\}$$

The positive cone on V (here $\mathbb{R}^n$) that is generated by Ω is then also determined to be the cone generated by the extreme points of Ω. For the case of classical probability theory then we have that an element $v \in \mathbb{R}^n$ is positive iff each of it's entries is positive, i.e. $V_+ = \mathbb{R}^n_+$.

Characterising the state space Ω by extreme points gives us a useful way of determining the structure of the space of effects. First assume that a linear functional $e \in V^*$ satisfies $0 \leq e(\omega_i) \leq 1$, where the ω_i are the extreme points of Ω. Now consider an abitrary point $\omega \in \Omega$. Since we know that Ω is the convex hull of the ω_i, there exist $\lambda_i \geq 0$, with $\sum_i \lambda_i = 1$ and $\omega = \sum_i \lambda_i \omega_i$. However now we have

$$e(\omega) = e\left(\sum_i \lambda_i \omega_i\right) = \sum_i \lambda_i e(\omega_i) \geq \sum_i \lambda_i \inf_i[e(\omega_i)] = \inf_i[e(\omega_i)] \geq 0,$$

and

$$e(\omega) = e\left(\sum_i \lambda_i \omega_i\right) = \sum_i \lambda_i e(\omega_i) \leq \sum_i \lambda_i \sup_i[e(\omega_i)] = \sup_i[e(\omega_i)] \leq 1.$$

Here we clearly see that a functional is an effect (i.e. taking values in $[0,1]$ on Ω) iff it takes values in $[0,1]$ on the extreme points of Ω.

In this case then the effects are precisely those vectors whose canonical inner product with all vectors of the form $(0,\ldots,0,1,0,\ldots,0)$ lies between 0 and 1, and we have

- $V^* = \mathbb{R}^n$
- $\mathscr{E}(\Omega) = \{e \in \mathbb{R}^n \,|\, 0 \leq e_i \leq 1, \forall\, 1 \leq i \leq n\}$

Here we see that the positive cone generated by the set of effects is the same as that generated by the set of states, namely $V_+^* = \mathbb{R}_+^n$.

2.5.2 Quantum

The next example to look at comes from quantum theory. When dealing with many microscopic objects the laws of classical physics that apply to much larger objects may no longer apply. Instead, the main method of modelling systems such as electrons passing through Stern-Gerlach magnets or photons in optical circuits, is to use quantum mechanics.

The state of a quantum mechanical system is given by a density operator, that is, a positive operator on some complex Hilbert space $\mathscr{H}$, whose trace is equal to 1. Which Hilbert space exactly is used to model any given system will depend upon the system itself and the number of it's degrees of freedom. As mentioned above, we shall only be considering systems which we are modelling with finite dimensional state spaces. This means that, for example we shall not be attempting to model the position of an electron, but we shall be attempting to model its spin, since the former has 'infinite' degrees of freedom, whereas the latter does not.

The set of density operators naturally sits inside the real vector space of self-adjoint trace class operators $\mathcal{T}(\mathscr{H})_s$, which in the case of $\mathscr{H}$ being finite dimensional is just equal to the space of all self-adjoint nxn matrices, where n is the dimension of $\mathscr{H}$. So we have quantum mechanically

- $V = \mathcal{M}_n(\mathbb{C})_s$

- $\Omega = \{M \in \mathcal{M}_n(\mathbb{C}) \,|\, \mathrm{Tr}[M] = 1,\ M \geq 0\} = \mathcal{T}_+^1(\mathbb{C}^n)$

For the case of qubits, where the Hilbert space has dimension 2, there is an especially nice parameterisation of the state space. To see this we need to introduce a basis for the space of 2x2 self-adjoint matrices, often referred to as the Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These three matrices, along with the Identity matrix, form a linearly independent set, and thus a basis for the space of self-adjoint matrices, meaning that each state can be written as a linear combination of them. Note here that the Pauli matrices are all traceless, and since states have trace equal to 1, it means that all states must have the same coefficient for the identity, namely $\frac{1}{2}$. So any state $\rho \in \mathcal{T}(\mathbb{C}^2)_+^1$ can be written as

$$\rho = \frac{1}{2}(\mathbb{I} + r.\sigma),$$

with $\sigma = (\sigma_x, \sigma_y, \sigma_z)$, parameterising the state space by a single 3-dimensional vector r. The trace condition of being a state here is guaranteed, however a further restriction is needed in order to ensure positivity. Since the eigenvalues of ρ are $\frac{1}{2}(1 \pm |r|)$, we see that we must have $|r| \leq 1$ for these to be positive. This gives us the the Block sphere representation of the state space.

The vector space dual to the space of trace class operators, where the state space lives, consists of the bounded operators, with the action given by taking the trace of the product of the operators. In the case of finite dimensional Hilbert spaces all operators are bounded, as they are trace class, and so the real linear space containing the set of effects is again just $\mathcal{M}_n(\mathbb{C})_s$. The convex set of effects itself is then given by all operators lying between 0 and $\mathbb{I}$, which in the case of qubits, using the Pauli basis any $E \in \mathscr{E}$ can be written as

$$E = \frac{1}{2}(e_0\mathbb{I} + e.\sigma),$$

where the inequality $0 \leq E \leq \mathbb{I}$ is satisfied when $|e| \leq e_0, 2 - e_0 \leq 1$.

Fitting this back into the framework of GPMs we get

- $V^* = \mathcal{M}_n(\mathbb{C})_s$
- $\mathscr{E}(\Omega) = \{M \in \mathcal{M}_n(\mathbb{C}) \,|\, 0 \leq M \leq \mathbb{I}\}$.

2.5.3 Polygons

Not all examples of GPMs come from what has actually been observed in nature, and it would not be a great tool if that were the case. Here we look at an example of a simple and yet interesting class of state spaces for which there is not necessarily any way to physically realise them.

As mentioned above, a state space can be defined as the convex hull of its extreme points, and here we give a class of state spaces with n extreme points for all $n \geq 3$. To do this we define the points, in $\mathbb{R}^3$

$$\omega_i^{(n)} = \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}, \qquad i = 1, ..., n$$

with $r_n = \sqrt{\sec(\frac{\pi}{n})}$. The n points all lie in the same plane and form a regular polygon with n sides. As a GPM this then looks like

- $V = \mathbb{R}^3$
- $\Omega_n = \text{conv}\left\{\omega_1^{(n)}, \ldots, \omega_n^{(n)}\right\}$

One example that will turn out to be useful later on is the case when $n = 4$ and the state space has the shape of a square, also known as a *squit*.

For the polygon state spaces it is also useful to give the effect spaces as a convex hull living in $\mathbb{R}^3$ as well. Although for all n the effect space will always contain the 0 and 1 effects, $(0, 0, 0)$ and $(0, 0, 1)$, it turns out, rather interestingly, that the structures differ depending on whether n is odd or even.

For the case where n is even we define

$$e_i^{(n)} = \frac{1}{2}\begin{pmatrix} r_n \cos(\frac{(2i-1)\pi}{n}) \\ r_n \sin(\frac{(2i-1)\pi}{n}) \\ 1 \end{pmatrix}, \qquad i = 1, ..., n$$

And then the effect space $\mathscr{E}(\Omega_n)$ is then given by the convex hull of these points along with the 0 and 1 effects. In the case where n is odd however we make the definitions

$$e_i^{(n)} = \frac{1}{1+r_n^2}\begin{pmatrix} r_n\cos(\frac{2\pi i}{n}) \\ r_n\sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}, \qquad i = 1, ..., n$$

which, although seemingly similar to the even case, are not enough to define the whole effect space. For odd n the effect space is the convex hull of the 0 and 1 effects, the $e_i^{(n)}$, and their complement effects

$$e_i^{'(n)} = 1 - e_i^{(n)} = \frac{1}{1+r_n^2}\begin{pmatrix} -r_n\cos(\frac{(2i-1)\pi}{n}) \\ -r_n\sin(\frac{(2i-1)\pi}{n}) \\ r_n^2 \end{pmatrix}, \qquad i = 1, ..., n$$

This fact seems less surprising when noting that for even n we have the identity $e_j = e_i' = 1 - e_i$, when $j = i + \frac{n}{2} \bmod n$, and we have a more consistent definition for both even and odd n:

- $V^* = \mathbb{R}^3$
- $\mathscr{E}(\Omega_n) = \text{conv}\left\{0, 1, e_1^{(n)}, \ldots, e_n^{(n)}, e_1^{'(n)}, \ldots, e_n^{'(n)}\right\}$

2.6 Observables

So far we have covered the set of states of a GPM and how effects can tell us the probability of a given outcome. In this sense effects can be considered as representing the result of a two outcome measurement; they signify the answer to a yes-no question about whether the system in question is seen to have the property asscociated with that effect. In practice however, if we wish to consider any possible observation on a system, then we should be able to have a model for situations that cannot be expressed by one, or a series of, yes-no questions. The most simple example of this would be when throwing a die. In general we are not just interested in whether it came up 6, or if the number was odd, both of which are describable just with effects. More commonly we wish to consider it as an observation with 6 possible outcomes, corresponding to the sides of the die.

The most natural way of being able to consistently describe the set of outcomes, is for them to form a measure space. Since we will not be considering continuous outcome measurements, for simplicity we shall assume all our measure spaces to be discrete.

Given a state space of a GPM, Ω and some set O, the outcome space, the idea of an observable is to provide a description of the propensities of experimental outcomes. In order to achieve this an obervable M should be able to assign to each state ω a probability distribution $p_\omega^{(M)}$ on O. Then for each $o \in O$, we have $M(\omega)[o] = p_\omega^{(M)}[o]$ as being the probability of achieving outcome o when observable M is measured in the state ω.

Another perspective on observables is to fix an outcome o, and consider the expression $M(\omega)[o]$ as a function of ω. As discussed in the previous section, in order to preserve the probabilities upon blindly and randomly choosing between two observations to perform, we can assume that the observable must act in an affine way. This leaves us with $M[o] : \Omega \to [0,1]$, with $(M[o])(\omega) = p_\omega^{(M)}[o]$, as an affine functional on Ω taking values in $[0,1]$, hence it exactly fits the definition of an effect from earlier.

Then we can make the following definition

Definition 3. An *observable* M on a system with state space Ω, taking values in O, is an (Ω-)effect valued measure on O. That is, M is a σ-additive function $O \to \mathscr{E}(\Omega)$.

2.6.1 Joint Measurability

One of the main initial reasons people were motivated to look at GPMs in the first place, was to understand certain experimentally observed phenomena. Specifically there were many seemingly strange and novel ideas and effects that were coming out of quantum mechanics. The framework of GPMs gives a way to talk about these non-classical phenomena from a more objective, external perspeective. One new feature that this opened up was the idea of joint measurability.

The idea of observables being jointly measurable, as is strongly suggested by the terminology, is that they can both be measured, and give definite values at the

same time. When viewing the world with a purely classical mindset, the notion of joint measurability does not arise, since it is taken as implicit. Indeed in everyday language we talk about objects having values for different observables at the same time, a prime example being the position and momentum of a body. As we will see however this is a very classical notion.

Definition 4. Two observables $M^{(1)}$ and $M^{(2)}$ on Ω, taking values in $O^{(1)}$ and $O^{(2)}$ respectively, are *jointly measurable* if there exists an observable M on Ω, taking values in $O^{(1)} \times O^{(2)}$, such that $M^{(1)}[o^{(1)}] = M[o^{(1)}, O^{(2)}]$ and $M^{(2)}[o^{(2)}] = M[O^{(1)}, o^{(2)}]$ for all $o^{(1)} \in O^{(1)}$, $o^{(2)} \in O^{(2)}$.

Such an M is often referred to as a joint obervable for $M^{(1)}$ and $M^{(2)}$.

This concept can be illustrated by looking first at the situation of a pair of observables in classical probability theory. Suppose we have a classical system with state space $\Omega = \text{conv}\{\omega_1, \ldots, \omega_n\} \subset \mathbb{R}^n$, where each ω_i is the ith standard unit vector in $\mathbb{R}^n$. Also suppose that we have two observables on Ω given by sets of effects $\{e_1, \ldots, e_k\}$ and $\{f_1, \ldots, f_l\}$. In order to satisfy the definition of observables we must have $\sum_i e_i = \sum_j f_j = 1$, the identity effect on Ω, and also $e_i, f_j \geq 0$ where we remember that the positivity is determined by being positive on all the extremal points ω_i.

Now we can define another observable on Ω with effects g_{ij}, where $1 \leq i \leq k$, $1 \leq j \leq l$, whose mth component (equivalently the value on ω_m) is given as $g_{ij}^{(m)} = e_i^{(m)} f_j^{(m)}$. Clearly then we have $g_{ij}^{(m)} \geq 0, \forall m$, and hence $g_{ij} \geq 0$, but also $\sum_{i,j} g_{ij}^{(m)} = 1$, so $\sum_{i,j} g_{ij} = 1$, and we find that we have a joint observable.

It is worth noting here the relationship between the notion of incompatibility as introduced here when applied to quantum observables, and the non-commutativity of operators which would commonly be referred to as incompatible. Any self adjoin operator on a Hilbert space, has an ascociated projection vaued measure (PVM), namely it's spectral measure. This PVM fits the conditions of definition 3, and so to each self adjoint operator we can assign an observable. If two self adjoint operators are commutative, then so will be the PVMs defining the observables. Such observables will always be jointly measurable, with the elements of the joint observable simply being the products of the corresponding projections in the PVMs. From this it follows that for any two

PVMs that are not jointly measurable, the corresponding self adjoint operators will be non-comutative. However, because there are more quantum effects than just projections that can be used to form observables, commutativity is not a necessary condition for joint measurability.

2.7 Composition of Systems

Very often, considering systems as individual entities, completely isolated, it is necessary to consider multiple systems that are described by your theory, at the same time. It may be that there are two systems which can undergo mutual interaction, there could be many separated systems that may have interacted (or not) in the past, or possibly there are just two systems which exist side by side at the same time. Whatever the setup, it would be expected that a physical theory should be able to account for such scenarios.

Since any multipartite system can be built up by adding one system at a time, we shall only need to consider how to describe the composition of two systems. So we are in a situation where we have two state spaces Ω_1 and Ω_2, in vector spaces V_1 and V_2, and wish to find a third vector space V with convex subset Ω, which will represent the state space of the composition of the systems.

In order to determine exactly what form the description of the composite system should take, it is useful to consider the problem of what the dimension of the linear space V should be. For this we will need two fairly reasonable assumptions. The first assumption is that of tomographic locality ([7]). This postulates that the state of a compound system is completely determined by the probabilities of outcomes of obervations made on the constituent systems individually. Such observations may be carried out simultaneously, to account for possible correlations between the two systems, but observations that require the systems to be observed together are not necessary. Such an assumption is not unreasonable since it is hard to justify talking about such an object as being a compound system with constituents, if it must be treated as if they were just one single entity. The second assumption is one of the most fundamental principles in modern physics, namely that of no-signalling. In this context the no-signalling principle implies that the statistics of outcomes of measurements

on one of the constituent parts should not differ depending on any measurements carried out on the other constituent, or indeed if no measurement is performed at all. It can be shown ([7]) that under these assumptions, if V_1 and V_2 have dimensions n_1 and n_2 respectively, the dimension of V must be $n_1 n_2$.

The previous result tells us that, in effect, the state space of a composite system must lie in the tensor product of the underlying vector spaces of the constituent systems, $V \simeq V_1 \otimes V_2$. This fact does not determine the exact convex set that should represent the state space, however there are considerations that we can use to narrow down the choice. For this it is useful to consider the positive cones that are generated by Ω_1 and Ω_2, $(V^+)_1$ and $(V^+)_2$, and their relation to the corresponding cone V^+. Given that the notion of compound systems should include the ability to model completely independent systems, formally considered as one, it seems logical that any state, and by extension any positive element, that is prepared by simply preparing states separately on the two systems should be considered part of the state space, and positive cone, of the compound system. Along with the previously stated condition that any state space must be convex, this leads us to define the *minimal tensor product* of two ordered linear spaces with positive cones $(V_1)_+$ and $(V_2)_+$

$$(V_1 \otimes_{min} V_2)_+ = \text{conv}\,\{v_1 \otimes v_2 \,|\, \forall\, v_1 \in (V_1)_+, v_2 \in (V_2)_+\}$$

One can also look at the state space from the dual perspective of the space of effects. Each positive cone ascociated with a state space can in fact be seen as the dual to the positive cone generated by the effect space. However reasoning similarly to above we can stipulate that such a cone should contain all elements that represent yes-no observations performed on the constituents individually. From this we can give the *maximal tensor product* of two ordered linear spaces with cones

$$\begin{aligned}(V_1 \otimes_{max} V_2)_+ &= (V_1^* \otimes_{min} V_2^*)_+^* \\ &= \{v \in V_1 \otimes V_2 \,|\, (e_1 \otimes e_2)(v) \geq 0, \forall\, e_1 \in (V_1^*)_+,\, e_2 \in (V_2^*)_+\}\end{aligned}$$

Indeed these two conditions both put constraints on what we should consider as the state space in V but they are, in general, not equivalent. This leaves us in a situation where we have restrictions on what may, or may not, constitute

a compound state space, but there is still some ambiguity and choice available to be specified by any given theory. To this end we will consider a valid tensor product to be any convex set that lies between the maximum and minimum, including these extremes.

2.8 Correlations

Along with the notion of joint measurability, one of the big new conceptual considerations to come out of the study of quantum mechanics was that of, so called, 'non-local' correlations. Such considerations originated with the thoughts or, among others, Einstein, Podolsky and Rosen, when they considered quantum behaviours similar to those exhibited by a maximally entangled two qubit state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle).$$

Their reasoning phrased here in terms of qubit experiments, went as follows. If a sharp $|0\rangle, |1\rangle$ measurement is made on the first system, and the outcome obtained is, say $|1\rangle$, then this would change the state of the second system into one of being certain of obtaining $|0\rangle$ in a similar measurement, whereas before both outcomes would be equally likely. The 'non-local' part comes from the fact that these considerations do not involve the relationship between the systems in any way. Indeed the two systems could be separated by a great distance, so with no way to influence each other, and yet the change in state would seem to happen instantly.

Initially some found this troubling, and thought that it meant that the formulation of quantum mechanics was inaccurate or incomplete. However, starting with the work of Bell, this has become to be seen as a non-classical feature inherent in quantum mechanics. The argument is based on looking at certain types of bipartite correlations, represented by the functional

$$\mathbb{B} = \langle A_1 B_1 + A_1 B_2 + A_2 B_1 - A_2 B_2 \rangle,$$

where A_1 is the ± 1 valued observable defined by effect e etc. on system 1. It can be easily shown that for classical models $\mathbb{B} \leq 2$, whereas in quantum mechanics we can get as high as $\mathbb{B} = 2\sqrt{2}$. However considering the case of two squits,

joined by the maximal tensor product, gives a PR box that gives $\mathbb{B} = 4$, its maximum possible value. In fact for each GPM there will be a bound on that set of correlations, that will generally be different. One example of a different value comes from taking the maximal tensor product of two pentagons, where we get a bound of $4\sqrt{5} - 6$ putting it between the quantum and maximal cases.

2.9 Transformations

It can also be useful to consider the possibilty of some kind of operation on a system other than a preparation or a measurement. These operations that act by modifying a system, but do not lead to any specific outcomes are referred to as transformations. These can come in two types, namely those which transform one type of system into another, and those which leave the type of system unchanged, just in a different configuration.

Since we want transformations to map one convex set into another, the natural type of map to consider to represent them is an affine map. This can be confirmed as being necessary by demanding that convex mixtures of states be preserved under transformations, in a similar manner to obervables. Affine maps between state spaces naturally lift to positive linear maps between the asscociated ordered vector spaces, $A_+((V_1)_+, (V_2)_+)$.

Chapter 3

Comparing the degrees of incompatibility inherent in probabilistic physical theories

3.1 Abstract

We introduce a new way of quantifying the degrees of incompatibility of two observables in a probabilistic physical theory and, based on this, a global measure of the degree of incompatibility inherent in such theories, across all observable pairs. This opens up a novel and flexible way of comparing probabilistic theories with respect to the nonclassical feature of incompatibility, raising many interesting questions, some of which will be answered here. We show that quantum theory contains observables that are as incompatible as any probabilistic physical theory can have if arbitrary pairs of observables are considered. If one adopts a more refined measure of the degree of incompatibility, for instance, by restricting the comparison to binary observables, it turns out that there are probabilistic theories whose inherent degree of incompatibility is greater than that of quantum mechanics.

3.2 Introduction

Quantum theory has a number of important features not known in classical physics, ranging from the superposition and indeterminacy principles formulated by the pioneers to the more recently discovered no-cloning and no-broadcasting theorems. It is an old problem to identify operationally significant properties of quantum theory that distinguish it from other probabilistic theories. In recent years many features have been under intensive investigation from this perspective, including information processing [12], optimal state discrimination [13], entropy [14], purification [15] and discord [16]. It has been found that some properties are quite generally valid in any non-classical (no-signaling) probabilistic theories while others are specifically quantum.

The existence of pairs of incompatible observables marks one of the most striking distinctions between quantum theory and classical physical theories. There are many manifestations of incompatibility, perhaps the most famous being the Heisenberg uncertainty principle [17]. However, there are many nonclassical probabilistic theories which also possess incompatible observables, and it will be of interest to compare quantum theory with alternative theories with respect to the feature of incompatibility.

To this end, we define the joint measurability region of any given pair of observables in a probabilistic theory. The joint measurability region describes the amount of added noise needed to make the observables jointly measurable. The global joint measurability feature of a probabilistic theory can then be characterized as the intersection of all the joint measurability regions associated with the theory.

We demonstrate that quantum theory contains observables that are as incompatible as observables in any probabilistic theory can be. Hence, we can say that, in a global sense, quantum theory has as great a degree of incompatibility as any other probabilistic theory. But if only binary observables are considered, the degree of incompatibility inherent in quantum theory is limited and we give an example of a probabilistic theory that contains maximally incompatible binary observables.

Our aim is thus to compare the incompatibility of pairs of observables in dif-

ferent probabilistic physical theories. We first need to set some minimal constraints.

3.3 Probabilistic Theories

A probabilistic theory is a framework that provides a description of physical systems in terms of states and observables with the following general properties:

(i) The states of a system are represented by the elements of a convex subset of a real vector space.
(ii) An observable is represented as an affine mapping from the set of states into the set of probability distributions on some outcome space. For simplicity, we restrict ourselves here to observables with a finite or countable number of outcomes.
(iii) Any affine mapping from the set of states into the set of probability distributions is a valid observable.

We consider a particular probabilistic theory (PT) as given by a family of convex sets of states with associated sets of observables that share some properties specific to that PT. One may think of each pair consisting of a set of states with associated set of observables as an *instance* of a PT representing a particular type of physical system.

Given a PT, we denote by $\mathfrak{M}(j|\varrho)$ the probability of obtaining a measurement outcome j when an observable $\mathfrak{M}$ is measured in a state ϱ. Hence, $0 \leq \mathfrak{M}(j|\varrho) \leq 1$ and $\sum_j \mathfrak{M}(j|\varrho) = 1$. We will typically label the measurement outcomes by integers.

In quantum theory the states are described by density operators and observables correspond to POVMs [18]. Their duality is given by the trace formula (with ϱ a density operator and M a POVM)

$$\mathfrak{M}(j|\varrho) = \mathrm{tr}\left[\varrho \mathsf{M}(j)\right] . \tag{3.1}$$

Another example of a probabilistic theory is a classical theory, where the states are probability measures on a phase space Ω and observables are traditionally

represented as functions $m : \Omega \to \mathbb{R}$; the associated affine maps from states ϱ to probability distributions are then given by the formula

$$\mathfrak{M}(j|\varrho) = \varrho(\{x \in m^{-1}(j)\}) . \tag{3.2}$$

Continuing our discussion on general probabilistic theories, we note that it follows from the required properties (i)-(iii) that the set of observables is a convex set; a mixture of two observables is an observable. Physical mixing corresponds to an experiment where we switch between two measurement apparatuses with a random probability. We can directly write a mixture of two observables with the same set of measurement outcomes. If the sets of measurement outcomes differ, we can still write a mixture by first adding enough outcomes and then embedding both sets into $\mathbb{Z}$.

Another consequence of the basic requirements is that every constant mapping $\varrho \mapsto p$, where p is a fixed probability distribution, is an observable and we call it a *trivial observable*. A trivial observable $\mathfrak{T}$ corresponds to a dice rolling experiment, where we randomly pick the measurement outcome according to a given fixed probability distribution, without manipulating the state at all. In quantum theory, trivial observables are described by POVMs T such that each operator $\mathsf{T}(j)$ is a multiple of the identity operator, i.e., $\mathsf{T}(j) = t_j \mathrm{id}$ for some $0 \leq t_j \leq 1$ with $\sum_j t_j = 1$.

3.4 Joint Measurability

The concept of *joint measurement* can be defined in any probabilistic theory. Two observables $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are *jointly measurable* if there exists an observable $\mathfrak{M}$ such that

$$\sum_k \mathfrak{M}(j,k|\varrho) = \mathfrak{M}_1(j|\varrho) , \tag{3.3}$$

$$\sum_j \mathfrak{M}(j,k|\varrho) = \mathfrak{M}_2(k|\varrho) . \tag{3.4}$$

In this case $\mathfrak{M}$ is called a *joint observable* of $\mathfrak{M}_1$ and $\mathfrak{M}_2$. If $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are not jointly measurable, then we say that they are *incompatible*.

Any probabilistic theory contains jointly measurable pairs of observables. Namely, a trivial observable $\varrho \mapsto p$ is jointly measurable with any other observable; we can write a joint observable

$$\mathfrak{M}(j,k|\varrho) = \mathfrak{M}_1(j|\varrho)p(k) \tag{3.5}$$

for the trivial observable and any other observable $\mathfrak{M}_1$. This simply corresponds to an experiment where we measure $\mathfrak{M}_1$ and simultaneously roll a dice. It is a well known fact that, in quantum theory, an observable which is jointly measurable with all other observables is necessarily a trivial observable. Indeed, any POVM element of such an observable commutes with all projections and must therefore be a scalar multiple of the identity (e.g. [19, Theorem IV.1.3.1]).

The following simple observation is a key ingredient for our discussion.

Proposition 1. *Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be two observables and $0 \leq \lambda \leq 1$. Then $\lambda\mathfrak{M}_1 + (1-\lambda)\mathfrak{T}_1$ and $(1-\lambda)\mathfrak{M}_2 + \lambda\mathfrak{T}_2$ are jointly measurable for any choice of trivial observables $\mathfrak{T}_1$ and $\mathfrak{T}_2$.*

This proposition can be proved with the following construction. First, let p_1 and p_2 be the probability distributions related to $\mathfrak{T}_1$ and $\mathfrak{T}_2$. We define an observable $\mathfrak{M}$ by the formula

$$\mathfrak{M}(j,k|\varrho) = \lambda p_2(k)\ \mathfrak{M}_1(j|\varrho) + (1-\lambda)p_1(j)\ \mathfrak{M}_2(k|\varrho)\,. \tag{3.6}$$

For a fixed ϱ, the right hand side is clearly a probability distribution. Moreover, the right hand side is an affine mapping on ϱ; therefore $\mathfrak{M}$ is an observable. The marginal observables are

$$\sum_k \mathfrak{M}(j,k|\varrho) = \lambda\mathfrak{M}_1(j|\varrho) + (1-\lambda)p_1(j),$$
$$\sum_j \mathfrak{M}(j,k|\varrho) = (1-\lambda)\mathfrak{M}_2(k|\varrho) + \lambda p_2(k).$$

This proves Prop. 1.

The physical idea behind this construction is the following. In each measurement run we flip a biased coin and, depending on the result, we measure either $\mathfrak{M}_1$ or $\mathfrak{M}_2$ in the input state ϱ. In this way we get a measurement outcome for

either $\mathfrak{M}_1$ or $\mathfrak{M}_2$. In addition to this, we roll a dice and pretend that this is a measurement outcome for the other observable. In this way we get an outcome for both observables simultaneously. The overall observable is the one given in formula (3.6).

3.5 Joint Measurability Region

For two observables $\mathfrak{M}_1$ and $\mathfrak{M}_2$, we denote by $J(\mathfrak{M}_1, \mathfrak{M}_2)$ the set of all points $(\lambda, \mu) \in [0,1] \times [0,1]$ for which there exist trivial observables $\mathfrak{T}_1, \mathfrak{T}_2$ such that $\lambda\mathfrak{M}_1 + (1-\lambda)\mathfrak{T}_1$ and $\mu\mathfrak{M}_2 + (1-\mu)\mathfrak{T}_2$ are jointly measurable, and we call $J(\mathfrak{M}_1, \mathfrak{M}_2)$ the *joint measurability region* of $\mathfrak{M}_1$ and $\mathfrak{M}_2$. The joint measurability region thus characterizes how much noise (in terms of trivial observables) we need to add to obtain jointly measurable approximations of $\mathfrak{M}_1$ and $\mathfrak{M}_2$. Clearly, $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are jointly measurable if and only if $(1,1) \in J(\mathfrak{M}_1, \mathfrak{M}_2)$.

The joint measurability region $J(\mathfrak{M}_1, \mathfrak{M}_2)$ is a convex region which can be plotted in the plane. To see this, let $(\lambda', \mu') \in J(\mathfrak{M}_1, \mathfrak{M}_2)$ and $(\lambda'', \mu'') \in J(\mathfrak{M}_1, \mathfrak{M}_2)$, then we have to show that $(\lambda, \mu) \in J(\mathfrak{M}_1, \mathfrak{M}_2)$ for $(\lambda, \mu) = t(\lambda', \mu') + (1-t)(\lambda'', \mu'')$. Thus let $\mathfrak{M}_1' = \lambda'\mathfrak{M}_1 + (1-\lambda')\mathfrak{T}_1'$ and $\mathfrak{M}_2' = \mu'\mathfrak{M}_2' + (1-\mu')\mathfrak{T}_2'$ be jointly measurable, and similarly for $\mathfrak{M}_1'' = \lambda''\mathfrak{M}_1 + (1-\lambda'')\mathfrak{T}_1''$ and $\mathfrak{M}_2'' = \mu''\mathfrak{M}_2' + (1-\mu'')\mathfrak{T}_2''$, with suitable choices of trivial observables. Then the observables $t\mathfrak{M}_1' + (1-t)\mathfrak{M}_1''$ and $t\mathfrak{M}_2' + (1-t)\mathfrak{M}_2''$ are jointly measurable [62, Prop. 2].

Note that according to Prop. 1 the line $\left\{(\lambda, (1-\lambda)) : 0 \leq \lambda \leq 1\right\} \subseteq J(\mathfrak{M}_1, \mathfrak{M}_2)$. Moreover, it is trivially the case that $(0,0) \in J(\mathfrak{M}_1, \mathfrak{M}_2)$. The convexity of $J(\mathfrak{M}_1, \mathfrak{M}_2)$ then entails that the convex hull of the three points $(1,0)$, $(0,1)$ and $(0,0)$ is in $J(\mathfrak{M}_1, \mathfrak{M}_2)$, hence we have:

$$\triangle \equiv \{(\lambda, \mu) \in [0,1] \times [0,1] : \lambda + \mu \leq 1\} \subseteq J(\mathfrak{M}_1, \mathfrak{M}_2)\,.$$

As an example, suppose that we are within quantum theory and $\mathfrak{M}_1$ and $\mathfrak{M}_2$ correspond to spin-$\frac{1}{2}$ measurements in two orthogonal directions, say x and y -axes. We then describe them with two POVMs M_x and M_y, where

$$\mathsf{M}_x(\pm 1) = \tfrac{1}{2}(\mathrm{id} \pm \sigma_x)\,, \quad \mathsf{M}_y(\pm 1) = \tfrac{1}{2}(\mathrm{id} \pm \sigma_y)\,, \tag{3.7}$$

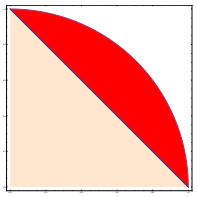


Figure 3.1: (Color online) The region $J(\mathsf{M}_x, \mathsf{M}_y)$ for two orthogonal spin-$\frac{1}{2}$ measurements is a quadrant of the unit disk. The region $\triangle$ (light) is a subset of $J(\mathfrak{M}_1, \mathfrak{M}_2)$ for any pair $\mathfrak{M}_1, \mathfrak{M}_2$, while the surplus region (dark) depends on the specific pair under consideration.

and σ_x, σ_y are the usual Pauli matrices in $\mathbb{C}^2$. It has been shown in [21] that for the uniformly distributed trivial observable $\pm 1 \mapsto \frac{1}{2}\mathrm{id}$ (hence describing an unbiased coin), the two observables $\lambda\mathsf{M}_x + (1-\lambda)\frac{1}{2}1$ and $\mu\mathsf{M}_y + (1-\mu)\frac{1}{2}1$ are jointly measurable if and only if $\lambda^2 + \mu^2 \leq 1$. It is also known [62, Prop. 3] that this inequality is a necessary condition for the joint measurability of any pair $\lambda\mathsf{M}_x + (1-\lambda)\mathsf{T}_1$ and $\mu\mathsf{M}_y + (1-\mu)\mathsf{T}_2$, where $\mathsf{T}_1, \mathsf{T}_2$ are arbitrary trivial observables. Therefore, we conclude that

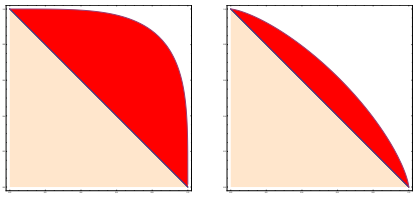
$$J(\mathsf{M}_x, \mathsf{M}_y) = \{(\lambda, \mu) \in [0,1] \times [0,1] : \lambda^2 + \mu^2 \leq 1\}\,. \tag{3.8}$$

This region is depicted in Fig. 3.1.

In addition to describing the incompatibility of pairs of observables, the concept of a joint measurability region also provides a means to compare the degrees of incompatibility inherent in entire theories. A global joint measurability feature of a probabilistic theory PT is characterized by the intersection of all the sets $J(\mathfrak{M}_1, \mathfrak{M}_2)$ across all instances of PT, and we denote

$$\begin{aligned} J_{PT} = \{&(\lambda, \mu) \in [0,1] \times [0,1] : (\lambda, \mu) \in J(\mathfrak{M}_1, \mathfrak{M}_2) \\ &\text{for all pairs of observables } \mathfrak{M}_1 \text{ and } \mathfrak{M}_2 \\ &\text{in all instances of PT}\}. \end{aligned}$$

We call J_{PT} the *joint measurability region* for PT. We always have $\triangle \subseteq J_{PT}$, but J_{PT} can be larger than $\triangle$. The larger the surplus region is, the more jointly measurable the theory is globally; see Fig. 3.2. If $(\lambda, \mu) \notin J_{PT}$, this



(a) (b)

Figure 3.2: (Color online) The region $\triangle$ (light) is a subset of the joint measurability region J_{PT} (colored) for any probabilistic theory. The larger the surplus region (dark) is, the more jointly measurable the theory globally is. If (a) and (b) are joint measurability regions for two different probabilistic theories, then we can conclude that (b) represents a greater degree of incompatibility than (a).

means that there is a pair of observables $\mathfrak{M}_1$ and $\mathfrak{M}_2$ such that the mixtures $\lambda\mathfrak{M}_1 + (1-\lambda)\mathfrak{T}_1$ and $\mu\mathfrak{M}_2 + (1-\mu)\mathfrak{T}_2$ are incompatible with any choice of trivial observables $\mathfrak{T}_1$ and $\mathfrak{T}_2$.

3.6 Quantum is Maximal

Since J_{PT} can be defined in any probabilistic theory, we can compare the joint measurability regions for different theories. We obviously have $J_{PT} = [0,1] \times [0,1]$ in any probabilistic theory where all measurements are jointly measurable, such as the classical probability theory. In the case of the greatest degree of incompatibility we would have $J_{PT} = \triangle$. We will next show that quantum theory incorporates, globally, as much incompatibility between pairs of observables as a probabilistic theory can do.

Theorem 1. *In quantum theory $J_{QT} = \triangle$. In particular, $J_{QT} \subseteq J_{PT}$ for any probabilistic theory PT.*

In quantum theory every observable $\mathfrak{M}$ corresponds to a unique POVM M by equation (3.1). We will prove that for any pair $(\lambda, \mu) \notin \triangle$, there are quantum observables M_1 and M_2 such that the mixtures $\lambda\mathsf{M}_1 + (1-\lambda)\mathsf{T}_1$ and $\mu\mathsf{M}_2 + (1-\mu)\mathsf{T}_2$ are incompatible with any choice of trivial observables $\mathsf{T}_1, \mathsf{T}_2$. Our proof is based on a recent result [22] on the joint measurability region for two

complementary observables, which is a generalization of the result illustrated in Fig. 3.1.

Proof of Theorem 1. We have earlier seen that $\triangle \subseteq J_{PT}$, so we need to show that $J_{QT} \subseteq \triangle$. Let $(\lambda, \mu) \notin \triangle$, i.e., $\lambda+\mu > 1$. Fix $\epsilon > 0$ such that $\lambda+\mu > 1+\epsilon$. We then choose d to be a positive integer satisfying

$$\frac{\sqrt{d}-1}{d-1} \leq \epsilon\,. \tag{3.9}$$

(This can be done since the left hand side $\to 0$ when $d \to \infty$.) We will consider a quantum system that is described by a d-dimensional Hilbert space $\mathcal{H}_d$. Let $\{\varphi_j\}_{j=0}^{d-1}$ be an orthonormal basis for $\mathcal{H}_d$. We define another orthonormal basis $\{\psi_k\}_{k=0}^{d-1}$ for $\mathcal{H}_d$ by

$$\psi_k = 1/\sqrt{d}\sum_j e^{-2\pi i \frac{jk}{d}}\varphi_j\,. \tag{3.10}$$

The orthonormal bases $\{\varphi_j\}_{j=0}^{d-1}$ and $\{\psi_k\}_{k=0}^{d-1}$ are mutually unbiased, i.e., $|\langle\, \varphi_j \,|\, \psi_k \,\rangle| = \mathit{constant} \quad \forall j,k$. We define two POVMs M_1 and M_2 by

$$\mathsf{M}_1(j) = |\varphi_j\rangle\langle\varphi_j|\,, \quad \mathsf{M}_2(k) = |\psi_k\rangle\langle\psi_k|\,. \tag{3.11}$$

We thus obtain a pair of d-outcome observables on $\mathcal{H}_d$. Since M_1 and M_2 consist of projections and $\mathsf{M}_1(j)\mathsf{M}_2(k) \neq \mathsf{M}_2(k)\mathsf{M}_1(j)$, it follows that they are incompatible.

As proved in [22], the observables $\lambda'\mathsf{M}_1 + (1-\lambda')\mathsf{T}_1$ and $\mu'\mathsf{M}_2 + (1-\mu')\mathsf{T}_2$ are incompatible for any choice of trivial observables $\mathsf{T}_1, \mathsf{T}_2$ whenever

$$\lambda' + \mu' > 1 + \frac{\sqrt{d}-1}{d-1}\,. \tag{3.12}$$

Since

$$\lambda + \mu > 1 + \epsilon \geq 1 + \frac{\sqrt{d}-1}{d-1}\,, \tag{3.13}$$

we conclude that $(\lambda, \mu) \notin J_{QT}$. □

Using the ideas of the proof of Theorem 1, we can also show that the conclusion $J_{QT} = \triangle$ can be reached by using a *single pair* of incompatible observables if we consider an infinite dimensional system and observables with a countably infinite number of outcomes.

Let $\mathcal{H}$ be an infinite dimensional Hilbert space and write it as a direct sum of finite d-dimensional Hilbert spaces $\mathcal{H}_d$, $\mathcal{H} = \bigoplus_{d=2}^{\infty} \mathcal{H}_d$. In each $\mathcal{H}_d$ consider a pair of mutually unbiased orthonormal bases $\{\varphi_j^d\}_{j=0}^{d-1}$ and $\{\psi_k^d\}_{k=0}^{d-1}$, where the latter is obtained from the first one by the formula (3.10). We define two POVMs N_1 and N_2 via

$$\mathsf{N}_1(d,j) = |\varphi_j^d\rangle\langle\varphi_j^d|\,, \quad \mathsf{N}_2(d,k) = |\psi_k^d\rangle\langle\psi_k^d|\,. \tag{3.14}$$

These observables act in the infinite dimensional Hilbert space $\mathcal{H}$ and d in (3.14) is an index labeling the different outcomes. The outcome space of N_1 and N_2 is $\Omega_\infty \equiv \{(d,j) : d = 2,3,\ldots, j = 0,\ldots,d-1\}$.

Theorem 2. *The observables* N_1 *and* N_2 *defined in* (3.14) *satisfy* $J(\mathsf{N}_1,\mathsf{N}_2) = \triangle$.

Proof of Theorem 2. Let p_1 and p_2 be two probability distributions defined on Ω_∞. Assume that $\lambda + \mu > 1$ and define two observables $\mathsf{N}_{1,\lambda}, \mathsf{N}_{2,\mu}$ via

$$\begin{aligned} \mathsf{N}_{1,\lambda}(d,j) &= \lambda\,|\varphi_j^d\rangle\langle\varphi_j^d| + (1-\lambda)\,p_1(d,j)I, \\ \mathsf{N}_{2,\mu}(d,k) &= \mu\,|\psi_k^d\rangle\langle\psi_k^d| + (1-\mu)\,p_2(d,k)I\,. \end{aligned} \tag{3.15}$$

We need to show that $\mathsf{N}_{1,\lambda}$ and $\mathsf{N}_{2,\mu}$ are incompatible. To prove this, we make the counter assumption that $\mathsf{N}_{1,\lambda}, \mathsf{N}_{2,\mu}$ are jointly measurable. This implies that for any projection P on $\mathcal{H}$, the projected observables $P\mathsf{N}_{1,\lambda}P$ and $P\mathsf{N}_{2,\mu}P$ acting on a subspace $P\mathcal{H}$ are jointly measurable. (If G is a joint observable of two observables $\mathsf{M}_1, \mathsf{M}_2$, then $P\mathsf{G}P$ is a joint observable of $P\mathsf{M}_1P, P\mathsf{M}_2P$ in $P\mathcal{H}$.) Especially, the projections of $\mathsf{N}_{1,\lambda}$ and $\mathsf{N}_{2,\mu}$ to any subspace $\mathcal{H}_d$ should be jointly measurable. But from the result cited in the proof of Theorem 1 we know that for d large enough, the projections to $\mathcal{H}_d$ are incompatible. Hence, $\mathsf{N}_{1,\lambda}$ and $\mathsf{N}_{2,\mu}$ are incompatible. □

We note that the observables N_1 and N_2 defined in (3.14) are not the only pair satisfying $J(\mathsf{N}_1,\mathsf{N}_2) = \triangle$. Namely, we can modify N_1 and N_2 in any chosen subspace $\mathcal{H}_d$ but the conclusion $J(\mathsf{N}_1,\mathsf{N}_2) = \triangle$ is still true since it depends on the fact that N_1 and N_2 contain mutually unbiased bases in arbitrarily high dimension.

An interesting problem within quantum theory would be to try to find a characterization of all pairs of quantum observables $\mathsf{M}_1, \mathsf{M}_2$ that satisfy $J(\mathsf{M}_1, \mathsf{M}_2) = \triangle$. In particular, we may ask if maximally incompatible observables can exist in a finite dimensional Hilbert space, or if they can have a finite number of outcomes. Since two mutually unbiased bases are expected to be among the most incompatible observable pairs in a fixed dimension d, our construction in the proof of Theorem 1 suggests that the answer to the first question would be negative. A proof of this claim is, however, lacking.

3.7 The Binary Case

As for the second question, we can present a partial answer by investigating the joint measurability region in the case of pairs of binary quantum observables. Our aim is to show that

$$\{(\lambda, \mu) \in [0,1] \times [0,1] : \lambda^2 + \mu^2 \leq 1\} \subseteq J(\mathsf{M}_1, \mathsf{M}_2)$$

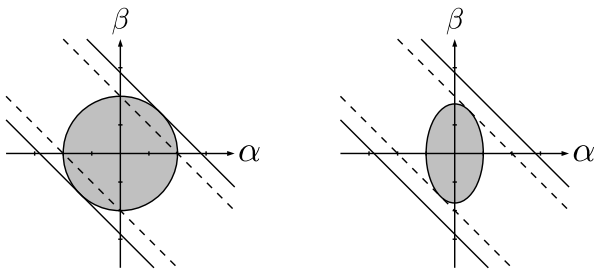
for any binary observables M_1 and M_2, regardless of the dimension of the Hilbert space. In other words, we will show that two orthogonal spin observables are as incompatible as any binary observables can be.

To this end, let us note that two binary quantum observables are incompatible if and only if they enable a violation of the Bell-CHSH inequality [35]. We must therefore look at the Bell expression

$$\mathcal{B} = |\langle \mathsf{M}_1\mathsf{N}_1\rangle + \langle \mathsf{M}_1\mathsf{N}_2\rangle + \langle \mathsf{M}_2\mathsf{N}_1\rangle - \langle \mathsf{M}_2\mathsf{N}_2\rangle| \,.$$

Let us denote $\alpha = \langle \mathsf{M}_1\mathsf{N}_1\rangle + \langle \mathsf{M}_1\mathsf{N}_2\rangle$ and $\beta = \langle \mathsf{M}_2\mathsf{N}_1\rangle - \langle \mathsf{M}_2\mathsf{N}_2\rangle$. By [23, Theorem 1], there exist unit vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y_1}, \mathbf{y}_2 \in \mathbb{R}^4$ such that $\langle \mathsf{M}_j\mathsf{N}_k\rangle = \mathbf{x}_j \cdot \mathbf{y}_k$ for $j, k = 1, 2$; and conversely, given any quadruple of unit vectors there exist a corresponding set of binary observables and a bipartite state such that this equality holds. In particular, we have $\alpha = \mathbf{x}_1 \cdot (\mathbf{y}_1 + \mathbf{y}_2)$ and $\beta = \mathbf{x}_2 \cdot (\mathbf{y}_1 - \mathbf{y}_2)$ so that an application of the Cauchy-Schwarz inequality along with the parallelogram law yields

$$\begin{aligned} \alpha^2 + \beta^2 &\leq \|\mathbf{x}_1\|^2\|\mathbf{y}_1 + \mathbf{y}_2\|^2 + \|\mathbf{x}_2\|^2\|\mathbf{y}_1 - \mathbf{y}_2\|^2 \\ &= 2\|\mathbf{y}_1\|^2 + 2\|\mathbf{y}_2\|^2 = 4. \end{aligned}$$



(a) (b)

Figure 3.3: In (a) the grey area represents the possible values that α and β can obtain by varying the observables and the state in the Bell expression $\mathcal{B} = |\alpha + \beta|$. The solid lines represent the Tsirelson bound $\mathcal{B} = 2\sqrt{2}$ and the dashed lines represent the bound $\mathcal{B} = 2$. By considering only observables which are mixtures with the uniformly distributed trivial observable with fixed λ and μ, the area becomes smaller as depicted in (b), and a suitable choice of weights makes the violation of the Bell-CHSH inequality impossible.

By choosing the unit vectors appropriately we also see that any pair (α, β) satisfying this condition can be obtained.

If we now mix the observables M_j with the trivial observable $\mathsf{T}(\pm 1) = \frac{1}{2}I$ with some weights λ and μ we see that the pair (α, β) turns into $(\lambda\alpha, \mu\beta)$, thus changing the Bell expression from $|\alpha + \beta|$ to $|\lambda\alpha + \mu\beta|$. We must therefore determine those (λ, μ) for which $|\lambda\alpha + \mu\beta| \leq 2$ for all (α, β) satisfying $\alpha^2 + \beta^2 \leq 4$ (see Fig. 3.3). But the boundary curve for this region is obtained when the equations $(\alpha/\lambda)^2 + (\beta/\mu)^2 = 4$ and $\alpha + \beta = 2$ have at most one common solution. By inserting $\beta = 2 - \alpha$ into the first equation the problem reduces to determining when the discriminant is negative or zero, and one readily verifies that this is the case exactly when $\lambda^2 + \mu^2 \leq 1$.

In conclusion, given any pair of binary observables M_1 and M_2, and weights λ and μ with $\lambda^2 + \mu^2 \leq 1$, the mixtures $\lambda\mathsf{M}_1 + (1 - \lambda)\mathsf{T}$ and $\mu\mathsf{M}_2 + (1 - \mu)\mathsf{T}$ can not be used to violate the Bell-CSHS inequality and must therefore be jointly measurable. We note that in the case $\mu = \lambda$ the same result using a different technique has been obtained by Banik *et al.* [36].

Although Theorem 2 shows that quantum theory contains pairs of observables that are maximally incompatible, the strictly larger joint measurability region

when restricting to binary observables suggests that more fine grained quantifications of the global degree of incompatibility between observables might not rank quantum theory among the most extreme theories in this respect. The example below will show that when restricting to just binary observables, it is indeed possible for a theory to have the smallest possible joint measurability region. In that sense such a theory must be considered to embody a strictly greater degree of incompatibility than quantum theory.

Consider any probabilistic theory, which contains a state space isomorphic to a square, by which we mean the convex hull of four different points s_1, s_2, s_3, s_4 in $\mathbb{R}^2$ satisfying $s_1 + s_4 = s_2 + s_3$, for instance $s_1 = (0,0), s_2 = (0,1), s_3 = (1,0)$ and $s_4 = (1,1)$. We will show that there is a pair of binary observables which are maximally incompatible. Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be binary observables that pick out the right and top sides of the square respectively, i.e.

$$\begin{aligned} \mathfrak{M}_1(+|s_1) &= \mathfrak{M}_1(+|s_2) = 0, \\ \mathfrak{M}_1(+|s_3) &= \mathfrak{M}_1(+|s_4) = 1, \\ \mathfrak{M}_2(+|s_1) &= \mathfrak{M}_2(+|s_3) = 0, \\ \mathfrak{M}_2(+|s_2) &= \mathfrak{M}_2(+|s_4) = 1. \end{aligned} \tag{3.16}$$

Proposition 2. *For the binary observables $\mathfrak{M}_1$ and $\mathfrak{M}_2$ defined in* (3.16)*, $J(\mathfrak{M}_1, \mathfrak{M}_2) = \triangle$.*

Proof. Suppose that there exists a joint observable $\mathfrak{M}$ for $\lambda\mathfrak{M}_1 + (1-\lambda)\mathfrak{T}_1$ and $\mu\mathfrak{M}_2 + (1-\mu)\mathfrak{T}_2$ where $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are trivial observables. Let p_1 and p_2 be the probability distributions associated to $\mathfrak{T}_1$ and $\mathfrak{T}_2$ so that we have for any state ρ

$$\begin{aligned} \mathfrak{M}(+,+|\varrho) + \mathfrak{M}(+,-|\varrho) &= \lambda\mathfrak{M}_1(+|\varrho) + (1-\lambda)p_1(+) \\ \mathfrak{M}(-,+|\varrho) + \mathfrak{M}(-,-|\varrho) &= \lambda\mathfrak{M}_1(-|\varrho) + (1-\lambda)p_1(-) \\ \mathfrak{M}(+,+|\varrho) + \mathfrak{M}(-,+|\varrho) &= \mu\mathfrak{M}_2(+|\varrho) + (1-\mu)p_2(+) \\ \mathfrak{M}(-,-|\varrho) + \mathfrak{M}(+,-|\varrho) &= \mu\mathfrak{M}_2(-|\varrho) + (1-\mu)p_2(-) \end{aligned}$$

Any $\mathfrak{M}$ satisfying such marginal properies will be correctly normalised, but to be a valid observable, all the components of $\mathfrak{M}$ must take positive values on

the points s_i. In particular, we must have

$$\begin{aligned}\mathfrak{M}(+,-|s_2) =& (1-\lambda)p_1(+) - \mathfrak{M}(+,+|s_2) \geq 0,\\ \mathfrak{M}(-,+|s_3) =& (1-\mu)p_2(+) - \mathfrak{M}(+,+|s_3) \geq 0,\\ \mathfrak{M}(-,-|s_4) =& 1 + \mathfrak{M}(+,+|s_4) - \lambda - (1-\lambda)p_1(+)\\ & - \mu - (1-\mu)p_2(+) \geq 0.\end{aligned}$$

Rewriting the last of these inequalities and invoking the defining property on the s_i gives

$$\begin{aligned}\lambda + \mu \leq& 1 - (1-\lambda)p_1(+) - (1-\mu)p_2(+) + \mathfrak{M}(+,+|s_4)\\ \leq& -(1-\lambda)p_1(+) + \mathfrak{M}(+,+|s_2)\\ & -(1-\mu)p_2(+) + \mathfrak{M}(+,+|s_3)\\ & + 1 - \mathfrak{M}(+,+|s_1) \leq 1,\end{aligned}$$

where the final step comes about from invoking the positivity of $\mathfrak{M}$ on s_1. □

The result of Proposition 2 does not come as a surprise in light of the fact that the barrier to maximal incompatibility of binary quantum observables comes from the connection with a Bell-CHSH inequality. Indeed, square shaped state spaces have been used in a model of a probabilistic theory containing the PR boxes which violate such an inequality to its maximal possible value.

We note that the conclusion of Proposition 2 is not restricted to the square state space. Consider any state space containing a square whose vertices s_i are extreme points of the state space and whose boundary lines lie on the boundary of the state space; assume further that opposite sides of the square are contained in parallel hyperplanes that do not intersect with the interior of the state space. These two pairs of hyperplanes define effects whose values on the s_i satisfy Eq. (3.16). It follows that the proof of Proposition 2 can be adopted in such cases. Examples are given by state space of the following shapes: pyramid, double pyramid, cube, cylinder.

The fact that the restriction to just binary observables allows one to differentiate between probabilistic theories that both contain maximally incompatible observables suggests that a more fine grained global measure of the degree of

incompatibility is needed if the aim is to pick out a single theory as the one containing overall the most incompatible pairs of observables. For instance, for a given probabilistic theory PT we may define $J_{PT}^{(d)}$ to be the joint measurability region for all possible d-outcome observables in PT. Since increasing the number of outcomes of observables by simply adding outcomes that never occur does not change the properties of incompatibility, we immediately have $J_{PT}^{(d+1)} \subseteq J_{PT}^{(d)}$. By comparing the regions in different theories for different values of d we obtain a more fine grained way of comparing the degrees of incompatibility within the theories. It may even turn out that in this sense quantum theory embodies globally the least amount of incompatibility among the theories containing maximally incompatible observables. However, this is still an open question and a topic for future investigations.

Chapter 4

Steering, incompatibility, and Bell inequality violations in a class of probabilistic theories

4.1 Abstract

We show that connections between a degree of incompatibility of pairs of observables and the strength of violations of Bell's inequality found in recent investigations can be extended to a general class of probabilistic physical models. It turns out that the property of universal uniform steering is sufficient for the saturation of a generalised Tsirelson bound, corresponding to maximal violations of Bell's inequality. It is also found that a limited form of steering is still available and sufficient for such saturation in some state spaces where universal uniform steering is not given. The techniques developed here are applied to the class of regular polygon state spaces, strengthening known results. We also find indications that the link between incompatibility and Bell inequality violation may be more complex than originally envisaged.

4.2 Introduction

The Bell inequalities [24] provide constraints that certain families of joint probability distributions must satisfy to admit a common joint distribution. It is known that the satisfaction of a full set of Bell inequalities in a probabilistic system is equivalent to the existence of such a joint probability[25, 26].[1] It was observed subsequently that joint measurability (in the sense that there exist joint probabilities of the usual quantum mechanical form for every state) entails an operator form of Bell inequalities; therefore, the Bell inequalities are satisfied whenever the observables involved in an EPR-Bell type experiment are mutually commutative [30]. In the case of unsharp observables, commutativity is not required for joint measurability and the degree of unsharpness of the observables required for joint measurability can be determined; this value is more restrictive than is needed for violations of the Bell inequalities to be eliminated in the case of the singlet state [31, 21, 32, 33].

The connection between joint measurability and Bell inequalities – in the specific form of the CHSH inequalities [34], which apply to experiments involving runs of measurements of two pairs of dichotomic observables on a bipartite system – has been further elucidated in two interesting recent publications by Wolf *et al* [35] and Banik *et al* [36]. The former have shown that for any pair of incompatible dichotomic observables in a finite dimensional quantum system a violation of a CHSH inequality will be obtained. Hence, incompatibility is not only necessary but also sufficient for obtaining Bell inequality violations. Wolf *et al* [35] conclude that "if a hypothetical no-signaling theory is a refinement of quantum mechanics (but otherwise consistent with it), it cannot render possible the joint measurability of observables which are incompatible within quantum mechanics". With this result a tight link has been established between the availability of incompatible observables and the possibility of violating a CHSH inequality. It is natural to ask whether a quantitative connection can be found between a degree of incompatibility and the strength of these violations, and whether such a connection is specific to quantum mechanics or holds in a wider

[1] As observed by Pitowsky [27], Bell-type inequalities had already been formulated as early as 1854 by George Boole, who deduced them as conditions for the possibility of objective experience [28, 29].

class of probabilistic physical theories.

It is a well known fact that two incompatible quantum observables can be *approximately* measured together if some unsharpness in the measurement is allowed. A measure of the incompatibility of two observables can then be obtained by quantifying the degree of unsharpness required to obtain an approximate joint measurement. In the case of dichotomic observables this can be achieved by mixing each observable with a trivial observable (a POVM whose positive operators are multiples of the identity)[2], with relative weights λ, $1-\lambda$. The mixing weight determines the degree of unsharpness of the resulting smeared observable.

Banik *et al* have shown that the degree of incompatibility (they use the term complementarity) of two dichotomic observables, quantified by the largest smearing parameter, λ, for which the smeared versions are compatible, puts limitations on the maximum strength of CHSH inequality violations available in such a theory [36]. The Bell functional, $\mathbb{B}$, a generalisation of what is known as the Bell operator in the quantum case, then is bounded by the parameter λ_{opt} associated with the "most incompatible" pair of observables, so that $\mathbb{B} \leq 2/\lambda_{\text{opt}}$.

Here we study the connection between degrees of incompatibility and CHSH inequality violation in the context of general probabilistic physical theories by way of unifying the approaches of [35] and [36]. We will see that the degree of incompatibility used by Banik *et al* is closely linked with an unnamed parameter used in [35] to characterise the joint measurability of two dichotomic observables. Under an additional assumption on the physical theory, namely that it supports a sufficient degree of steering, the construction used to violate the CHSH inequality generalises. This gives a sufficient condition under which the maximal violation can be saturated. This result can be rephrased by saying that probabilistic theories can be classified according to the value of the *generalised Tsirelson bound*, defined as the maximum value of the Bell functional, and this bound can (under said assumptions) be realised by suitable maximally incompatible observables (see Theorem 1).

Finally we illustrate the link between incompatibility and Bell violation in the

[2] Such mixing procedures and their connection with goal of achieving joint measurability are investigated systematically in [62]).

class of regular polygon state spaces. It turns out that this connection appears to hold generally in the case of even-sided polygons but not, at least in the same form, for odd-sided cases.

4.3 General Probabilistic Models

We begin by presenting the basic elements of the standard framework of probabilistic models. The framework was introduced in the 1960s by researchers in quantum foundations who used it to investigate axiomatic derivations of the Hilbert space formalism of quantum mechanics from operational postulates. Due to the emphasis on the convex structure of the set of states and the use of operations to model state transformations, the approach was called *convex state approach* or *operational approach*. Some pioneering references are [1, 2, 3, 4, 5]. An overview of the literature and of relevant monographs can be obtained from [6] and [18]. Recently the approach has gained renewed interest from researchers in quantum information exploring the information theoretic foundations of quantum mechanics. Accessible recent introductions can be found in e.g. [8, 9, 38].

The set of states Ω of a general probabilistic model is taken to be a compact convex subset of a finite dimensional vector space V, where the convexity corresponds to the ability to define a preparation procedure as a probabilistic mixture of preparation procedures corresponding to other states. We write $A(\Omega)$ for the ordered linear space of affine functionals on Ω, with the (partial) ordering given pointwise: $f \geq 0$ if $f(\omega) \geq 0$ for all $\omega \in \Omega$. $A(\Omega)$ is also canonically an order unit space, with order unit u defined by $u(\omega) = 1$ for all states $\omega \in \Omega$. The (convex) set of effects on Ω is then taken to be the unit interval $[0, u]$ inside $A(\Omega)$, i.e.

$$\mathscr{E}(\Omega) = \{e \in A(\Omega) | 0 \leq e(\omega) \leq 1, \forall\, \omega \in \Omega\}. \tag{4.1}$$

A discrete observable O is then a function from an outcome set X into $\mathscr{E}(\Omega)$, that satisfies the normalisation condition $\sum_{x \in X} \mathsf{O}[x] = u$. The value (lying between 0 and 1) of $\mathsf{O}[x](\omega)$ denotes the probability of getting outcome x for a measurement of the observable O in state ω.

Under the assumption of tomographic locality [10], the state space of a compos-

ite system with local state spaces Ω_1 and Ω_2 naturally lives in the vector space $V_1 \otimes V_2$. We then write $\Omega = \Omega_1 \otimes \Omega_2 = (V_1 \otimes V_2)^1_+$, where the normalisation is given by the order unit $u_1 \otimes u_2 \in V_1^* \otimes V_2^*$, but in general the positive cone is not unique [39].

Although there is much choice in general for the ordering on $V_1 \otimes V_2$, there are two canonical choices, the *maximal* and *minimal*. As a minimal demand it is reasonable to expect $v_1 \otimes v_2 \geq 0$ whenever $v_1, v_2 \geq 0$, therefore we make the definition

$$(V_1 \otimes_{min} V_2)_+ = \left\{ \sum_{i,j} \lambda_{ij} v_1^{(i)} \otimes v_2^{(j)} \middle| \lambda_{ij} \in \mathbb{R}_+, v_k^{(i)} \in (V_k)_+ \right\}. \tag{4.2}$$

We can similarly make such demands on the order structure on $V_1^* \otimes V_2^*$ leading to the converse definition

$$(V_1 \otimes_{max} V_2)_+ = (V_1^* \otimes_{min} V_2^*)^*_+. \tag{4.3}$$

Any cone on $V_1 \otimes V_2$ which lies between the maximal and minimal cones is then admissible as a viable order structure. In general the tensor product chosen is an important part in defining a theory; the only time when there is no choice (since maximal and minimal are the same) is when the local state spaces are simplexes [39]. The case where both Ω_1 and Ω_2 are quantum state spaces provides a prime example of a nonminimal, nonmaximal order structure, namely the standard quantum mechanical tensor product. By definition $\Omega_1 \otimes_{min} \Omega_2$ contains only separable states, which form a proper subset of all bipartite states; by contrast, $\Omega_1 \otimes_{max} \Omega_2$ contains not only the usual quantum states, but also all normalised entanglement witnesses.

A bipartite state $\omega \in \Omega_1 \otimes \Omega_2$ can also be viewed as a way to prepare states in Ω_1, via the measurement of an observable on Ω_2. In this way, for each state ω, we can define the corresponding linear map $\hat{\omega} : V_2^* \to V_1$ by

$$a(\hat{\omega}(b)) = \omega(a, b), \quad a \in V_1^*,\ b \in V_2^*.$$

4.4 Fuzziness and joint measurability

Consider a system represented by a probabilistic model, whose state space is given by the convex set Ω. Any dichotomic (or two-outcome) observable O on

Ω is determined by an effect $e =: \mathsf{O}[+1] \in \mathscr{E}(\Omega)$, where for any $\omega \in \Omega$, the probability of getting the outcome labelled by '+1' in the state ω is given by $e(\omega)$, and similarly for the outcome '-1' associated with the complement effect $e' := u - e = \mathsf{O}[-1]$.

Two effects e and f are said to be jointly measurable if there exists $g \in A(\Omega)$ satisfying

$$\begin{aligned} 0 &\leq g, \\ g &\leq e, \\ g &\leq f, \\ e + f &\leq g + u, \end{aligned} \tag{4.4}$$

where u is the order unit on Ω. The existence of such a g is equivalent to the existence of a joint observable for the dichotomic observables corresponding to e and f. In fact, if the system of inequalities (4.4) is satisfied for some effect g then the set of effects $g_{++} := g,\ g_{+-} := e-g,\ g_{-+} := f-g,\ g_{--} := u-e-f+g$ defines an observable that comprises e, e' and f, f' as marginals, in the sense that $e = g_{++} + g_{+-}$, $f = g_{++} + g_{-+}$, etc.[3]

Given a two-outcome observable A determined by effect e, one can introduce a corresponding fuzzy observable $\mathsf{A}^{(\lambda)}$ as a smearing (or fuzzy version) of A, whose defining effect is given by

$$e^{(\lambda)} = \frac{1+\lambda}{2} e + \frac{1-\lambda}{2} e' = \lambda e + \frac{1-\lambda}{2} u, \tag{4.5}$$

with smearing parameter $\lambda \in [0, 1]$, and complement effect $e^{(\lambda)\prime} = e'^{(\lambda)}$.

Given any pair of two-outcome observables $\mathsf{A}_1, \mathsf{A}_2$, with corresponding effects e, f, we can use the parameter λ to give a measure of how incompatible they are. First we note that for $\lambda = \frac{1}{2}$, the choice of effect $g = \frac{1}{4}(e + f)$ generates a joint observable for e and f since it satisfies (4.4), as is readily verified. Thus the set of values of λ which make $e^{(\lambda)}$ and $f^{(\lambda)}$ jointly measurable contains $\frac{1}{2}$. Further, if $e^{(\lambda)}$ and $f^{(\lambda)}$ are jointly measurable, then for any $\lambda' \leq \lambda$ so are $e^{(\lambda')}$ and $f^{(\lambda')}$. Hence the set lies inside the interval $[0, \lambda_{e,f}]$, where we define $\lambda_{e,f}$ to

[3] For more detail on the notion of joint observable in probabilistic theories we refer the reader to [40], where further relevant references can be found.

be the solution to the cone-linear program

$$\begin{aligned} \text{maximise:} \quad & \lambda \\ \text{subject to:} \quad & g \leq e^{(\lambda)} \\ & g \leq f^{(\lambda)} \\ & 0 \leq g \\ & e^{(\lambda)} + f^{(\lambda)} - u \leq g. \end{aligned} \tag{4.6}$$

This measure of incompatibility of a pair of effects in turn leads to a measure of the degree of incompatibility of a given model by looking for the most incompatible pair:

$$\lambda_{\text{opt}} = \inf_{e,f \in \mathscr{E}(\Omega)} \lambda_{e,f}. \tag{4.7}$$

Following a path similar to [35], we can define a different parameter $t_{e.f}$, which we will see is closely linked with $\lambda_{e,f}$. For a given pair of effects e and f, we define $t_{e,f}$ to be the solution to the cone-linear program:

$$\begin{aligned} \text{minimise:} \quad & t \\ \text{subject to:} \quad & g \leq e + tu \\ & g \leq f + tu \\ & 0 \leq g \\ & e + f - u \leq g. \end{aligned} \tag{4.8}$$

As shown in [44], the optimal set for (4.8) is nonempty, so the minimum can be achieved, hence e and f are incompatible if and only if $t_{e,f} > 0$. Here we notice that the pair (λ, g) being feasible for the problem (4.6) is equivalent to the pair $\left(\frac{1-\lambda}{2\lambda}, \frac{g}{\lambda}\right)$ being feasible for the problem (4.8). Combining this with the fact that the function $\frac{1-\lambda}{2\lambda}$ is monotonically decreasing for $\lambda \in [0, 1]$ brings us to the promised link

$$t_{e,f} = \frac{1 - \lambda_{e,f}}{2\lambda_{e,f}}. \tag{4.9}$$

Examples

In a model of discrete classical probability theory we take the state space to be the set of all probability measures on some countable set X, i.e.

$$\Omega = \left\{ (\omega_x)_{x\in X} \,\middle|\, \omega_x \geq 0 \ \forall x \in X,\ \sum_x \omega_x = 1 \right\}. \tag{4.10}$$

A functional e on Ω with action $e(\omega) = \sum_x e_x \omega_x$ is easily seen to be positive iff $e_x \geq 0$ for all $x \in X$, and the order unit satisfies $u_x = 1$ for all $x \in X$.

Suppose we now have two effects $e, f \in \mathscr{E}(\Omega)$. Taking g to have components $g_x = \min\{e_x, f_x\}$, then since positivity is determined componentwise the inequalities (4.4) are immediately satisfied, and hence e and f are jointly measurable. Since this holds for arbitrary e and f in this case we have $\lambda_{opt} = 1$.

As shown in [36], in any finite dimensional Hilbert space the value of the joint measurability parameter for a pair of dichotomic observables is $\lambda_{\text{opt}} = 1/\sqrt{2}$.

A simple non-classical, non-quantum example is that of the *squit*. The two dimensional state space is given by a square, denoted $\square$; it contains all points $(x, y, 1)$ with $-1 \leq x + y \leq 1$, $-1 \leq x - y \leq 1$, and takes the shape of a square. As we will see, the squit leads to maximally incompatible effects in the sense that it leads to the smallest possible value of λ_{opt}.

Firstly we note that for any probabilistic model $\lambda = \frac{1}{2}$ provides a lower bound for λ_{opt}, since $e^{(\frac{1}{2})} = \frac{1}{2}e + \frac{1}{4}u$ and $f^{(\frac{1}{2})} = \frac{1}{2}f + \frac{1}{4}u$ are always jointly measurable. This can be seen explicitly by setting $g = \frac{1}{4}e + \frac{1}{4}f$, then the corresponding equations (4.4) are satisfied.

As a convenient parametrisation we can write a generic affine functional $g \in A(\square)$ as a vector $g = (a, b, c)$, with action given by the canonical inner product scaled by a factor of $\frac{1}{2}$. In this case the order unit is given by $u = (0, 0, 2)$. Since the positivity of a functional g on a compact convex set is equivalent to positivity on its extreme points, we can determine the structure of the set of effects by demanding that its elements g take values between 0 and 1 on the extreme points of the set of states. In the case of the squit, $\mathscr{E}(\square)$ is a convex polytope with defining inequalities given by

$$u \geq g \geq 0 \iff \begin{cases} 2 \geq c + a \geq 0, & 2 \geq c + b \geq 0, \\ 2 \geq c - a \geq 0, & 2 \geq c - b \geq 0. \end{cases} \tag{4.11}$$

We note the extreme points: $(0,0,2) = u$, $(0,0,0)$, $(1,1,1)$, $(1,-1,1)$, $(-1,1,1)$, $(-1,-1,1)$.

In an attempt to find the lowest possible value of $\lambda_{e,f}$ we consider the case of the two orthogonal extremal effects $e = (1,1,1)$ and $f = (1,-1,1)$. In order for $e^{(\lambda)}$ and $f^{(\lambda)}$ to be jointly measurable we need to be able to find a g that satisfies all the inequalities in (4.4). This entails, in particular:

$$\begin{aligned} g - e^{(\lambda)} - f^{(\lambda)} + u = (a - 2\lambda, b, c) \geq 0, \\ \text{giving} \quad 2\lambda \leq a + c; \\ e^{(\lambda)} - g = (\lambda - a, \lambda - b, 1 - c) \geq 0, \\ \text{giving} \quad \lambda \leq 1 + a - c; \\ f^{(\lambda)} - g = (\lambda - a, -\lambda - b, 1 - c) \geq 0, \\ \text{giving} \quad \lambda \leq 1 - a - c; \\ g = (a, b, c) \geq 0, \\ \text{giving} \quad a \leq c. \end{aligned}$$

Combining these inequalities leads to $4\lambda \leq 2 + a - c \leq 2$, so for this choice of e and f we must have $\lambda_{e,f} \leq \frac{1}{2}$. Given that $\frac{1}{2}$ is the lowest possible value, we conclude that in the case of the squit $\lambda_{opt} = \frac{1}{2}$.

4.5 Steering and saturation of the generalised Tsirelson bound

In order to give conditions on a generalised probabilistic model under which the bound on CHSH violations given in [36] can be achieved we need to introduce the notion of steering, as given in [41].

Given two systems A and B, with state spaces Ω_A and Ω_B respectively, for any bipartite state $\omega \in \Omega_A \otimes \Omega_B$ we can define its A *marginal*, living in Ω_A in an analogue to the quantum mechanical partial trace:

$$\omega^A = \hat{\omega}(u_B), \tag{4.12}$$

where u_B is the order unit on B, with a similar definition for ω^B.

Following this we say that a state $\omega \in \Omega_A \otimes \Omega_B$ is *steering* for its A marginal if for any collection of sub-normalised states that form a decomposition of that marginal, i.e., $\{\alpha_1, ..., \alpha_n | \sum_i \alpha_i = \omega^A, 0 \leq u_A(\alpha_i) \leq 1\}$, there exists an observable $\{e_1, ..., e_n\} \subset \mathscr{E}(\Omega_B)$ with $\alpha_i = \hat{\omega}(e_i)$.

It was observed by Schrödinger that this property holds in quantum mechanics for all pure bipartite states [42], originally coining the term steering, which we generalise now, following [41]: A general probabilistic model of a system A with state space Ω_A supports *uniform universal steering* if there is another system B with state space Ω_B, such that for any $\alpha \in \Omega_A$, there is a state $\omega_\alpha \in \Omega_A \otimes \Omega_B$, with $\omega_\alpha^A = \alpha$ that is steering for its A marginal, and supports *universal self-steering* if the above is satisfied with $B = A$. The existence of steering in this manner is similar to the idea of purification to be found, for example, in [43]. Indeed any purification of a state will be steering for its marginals; however steering states being pure is not required here.

The magnitude of maximal CHSH violations is quantified in quantum mechanics by the norm of the *Bell operator*. We take $\mathsf{A}_1, \mathsf{A}_2, \mathsf{B}_1$ and B_2 to be ± 1-valued observables, and define following [36]

$$\mathbb{B} := \langle A_1 B_1 + A_1 B_2 + A_2 B_1 - A_2 B_2 \rangle_\omega,$$

where $A_1 := \mathsf{A}_1[+1] - \mathsf{A}_1[-1]$, etc., and $\langle X \rangle_\omega := X(\omega)$ for any affine functional X. We will call the map $\omega \mapsto \mathbb{B}$ the *Bell functional* and refer to $\sup_\omega \mathbb{B}$ as the (generalised) *Tsirelson bound*.

In order to see where steering enters the picture, we follow [36] to get a simple bound on the norm of $\mathbb{B}$. In order to do this we consider what effect smearing the observables of one party has by defining

$$\mathbb{B}^{(\lambda)} = \langle A_1^{(\lambda)} B_1 + A_1^{(\lambda)} B_2 + A_2^{(\lambda)} B_1 - A_2^{(\lambda)} B_2 \rangle, \tag{4.13}$$

where $A_1^{(\lambda)} = \mathsf{A}_1^{(\lambda)}[+1] - \mathsf{A}_1^{(\lambda)}[-1]$etc., with the smearing of the effects as defined as in (4.5). Due to the fact that the choice of observable that is mixed to form the smearing is an unbiased trivial observable, the resulting expectation scales with the smearing parameter:

$$A_1^{(\lambda)} = \lambda \mathsf{A}_1[+1] + \frac{1-\lambda}{2} u - \lambda \mathsf{A}_1[-1] - \frac{1-\lambda}{2} u = \lambda A_1. \tag{4.14}$$

Now since the Bell functional is bilinear, and the same smearing parameter is being used on all functionals on the first system, the linear scaling carries over and we get $\mathbb{B}^{(\lambda)} = \lambda\mathbb{B}$.

As shown in the previous chapter, there always exist jointly measurable fuzzy versions of any pair of observables, so long as the value of the smearing parameter is small enough. Now if we take any λ such that $A_1^{(\lambda)}$ and $A_2^{(\lambda)}$ are jointly measurable, then we know that the corresponding Bell functional satisfies the usual Bell inequality, and thus its value is bounded by $\mathbb{B}^{(\lambda)} \leq 2$. Consequently, each such value of λ gives a bound on on the Bell functional of $\mathbb{B} \leq \frac{2}{\lambda}$, and in order to obtain the lowest such upper bound we take the largest smearing parameter which still results in joint measurability, to get

$$\mathbb{B} \leq \frac{2}{\lambda_{\mathsf{A}_1[+1],\mathsf{A}_2[+1]}}. \tag{4.15}$$

Since every probabilistic model contains observables which are jointly measurable with no smearing, and thus satisfying the usual Bell inequality, knowing the above bound for a single pair of observables will not necessarily yield information about the structure of the system itself. A more general bound however can be written down by simply taking the most incompatible pair of observables:

$$\mathbb{B} \leq \frac{2}{\lambda_{\text{opt}}}. \tag{4.16}$$

Theorem 3. *In any probabilistic model of a systen A that supports uniform universal steering, the Tsirelson bound is given by the tight inequality that can be saturated:*

$$\mathbb{B} \leq \frac{2}{\lambda_{\text{opt}}}, \tag{4.17}$$

with λ_{opt} defined in Eq. (4.7).

Proof. Suppose we have a model of a system A that supports uniform universal steering, and that we have two effects $e, f \in \mathscr{E}(\Omega_A)$. The parameter introduced earlier, $t_{e,f}$ can now also be calculated from the program dual to (4.8), which

can be given as [44]

$$\begin{aligned} \text{maximise:} \quad & \mu_3(e+f-u_A)-\mu_1(e)-\mu_2(f) \\ \text{subject to:} \quad & (\mu_1+\mu_2)(u_A)=1 \\ & \mu_1+\mu_2=\mu_3+\mu_4 \\ & 0\leq \mu_1,\mu_2,\mu_3,\mu_4 \end{aligned} \tag{4.18}$$

with the $\mu_i \in A(\Omega_A)^*$.

Writing $\mu_1+\mu_2=\rho$, for the μ_i that achieve the optimal value for (4.18), we find that $\rho \geq 0$ and $u_A(\rho)=1$, so $\rho \in \Omega_A$. By the assumption of uniform universal steering therefore we can find a state $\omega \in \Omega_A \otimes \Omega_B$ with $\omega^A = \hat{\omega}(u_B)=\rho$; moreover, in $\{\mu_1,\mu_2\}$ and $\{\mu_3,\mu_4\}$ we have two different decompositions of ρ, and we can thus find effects $\tilde{e},\tilde{f} \in \mathscr{E}(\Omega_B)$ satisfying

$$\hat{\omega}(\tilde{e})=\mu_1, \qquad \hat{\omega}(\tilde{f})=\mu_3. \tag{4.19}$$

To achieve the maximum CHSH violations we take $\mathsf{A}_1,\mathsf{A}_2,\mathsf{B}_1$ and B_2 to be ± 1-valued observables defined by effects $f', e, \tilde{e}'$ and $\tilde{f}'$ respectively; we then have

$$\begin{aligned} A_1 &= u_A - 2f, \quad & B_1 &= u_B - 2\tilde{e}, \\ A_2 &= 2e - u_A, \quad & B_2 &= u_B - 2\tilde{f}. \end{aligned} \tag{4.20}$$

The value of the Bell functional can now be evaluated:

$$\begin{aligned} \mathbb{B} &= \omega(u_A-2f, 2u_B-2\tilde{e}-2\tilde{f}) + \omega(2e-u_A, 2\tilde{f}-2\tilde{e}) \\ &= 2\hat{\omega}(u_B)(u_A-2f) \\ &\qquad + 4\hat{\omega}(\tilde{e})(f-e) + 4\hat{\omega}(\tilde{f})(f+e-u_A) \\ &= 2+4[(\mu_1+\mu_2)(-f) \\ &\qquad + \mu_1(f)-\mu_1(e)+\mu_3(f+e-u_A)] \\ &= 2+4[\mu_3(e+f-u_A)-\mu_1(e)-\mu_2(f)] \\ &= 2(2t_{e,f}+1) = \frac{2}{\lambda_{e,f}}, \end{aligned}$$

thus saturating the generalised Tsirelson bound as claimed. □

Not every probabilistic model may possess the property of supporting uniform universal steering, and although it is a sufficient condition to obtain the conclusion of the above theorem, as the following example will show, it is not a

necessary one. Indeed a model of 'boxworld', which contains Popescu-Rohrlich (PR) box states exhibiting the maximum possible CHSH violations, uses local state spaces that are the squits introduced earlier, and composition is given by the maximal tensor product. Despite the saturation of the generalised Tsirelson bound, such a state space does not admit uniform universal steering.

To see this, we consider a bipartite state $\omega \in \square \otimes_{max} \square$ with the corresponding map $\hat{\omega}$. Note that from the definition of ω being a state, $\hat{\omega}$ will automatically be a positive map sending V_+^* into V_+. Now suppose ω is steering for its marginal ρ, i.e. $\hat{\omega}(u) = \rho$, and choose a decomposition of ρ into pure states: $\rho = \sum_i \alpha_i$. Since the subnormalised states in the decomposition are pure, and $\hat{\omega}$ is positive, the inverse images $\hat{\omega}^{-1}(\alpha_i)$ must lie on extremal rays of the cone V_+^*. Consider the extremal ray effect $e = (1, 1, 1)$ with its complement $e' = (-1, -1, 1)$ (which is again extremal). With appropriate labelling of the α_i we can then write $\alpha_1 = \hat{\omega}(e)$ and $\alpha_2 = \hat{\omega}(e')$; however since we have $e + e' = u$,

$$\alpha_1 + \alpha_2 = \hat{\omega}(e + e') = \hat{\omega}(u) = \rho,$$

and hence ρ can be written as a mixture of just two pure states. Since there are many points in a square that can only be written as a convex combination of a minimum of three extreme points, we conclude that such a model of 'boxworld' does not support universal uniform steering.

Remark 1. It is interesting to note that there is another set of conditions sufficient to obtain the conclusion of the above theorem. We say that a positive cone V_+ is *homogeneous* if the space of order automorphisms of V acts transitively on the interior of V_+, and (weakly) *self dual* if there exists a linear map $\eta : V \to V^*$ that is an isomorphism of ordered linear spaces i.e. $\eta(V_+) = V_+^*$. It is known that homogeneity follows from uniform universal steering. Conversely, if the positive cone V_+ generated by the state space Ω of the probabilistic model of a system A is homogeneous and weakly self-dual, then uniform universal self-steering follows if the maximal tensor product is adopted. Hence the conditions of Theorem 1 are fulfilled [41] and the Tsirelson bound in the inequality $\mathbb{B} \leq 2/\lambda_{\text{opt}}$ can be saturated.

In the quantum probabilistic model, the tensor product is not maximal but still uniform universal steering holds. The classical model (trivially) satisfies

the conditions of weak self-duality and homogeneity, and the tensor product is maximal. The squit is weakly self-dual but does not satisfy uniform universal steering, so that homogeneity fails; but it allows enough self-steering so that the maximal Bell-Tsirelson bound of 4 can be realised.

4.6 Generalised Tsirelson bounds for polygon state spaces

Work in [45] suggests that there is a spectrum of values for the generalised Tsirelson bound in the case of 2-dimensional polygon state spaces (given as the convex hulls of regular polygons). It is shown there that for a system composed of two identical polygon state spaces with an odd number of vertices, the maximally entangled state does not lead to a violation of the standard Tsirelson bound of $2\sqrt{2}$, whereas in the case of an even number of vertices this bound can be exceeded. This suggests that among the class of polygon state spaces, the generalised Tsirelson bound can be either smaller or greater than the standard Tsirelson bound.

Remark 2. *We note that of the polygon state spaces, the only cases in which homogeneity holds are the $n = 3$ triangle, and the $n \to \infty$ circle. Hence in general uniform universal steering is not available, however it may still be possible to saturate the generalised Tsirelson bound in some cases, but in others this may not be possible.*

As shown in [45], in the case of 'boxworld', where each local state space is a square, the maximally entangled state is a PR box; it takes the maximum possible value for the Bell functional of 4. This agrees with the result that the squit does indeed lead to the maximum amount of incompatibility, and shows that in this case the generalised Tsirelson bound can be saturated. We have been able to show that this conclusion holds also in regular polygon state spaces where the number of vertices is a multiple of 8. We expect this result to extend to all even-sided cases. This strengthens the expectation, expressed in [45], that in these cases the Tsirelson bound is saturated with the maximally entangled state.

Moving to the $n = 5$ case makes things a lot more interesting however. To see this we follow the notation in [45] and define the family of state spaces Ω_n to be the convex hull of the points

$$\omega_i = \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}, \qquad i = 1, ..., n$$

with $r_n = \sqrt{\sec(\frac{\pi}{n})}$.

The qualitative difference between the state spaces of odd and even sided polygons first appears in the structure of the set of effects. For the case of even n, along with 0 and u, there are n extremal effects:

$$e_i = \frac{1}{2} \begin{pmatrix} r_n \cos(\frac{(2i-1)\pi}{n}) \\ r_n \sin(\frac{(2i-1)\pi}{n}) \\ 1 \end{pmatrix}, \qquad i = 1, ..., n$$

and in this case all the e_i lie on extremal rays of the cone V_+^*. This important fact occurs since for each of the e_i we can find another effect e_j, also extremal, which is its complement, i.e. $e_j = e_i' = u - e_i$, namely for $j = i + \frac{n}{2} \bmod n$. For the case of odd n, a seemingly similar expression arises for the ray extremal effects:

$$e_i = \frac{1}{1 + r_n^2} \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix}, \qquad i = 1, ..., n$$

On this occasion however, the complements of the e_i are given by

$$e_i' = u - e_i = \frac{1}{1 + r_n^2} \begin{pmatrix} -r_n \cos(\frac{2\pi i}{n}) \\ -r_n \sin(\frac{2\pi i}{n}) \\ r_n^2 \end{pmatrix}, \qquad i = 1, ..., n$$

which do not coincide with the e_i, and thus there are $2n$ non-trivial extreme points of $\mathscr{E}(\Omega_n)$.

Now we can pose the question of what the value is for λ_{opt} when the state space is Ω_5, and whether is it possible to achieve the corresponding Bell value $\mathbb{B} = 2/\lambda_{opt}$. Since each extreme two valued observable is determined by a ray

effect, the largest value of incompatibility will come from one of the possible pairs of the e_i. However due to the symmetry of the state space, the affine transformation of rotating by $\pi/5$ serves only to cyclically permute the indices of the e_i modulo 5. This means that there are only two possible values of λ_{e_i,e_j}, those for nearest neighbors, and those for next-nearest neighbors. Calculation shows that these values are, for example

$$\lambda_{e_1,e_2} = \frac{3+2\sqrt{5}}{11} \approx 0.67928,$$

$$\lambda_{e_1,e_3} = \frac{8+3\sqrt{5}}{19} \approx 0.77416.$$

hence the value of λ_{opt} for the pentagon is $\frac{3+2\sqrt{5}}{11}$. From (4.16) this gives the bound on the Bell functional as $\mathbb{B} \leq 4\sqrt{5}-6$, however unlike in the case of the tensor product of two squits, the maximally entangled state between two pentagonal state spaces does not saturate the corresponding bound; instead we get a value of $\mathbb{B} = \frac{6}{\sqrt{5}}$, strictly below that coming from the level of incompatibility on one state space. This fact suggests that either the chosen way of evaluating the level of incompatibility in a system used does not capture everything, or that there is some structural obstruction that prevents such a link holding, that does not exist in other cases. Here we present some evidence towards the former.

In order to improve the measure of incompatibility used, we wish to modify the program used in eqn. (4.6). To do this we relax the method of smearing used, still mixing in multiples of the order unit, corresponding to trivial observables; but we now allow them to be possibly biased as follows:

$$e^{(\lambda,p)} = \lambda e + p(1-\lambda)u. \tag{4.21}$$

This definition encompasses the old, with $e^{(\lambda)} = e^{(\lambda,\frac{1}{2})}$.

The updated measure of incompatibility of a pair of effects e and f, which we

denote $\bar{\lambda}_{e,f}$, is now given by the optimal value of the optimisation program

$$\begin{aligned} \text{maximise:} \quad & \lambda \\ \text{subject to:} \quad & g \leq e^{(\lambda,p)} \\ & g \leq f^{(\lambda,q)} \\ & 0 \leq g \\ e^{(\lambda,p)} + f^{(\lambda,q)} - u & \leq g \\ & 0 \leq p, q \leq 1. \end{aligned} \tag{4.22}$$

Solving this updated problem in the case of the pentagon again gives the optimal value on e.g. e_1 and e_2, with

$$\bar{\lambda}_{opt} = \frac{5+\sqrt{5}}{10} \approx 0.72361,$$

which occurs for the values $p = q = 1$.

This is indeed a different value from earlier, but still we have that $\frac{2}{\bar{\lambda}_{opt}} \neq \frac{6}{\sqrt{5}}$, however in this case, the unbiased nature of the observables mixed in means such a simple link is no longer expected, and indeed we can see that there is a link to the Bell value on the maximally entangled state as follows. As in the previous, we can define a smeared version of the Bell functional, where the smearing is all done on the functionals of one party:

$$\mathbb{B}^{(\lambda,1)} = \langle A_1^{(\lambda,1)} B_1 + A_1^{(\lambda,1)} B_2 + A_2^{(\lambda,1)} B_1 - A_2^{(\lambda,1)} B_2 \rangle, \tag{4.23}$$

but now instead of having the linear scaling in λ, we gain an extra expectation term $\mathbb{B}^{(\lambda,1)} = \lambda\mathbb{B} + 2(1-\lambda)\langle B_1 \rangle$, and again under the assumption that λ is small enough to ensure joint measurability, and then taking the largest such value we can write the inequality

$$\mathbb{B} \leq \frac{2\left[1 - (1 - \bar{\lambda}_{opt})\langle B_1 \rangle\right]}{\bar{\lambda}_{opt}}. \tag{4.24}$$

The link to the maximally entangled state on two pentagons now comes from noting that the expectation of any observable B_1 defined by an extreme effect on the maximally entangled state is $\langle B_1 \rangle = \frac{5-2\sqrt{5}}{5}$. This means that if evaluated in the maximally entangled state, the inequality in (4.24), for the value of $\bar{\lambda}_{opt}$ given above, is indeed saturated.

4.7 Conclusion

By combining and developing ideas from the works of Wolf *et al* [35] and Banik *et al* [36], we have shown that probabilistic models can be classified according to their associated value of the generalised Tsirelson bound, which specifies the maximum possible violation of CHSH inequalities. We have given conditions (defined and studied in [41]), that probabilistic models may or may not satisfy, under which the maximal CHSH violations are attained for appropriate choices of maximally incompatible dichotomic observables. Here the degree of the incompatibility of two observables is defined by the minimum amount of smearing of these observables necessary to turn them into jointly measurable observables.

The authors of [35] concluded that observables which are incompatible in quantum mechanics remain incompatible in any probabilistic model that serves as an extension of quantum mechanics. Here we have shown that this conclusion applies to extensions of any probabilistic model which allows for sufficient steering.

As an illustration of the general results we have considered the squit system which underlies the PR box model, and have identified the pair of maximally incompatible extremal effects of the squit that give rise to the saturation of the largest possible value (i.e., 4) of the Tsirelson bound. In addition, we have obtained partial confirmation of the conjectured maximality of the Bell functional if evaluated on the maximally entangled state in the class of regular polygon state spaces considered in [45].

In the case of the pentagon state space we discovered that the connection between incompatibility and Bell violation is not always of the simple form envisaged originally and used through most of this paper; this suggests that the definitive universal expression of this connection remains yet to be found.

The methods used here are taken from amongst some of the standard tools of quantum measurement and information theory used in [35] and [36], and we have shown that they apply equally well in a wide class of probabilistic models. This insight may prove valuable in future investigations into the characterisation of quantum mechanics among all probabilistic models.

Chapter 5

Direct tests of measurement uncertainty relations: what it takes

5.1 Background

Recently there have been claims of experimental violations of Heisenberg's error-disturbance relation [Rozema *et al*, PRL 109, 200404 (2012), Erhard *et al*, Nature Phys. 8, 185 (2012)]. These experiments may well be considered the first tests of measurement uncertainty relations ever attempted; they do confirm inequalities due to Ozawa and Branciard, which constitute a trade-off for certain *state-dependent* measures of error and disturbance. However, the reliability of these measures was shown to be limited to a restricted class of measurements, which casts doubts on the universality of the Ozawa-Branciard inequalities. This raises the general question of what it takes for an experimental investigation to constitute a *direct test* of measurement uncertainty relations. Here we argue that the state-dependent error and disturbance quantities in question are not in general amenable to direct testing as their values cannot always be compared with an appropriate experimental error analysis. Such direct comparisons are shown to be possible, but only for the said restricted class of measurements. The existing qubit experiments are found to be best understood as tests of state-independent measurement uncertainty relations. We conclude that directly testable and universal state-specific measurement uncertainty relations must be based on alternative state-dependent error and

disturbance measures.

5.2 Introduction

Heisenberg's uncertainty principle is a cornerstone in our understanding of quantum mechanics. It is therefore remarkable that the important quest for tests of measurement uncertainty relations has only been addressed very recently, when theoretical and experimental work surprisingly led to claims of a violation of Heisenberg's error-disturbance relation (e.g. [46, 47, 48, 49, 50, 51, 52]). The experiments confirm an inequality due to Ozawa and a strengthened form of it due to Branciard, which are trade-offs between quantities $\varepsilon(A)$ and $\eta(B)$, taken to quantify the state-specific error of an approximate measurement of an observable A and the ensuing state specific disturbance imparted on an observable B.

These violation claims have been contrasted with proofs of Heisenberg-type error-disturbance relations for position and momentum [53] and for qubit observables [54], which are based on alternative, state-independent "worst-case" measures $\Delta(A), \Delta(B)$. A comparison and reconciliation of both approaches were given in [55].

A detailed analysis of the quantities ε, η carried out in [55] has shown that these quantities can be unreliable as indicators of error and disturbance if applied outside a limited range of applicability (which will be specified below) [56]. This casts doubts on the *universality* of Ozawa's and Branciard's inequalities as error-disturbance trade-offs. In addition, it was noted in this study that the experimental procedures proposed and used so far to determine the values of ε, η are rather indirect; these are the so-called *three-state* and *weak measurement* methods. This raises the general question of what constitutes a *direct test* of a *universal* measurement uncertainty relation. A necessary requirement for a direct test is that the values of the measures of error and disturbance used can be estimated by way of an *error analysis* based on the data of the experiment at hand. We argue that this condition is not met unconditionally by the existing experiments, and propose alternative procedures that do meet this requirement, demonstrating that they can be fulfilled in principle.

However, these procedures are feasible only within the specific class of measurements in which the quantities ε, η are in fact reliable as error and disturbance measures. Consequently, universal, directly testable, state-dependent error-disturbance relations must be formulated in terms of measures other than ε, η (for an interesting recent proposal, see [57]). We find that the existing experiments are appropriately interpreted as direct confirmations of Heisenberg-type measurement uncertainty relations for *state-independent*, worst-case error and disturbance measures.

5.3 Error and disturbance and their determination

We consider the following generic scenario. An observable, represented by a selfadjoint operator A, is to be measured approximately by a scheme actually measuring some general observable, described by the positive operator valued measure (POVM) C. The measurement will generally disturb any other observable, represented by operator B, and distort it into some observable (POVM) D. It is known that a measurement of C followed by an accurate measurement of B constitutes a joint measurement of C and the "distorted" observable D.

An *error analysis* based on *state-specific* measures would reveal the error of the A measurement in the difference between the A and C distributions, for example in a measure of root-mean-square (rms) deviation of the measured values of these quantities; likewise the disturbance of B is manifest in the difference between the B and D distributions or values. This shows that disturbance is itself a form of approximation error in a joint measurement, and error-disturbance relations are a special form of joint measurement error relations [46, 55].

Alternatively, one may be interested in specifying errors as figures of merit for the devices used, that is, applicable to all states. In this case, an error analysis simply consists of using the relevant statistics to assess the *distance* between the approximating and target observables.

The different ways of measuring ε, η suggest themselves from the different expressions of these quantities. First, ε, η can be written in terms of the first and

second moment operators of C, D (where, say, $\mathsf{C}[x^n] = \int x^n \mathsf{C}(dx)$):

$$\varepsilon(A)^2 = \mathrm{tr}\left[\rho A^2\right] + \mathrm{tr}\left[\rho \mathsf{C}[x^2]\right] - 2\mathrm{Re}\,\mathrm{tr}\left[\rho A \mathsf{C}[x]\right], \tag{5.1}$$

$$\eta(B)^2 = \mathrm{tr}\left[\rho B^2\right] + \mathrm{tr}\left[\rho \mathsf{D}[x^2]\right] - 2\mathrm{Re}\,\mathrm{tr}\left[\rho B \mathsf{D}[x]\right]. \tag{5.2}$$

Here ρ is a general density operator of the object. This formulation leads to the *three-state method* [46], based on the following rewriting of, say, $\varepsilon(A)$:

$$\begin{aligned}\varepsilon_{\mathrm{NO}}(A, \mathcal{M}, \rho)^2 = \mathrm{tr}\left[\rho A^2\right] + \mathrm{tr}\left[\rho \mathsf{C}[x^2]\right] \\ + \mathrm{tr}\left[\rho \mathsf{C}[x]\right] + \mathrm{tr}\left[\rho_1 \mathsf{C}[x]\right] - \mathrm{tr}\left[\rho_2 \mathsf{C}[x]\right];\end{aligned} \tag{5.3}$$

here the (non-normalized) states ρ_1, ρ_2 are given by $\rho_1 = A\rho A$, $\rho_2 = (A + \mathbf{1})\rho(A + \mathbf{1})$. While now the quantity $\varepsilon(A)$ is manifestly determined by the statistics of A and C, one can no longer claim it to be state-specific. This is because $\varepsilon(A)^2$ is a combination of numbers that are obtained from measurements performed on three distinct states ρ, ρ_1, ρ_2. This method has been applied in the Vienna experiment [48].

Interestingly, $\varepsilon(A)$ and $\eta(B)$ happen to become entirely state-independent for optimal approximate joint measurements of qubit observables as they are investigated in the experiments cited above, and it has been shown in [54] that they then relate closely to the alternative measures $\Delta(A), \Delta(B)$ of worst-case error and disturbance used in [54]. In this case a probing of three states turns out sufficient to obtain the values of $\varepsilon(A), \eta(B)$ maximized over all states. This explains why qubit experiments utilizing the three-state method can serve as *direct test* of any trade-off for these *maximized* error and disturbance quantities: the error analysis consists here of determining the distance between the *observables* A and C, or B and D, rather than facing the generally impossible task of determining the rms deviation of values of incompatible observables in any particular state.

Another rewriting of (5.1), (5.2) is as follows:

$$\varepsilon(A)^2 = \iint (x - y)^2 \mathrm{Re}\,\mathrm{tr}\left[\mathsf{A}(dx)\mathsf{C}(dy)\right], \tag{5.4}$$

$$\eta(B)^2 = \iint (x - y)^2 \mathrm{Re}\,\mathrm{tr}\left[\mathsf{B}(dx)\mathsf{D}(dy)\right], \tag{5.5}$$

where A, B denote the spectral measures of A, B.

For an observable B with discrete values b_k and spectral projections B_k and a distorted B observable D with the same values and positive operators D_ℓ (such that $\sum_k D_k = \mathbf{1}$), the second equation becomes

$$\eta(B)^2 = \sum_{k\ell}(b_k - b_\ell)^2 \mathrm{Re}\,\mathrm{tr}\,[\rho B_k D_\ell]\,. \tag{5.6}$$

The map $(b_k, b_\ell) \mapsto \mathrm{Re}\,\mathrm{tr}\,[\rho B_k D_\ell]$ is a probability distribution if the operators B_k and D_ℓ commute [58]. But this will not be the case in general, and then the "rms" interpretation of η becomes problematical [55].

To overcome this problem, Lund and Wiseman [59] proposed an indirect determination of $\varepsilon(A), \eta(B)$ via reconstruction formulas that render them as *weak values*. Implementations of this method were realized for η in the case of qubit observables in [47, 51, 52]. The following detailed discussion will lead us to envisage our alternative, direct method for obtaining error and disturbance via error analysis.

5.4 Weak measurement vs strong measurement method

Lund and Wiseman proposed to consider the quantities
$\mathrm{Re}\,\mathrm{tr}\,[\rho B_k D_\ell] \equiv P_{WV}(b_k, b_\ell)$ as "weak valued probabilities", which led them to rewrite (5.6) formally as

$$\eta(B)^2 = \sum_{\delta b}(\delta b)^2 P_{WV}(\delta b), \tag{5.7}$$

with $P_{WV}(\delta b) = \sum_{k,\ell:b_\ell = b_k + \delta b} P_{WV}(b_k, b_\ell)$.

This equation is then taken *at face value* in [59] and also [47], *as if* it had an immediate operational meaning. Yet, however suggestive the form of the above expression for $\eta(B)$ may be, since in general the so-called weak-valued probabilities are not probabilities at all, there is in general no justification for calling (5.7) a rms deviation.

As will be seen below, the example proposed by Lund and Wiseman, a model experimental determination of $\eta(B)$ for qubit observables, does fall into the class of schemes where the commutativity of B with the distorted observable D is

given. In this model (Fig. 5.1) an initial approximate (or weak) measurement of the qubit observable $B = X$ [60] is done, with strength $2\gamma^2 - 1$. This is then followed by an approximate measurement of Z on the resulting state, with strength $\cos 2\theta$. Finally there is an accurate X measurement (denoted X_f). The initial and final X measurements are intended to provide information about the disturbance of X by the approximate Z measurement. The probe and measurement system performing the first X measurement and the approximate Z measurement are again qubit observables, and their readout observables are Z_p and Z_m, respectively.

The scheme thus realises a joint (sequential) measurement of three ± 1 valued observables, with probabilities

$$P_{k,\ell,n} := P(Z_p = k, X_f = \ell, Z_m = n), \quad k, \ell, n \in \{+, -\},$$

which are determined together the associated POVMs in the Supplemental Material [?] (see also [61]).

It is important to note that the weak-valued probabilities $P_{WV}(b_k, b_\ell)$ required for $P_{WV}(\delta b)$ do *not* coincide with the operational joint probabilities $P(Z_p = k, X_f = \ell)$ of the proposed experiment. Indeed, the reconstruction of the value of $\eta(B)$ from these operational joint probabilities is rather indirect in the proposed setup and does not suggest any relation with the rms value deviation interpretation.

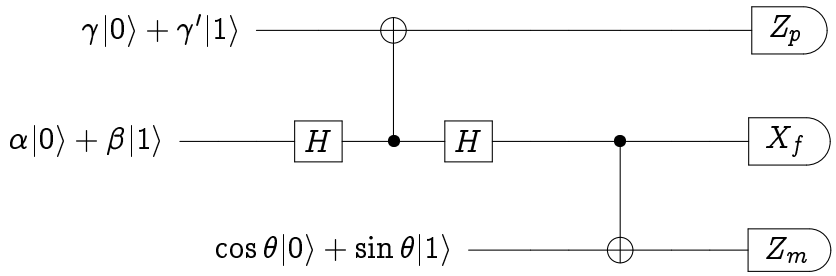


Figure 5.1: Model implementation of a determination of $\eta(X)$. The top and bottom wires represent the probe and measuring system while the middle wire corresponds to the observed qubit. As shown in the text, the value of $\eta(X)$ can be extracted from the joint distribution of the initial and final X measurements, obtained by reading the outputs Z_p and X_f.

As observed in [59], the "weak-valued" probabilities $P_{WV}(\delta b) = P_{WV}(\pm 2)$ can

be expressed in terms of the operational joint probabilities $\sum_n P_{k,\ell,n} = P(Z_p = k, X_f = \ell)$ in a rather involved (see the Supplement).

Using the explicit expressions for these joint probabilities and inserting the resulting values of $P_{WV}(\delta X = \pm 2)$ into (5.7) yields the same result *for all values of the strength parameter*, $2\gamma^2 - 1$:

$$\eta(X) = \sqrt{2}|\cos\theta - \sin\theta|,$$

There is no need to perform the limit to vanishing strength, $\gamma \to 1/\sqrt{2}$ for the determination of $\eta(X)$.

Curiously, not much was made of the fact that the coupling strength parameter for the weak measurement applied dropped out of the calculations *before* the weak limit was taken. We are thus led to consider a change of perspective – away from the focus on weak values to considering the *strong-measurement* limit of the scheme ($\gamma = 1$). We will see that this yields at once both a much simpler way of evaluating η and a more direct operational interpretation of this quantity. We also show that this alternative method is applicable beyond the qubit case, to joint measurement schemes where the measured approximating observables commute with their sharp target observables. To be sure, the weak-value method remains applicable also in the case of noncommuting approximators where the strong-measurement method is not available; but, as we noted above, the measures ε, η become unreliable if used beyond the realm of commuting approximators.

We can thus use the freedom of choice of initial interaction strength to explore what happens if we set γ to maximum strength, $\gamma = 1$, so that we are in fact performing a sharp (or *strong*) X measurement. Thus we are led to taking seriously the fact that here the numbers $\mathrm{tr}\,[\rho B_k D_\ell]$ are *bona fide* probabilities: they are simply the operational joint probabilities of the outcomes of the initial and final X measurements, and directly yield the probabilities for these values differing by $\delta X = \pm 2$:

$$P_{WV}(\delta X = \pm 2) = P(Z_p = \mp 1, X_f = \pm 1).$$

Correspondingly, the value of the disturbance quantity $\eta(X)$ is given by the

actual squared deviation of the values of the two measurements:

$$4P(Z_p = +1, X_f = -1) + 4P(Z_p = -1, X_f = +1) \equiv \eta(X)^2.$$

This surprising result becomes understandable when one considers that the X_f measurement can be viewed as an approximate repetition of the initial sharp measurement of X; the measurement of X_f is distorted into a measurement of a POVM D that is also compatible with approximate Z measurement performed between the X measurements. In this case X_f, or rather D, acts as a smeared version of X on the initial state, and thus commutes with the initial sharp X measurement. When the observable D commutes with X, the interpretation of $\eta(X)$ as disturbance is unproblematic [54].

It is therefore the strong measurement limit that provides a direct operational scheme for determining $\eta(X)$ *directly* as the rms deviation of the values of two X measurements performed on the same system before and after the Z measurement. It is important to note that this interpretation works under the assumption that the initial X measurement is of the Lüders type, which projects into X eigenstates. In this case, the quantities $\mathrm{tr}\,[\rho B_k D_\ell]$ appearing in (5.6) represent exactly the joint probabilities for the initial and final X measurements. In particular, the final marginal of the scheme of Fig. (5.1) is not affected by the presence of the inital X measurement.

5.5 Generalization

The alternative, “strong measurement", perspective on the disturbance measure η presented in the above model can be generalized to a rather wider class of sequential joint measurement scenarios. Let B be a sharp observable with values b_k and spectral projections B_k. Suppose an approximate measurement of A represented by POVM C is followed by a sharp measurement of observable B. This sequential scheme defines a joint measurement of C and some POVM D, which is an approximation of B. Assume that the disturbance is *benign*, in the sense that the D_ℓ commute with the B_k, which occurs typically when D is a smearing of B by means of a stochastic matrix $(\lambda_{\ell m})$, i.e., $D_\ell = \sum_m \lambda_{\ell m} B_m$. Now assume

that the measurement of C is preceded by a projective measurement of B. It follows that the operational joint probabilities are

$$\begin{aligned} P(B_i = b_k, B_f = b_\ell, \mathsf{C} = c_n) &= \mathrm{tr}\left[\mathcal{I}_n^{\mathsf{C}}(B_k \rho B_k) B_\ell\right] \\ &= \mathrm{tr}\left[B_k \rho B_k (\mathcal{I}_n^{\mathsf{C}})^*(B_\ell)\right]. \end{aligned}$$

Here $n \mapsto \mathcal{I}_n^{\mathsf{C}}$ denotes the instrument associated with C, giving the state change conditional on the outcome n, and $(\mathcal{I}_n^{\mathsf{C}})^*$ is the dual of the operation $\mathcal{I}_n^{\mathsf{C}}$. Disregarding the outcomes of the C measurement and noting that $D_\ell = \sum_n (\mathcal{I}_n^{\mathsf{C}})^*(B_\ell)$, we obtain the marginal probability

$$\begin{aligned} P(B_i = b_k, B_f = b_\ell) &= \mathrm{tr}\left[B_k \rho B_k D_\ell\right] \\ &= \mathrm{tr}\left[\rho B_k D_\ell\right] \equiv P_{WV}(b_k, b_\ell), \end{aligned}$$

since B_k commutes with D_ℓ. Therefore,

$$\begin{aligned} \eta(B)^2 &= \sum_{k,\ell} (b_k - b_\ell)^2 P_{WV}(b_k, b_\ell) \\ &= \sum_{k,\ell} (b_k - b_\ell)^2 P(B_i = b_k, B_f = b_\ell). \end{aligned}$$

A similar, even simpler consideration leads to a strong-measurement procedure for the error analysis required for a comparison of the measured observable C and the target observable A. In fact, assume A to have discrete values a_k with associated spectral projections A_k, and let C be discrete with values c_ℓ, where the associated positive operators C_ℓ are assumed to commute with the A_k. If the measurement of C is preceded by a Lüders measurement of A, the joint probability for an outcome pair (a_k, c_ℓ) is

$$P(A = a_k, \mathsf{C} = c_\ell) = \mathrm{tr}\left[\rho A_k C_\ell\right] \equiv P_{WV}(a_k, c_\ell).$$

Thus, in analogy to (5.6) one can write $\varepsilon(A)$ as a true value-comparison error, testable by preceding the C measurement with a strong (Lüders) measurement of A:

$$\varepsilon(A)^2 = \sum_{k\ell} (a_k - c_\ell)^2 \mathrm{tr}\left[\rho A_k C_\ell\right]. \tag{5.8}$$

5.6 Discussion

Considering the strong coupling limit of the measurement scheme designed originally for a weak-value determination of the disturbance η, we discovered a procedure for a *direct* determination of η via *error analysis*, that is, as the rms deviation of the values of the disturbed observable and the values of an accurate control measurement of the same observable. We showed that this error analysis method applies to joint measurements of compatible observables C, D as approximations to two discrete observables A, B, provided C, D commute with A, B, respectively.

We have thus obtained a demonstration of the possibility of determining ε, η via direct error analysis, together with a model realization arising from a modification of the Lund-Wiseman scheme, which can be implemented by adapting the Toronto experiment.

However, the restriction of the error analysis method to approximators that commute with the taget observables underlines the limitation of the measures ε, η pointed out in [55], namely, that they become unreliable where this commutativity is not given. This means that the Ozawa-Branciard inequalities lack universality as they cannot be interpreted safely as error-disturbance relations for noncommuting approximator and target observables. We conclude that the formulation of universal and directly testable state-dependent error-disturbance relations is to be based on alternative, generally applicable, measures of error and disturbance.

This leaves us with the question about the status of the existing tests of the Ozawa-Branciard inequalities for qubit observables. As was shown in [55], these experiments realize approximating observables C, D of the form $C_\pm = \frac{1}{2}(\mathbf{1} \pm \boldsymbol{c} \cdot \boldsymbol{\sigma})$ and $D_\pm = \frac{1}{2}(\mathbf{1} \pm \boldsymbol{d} \cdot \boldsymbol{\sigma})$, where the target observables are $A = \boldsymbol{a} \cdot \boldsymbol{\sigma}$ and $B = \boldsymbol{b} \cdot \boldsymbol{\sigma}$, respectively. Observables C, D of this kind are known to give optimal approximations, in the sense that for any general approximating observable one can always find a better approximator from this class [62]. Here the quality of approximations is judged not by state-dependent errors but by a *distance of observables*. This is the state-independent error, $\Delta(A)$, referred to in the Introduction, which is really a distance, $\Delta(\mathsf{A}, \mathsf{C})$, between the spectral measure A of A and the approximating POVM C. In [54], this distance was defined and

evaluated as

$$\Delta(\mathsf{A},\mathsf{C})^2 = 2\|\boldsymbol{a}-\boldsymbol{c}\|.$$

Moreover, it was shown that the quantity $\varepsilon(A)$ is state-independent and directly related to this distance:

$$\varepsilon(A)^2 \;=\; 1-\|\boldsymbol{c}\|^2+\tfrac{1}{4}\Delta(\mathsf{A},\mathsf{C})^4 \;\leq\; \Delta(\mathsf{A},\mathsf{C})^2.$$

In the Vienna experiment [48], the approximators are misaligned sharp observables ($\|\boldsymbol{c}\| = 1$), giving $\varepsilon(A) = \frac{1}{2}\Delta(\mathsf{A},\mathsf{C})^2$. In the Toronto experiment [47], they are smearings of the target with $\boldsymbol{c} = \lambda\boldsymbol{a}$, hence commuting, and one has $\varepsilon(A) = \Delta(\mathsf{A},\mathsf{C})$.

$\Delta(\mathsf{A},\mathsf{C})$ is directly obtained from the statistics of the A and C measurements for sufficiently many states [62, 54]; this is confirmed by the fact that $\varepsilon(A)$ can be determined by the three-state method. It follows that the experiments are appropriately interpreted as direct tests of a universal error-disturbance relation for worst-case errors and disturbances, namely Branciard's inequality in the form [49, Eq. (12)], evaluated for the observables $A = Z, B = X$ and a Y eigenstate. Using the scaling $\varepsilon(Z)^2 = 2d_Z$, $\eta(X)^2 = 2d_X$, this inequality reads simply

$$(d_Z-1)^2+(d_X-1)^2 \leq 1,$$

with values of interest being $d_Z, d_X \leq 1$. In the case of commuting approximators this strengthens the inequality $d_Z + d_X \geq 2-\sqrt{2}$ obtained in [62, 54], with d_Z, d_X now equal to $\|\boldsymbol{a}-\boldsymbol{c}\|, \|\boldsymbol{b}-\boldsymbol{d}\|$. Rather than being violations of Heisenberg's principle, the experiments thus confirm inequalities that are very much in the spirit of Heisenberg's uncertainty ideas.

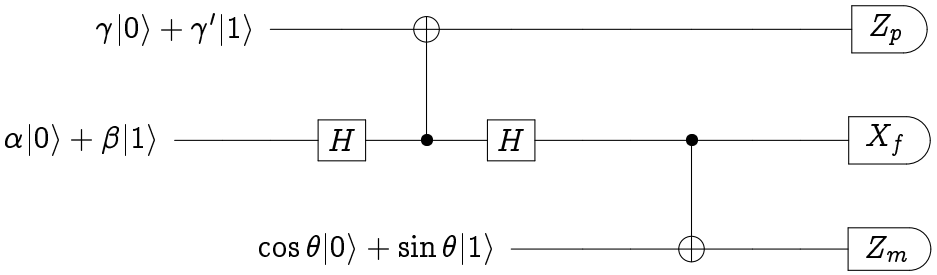
To conclude, it remains an interesting open problem whether *state-specific* error measures and associated *universal* and directly testable error-disturbance relations with nontrivial trade-off bounds can be found.

Supplemental Material

The experimental setup displayed in Figure 1 in the main text (reproduced below) consists of a three-qubit system, the object in initial state $\alpha|0\rangle+\beta|1\rangle$, a

"weak measurement probe (p)" initially in state $\gamma|0\rangle+\gamma'|1\rangle$, and the apparatus m with initial state $\cos\theta|0\rangle+\sin\theta|1\rangle$, all in their respective 2-dimensional Hilbert spaces $\mathscr{H}$, $\mathscr{H}_p$ and$\mathscr{H}_m$, respectively.

In the scenario when the disturbance measure for the observable X is to be determined, the initial approximate X measurement is enacted by first applying a Hadamard gate on the object system $\mathscr{H}$, followed by a C_{NOT} gate acting on $\mathscr{H}_p$, controlled on $\mathscr{H}$, and finally with another Hadamard gate performed on $\mathscr{H}$. This is followed by the device whose disturbance is being measured, wherein a C_{NOT} gate acts on $\mathscr{H}_m$, again controlled on $\mathscr{H}$. Sharp Z measurements are then performed on $\mathscr{H}_p$ and $\mathscr{H}_m$, (denoted Z_p and Z_m respectively), along with a sharp X measurement (X_f) on $\mathscr{H}$:



The state of the object and weak probe combined $|\psi_1\rangle$, after the the initial interaction is then given by:

$$\begin{aligned}
|\psi_1\rangle &= (\mathbb{I}\otimes H)C_{NOT}(\mathbb{I}\otimes H)(\gamma|0\rangle+\gamma'|1\rangle)\otimes(\alpha|0\rangle+\beta|1\rangle)\\
&= \frac{1}{\sqrt{2}}(\mathbb{I}\otimes H)C_{NOT}(\gamma|0\rangle+\gamma'|1\rangle)\otimes((\alpha+\beta)|0\rangle+(\alpha-\beta)|1\rangle)\\
&= \frac{1}{\sqrt{2}}(\mathbb{I}\otimes H)\left[(\gamma|0\rangle+\gamma'|1\rangle)\otimes(\alpha+\beta)|0\rangle+(\gamma'|0\rangle+\gamma|1\rangle)\otimes(\alpha-\beta)|1\rangle\right]\\
&= \frac{1}{2}[(\gamma(\alpha+\beta)+\gamma'(\alpha-\beta))|0\rangle\otimes|0\rangle+(\gamma'(\alpha+\beta)+\gamma(\alpha-\beta))|1\rangle\otimes|0\rangle\\
&\qquad +(\gamma(\alpha+\beta)-\gamma'(\alpha-\beta))|0\rangle\otimes|1\rangle+(\gamma'(\alpha+\beta)-\gamma(\alpha-\beta))|1\rangle\otimes|1\rangle]\\
&= |p_0\rangle\otimes|0\rangle+|p_1\rangle\otimes|1\rangle,
\end{aligned}$$

where

$$\begin{aligned}
|p_0\rangle &= \frac{1}{2}(\gamma(\alpha+\beta)+\gamma'(\alpha-\beta))|0\rangle+(\gamma'(\alpha+\beta)+\gamma(\alpha-\beta))|1\rangle\\
|p_1\rangle &= \frac{1}{2}(\gamma(\alpha+\beta)-\gamma'(\alpha-\beta))|0\rangle+(\gamma'(\alpha+\beta)-\gamma(\alpha-\beta))|1\rangle.
\end{aligned}$$

The state of the whole system after the measuring device, $|\psi_f\rangle$ is then

$$\begin{aligned}|\psi_f\rangle &= (\mathbb{I}\otimes C_{NOT})(|p_0\rangle\otimes|0\rangle + |p_1\rangle|1\rangle)\otimes(\cos\theta|0\rangle+\sin\theta|1\rangle)\\ &= |p_0\rangle\otimes|0\rangle\otimes(\cos\theta|0\rangle+\sin\theta|1\rangle) + |p_1\rangle\otimes|1\rangle\otimes(\sin\theta|0\rangle+\cos\theta|1\rangle)\\ &= |p_0\rangle\otimes|0\rangle\otimes|m_0\rangle + |p_1\rangle\otimes|1\rangle\otimes|m_1\rangle,\end{aligned}$$

with

$$\begin{aligned}|m_0\rangle &= \cos\theta|0\rangle+\sin\theta|1\rangle\\ |m_1\rangle &= \sin\theta|0\rangle+\cos\theta|1\rangle.\end{aligned}$$

Now writing $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)$, the eigenstates of X, we have

$$\begin{aligned}|\psi_f\rangle =& \frac{1}{\sqrt{2}}[|p_0\rangle\otimes|+\rangle\otimes|m_0\rangle + |p_1\rangle\otimes|+\rangle\otimes|m_1\rangle\\ &\qquad + |p_0\rangle\otimes|-\rangle\otimes|m_0\rangle - |p_1\rangle\otimes|-\rangle\otimes|m_1\rangle]\\ =& \frac{1}{2\sqrt{2}}[([\gamma(\alpha+\beta)+\gamma'(\alpha-\beta)]\cos\theta\\ &\qquad + [\gamma(\alpha+\beta)-\gamma'(\alpha-\beta)]\sin\theta)|0\rangle\otimes|+\rangle\otimes|0\rangle\\ &+ ([\gamma(\alpha+\beta)+\gamma'(\alpha-\beta)]\sin\theta\\ &\qquad + [\gamma(\alpha+\beta)-\gamma'(\alpha-\beta)]\cos\theta)|0\rangle\otimes|+\rangle\otimes|1\rangle\\ &+ ([\gamma'(\alpha+\beta)+\gamma(\alpha-\beta)]\cos\theta\\ &\qquad + [\gamma'(\alpha+\beta)-\gamma(\alpha-\beta)]\sin\theta)|1\rangle\otimes|+\rangle\otimes|0\rangle\\ &+ ([\gamma'(\alpha+\beta)+\gamma(\alpha-\beta)]\sin\theta\\ &\qquad + [\gamma'(\alpha+\beta)-\gamma(\alpha-\beta)]\cos\theta)|1\rangle\otimes|+\rangle\otimes|1\rangle\\ &+ ([\gamma(\alpha+\beta)+\gamma'(\alpha-\beta)]\cos\theta\\ &\qquad - [\gamma(\alpha+\beta)-\gamma'(\alpha-\beta)]\sin\theta)|0\rangle\otimes|-\rangle\otimes|0\rangle\\ &+ ([\gamma(\alpha+\beta)+\gamma'(\alpha-\beta)]\sin\theta\\ &\qquad - [\gamma(\alpha+\beta)-\gamma'(\alpha-\beta)]\cos\theta)|0\rangle\otimes|-\rangle\otimes|1\rangle\\ &+ ([\gamma'(\alpha+\beta)+\gamma(\alpha-\beta)]\cos\theta\\ &\qquad - [\gamma'(\alpha+\beta)-\gamma(\alpha-\beta)]\sin\theta)|1\rangle\otimes|-\rangle\otimes|0\rangle\\ &+ ([\gamma'(\alpha+\beta)+\gamma(\alpha-\beta)]\sin\theta\\ &- \qquad [\gamma'(\alpha+\beta)-\gamma(\alpha-\beta)]\cos\theta)|1\rangle\otimes|-\rangle\otimes|1\rangle].\end{aligned}$$

From here the probabilities of the various outcomes can be read off; writing, say P_{+-+} for the probability $P(Z_p = +1, X_f = -1, Z_m = +1)$, we have:

$$\begin{aligned}
8P_{+++} &= 1 + (2\gamma^2 - 1)(\alpha\bar{\beta} + \bar{\alpha}\beta) + \sin(2\theta)[(2\gamma^2 - 1) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] + 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{++-} &= 1 + (2\gamma^2 - 1)(\alpha\bar{\beta} + \bar{\alpha}\beta) + \sin(2\theta)[(2\gamma^2 - 1) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] - 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{-++} &= 1 + (1 - 2\gamma^2)(\alpha\bar{\beta} + \bar{\alpha}\beta) + \sin(2\theta)[(1 - 2\gamma^2) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] + 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{-+-} &= 1 + (1 - 2\gamma^2)(\alpha\bar{\beta} + \bar{\alpha}\beta) + \sin(2\theta)[(1 - 2\gamma^2) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] - 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{+-+} &= 1 + (2\gamma^2 - 1)(\alpha\bar{\beta} + \bar{\alpha}\beta) - \sin(2\theta)[(2\gamma^2 - 1) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] + 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{+--} &= 1 + (2\gamma^2 - 1)(\alpha\bar{\beta} + \bar{\alpha}\beta) - \sin(2\theta)[(2\gamma^2 - 1) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] - 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{--+} &= 1 + (1 - 2\gamma^2)(\alpha\bar{\beta} + \bar{\alpha}\beta) - \sin(2\theta)[(1 - 2\gamma^2) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] + 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta) \\
8P_{---} &= 1 + (1 - 2\gamma^2)(\alpha\bar{\beta} + \bar{\alpha}\beta) - \sin(2\theta)[(1 - 2\gamma^2) \\
&\quad + (\alpha\bar{\beta} + \bar{\alpha}\beta)] - 2\gamma\gamma'(|\alpha|^2 - |\beta|^2)\cos(2\theta).
\end{aligned}$$

This gives the respective 8-outcome POVM with positive operators $E_{k\ell m}$ on the target system:

$$\begin{aligned}
8E_{+++} &= (1 + \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} + (2\gamma^2 - 1 + \sin(2\theta))X + 2\gamma\gamma'\cos(2\theta)Z \\
8E_{++-} &= (1 + \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} + (2\gamma^2 - 1 + \sin(2\theta))X - 2\gamma\gamma'\cos(2\theta)Z \\
8E_{-++} &= (1 + \sin(2\theta)(1 - 2\gamma^2))\mathbb{I} + (1 - 2\gamma^2 + \sin(2\theta))X + 2\gamma\gamma'\cos(2\theta)Z \\
8E_{-+-} &= (1 + \sin(2\theta)(1 - 2\gamma^2))\mathbb{I} + (1 - 2\gamma^2 + \sin(2\theta))X - 2\gamma\gamma'\cos(2\theta)Z \\
8E_{+-+} &= (1 - \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} + (2\gamma^2 - 1 - \sin(2\theta))X + 2\gamma\gamma'\cos(2\theta)Z \\
8E_{+--} &= (1 - \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} + (2\gamma^2 - 1 - \sin(2\theta))X - 2\gamma\gamma'\cos(2\theta)Z \\
8E_{--+} &= (1 - \sin(2\theta)(1 - 2\gamma^2))\mathbb{I} + (1 - 2\gamma^2 - \sin(2\theta))X + 2\gamma\gamma'\cos(2\theta)Z \\
8E_{---} &= (1 - \sin(2\theta)(1 - 2\gamma^2))\mathbb{I} + (1 - 2\gamma^2 - \sin(2\theta))X - 2\gamma\gamma'\cos(2\theta)Z.
\end{aligned}$$

From here we can read off the actual (marginal) 2-outcome POVMs that are being measured on the system at the three stages. Firstly the Z_p measurement

defines the positive operators $B_k = \sum_{\ell m} E_{k\ell m}$ representing the initial weak X measurement:

$$\begin{aligned} B_+ &= \tfrac{1}{2}\Big[\, \mathbb{I} + (2\gamma^2 - 1)X \,\Big] \\ B_- &= \tfrac{1}{2}\Big[\, \mathbb{I} - (2\gamma^2 - 1)X \,\Big], \end{aligned}$$

the final sharp X_f corresponds to measuring the POVM $D_\ell = \sum_{km} E_{k\ell m}$:

$$\begin{aligned} D_+ &= \tfrac{1}{2}\Big[\, \mathbb{I} + \sin(2\theta)X \,\Big] \\ D_- &= \tfrac{1}{2}\Big[\, \mathbb{I} - \sin(2\theta)X \,\Big], \end{aligned}$$

and the observable actually being measured by the measurement device whose disturbance power is being assessed is $C_m = \sum_{k\ell} E_{k\ell m}$:

$$\begin{aligned} C_+ &= \tfrac{1}{2}\Big[\, \mathbb{I} + 2\gamma\gamma'\cos(2\theta)Z \,\Big] \\ C_- &= \tfrac{1}{2}\Big[\, \mathbb{I} - 2\gamma\gamma'\cos(2\theta)Z \,\Big] \end{aligned}$$

We also note down the POVM, $F_{k\ell} = \sum_m E_{k\ell m}$, representing the joint measurement of the initial weak X observable and the final X_f measurement, which is used to calculate the disturbance quantity:

$$\begin{aligned} F_{++} &= \tfrac{1}{4}\Big[\, (1 + \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} + (2\gamma^2 - 1 + \sin(2\theta))X \,\Big] \\ F_{-+} &= \tfrac{1}{4}\Big[\, (1 - \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} - (2\gamma^2 - 1 - \sin(2\theta))X \,\Big] \\ F_{+-} &= \tfrac{1}{4}\Big[\, (1 - \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} + (2\gamma^2 - 1 - \sin(2\theta))X \,\Big] \\ F_{--} &= \tfrac{1}{4}\Big[\, (1 + \sin(2\theta)(2\gamma^2 - 1))\mathbb{I} - (2\gamma^2 - 1 + \sin(2\theta))X \,\Big]. \end{aligned}$$

The associated operational joint probabilities in the state $\alpha|0\rangle + \beta|1\rangle$ are (putting $\langle X \rangle = \alpha\bar{\beta} + \bar{\alpha}\beta$):

$$\begin{aligned} P(Z_p = +1, X_f = +1) &= \tfrac{1}{4}\Big[\, (1 + \sin(2\theta)(2\gamma^2 - 1)) + (2\gamma^2 - 1 + \sin(2\theta))\langle X \rangle \,\Big] \\ P(Z_p = -1, X_f = +1) &= \tfrac{1}{4}\Big[\, (1 - \sin(2\theta)(2\gamma^2 - 1)) - (2\gamma^2 - 1 - \sin(2\theta))\langle X \rangle \,\Big] \\ P(Z_p = +1, X_f = -1) &= \tfrac{1}{4}\Big[\, (1 - \sin(2\theta)(2\gamma^2 - 1)) + (2\gamma^2 - 1 - \sin(2\theta))\langle X \rangle \,\Big] \\ P(Z_p = -1, X_f = -1) &= \tfrac{1}{4}\Big[\, (1 + \sin(2\theta)(2\gamma^2 - 1)) - (2\gamma^2 - 1 + \sin(2\theta))\langle X \rangle \,\Big]. \end{aligned}$$

With these expressions it is straightforward to verify Eq. (3) of the main text,

$$\begin{aligned} 2P_{WV}(\delta X = \pm 2) &= 2P_{WV}(X_i = \mp 1 | X_f = \pm 1) P(X_f = \pm 1) \\ &= P(Z_p = 1, X_f = \pm 1) + P(Z_p = -1, X_f = \pm 1) \\ &\quad \mp \frac{P(Z_p = 1, X_f = \pm 1) - P(Z_p = -1, X_f = \pm 1)}{2\gamma^2 - 1}. \end{aligned}$$

The last expression, which can be directly evaluated using the above probabilities, is to be compared with the weak-valued probability on the left hand side:

$$\begin{aligned} P_{WV}(\delta X = \pm 2) = P_{WV}(X_i = \mp 1, X_f = \pm 1) &= \left\langle \frac{1}{2}(\mathbb{I} \mp X) \frac{1}{2}\Big(\mathbb{I} \pm \sin(2\theta) X\Big) \right\rangle \\ &= \tfrac{1}{2}\Big(1 - \sin(2\theta)\Big) \tfrac{1}{2}\Big(1 \mp \langle X \rangle\Big). \end{aligned}$$

We observe that these weak-valued joint probabilities do not coincide with the operational probabilities, $P(Z_p = \mp 1, X_f = \pm 1)$, *except* in the strong measurement case, $\gamma = 1$.

Finally we verify the strong measurement realization of $\eta(X)$.

$$4P(Z_p = +1, X_f = -1) + 4P(Z_p = -1, X_f = +1) = 2 - 2\sin(2\theta)(2\gamma^2 - 1).$$

Note that this is already state-independent. On putting $\gamma = 1$, we finally obtain

$$4P(Z_p = +1, X_f = -1) + 4P(Z_p = -1, X_f = +1) = 2 - 2\sin(2\theta) = \eta(X)^2.$$

Chapter 6

Summary

It is worth summarising briefly some of the main results of the thesis here. The first part of the thesis had a lot to do with general probabilistic models. The idea of 'smearing' two observables, by mixing them with trivial observables enough to make them jointly measurable was introduced. This led to the idea of the joint measurability region for a pair of obervables, all observables, and all n-outcome observables. These concepts gave us the tools to compare theories based on the degrees on incompatibility inherent in them. We saw that for the measure that considers all observables in a given theory, quantum theory comes out as containing the most incompatibility that a theory theoretically can. This in this sense quantum theory is maximally incompatible. Howver we also saw that when the joint measurability regions for n-outcome observables are used, quantum theory is a long way off containing the maximum amount of incompatibility, with an example given of a theory that is maximally incompatible at only the 2-outcome level.

Furthering the investigation into GPMs, a single numerical measure of the degree of incompatibility of a pair of observables was then given. This measure is also derived from the idea of mixing the observables with trivial observables, this time unbiased ones. This measure was then phrased as a conic optimisation problem, whose dual problem is directly related to the notion of non-local correlations. This led to the notion of a generalised Tsirelson bound for a given theory, which limits the strength of certain correlations based on the degree of incompatibility in the theory. The condition of a theory supporting the con-

cept of steering was shown to be a sufficient condition for this bound to be saturated. Examples were given where the general results do and do not apply, and discussion was given about possible modifications to the used measure of incompatibility.

Finally the topic of measurement uncertainty relations in quantum theory was visited. Analysis was presented of certain proposed state-dependent measures of error and disturbance. The concept of a direct test of error or disturbance was formalised, and a potential experimental test was given for such a direct test. It was pointed out that such tests can only happen under certain conditions, which is consistent with other results about these measures. It was concluded that the measures presented so far do not meet the demands that would be necessary to be used in universally valid relations.

Bibliography

[1] G. Ludwig. *Commun. Math. Phys.*, 4:331–348, 1967.

[2] G. Ludwig. *Commun. Math. Phys.*, 9:1–12, 1968.

[3] B Mielnik. *Communications in Mathematical Physics*, 15:1–46, 1969. 10.1007/BF01645423.

[4] E.B. Davies. *J. Funct. Anal.*, 6:318–346, 1970.

[5] E.B. Davies. *Quantum Theory of Open Systems*. Academic Press, London, 1976.

[6] S.P. Gudder. *Stochastic Methods in Quantum Mechanics*. North Holland, New York, 1979.

[7] Lucien Hardy *arXiv:quant-ph/0101012*, 2001

[8] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. *arXiv:quant-ph/0611295*, 2006.

[9] H. Barnum and A. Wilce. *Electronic Notes in Theoretical Computer Science*, 270(1):3–15, Feb. 2011.

[10] L. Hardy. *arXiv:1104.2066*, 2011.

[11] Peter Janotta, Haye Hinrichsen *J. Phys. A: Math. Theor.* 47 (2014) 323001

[12] J. Barrett, *Phys. Rev. A*, **75** (2007) 032304.

[13] G. Kimura, T. Miyadera, H. Imai, *Phys. Rev. A*, **79** (2009) 062306.

[14] H. Barnum, J. Barrett, L. Orloff Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, R. Wilke, *New J. Phys.*, **12** (2010) 033024.

[15] G. Chiribella, G.M. D'Ariano, P. Perinotti, *Phys. Rev. A*, **84** (2011) 012311.

[16] P. Perinotti, *Phys. Rev. Lett.*, **108** (2012) 120502.

[17] P. Busch, T. Heinonen, P. Lahti. *Phys. Rep.*, **452** (2007) 155.

[18] P. Busch, M. Grabowski, P.J. Lahti, *Operational Quantum Physics*, Springer-Verlag, Berlin, 1997. 2nd. corr. printing.

[19] G. Ludwig, *Foundations of Quantum Mechanics I*, Springer-Verlag, Berlin, 1983.

[20] P. Busch T. Heinosaari, *Quant. Inf. Comp.*, **8** (2008) 0797.

[21] P. Busch, *Phys. Rev. D*, **33** (1986) 2253.

[22] C. Carmeli, T. Heinosaari, A. Toigo, *Phys. Rev. A*, **85** (2012) 012109.

[23] B.S. Cirel'son, *Lett. Math. Phys.*,**4** (1980) 93.

[24] J.S. Bell. *Physics*, 1:195–200, 1964.

[25] Arthur Fine. *Phys. Rev. Lett.*, 48:291–295, Feb 1982.

[26] Arthur Fine. *Journal of Mathematical Physics*, 23(7):1306–1310, 1982.

[27] I. Pitowsky. *British J. Philos. Sci.*, 45:95 – 125, 1994.

[28] G. Boole. *The Laws of Thought.* Macmillan, Cambridge, 1854.

[29] G. Boole. *Philosophical Transactions of the Royal Society of London*, 152:225 – 252, 1862.

[30] W.M. de Muynck and O. Abu Zeid. *Phys. Lett. A*, 100:486 – 489, 1984.

[31] P. Busch. In P.J. Lahti and P. Mittelstaedt, editors, *Symposium on the Foundations of Modern Physics*, pages 343–357, Singapore, 1985. World Scientific.

[32] G. Kar and S. Roy. *Phys. Lett. A*, 199(1–2):12–14, 1995.

[33] E. Andersson, S.M. Barnett, and A. Aspect. *Phys. Rev. A*, 72:042104, Oct 2005.

[34] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[35] Michael M. Wolf, David Perez-Garcia, and Carlos Fernandez. *Phys. Rev. Lett.*, 103:230402, Dec 2009.

[36] M. Banik, R. Gazi, S. Ghosh, and G. Kar. *Phys. Rev. A* **87**, 052125 (2013).

[37] P. Busch and T. Heinosaari. *Quantum Inf. & Comput.*, 8:0797–0818, 2008.

[38] C. Pfister. *arXiv:1203.5622*, 2012.

[39] I. Namioka and R.R. Phelps. *Pac. J. Math.*, 31:469–480, 1969.

[40] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens. *EPL*. 103:10002, 2013.

[41] H. Barnum, C.P. Gaebler, and A. Wilce. *arXiv:0912.5532*, 2009.

[42] E. Schrödinger. *Proc. Camb. Phil. Soc.*, 32:446, 1936.

[43] G. Chiribella, G. D'Ariano, P. Perinotti *Phys. Rev. A*, 81, 062348, 2010.

[44] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[45] P. Janotta, C. Gogolin, J. Barrett, N. Brunner. *New J. Phys.*, 13:063024, 2011.

[46] M. Ozawa, *Ann. Phys.* **311**, 350, 2004.

[47] L.A. Rozema, A. Darabi, D.H. Mahler, A. Hayat, Y. Soudagar, A.M. Steinberg, *Phys. Rev. Lett.***109**, 100404, 2012.

[48] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, Y. Hasegawa, *Nature Phys.* **8**, 185, 2012.

[49] C. Branciard, *Proc. Nat. Acad. Sci.* **110**, 6742, 2013.

[50] S.-Y. Baek, F. Kaneda, M. Ozawa, K. Edamatsu, *Scientific Reports* **3**, 2221, 2013.

[51] M. Ringbauer, D.N. Biggerstaff, M.A. Broome, A. Fedrizzi, C. Branciard, A.G. White, *Phys. Rev. Lett.* **112**, 020401, 2014.

[52] F. Kaneda, S.-Y. Baek, M. Ozawa, K. Edamatsu, *Phys. Rev. Lett.***112**, 020402, 2014.

[53] P. Busch, P. Lahti, R.F. Werner, *Phys. Rev. Lett.* **111**, 160405, 2013.

[54] P. Busch, P.J. Lahti, R.F. Werner, *Phys. Rev. A* **89**, 012129, 2014.

[55] P. Busch, P.J. Lahti, R.F. Werner, Quantum root-mean-square error and measurement uncertainty relations, arXiv:1312.4393v2, 2014.

[56] Specifically, it was shown in [55] that $\varepsilon(A)$ can be zero although the approximation is evidently bad in the given state. This can even happen for both error measures in an approximate joint measurement, rendering a "tight" instance, 0=0, of the Ozawa and Branicard inequalities, while the "true" accuracy is definitely not perfect. That something was wrong with the quantity ε was noted by Branciard [49], who also observed that ε can be zero while the approximation does not appear to be perfect. He considered this an artefact of the choice of measure and remarked that it would be interesting to consider alternative measures.

[57] J.M. Renes, V.B. Scholz, arXiv:1402.6711, 2014.

[58] To be sure, a weaker and still sufficient condition for the numbers $\mathrm{tr}\,[\rho B_k \mathsf{D}_\ell]$ to comprise a probability distribution is that the operators B_k and D_ℓ

commute on the state ρ, that is, $B_k D_\ell \varphi = D_\ell B_k \varphi$ for all states in the range of ρ.

[59] A.P. Lund, H.M. Wiseman, *New J. Phys.* **12**, 093011, 2010.

[60] The observables X, Y, Z are identical to the Pauli operators $\sigma_x, \sigma_y, \sigma_z$.

[61] M. Tukiainen, Masters, Dissertation, Univerity of Turku, 2013 (https://www.doria.fi/handle/10024/91949).

[62] P. Busch, T. Heinosaari, *Quantum Inf. Comput.* **8**, 797, 2008.