Julius-Maximilians-Universität Würzburg
Institut für theoretische Physik und Astronomie

Nonlocality and entanglement in Generalized Probabilistic Theories and beyond

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

Peter Janotta

aus Würzburg

Würzburg 2014

Eingereicht am: 08.07.2014
bei der Fakultät für Physik und Astronomie

1. Gutachter: Prof. Dr. Haye Hinrichsen
2. Gutachter: Prof. Dr. Ewelina M. Hankiewicz
3. Gutachter:
der Dissertation.

1. Prüfer: Prof. Dr. Haye Hinrichsen
2. Prüfer: Prof. Dr. Ewelina M. Hankiewicz
3. Prüfer: Prof. Dr. Hartmut Buhmann
im Promotionskolloquium

Tag des Promotionskolloquiums: 17.10.2014

Doktorurkunde ausgehändigt am:

Zusammenfassung

Die Quantentheorie wird als eine der grundlegensten und präzisesten physikalischen Theorien unserer Zeit angesehen. Auch wenn die Theorie konzeptionell schwierig zu verstehen ist, so ist die eigentliche mathematische Struktur überraschend einfach. Auf welcher physikalischen Grundlage basiert diese besonders einfache und elegante mathematische Struktur? Oder anders ausgedrückt: Warum ist die Quantentheorie so wie sie ist?

Das Gebiet der "Grundlagen der Quantentheorie" versucht, auf diese Fragen Antworten zu finden. In der Vergangenheit wurde dieses Forschungsgebiet als überwiegend akademisch mit wenig praktischen Nutzen angesehen. Mit dem Fortschritt der Quanteninformationstheorie hat sich diese Sicht aber grundsätzlich gewandelt, da beide Gebiete einander stetig fruchtbar beeinflussen [1, 2]. Dadurch stößt die Grundlagenforschung zur Quantentheorie derzeit auf wachsendes Interesse, so dass das Gebiet nun wohl zu den spannensten Themenbereichen der theoretischen Physik gezählt werden darf. Diese Arbeit beschäftigt sich mit einer bestimmten Richtung in diesem Feld – den sogenannten "Verallgemeinerten Wahrscheinlichkeitstheorien" (GPTs). Diese bilden einen umfassenden theoretischen Rahmen zur Beschreibung physikalischer Theorien, wobei die klassische Wahrscheinlichkeitstheorie und die Quantentheorie als Spezialfälle enthalten sind. Wir nutzen diesen Ansatz in dieser Arbeit, um nicht-lokale Eigenschaften zu untersuchen, die helfen die Sonderrolle der Quantentheorie gegenüber nicht realisierter Alternativen zu verstehen. Um die Anwendungsbereich dieses Ansatzes zu vergrößern, führen wir verschiedene Verallgemeinerungen ein. Dadurch wird es möglich die Auswirkungen von Annahmen zu untersuchen, die typischerweise beim GPT-Ansatz gemacht werden.

Basierend auf einem Übersichtsartikel des Autors, beginnen wir in Kapitel 1 zunächst mit einer Einführung des üblichen GPT-Ansatzes und fassen bisherige Ergebnisse zusammen. Um die Einführung möglichst verständlich zu halten, verfolgen wir dabei einen konstruktiven Ansatz. Beginnend mit wenigen, physikalisch wohlmotivierten Annahmen, zeigen wir wie beliebige experimentelle Beobachtungen in einer operationalen Theorie festgehalten werden

können. Desweiteren, charakterisieren wir die Konsistenzbedingungen, die bei darauf aufbauenden Erweiterungen der Theorie beachtet werden müssen. Wir zeigen auf, dass nicht-klassische Eigenschaften eines Einzelsystems in gleicher Weise auch in einem höherdimensionalen klassischen System auftreten können, wenn man die Menge möglicher Messungen beschränkt. Hingegen wird gezeigt, dass Verschränkung und Nicht-Lokalität echt nicht-klassische Eigenschaften darstellen. Besondere Eigenschaften, die spezifisch für die Quantentheorie sind, werden separat für Einzelsysteme und zusammengesetzte Systeme besprochen.

Kapitel 2 enthält Ergebnisse, die wir in [3] und [4] veröffentlicht haben. Der GPT-Ansatz wird dort dazu benutzt, um zu zeigen, wie die Struktur lokaler Zustandsräume indirekt die möglichen nicht-lokalen Korrelationen beeinflusst, die selbst eigentlich globale Eigenschaften darstellen. Solche Korrelationen sind stärker als jene, die in der klassischen Wahrscheinlichkeitstheorie möglich sind, zeigen jedoch andere Beschränkungen, die wir auf die Struktur der Subsysteme zurück führen können. Zunächst illustrieren wir dieses Phänomen mit Spieltheorien mit bestimmten lokalen Zustandsräumen. Danach zeigen wir, dass für eine besondere Klasse von zusammengesetzten Zuständen (inner product states), deren Existenz von geometrischen Eigenschaften der lokalen Subsysteme abhängt, Korrelationen im Allgemeinen auf eine Menge beschränkt sind, die als $\mathbb{Q}_1$ bekannt ist. Alle bipartiten Korrelationen von Quantentheorie und klassischer Wahrscheinlichkeitstheorie können auf die Messstatistiken dieser Zusände zurückgeführt werden.

Kapitel 3 beinhaltet größtenteils unpublizierte Ergebnisse zu Verschränkungstausch (*entanglement swapping*) in GPTs. Diese Protokoll, das aus der Quanteninformationstheorie bekannt ist, erlaubt den nicht-lokalen Transfer von Verschränkung zu anfangs unverschränkten Parteien mit Hilfe eines Dritten, der verschränkte Zustände mit Beiden teilt. Wir stellen zunächst unseren in [4] eingeführten Ansatz vor, der die Struktur zusammengesetzter Systeme in der Quantentheorie nachahmt. Dafür modifizieren wir eine populäre Spieltheorie, die unter dem Namen *boxworld* bekannt ist. Es stellt sich jedoch heraus, dass dieser Ansatz für größere multipartite Systeme fehlschlägt, da die An-

wendung des Verschränkungstausch-Protokolls zu Inkonsistenzen führt. Wir zeigen dann, dass der GPT-Ansatz generell konsistenten Verschränkungstausch für solche Systeme verbietet, die Subsysteme mit zwei-dimensionalen Zustandsräumen haben, wo die reinen Zustände reversibel in einander überführbar sind. Ändern wir den GPT-Ansatz jedoch insofern, dass rein globale Freiheitsgrade zugelassen sind, zeigt sich das Verschränkungstausch auch für diese Systeme möglich wird. Dabei kommt eine Konstruktion zum Einsatz, die die Situation nachahmt, wie sie in der Quantentheorie auf einem reellen Hilbertraum herrscht.

Normalerweise geht der GPT-Ansatz von der sogenannten *no-restriction*-Hypothese aus, bei der der Zustandsraum einer physikalischen Theorie auch die Menge möglicher Messungen bestimmt. Allerdings scheint diese Annahme nicht physikalisch motiviert. Wir verallgemeinern daher in Kapitel 4 den Ansatz auf Systeme, die nicht der no-restriction-Hypothese gehorchen unter Verwendung von Resultaten aus [5] und [6]. Wir zeigen, wie unser so erweiterte Ansatz dazu genutzt werden kann, neue Klassen von Wahrscheinlichkeitstheorien zu beschreiben. Dadurch lässt sich beispielsweise in eine Theorie intrinsisches Rauschen fest einbauen. Das Aufheben der no-restriction-Hypothese erlaubt es uns außerdem eine *Selbstualisierungsprozedur* einzuführen. Dadurch lassen sich eine neue Klasse von Theorien definieren, die der Quantentheorie ähnelnde Eigenschaften aufweisen. Beispielsweise sind die Korrelationen durch Messungen auf den maximal verschränkten Zustands durch die Tsirelson-Schranke beschränkt. Schließlich charakterisieren wir die maximale Menge zusammengesetzter Zustände, die sich allgemein konsistent für gegebene Subsysteme definieren lassen. Dies verallgemeinert das aus dem normalen GPT-Ansatz bekannte, sogenannte maximale Tensorprodukt.

Abstract

Quantum theory is considered to be the most fundamental and most accurate physical theory of today. Although quantum theory is conceptually difficult to understand, its mathematical structure is quite simple. What determines this particularly simple and elegant mathematical structure? In short: Why is quantum theory as it is?

Addressing such questions is the aim of investigating the *foundations of quantum theory*. In the past this field of research was sometimes considered as an academic subject without much practical impact. However, with the emergence of quantum information theory this perception has changed significantly and both fields started to fruitfully influence each other [7, 2]. Today fundamental aspects of quantum theory attract increasing attention and the field belongs to the most exciting subjects of theoretical physics. This thesis is concerned with a particular branch in this field, namely, with so-called *Generalized Probabilistic Theories* (GPTs), which provide a unified theoretical framework in which classical and quantum theory emerge as special cases. This is used to examine nonlocal features that help to distinguish quantum theory from alternative toy theories. In order to extend the scope of theories that can be examined with the framework, we also introduce several generalizations to the framework itself.

We start in Chapter 1 with introducing the standard GPT framework and summarize previous results, based on a review paper of the author [8]. To keep the introduction accessible to a broad readership, we follow a constructive approach. Starting from few basic physically motivated assumptions we show how a given set of observations can be manifested in an operational theory. Furthermore, we characterize consistency conditions limiting the range of possible extensions. We point out that non-classical features of single systems can equivalently result from higher dimensional classical theories that have been restricted. Entanglement and non-locality, however, are shown to be genuine non-classical features. We review features that have been found to be specific for quantum theory separably or single and joint systems.

Chapter 2 incorporates results published in [3] and [4]. The GPT framework

is applied to show how the structure of local state spaces indirectly affects possible nonlocal correlations, which are global properties of a theory. These correlations are stronger than those possible in a classical theory, but happen to show different restrictions that can be linked to the structure of subsystems. We first illustrate the phenomenon with toy theories with particular local state spaces. We than show that a particular class of joint states (inner product states), whose existence depends on geometrical properties of the local subsystems, can only have correlations for a known limited set called $\mathbb{Q}_1$. All bipartite correlations of both, quantum and classical correlations, can be mapped to measurement statistics from such joint states.

Chapter 3 shows unpublished results on entanglement swapping in GPTs. This protocol, which is well known in quantum information theory, allows to nonlocally transfer entanglement to initially unentangled parties with the help of a third party that shares entanglement with each. We review our approach published in [4], which mimics the joint systems' structure of quantum theory by modifying a popular toy theory known as boxworld. However, it is illustrated that this approach fails for bigger multipartite systems due to inconsistencies evoked by entanglement swapping. It turns out that the GPT framework does not allow entanglement swapping for general subsystems with two-dimensional state spaces with transitive pure states. Altering the GPT framework to allow completely globally degrees of freedom, however, enables us to construct consistent entanglement swapping for these subsystems. This construction resembles the situation in quantum theory on a real Hilbert space.

A questionable assumption usually taken in the standard GPT framework is the so-called no-restriction hypothesis. It states that the measurement that are possible in a theory can be derived from the state space. In fact, this assumption seems to exist for reasons of mathematical convenience, but it seems to lack physical motivation. We generalize the GPT framework to also account for systems that do not obey the no-restriction hypothesis in Chapter 4, which presents results published in [5] and [6]. The extended framework includes new classes of probabilistic theories. As an example, we show how to construct theories that include intrinsic noise. We also provide a "self-dualization" procedure

that requires the violation of the no-restriction hypothesis. This procedure restricts the measurement of arbitrary theories such that the theories act as if they were self-dual. Self-duality has recently gathered lots of interest, since such theories share many features of quantum theory. For example Tsirelson's bound holds for correlations on the maximally entangled state in these theories. Finally, we characterize the maximal set of joint states that can be consistently defined for given subsystems. This generalizes the maximal tensor product of the standard GPT framework.

Contents

1. The GPT framework

1.1. Introduction

This chapter introduces Generlized Probabilistic Theories (GPTs) based on review article of the author [8]. It presents the concept of GPTs in the language of statistical physics, with the aim to establish a bridge between the communities of classical statistical physics and quantum information science.

The early pioneers of quantum theory were strongly influenced by positivism, a philosophy postulating that a physical theory should be built and verified entirely on the basis of accessible sensory experience. Nevertheless the standard formulation of quantum theory involves additional concepts such as global complex phases which are not directly accessible. The GPT framework, which is rooted in the pioneering works by Mackey, Ludwig and Kraus [9, 10, 11, 12], tries to avoid such concepts as much as possible by defining a theory operationally in terms of preparation procedures and measurements.

As measurement apparatuses yield classical results, GPTs are exclusively concerned with the classical *probabilities* of measurement outcomes for a given preparation procedure. As we will see below, classical probability theory and quantum theory can both be formulated within this unified framework. Surprisingly, starting with a small set of basic physical principles, one can construct a large variety of other consistent theories with different measurement statistics. This generates a whole spectrum of possible theories, in which classical and quantum theory emerge just as two special cases. Most astonishingly, various properties thought to be special for quantum theory turn out to be quite general in this space of theories. As will be discussed below, this includes the phenomenon of entanglement, the no-signaling theorem, and the impossibility to decompose a mixed state into unique ensemble of pure states.

Although GPTs are defined operationally in terms of probabilities for measurement outcomes, it is not immediately obvious how such a theory can be constructed from existing measurement data. This introduction tries to shed some light on the process of building theories in the GPT framework on the basis of a set of given experimental observations, making the attempt to provide step-by-step instructions how a theory can be constructed. We try to explain the essential concepts of the GPT framework in terms of simple examples, avoiding mathematical details whenever it is possible. We present the subject from the perspective of model building. To this end we start in Sect. 1.2.1 with a data table that contains all the available statistical information of measurement outcomes. In Sect. 1.2.3 the full space of possible experimental settings is then grouped into equivalence classes of observations, reducing the size of the table and leading to a simple prototype model. As shown in Sect. 1.2.5 this prototype model has to be extended in order to reflect possible deficiencies of preparations and measurements, leading in turn to new suitable representations of the theory. Further extensions can be chosen freely within a certain range limited by consistency conditions (see Sect. 1.2.9). Depending on this choice the extended theory finally allows one to make new predictions in situations that have not been examined so far. Within this framework we discuss three important minimal systems, namely, the classical *bit*, then quantum bit (*qubit*), as well as the so-called *gbit*.

In Sect. 1.4 we devote our attention to the fact that any non-classical system is equivalent to a classical system in a higher-dimensional state space combined with certain constraints. However, this equivalence is only valid as long as non-composite (single) systems are considered. Turning to bipartite and multipartite systems the theory has to be complemented by a set of composition rules in the form of a suitable tensor product (see Sect. 1.5). Again it turns out that there is some freedom in choosing the tensor product, which determines the structure of a GPT to a large extent. Finally, in Sect. 1.6 we discuss nonlocal correlations as a practical concept that can be used to experimentally prove the existence of non-classical entanglement in composite systems without the need to rely on a particular theory.

For beginners it is often difficult to understand the construction of a non-classical theory without introducing concepts such as Hilbert spaces and state vectors. For this reason we demonstrate how ordinary quantum mechanics fits into the GPT framework, both for single systems in Sect. 1.3.6 and for composite systems in Sect. 1.8.

1.2. Generalized probabilistic theories: Single systems

1.2.1. Preparation procedures and measurements

As sketched schematically in Fig. 1.1, a typical experimental setup in physics consists of a preparation procedure, possibly followed by a sequence of manipulations or transformations, and a final measurement. For example, a particle accelerator produces particles in a certain physical state which are then manipulated in a collision and finally measured by detectors. Since the intermediate manipulations can be thought of as being part of either the preparation procedure or the measurement, the setup can be further abstracted to preparations and measurements only[1].

We can think of a measurement apparatus as a physical device which is prepared in a spring-loaded non-equilibrium idle state. During the measurement process the interaction of the physical system with the device releases a cascade of secondary interactions, amplifying the interaction and eventually leading to a *classical* response that can be read off by the experimentalist. This could be, for example, an audible 'click' of a Geiger counter or the displayed value of a voltmeter.

In practice a measurement device produces either digital or analog results. For analog devices there are in principle infinitely many possible outcomes, but due to the final resolution the amount of information obtained during the

[1] In standard quantum theory the absorption of intermediate transformations into the preparation procedure corresponds to the Schrödinger picture, the absorption into the measurement to the Heisenberg picture.

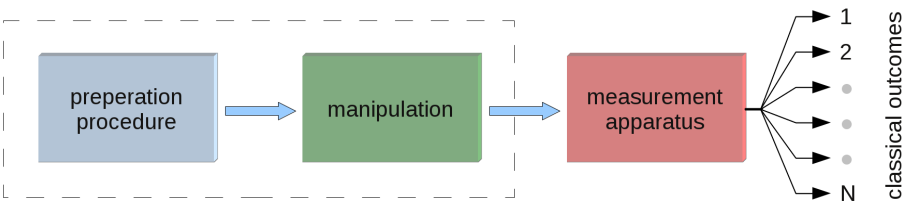


Figure 1.1.: Typical experimental setup consisting of a preparation procedure, a sequence of intermediate manipulations, and a final measurement with a certain set of possible classical outcomes (see text). The intermediate manipulations can be thought of as being part of either the preparation procedure (dashed box) or the measurement.

measurement is nevertheless finite. Thus, for the sake of simplicity, we will assume that the number of possible outcomes in a measurement is finite.

For an individual measurement apparatus we may associate with each of the possible outcomes a characteristic one-bit quantity which is '1' if the result occurred and '0' otherwise. In this way a measurement can be decomposed into mutually excluding classical bits, as sketched in Fig. 1.1. Conversely every single measurement can be interpreted as a joint application of such fundamental 1-bit measurements.

If we are dealing with several different measurement devices the associated classical bits are of course not necessarily mutually excluding. This raises subtle issues about coexistence, joint measurability, mutual disturbance and commutativity [13, 14], the meaning of a '0' if the measurement fails, and the possibility to compose measurement devices out of a given set of 1-bit measurements. For simplicity let us for now neglect these issues and return to some of the points later in the introduction.

1.2.2. Toolbox and probability table

In practice we have only a limited number of preparation procedures and measurement devices at our disposal. It is therefore meaningful to think of some kind of 'toolbox' containing a finite number of 1-bit measurements labeled by $k = 1 \ldots K$ and a finite number of preparation procedures labeled by $\ell = 1 \ldots L$. As mentioned before, if the range of preparations and measurements is con-

tinuous, we assume for simplicity that the finite accuracy of the devices will essentially lead to the same situation with a finite number of elements. Our aim is to demonstrate how the GPT approach can be used to construct a physical theory on the basis of such a toolbox containing K measurement devices and L preparation methods.

With each pair of a 1-bit measurement k and a preparation procedure ℓ we can set up an experiment which produces an outcome $\chi_{k\ell} \in \{0, 1\}$. An important basic assumption of the GPT framework is that experiments can be repeated under identical conditions in such a way that the outcomes are statistically independent. Repeating the experiment the specific outcome $\chi_{k\ell}$ is usually not reproducible, instead one can only reproducibly estimate the probability

$$p_{k\ell} = \langle \chi_{k\ell} \rangle \tag{1.1}$$

to obtain the result $\chi_{k\ell} = 1$ in the limit of infinitely many experiments. For a given toolbox the values of $p_{k\ell}$ can be listed in a probability table. This data table itself can already be seen as a precursor of a physical model. However, it just reproduces the observable statistics and apart from the known probabilities it has no predictive power at all. Moreover, the table may grow as we add more preparation and measurement devices. In order to arrive at a meaningful physical theory, we thus have to implement two important steps, namely,

1. to remove all possible redundancies in the probability table, and

2. to make reasonable assumptions which allow us to predict the behavior of elements which are not yet part of our toolbox.

1.2.3. Operational equivalence, states and effects

In order to remove redundancies in the probability table let us first introduce the notion of operational equivalence. Two preparation procedures are called *operationally equivalent* if it is impossible to distinguish them experimentally, meaning that any of the available measurement devices responds to both of them with the same probability. Likewise two one-bit measurements are called

operationally equivalent if they both respond with the same probability to any of the available preparation procedures.

The notion of operational equivalence allows one to define equivalence classes of preparations and one-bit measurements. Following the terminology introduced by Ludwig and Kraus [10, 12] we will denote these classes as states and effects:

- A *state* ω is a class of operationally equivalent preparation procedures.
- An *effect* e is a class of operationally equivalent 1-bit measurements.

This allows us to rewrite the probability table in terms of states and effects, which in practice means to eliminate identical rows and columns in the data table. Enumerating effects by $\{e_1, e_2, \ldots, e_M\}$ and states by $\{\omega_1, \omega_2, \ldots, \omega_N\}$ one is led to a reduced table of size $M \times N$, the so-called *fundamental probability table*.

If we denote by $e(\omega) = p(e|\omega)$ the probability that an experiment chosen from the equivalence classes e and ω produces a '1', the matrix elements elements of the fundamental probability table can be written as

$$p_{ij} = \langle \chi_{ij} \rangle = e_i(\omega_j)\,. \tag{1.2}$$

Obviously, this table contains all the experimentally available information. Since effects and states are defined as equivalence classes, it is ensured that no column (and likewise no row) of the table appears twice.

Note that the later inclusion of additional measurement apparatuses might allow the experimentalist to distinguish preparation procedures which were operationally equivalent before, splitting the equivalence class into smaller ones. This means that a state may split into several states if a new measurement device is added to the toolbox. The same applies to effects when additional preparation procedures are included.

As the introduction of equivalence classes described above eliminates only identical rows and columns, the fundamental probability table can be still very large. In addition, there may be still linear dependencies among rows and

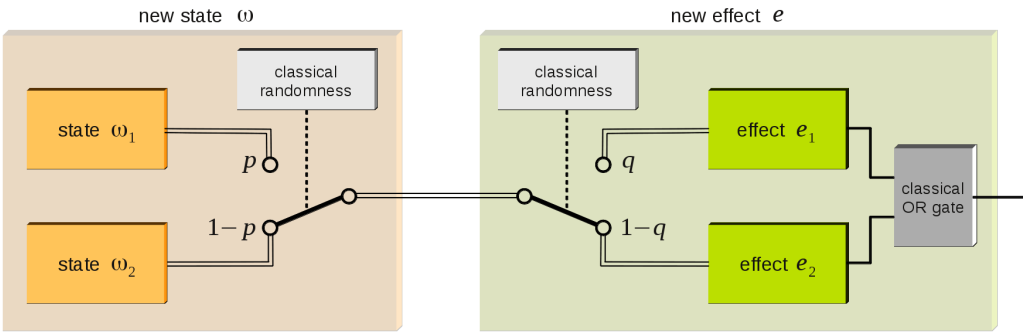


Figure 1.2.: New states and effects can be generated by probabilistically mixing the existing ones, illustrated here in a simple example (see text).

columns. As we will see below, these linear dependencies can partly be eliminated, leading to an even more compact representation, but they also play an important role as they define the particular type of the theory.

1.2.4. Noisy experiments and probabilistic mixtures of states and effects

Realistic experiments are noisy. This means that a preparation procedure does not always create the physical object in the same way, rather the preparation procedure itself may randomly vary in a certain range. Similarly, a measurement is noisy in the sense that the measurement procedure itself may vary upon repetition, even when the input is identical. In the GPT framework this kind of classical randomness is taken into account by introducing the notion of *mixed* states and effects.

The meaning of probabilistic mixtures is illustrated for the special case of bimodal noise in Fig. 1.2. On the left side of the figure a classical random number generator selects the preparation procedure ω_1 with probability p and another preparation procedure ω_2 with probability $1-p$. Similarly, on the right side another independent random number generator selects the effect e_1 with probability q and the effect e_2 otherwise, modeling a noisy measurement device.

If we apply such a noisy measurement to a randomly selected state, all what we get in the end is again a 'click' with certain probability P. In the example

shown in Fig. 1.2, this probability is given by

$$P = p\,q\,e_1(\omega_1) + p\,(1-q)\,e_2(\omega_1) + (1-p)\,q\,e_1(\omega_2) + (1-p)\,(1-q)\,e_2(\omega_2)\,, \quad (1.3)$$

where we used the obvious assumption that the intrinsic probabilities $p_{ij} = e_i(\omega_j)$ are independent of p and q.

It is intuitively clear that a machine which randomly selects one of various preparation procedures can be considered as a preparation procedure in itself, thus defining a new state ω. Similarly, a device of randomly selected effects can be interpreted in itself as a new effect e. Using these new effects and states the probability (1.3) to obtain a 'click' can simply be expressed as $P = e(\omega)$

It is easy to see that probability (1.3) can be recovered by conveniently describing the new states and effects by the following linear combinations

$$\omega := p\,\omega_1 + (1-p)\,\omega_2\,, \qquad e := q\,e_1 + (1-q)\,e_2, \quad (1.4)$$

when we define them to act linear on each other. We call these new objects *mixed states* and *mixed effects* respectively.

We introduced the probabilistic mixing of devices as a consequence of noise. Of course, the experimenter could also chose the set-up in Fig. 1.2 on purpose with any probabilities p and q. As p, q are continuous, probabilistic mixing yields a continuous variety of states and effects that can be realized.

1.2.5. Linear spaces, convex combinations, and extremal states and effects

The previous example shows that it is useful to represent probabilistically mixed states and effects as linear combinations. It is therefore meaningful to represent them as vectors in suitable vector spaces, whose structure, dimension and the choice of the basis we will discuss further below. For now, let us assume that each state ω_i is represented by a vector in a linear space V and similarly each effect e_i by a vector in another linear space V^*, which is called the *dual space* of V.

The embedding of states and effects in linear spaces allows us to consider arbitrary linear combinations

$$e = \sum_i \lambda_i \, e_i \,, \qquad \omega = \sum_j \mu_j \, \omega_j \tag{1.5}$$

with certain coefficients λ_i and μ_j. Moreover, the fundamental probability table $p_{ij} = e_i(\omega_j)$ induces a bilinear map $V^* \times V \to \mathbb{R}$ by

$$e(\omega) = \Big[\sum_{i=1}^{M} \lambda_i \, e_i\Big] \Big(\sum_{j=1}^{N} \mu_j \omega_j\Big) = \sum_{i=1}^{M} \sum_{j=1}^{N} \lambda_i \mu_j \underbrace{e_i(\omega_j)}_{=p_{ij}}, \tag{1.6}$$

generalizing Eq. (1.3) in the previous example. Note that this bilinear map on $V^* \times V$ should not be confused with an inner scalar product on either $V \times V$ or $V^* \times V^*$. In particular, it does not induce the notion of length, norm, and angles.

Having introduced linear combinations we also need to define what it means when two vectors coincide. Obviously, for states (effects) the same vector yield the same probability distribution with respect to all effects (states). Moreover, the GPT framework also assumes that the opposite is true. That is, each state/effect is represented by an unique vector. This *equivalence principle* ensures that the vectors reflect the operational equivalence defining states and effects[2]. The equivalence principle thus allows us to identify an element with the linear combinations of other elements.

At this point it is not yet clear which of the linear combinations in (1.5) represent physically meaningful objects. However, as shown above, the set of physically meaningful objects will at least include all probabilistic mixtures of the existing states and effects, which are mathematically expressed as *convex combinations* with non-negative coefficients adding up to 1.

States which can be written as convex combinations of other states are re-

[2]Note that bra vectors in the Dirac formulation of quantum mechanics do not satisfy the equivalence principle, since global phases do not affect operational equivalence. Consequently, in the GPT terminology a "state" is not a particular bra vector, but refers to the equivalence class of bra vectors that only differ by a global phase.

	e_1	e_2	e_3	e_4	e_5
ω_1	1	0	1	1	1
ω_2	$\frac{1}{2}$	0	1	$\frac{2}{3}$	$\frac{3}{4}$
ω_3	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{3}{4}$
ω_4	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$

Table 1.1.: Example of a probability table after removing identical columns and rows.

ferred to as *mixed states*. Conversely, states which cannot be expressed as convex combinations of other states are called *extremal states*. As any convex set is fully characterized by its extremal points, we can reduce the probability table even further by listing only the extremal states, tacitly assuming that all convex combinations are included as well. The same applies to effects.

1.2.6. Linear dependencies among extremal states and effects

What is the dimension of the spaces V and V^* and how can we choose a suitable basis? To address these questions it is important to note that the extremal vectors of the convex set of states (or effects) are not necessarily linearly independent. As we shall see below, linear independence is in fact a rare exception that emerges only in classical theories, while any non-classicality will be encoded in certain linear dependencies among the extremal states and effects.

Let us illustrate the construction of a suitable basis in the example of a fictitious model with probabilities listed in Table 1.1. As states and effects are defined as equivalence classes, multiple rows and columns have already been eliminated. However, there are still linear dependencies among the rows and the columns. For example, the effect e_5 is related to the other ones by

$$e_5 \;=\; \frac{1}{2}\,(e_1 + e_3)\,. \tag{1.7}$$

Since the expression on the r.h.s. is a convex combination it is automatically

assumed to be part of the toolbox so that we can remove the rightmost column from the probability table, obtaining a reduced table in form of a 4×4 matrix. The remaining (non-convex) linear dependencies are

$$e_4 \;=\; \frac{2}{3}\, e_1 - \frac{2}{3}\, e_2 + \frac{1}{3}\, e_3\,, \qquad \omega_4 \;=\; -\omega_1 + \omega_2 - \omega_3\,. \tag{1.8}$$

so that the rank of the matrix is 3. Since row and column rank of a matrix coincide, the vector spaces V and V^* always have the same dimension

$$n := \dim V = \dim V^* = \mathrm{rank}[\{p_{ij}\}]. \tag{1.9}$$

In other words, the number of different states needed to identify an effect is always equal to the number of different effects needed to identify a state.

As for any vector space representation, there is some freedom in choosing a suitable basis. As for the effects, we may simply choose the first n linearly independent effects $e_1, \ldots, e_n$ as a basis of V^*, assigning to them the canonical coordinate representation

$$e_1 = (1,0,0), \quad e_2 = (0,1,0), \quad e_3 = (0,0,1)\,. \tag{1.10}$$

Likewise we could proceed with the states, choosing $\omega_1, \omega_2, \omega_3$ as a basis of V, but then the matrix p_{ij} would be quite complicated whenever we compute $e(\omega)$ according to Eq. (1.5). Therefore it is more convenient to use the so-called *conjugate basis* $\{\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3\}$ which is chosen in such a way that the extremal states are just represented by the corresponding lines in the probability table. In the example given above this means that the states have the coordinate representation

$$\omega_1 = (1,0,1)\,, \quad \omega_2 = \left(\frac{1}{2}, 0, 1\right), \quad \omega_3 = \left(\frac{1}{2}, \frac{1}{2}, 1\right). \tag{1.11}$$

The basis vectors $\{\hat{\omega}_i\}$ can be determined by solving the corresponding linear equations. In the present example, one can easily show that these basis vectors

are given by the (non-convex) linear combinations

$$\hat{\omega}_1 = 2\omega_1 - 2\omega_2, \quad \hat{\omega}_2 = -2\omega_2 + 2\omega_3, \quad \hat{\omega}_3 = -\omega_1 + 2\omega_2\,. \tag{1.12}$$

Using the conjugate basis the bilinear map $e(\omega)$ can be computed simply by adding the products of the corresponding components like in an ordinary Euclidean scalar product.

Recall that the vector spaces V and V^* are *probabilistic* vector spaces which should not be confused with the Hilbert space of a quantum system. For example, probabilistic mixtures cannot be represented by Hilbert space vectors. We will return to this point when discussing specific examples.

1.2.7. Reliability

Realistic experiments are not only noisy but also unreliable in the sense that they sometimes fail to produce a result. For example, a preparation procedure may occasionally fail to create a physical object. Similarly, a detector may sometimes fail to detect an incident particle.

Preparation procedures which create a physical state with certainty are called *reliable*. The same applies to measurement devices which respond to an incident particle with certainty.

An unreliable effect may be thought of as a reliable one that is randomly switched on and off with probability q and $1-q$, as sketched in Fig. 1.3. Applying this effect to a state ω, the probability to obtain a 'click' would be given by $q\,e(\omega)$. This example demonstrates that unreliable effects can consistently be represented as sub-normalized vectors $q\,e \in V^*$ with $0 \leq q < 1$, extending the set of physical effects to a truncated *convex cone* which is shown as a shaded region in in the right panel of Fig. 1.3. The zero vectors of V and V^* represent the extreme cases of preparation procedures and a measurement apparatuses which always fail to work.

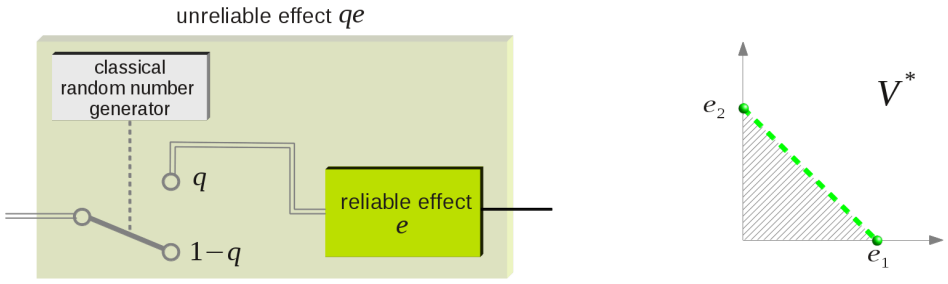


Figure 1.3.: Unreliable effects. Left: A reliable effect e can be made unreliable by randomly switching it on and off, constituting a new effect $q\,e$. Right: Reliable effects are represented by points of the convex set in V^* (green dashed line). Including unreliable effects this set is extended to a truncated convex cone (the hatched region) spanned by the extremal effects.

1.2.8. Unit measure and normalization

If a given effect e responds to a specific state ω with the probability $e(\omega) = 1$, then it is of course clear that both the state and the effect are reliable. However, if $e(\omega) < 1$, there is no way to decide whether the reduced probability for a 'click' is due to the unreliability of the state, the unreliability of the effect, or caused by the corresponding entry in the probability table.

To circumvent this problem, it is usually assumed that the toolbox contains a special reliable effect which is able to validate whether a preparation was successful, i.e. it 'clicks' exactly in case of a successful preparation. This effect is called *unit measure* and is denoted by u. The unit measure allows us to quantify the reliability of states: If $u(\omega) = 1$ the state ω is reliable, otherwise its rate of failure is given by $1 - u(\omega)$.

The unit measure can be interpreted as a norm

$$||\omega|| = u(\omega) \tag{1.13}$$

defined on states in the convex cone of V. By definition, the normalized states with $u(\omega) = 1$ are just the reliable ones. The corresponding set (the green dashed line in Fig. 1.3) is usually referred to as the *state space* Ω of the theory.

In the example of Table 1.1 it is easy to see that the effect e_3 plays the role of the unit measure. Since the unit measure cannot be represented as a convex combination of other effects, it is by itself an extremal effect and thus may be used as a basis vector of V^*. Here we use the convention to sort the Euclidean basis of V^* in such a way that the unit measure appears in the last place, i.e. $e_M \equiv u$. Using this convention the norm of a state is just given by its last component. For example, in Table 1.1, where $e_3 = u$, the third component of all states $\omega_1, \ldots, \omega_4$ is equal to 1, hence all states listed in the table are normalized and thus represent reliable preparation procedures.

The unit measure also induces a norm on effects defined by

$$||e|| = \max_{\omega \in \Omega} e(\omega)\,. \tag{1.14}$$

Since $e(\omega) \leq 1$ an effect is normalized (i.e. $||e|| = 1$) if and only if there exists a state ω for which $\omega(e) = 1$. By definition, such an effect is always reliable. The opposite is not necessarily true, i.e. reliable effects may be non-normalized with respect to the definition in (1.14).

Note that a 'unit state', analogous to the unit effect u, is usually not introduced since this would correspond to a preparation procedure to which *every* reliable effect of the toolbox responds with a '1' with certainty, which is highly unphysical. If we had introduced such a 'unit state', it would have allowed us to define a norm on effects analogous to Eq. (1.13), preserving the symmetry between states and effects. Using instead the norm (1.14) breaks the symmetry between the spaces V and V^*.

As we will see in the following, the unit measure u plays a central role in the context of consistency conditions and it is also needed to define measurements with multiple outcomes. Moreover, the definition of subsystems in Sect. 1.5.4 relies on the unit measure.

1.2.9. General consistency conditions

The concepts introduced so far represent only the factual experimental observations and immediate probabilistic consequences. However, the purpose of a

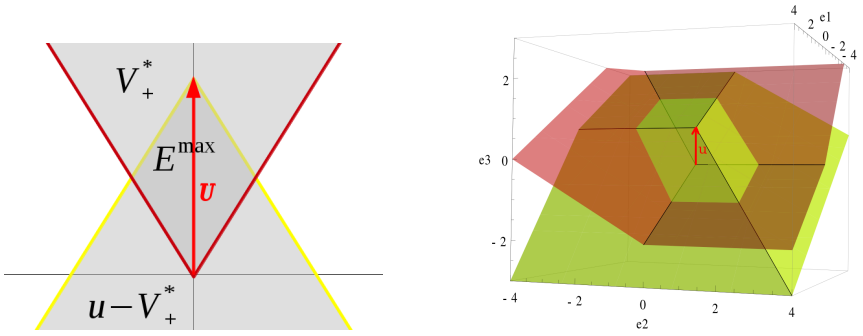


Figure 1.4.: Consistency conditions. Left: Schematic illustration of the lower and the upper bound, defining the intersection $E^{\max}$. Right: The same construction for the probabilities listed in Table 1.1 in the three-dimensional representation (1.10). The red (yellow) planes indicate the lower (upper) bound. The maximal set of effects $E^{\max}$ is the enclosed parallelepiped in the center.

physical model is not only to reproduce the existing data but rather to make new predictions, eventually leading to a set of hypotheses that can be tested experimentally.

In order to give a GPT the capability of making new predictions one has to postulate additional extremal states and effects which are not yet part of the existing toolbox. Such an extension is of course not unique, rather there are various possibilities which can be justified in different ways. For example, a particular extension might be reasonable in view of the underlying structure and the expected symmetries of the physical laws. Moreover, certain expectations regarding the relationship between the parameters of the apparatuses and the corresponding states and effects as well as analogies to other models could inspire one to postulate a specific structure of the state space and the set of effects. This includes dynamical aspects of the systems, which are absorbed into preparations and measurements in the present framework.

However, not every extension of states and effects gives a consistent theory. First of all, the extension should be introduced in such a way that any

combination of effects and states yields a probability-valued result, i.e.,

$$0 \leq e(\omega) \leq 1 \quad \forall e \in E, \omega \in \Omega. \tag{1.15}$$

This restriction consists of a lower and an upper bound. The lower bond $0 \leq e(\omega)$, the so-called *non-negativity constraint*, remains invariant if we rescale the effect e by a positive number. In other words, for any effect e satisfying the non-negativity constraint, the whole positive ray $\lambda\, e$ with $\lambda \geq 0$ will satisfy this constraint as well. The set of all rays spanned by the non-negative effects is the so-called *dual cone*, denoted as

$$V_+^* := \{e \in V^* \,|\, e(\omega) \geq 0 \,\forall \omega \in \Omega\}\,. \tag{1.16}$$

The upper bound can be expressed conveniently with the help of the unit measure u. Since the unit measure is the unique effect giving 1 on all normalized states, it is clear that $e(\omega) \leq 1$ if and only if $u(\omega) - e(\omega) = (u - e)(\omega) \geq 0$, i.e., the complementary effect $u - e$ must be included in the dual cone given by (1.16). Note that this criterion is valid not only for normalized states but also for sub-normalized states. This means that the set of effects, which obey the upper bound $e(\omega) \leq 1$, is just $u - V_+^*$. Consequently, the set which satisfies both bounds in (1.15), is just the intersection of V_+^* and $u - V_+^*$, as illustrated in Fig. 1.4. This maximal set of effects is denoted by[3]

$$E^{\max} \;=\; V_+^* \cap \left(u - V_+^*\right). \tag{1.17}$$

Thus, if we extend the theory by including additional effects, the resulting set of effects E has to be a subset of this maximal set. Similarly we may extend the theory by including additional states. Here we have to specify the set of states which satisfy (1.15) for a given set of effects E. Generally the inclusion of additional states imposes additional restrictions on possible effects and vice versa. Consequently, there is a trade-off between states and effects whenever a

[3] In the literature this set is also denoted by $[\emptyset, u]$ because of a partial ordering induced by V_+^*, as we explain in more detail in appendix A.1.

theory is extended without changing the dimension of the vector spaces V and V^*.

The GPT framework always saturates this trade-off. That is, for a given state space it includes the full set $E^{\max}$. Consequently, state spaces and effect sets are completely determined by each other and cannot be chosen independently. This condition is referred to as the *no-restriction hypothesis* [5]. Classical probability theory and quantum theory both satisfy the no-restriction hypothesis. We will keep the no-restriction hypothesis for most of the work presented in this thesis.

Note, however, in general there seems no operational reason to assume that the preparations in our current toolbox should fully determine the range of possible measurements. This motivated us to study restricted systems and the consequences of such a generalization to the fundamental concepts of the GPT framework in Chapter 4.

A given GPT can also be generalized by increasing the dimension of V and V^*. In fact, as will be shown in Sect. 1.4, every non-composite system from an arbitrary GPT can equivalently be realized as a classical theory in a higher-dimensional state space combined with suitable restrictions on the effects. However, as we will see in Sect. 1.5, the treatment of multipartite systems leads to additional consistency conditions which cannot be fulfilled by restricted classical systems in higher dimensions, allowing us to distinguish classical from genuine non-classical theories.

1.2.10. Jointly measurable effects

A set of effects is said to be *jointly measurable* if all of them can be evaluated in a single measurement, meaning that there exists a measurement apparatus that contains all these effects. By definition, effects belonging to the same measurement apparatus are jointly measurable. However, a GPT may also include effects that cannot be measured jointly. Therefore, it is of interest to formulate a general criterion for joint measurability.

Before doing so, let us point out that joint measurability neither requires the effects to be evaluated at the same time nor does it mean that they do not in-

fluence each other. For example, let us consider a non-destructive measurement of effects $\{e_i^1\}$ with results $\{\chi_j^1\}$ followed by a second measurement. The results $\{\chi_j^2\}$ of the second measurement correspond to effects e_j^2 with the proviso that the first measurement has already been carried out. If the first measurement was not carried out, we would obtain potentially different effects. Nevertheless, the whole setup measures all effects $\{e_i^1\}$ and $\{e_j^2\}$ jointly, irrespective of the fact that the second group depends on the first one.

Joint measurability of effects is in fact a weaker requirement than non-disturbance and commutativity of measurements. In standard quantum theory these terms are often erroneously assumed to be synonyms. This is because in the special case of projective measurements they happen to coincide. However, as shown in [13, 14], they even differ in ordinary quantum theory in the case of non-projective measurements.

Let us now formally define what joint measurability means. Consider two effects e_i and e_j. Applied to a state ω each of them produces a classical one-bit result $\chi_i \in \{0,1\}$ and $\chi_j \in \{0,1\}$. Joint measurability means that there exists another single measurement apparatus in the toolbox that allows us to extract two bits $(\tilde{\chi}_i, \tilde{\chi}_j)$ by Boolean functions with the same measurement statistics as (χ_i, χ_j).

In other words, two effects e_i, e_j are jointly measurable if the toolbox already contains all effects which are necessary to set up the corresponding Boolean algebra, i.e. there are mutually excluding effects $e_{i\wedge j}, e_{i\wedge \bar{j}}, e_{\bar{i}\wedge j}, e_{\bar{i}\wedge \bar{j}}$ with the properties

$$\begin{aligned} e_i = e_{i\wedge j} + e_{i\wedge \bar{j}}\,, \qquad e_j = e_{i\wedge j} + e_{\bar{i}\wedge j} \\ u = e_{i\wedge j} + e_{i\wedge \bar{j}} + e_{\bar{i}\wedge j} + e_{\bar{i}\wedge \bar{j}} \\ e_{i\vee j} = e_i + e_j - e_{i\wedge j}\,. \end{aligned} \tag{1.18}$$

Let us use Eqs. (1.18) to rewrite $e_{i\wedge j}$ in three different ways:

$$\begin{aligned} e_{i\wedge j} &= e_i - e_{i\wedge\bar{j}} \\ &= e_j - e_{\bar{i}\wedge j} \\ &= e_i + e_j - u + e_{\bar{i}\wedge\bar{j}} \,. \end{aligned} \tag{1.19}$$

We can now translate the joint measurability condition to

$$\exists e_1, e_2, e_3, e_3 \in E: \;\; e_1 \,=\, e_i - e_2 \,=\, e_j - e_3 \,=\, e_i + e_j - u + e_4 \,. \tag{1.20}$$

This condition can be rewritten elegantly as an intersection of sets

$$E \cap (e_i - E) \cap (e_j - E) \cap (e_i + e_j - u + E) \neq \{\}. \tag{1.21}$$

For joint measurability of the effects e_i, e_j this set has to be non-empty. If this is the case, any choice of the AND effect $e_{i\wedge j}$ in the intersection (1.21) allows one to consistently construct all other effects by means of Eqs. (1.18). This means that joint measurability of two effects can be implemented in various ways with different $e_{i\wedge j}$. Note that the status of joint measurability of a given set of effects may even change when a theory is extended by including additional effects.

1.2.11. Complete and incomplete Measurements

A *measurement* is defined as a set of jointly measurable effects. If these effects have a non-trivial overlap $e_{i\wedge j} \neq 0$ we can further refine the measurement by including the corresponding AND effects. Thus, we can describe any measurement by a set of mutually excluding effects $\{e_k\}$, where only one of the outcomes χ_k occurs, as sketched in Fig. 1.1. These refined effects have no further overlap, i.e. $e_{k\wedge l} = 0$ for $k \neq l$. Moreover, these effects can be coarse-grained by computing their sum $e_{k\vee l} = e_k + e_l$.

A measurement is called *complete* if all mutually excluding effects sum up to the unit measure u. Obviously an incomplete measurement can be completed by including a failure effect $e_m = u - \sum_{i=1}^{m-1} e_i$ that is complementary to all

other effects. As a consequence a complete measurement maps a normalized state to a normalized probability distribution.

We will show in Chapter 4 in Theorem 4.2 that we need to treat every measurement fundamentally as such a complete measurement in order to prevent inconsistencies in joint systems.

1.2.12. Equivalent Representations

Consider applying arbitrary bijective linear maps L^T on all effects and the corresponding inverse map L^{-1} on all states. This leaves the results from any combination of effects and states invariant, since:

$$\big(L^T\cdot e\big)\big[L^{-1}\cdot\omega\big] = \big(L^T\cdot e\big)^T\cdot L^{-1}\cdot\omega = e^T\cdot L\cdot L^{-1}\cdot\omega = e^T\cdot\omega = e(\omega). \tag{1.22}$$

Now, a particular probabilistic theory is associated with a particular state space Ω and set of effects E. But theories are distinguished only by the different measurement statistics that are possible (as is guaranteed by using the equivalence principle). Hence if Ω and E are transformed according to (1.22), then the resulting Ω' and E' define the same theory, since this transformed state space and effect set yield the same measurement statistics.

1.3. Examples

1.3.1. Classical probability theory

Classical systems have properties that take definite perfectly distinguishable values that can be directly revealed via measurements. Probabilistic mixtures can be regarded as a mere consequence of subjective ignorance.

In the GPT framework the different possible definite values of a classical system are represented by the pure states ω_i. They are linearly independent and can be used as an Euclidean basis of the linear space V. The corresponding state space is a probability simplex (see Fig. 1.5). Probabilistic mixtures are represented by convex combinations of pure states. As the pure states form a

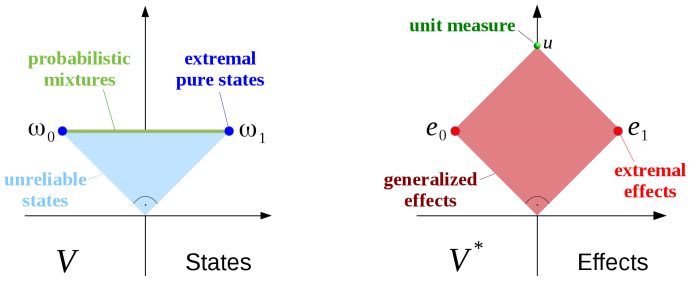


Figure 1.5.: State and effect space of a classical bit in the GPT formalism with the probability table $e_i(\omega_j) = \delta_{ij}$. In classical systems the extremal states and effects are linearly independent and can be used as an orthonormal basis of the vector spaces.

basis, any mixed state can be uniquely decomposed into pure states weighted by the probabilities of occurrence.

The perfect distinguishability of pure states means that the extremal effects e_j simply read out whether a particular value has been realized or not, i.e. $e_j(\omega_i) = \delta_{ij}$. Like the pure states in V these effects provide an Euclidean basis for V^*. Furthermore, the zero effect $\emptyset$, and coarse-grained basis effects e_j have to be included as additional extremal effects. In particular, this includes the unit measure u which is obtained by coarse-graining all basis effects e_j. The unit measure responds with a '1' to any successful preparation of a classical system, independent of its values. In classical systems all effects are jointly measurable.

1.3.2. Standard quantum theory: State space

Most textbooks on quantum theory introduce quantum states as vectors $|\Psi\rangle$ of a complex Hilbert space $\mathcal{H}$. These vectors represent pure quantum states. The existence of a Hilbert space representation is in fact a special feature of quantum mechanics. In particular, it allows one to combine any set of pure states $|\Psi_i\rangle$ linearly by coherent superpositions

$$|\phi\rangle = \sum_i \lambda_i |\psi_i\rangle\,, \qquad \lambda_i \in \mathbb{C}\,, \qquad \sum_i |\lambda_i|^2 = 1\,. \tag{1.23}$$

Note that the resulting state $|\phi\rangle$ is again a pure state, i.e. coherent superpositions are fundamentally different from probabilistic mixtures. In fact, Hilbert space vectors alone cannot account for probabilistic mixtures.

To describe mixed quantum states one has to resort to the density operator formalism. To this end the pure states $|\Psi\rangle$ are replaced by the corresponding projectors $\rho_\Psi = |\Psi\rangle\langle\Psi|$. Using this formulation one can express probabilistically mixed states as convex combinations of such projectors, i.e.

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i| \,, \qquad \sum_i p_i = 1 \,. \tag{1.24}$$

As the expectation value of any observable $\mathbf{A}$ is given by $\mathrm{tr}[\rho\mathbf{A}]$, it is clear that the density matrix includes all the available information about the quantum state that can be obtained by means of repeated measurements.

It is important to note that the density matrix itself does not uniquely determine the p_i and $|\psi\rangle_i$ in (1.24), rather there are many different statistical ensembles which are represented by the same density matrix. For example, a mixture of the pure qubit states $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ with equal probability, and a mixture $|+\rangle\langle +|$ and $|-\rangle\langle -|$ of the coherent superpositions $|\pm\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \pm |1\rangle\right)$ are represented by the same density matrix

$$\rho = \frac{1}{2}\Big(|0\rangle\langle 0| + |1\rangle\langle 1|\Big) = \frac{1}{2}\Big(|+\rangle\langle +| + |-\rangle\langle -|\Big) \,, \tag{1.25}$$

meaning that these two ensembles cannot be distinguished experimentally. Thus, in ordinary quantum mechanics the density matrices ρ label equivalence classes of indistinguishable ensembles and therefore correspond to the physical states ω in the GPT language. The set of all quantum states (including probabilistic mixtures) can be represented by Hermitean matrices with semi-definite positive eigenvalues. A state is normalized if $\mathrm{tr}[\rho] = 1$, reproducing the usual normalization condition $\langle\psi|\psi\rangle = 1$ for pure states.

Identifying the density operators as states, one faces the problem that these operators live in a complex-valued Hilbert space whereas the GPT framework introduced above involves only real-valued vector spaces. In order to embed

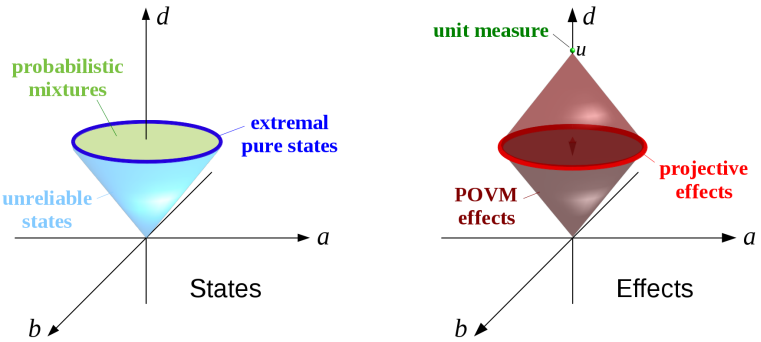


Figure 1.6.: State and effect spaces of a quantum-mechanical qubit in the GPT formalism. Since the vector space are four-dimensional the figure shows a three-dimensional projection, omitting the third coefficient c.

quantum theory in the GPT formalism, let us recall that a $n \times n$ density matrix can be parametrized in terms of $SU(n)$ generators with real coefficients. For example, the normalized density matrix of a qubit can be expressed in terms of $SU(2)$-generators (Pauli matrices) as

$$\rho = \frac{1}{2}\left(\mathbf{1} + a\,\sigma^x + b\,\sigma^y + c\,\sigma^z\right) \tag{1.26}$$

with real coefficients $a, b, c \in [-1, 1]$ obeying the inequality $a^2 + b^2 + c^2 \leq 1$. Regarding the coefficients (a, b, c) as vectors in $\mathbb{R}^3$, the normalized states of a qubit form a ball in three dimensions. The extremal pure states are located on the surface of this ball, the so-called *Bloch sphere*.

In order to include non-normalized states (e.g. unreliable preparation procedures), we have to append a forth coefficient d in front of the unit matrix, i.e.

$$\rho = \frac{1}{2}\left(d\,\mathbf{1} + a\,\sigma^x + b\,\sigma^y + c\,\sigma^z\right) \tag{1.27}$$

which is 1 for any normalized state and less than 1 if the preparation procedure is unreliable. The four coefficients (a, b, c, d) provide a full representation of the state space in $\mathbb{R}^4$ according to the GPT conventions introduced above. This state space is illustrated for the simplest case of a qubit in the left panel of Fig. 1.6.

1.3.3. Standard quantum theory: Effect space

As there are pure and mixed quantum states there are also two types of measurements. Most physics textbooks on quantum theory are restricted to 'pure' measurements, known as projective measurements. A projective measurement is represented by a Hermitean operator $\mathbf{A}$ with the spectral decomposition

$$\mathbf{A} = \sum_a a \, |a\rangle\langle a| \tag{1.28}$$

with real eigenvalues a and a set of orthonormal eigenvectors $|a\rangle$. If such a measurement is applied to a system in a pure state $|\psi\rangle$ it collapses onto the state $|a\rangle$ with probability $p_a = |\langle a|\psi\rangle|^2$. Introducing projection operators $E_a = |a\rangle\langle a|$ and representing the pure state by the density matrix $\rho = |\psi\rangle\langle\psi|$ this probability can also be expressed as

$$p_a = |\langle a|\psi\rangle|^2 = \langle a|\psi\rangle\langle\psi|a\rangle = \mathrm{tr}[E_a^\dagger \, \rho], \tag{1.29}$$

i.e. the absolute square of the inner product between bra-ket vectors is equivalent to the Hilbert-Schmidt inner product of operators E_a and ρ. Hence we can identify the projectors $E_a = |a\rangle\langle a|$ with extremal effects in the GPT framework, where $e_a(\omega) = \mathrm{tr}[E_a\rho]$. As the projectors E_a cannot be written as probabilistic combinations of other projectors, it is clear that they represent extremal effects. As all these effects sum up to $\sum_a E_a = \mathbb{1}$, the unit measure u is represented by the identity matrix.

Turning to generalized measurements, we may now extend the toolbox by including additional effects which are defined as probabilistic mixtures of projection operators of the form

$$E_a = \sum_i q_i |a_i\rangle\langle a_i| \, , \qquad 0 \leq q_i \leq 1. \tag{1.30}$$

As outlined above, such mixtures can be thought of as unreliable measurements. A general measurement, a so-called *positive operator valued measurement* (POVM), consists of a set of such effects that sum up to the identity.

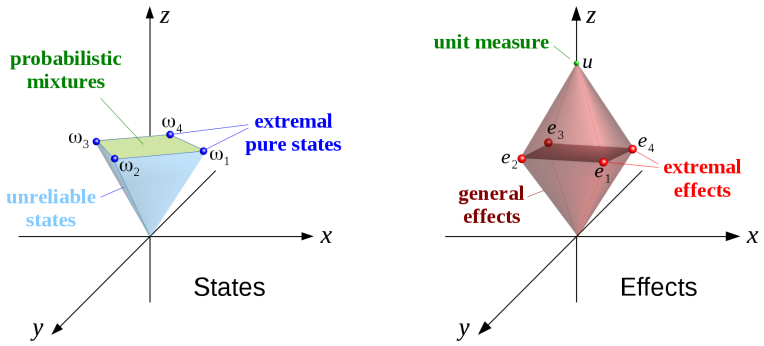


Figure 1.7.: State and effect spaces of a gbit.

Interestingly, the generalized effects in Eq. (1.30) are again positive operators. i.e. mixed effects and mixed quantum states are represented by the same type of mathematical object. Therefore, quantum theory has the remarkable property that the spaces of states and effects are isomorphic. In the GPT literature this special property is known as (strong) *self-duality*.

Note that for every given pure state $\rho = |\Psi\rangle\langle\Psi|$ there is a corresponding measurement operator $E = |\Psi\rangle\langle\Psi|$ that produces the result $\mathrm{tr}[E\,\rho] = 1$ with certainty on this state. In so far the situation is similar as in classical systems. However, in contrast to classical systems, it is also possible to obtain the same outcome on other pure states with some probability. This means that in quantum mechanics pure states are in general not perfectly distinguishable.

1.3.4. The gbit

A popular toy theory in the GPT community, which is neither classical nor quantum, is *boxworld*. The simplest system in this theory is given by the *generalized bit*, usually simply called gbit. It has a square-shaped state space defined by the convex hull of the following extremal states ω_i:

$$\omega_1 = (1,0,1), \quad \omega_2 = (0,1,1), \quad \omega_3 = (-1,0,1), \quad \omega_4 = (0,-1,1)\,. \tag{1.31}$$

The corresponding set of effects is determined by the no-restriction hypothesis. That is, it includes all linear functionals that give probability-valued results

when applied to states. This results in an effect space given by the convex hull of the following extremal vectors:

$$\begin{aligned} e_1 = \frac{1}{2}(1,1,1) \qquad & e_2 = \tfrac{1}{2}(-1,1,1), \\ e_3 = \frac{1}{2}(-1,-1,1), \qquad & e_4 = \tfrac{1}{2}(1,-1,1) \end{aligned} \tag{1.32}$$

together with the zero effect and the unit measure

$$\emptyset = (0,0,0), \quad u = (0,0,1). \tag{1.33}$$

Remarkably, in contrast to the classical and the quantum case, each extremal effect e_i gives certain outcomes on more than one extremal state.

1.3.5. Other toy theories

Further examples of state spaces discussed in the literature include a complicated three-dimensional cushion-like state space to model three-slit interference [15], a cylinder-shaped state space [16], hypersheres of arbitrary dimensions used as a generalization of the Bloch sphere of qubits [17], a three-dimensional state space with triangle-shaped and disc-shaped subspaces [18] and a three-dimensional approximation of the Bloch ball with finite extremal states [19].

The study of physical principle in this work deploys a variety of new toy theories. A whole class of toy theories with two-dimensional state spaces given by regular polygons will be discussed in section 2.1 and were originally introduced in the authors master thesis [20]. Remarkably, these include some of the above standard cases, which correspond to state spaces with a particular number n of vertices. The state space of a classical theory with three distinguishable pure states is given by a triangle-shaped state space, that is the regular polygon with three vertices ($n = 3$). The square shaped state space of the gbit corresponds to $n = 4$. In the limit $n \to \infty$, we get a two-dimensional subspace of a qubit, inheriting some of the quantum features. The polygon theories can therefore be used to compare the different standard cases. Furthermore, increasing the

number of vertices yields a transition from a classical theory and a gbit to a quantum-like theory in the limit of infinitely many vertices.

A house-shaped toy theory is studied in Section 2.4. Section 2.1.1 introduces another class of toy theories. This theories allow a continuous transition between a classical system and a gbit. Such a toy theory consists of two-dimensional state spaces with four vertices. The location of one of the vertices is parametrized, such that the square and the triangle appear as special cases for particular parameters.

All the theories mentioned so far, include the full set of potential effects. That is, they obey the no-restriction hypothesis. Chapter 4 studies the consequences of relaxing this condition and introduces toy theories with restricted effect sets. Particular interesting examples are theories with inherent noise and a construction that mimics the state-effect duality of quantum theory by restricting the effect set of general theories. Another example of a restricted GPT is the probabilistic version of Spekken's toy theory [21], given by octrahedron-shaped state space and effect set.

1.3.6. Special features of quantum theory

Having introduced some examples of GPTs let us now return to the question what distinguishes quantum mechanics as the fundamental theory realized in nature from other possible GPTs. Although a fully conclusive answer is not yet known, one can at least identify various features that characterize quantum mechanics as a particularly elegant theory.

Continuous state and effect spaces:

Comparing the state and effect spaces in Figs. 1.6 and 1.7 visually, one immediately recognizes that quantum theory is special in so far as extremal states and effects form *continuous manifolds* instead of isolated points. In the Hilbert space formulation this allows one to construct coherent superpositions and to perform reversible unitary transformations, giving the theory a high degree of symmetry. Note that coherent superpositions and probabilistic mixtures are

very different in character: While mixtures exist in all GPTs, coherent superpositions turn out to be a special property of quantum theory. GPTs do in general not admit reversible transformations between different extremal states which would be a prerequisite for the possibility of superpositions [21, 22].

Quantum theory does not only allow one to relate pure states by reversible unitary transformations (transitivity) [1, 23, 24], but even mixed states can be reversibly transformed into each other (homogeneity) [25]. Moreover, the continuous manifold of infinitely extremal quantum states does not require infinite-dimensional vector spaces. For example, the state space of a qubit (Bloch ball) is three-dimensional although it has infinitely many extremal points.

Distinguishability and sharpness:

The possibility of reversible transformations between extremal states has direct consequences in terms of the information processing capabilities of a theory [26]. As we have seen, in non-classical theories pairs of states are in general not perfectly distinguishable. Remarkably, quantum theory is also special in so far as it allows for a weaker notion of perfect distinguishability [27], namely, for any extremal state one can find a certain number of other perfectly distinguishable extremal states (the orthogonal ones in the Hilbert space formulation). This number is called the *information capacity* of the system which corresponds to the classical information that can be encoded in such a subsystem of distinguishable states. In quantum theory it is equal to the dimension of the Hilbert space.

Obviously, any GPT with given state and effect spaces has well-defined subsets of perfectly distinguishable states and therewith a well-defined information capacity. Remarkably, for quantum theory the opposite is also true, i.e., the information capacity of a system can be shown to determine its state space [1, 23]. As a consequence it turns out that a system of given information capacity includes non-classical ones with a lower information capacity as subspaces [24], which allows for an ideal compression of the encoded information [27]. This embedding is reflected by a rich geometrical structure of state spaces that is still rather unexplored. The interested reader may be referred to [28] for an

detailed discussion of the geometry of quantum state spaces. An example of a state space that is still rather low dimensional, but nevertheless has a highly non-trivial structure, is the *qutrit*, a quantum system with information capacity three. It has extremal points that lie on the surface of an eight-dimensional ball with radius $\sqrt{2/3}$. However, the sphere is only partially covered with extremal states. In particular, for any pure state of a qutrit there is a subspace with information capacity 2 including all states that can be perfectly distinguished from the first one. As quantum systems with information capacity 2 are represented by the three-dimensional Bloch ball, we can conclude that there is an opposing Bloch-ball-shaped facet for any extremal point of the qutrit state space.

Quantum theory is also special in so far as for any extremal state ω there exists a unique extremal effect e which gives $e(\omega) = 1$ while it renders a strictly lower probability for all other extremal states. Therefore, this effect allows one to unambiguously identify the state ω. The existence of such identifying effects is another special quantum feature known as sharpness [29].

Strong self-duality:

Another striking feature of quantum theory is the circumstance that extremal states and the corresponding identifying effects are represented by the same density operator. This is related to the fact that quantum theory is (strongly) self-dual, i.e. the cone of non-normalized states and its dual cone coincide [30] and obey the no-restriction hypothesis [5]. It was shown that this is a consequence of bit symmetry, i.e. all pairs of distinguishable states can be reversibly transformed into each other [31].

To summarize, from the perspective of GPTs quantum theory has remarkable characteristic properties which may give us an idea why this theory is the one realized in Nature. On the other hand, various other features that seem special for quantum theory turned out to be common for non-classical theories within the GPT framework. Examples are the operational equivalence of different ensembles and the impossibility to clone a state [32].

Note, that so far we have only discussed single (i.e. non-composite) systems. Already on this level quantum theory exhibits very special features that are hard to find in any other toy theories. Nevertheless, as we will show in the next section, any non-classical single system can be simulated by a higher dimensional classical system with appropriate restrictions.

1.4. Non-classicality by restriction

Any non-classical single (non-partitioned) system can be interpreted as a classical system with appropriately restricted effects in higher dimensions. This can easily be illustrated in the example of Table 1.1. Suppose we extend the table by one additional column for each preparation procedure ω_i which contains a '1' for ω_i and '0' otherwise (see Table 1.2). Obviously, these additional columns can be interpreted as additional effects that allow us to perfectly distinguish different preparation procedures, just in the same way as in a classical model. In other words, by adding these columns we have extended the model to a classical one in a higher-dimensional space, where each of the preparation procedures is represented by a different pure state. The original effects can simply be interpreted as coarse-grained mixtures of the additional effects.

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9
ω_1	1	0	1	1	1	1	0	0	0
ω_2	$\frac{1}{2}$	0	1	$\frac{2}{3}$	$\frac{3}{4}$	0	1	0	0
ω_3	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{3}{4}$	0	0	1	0
ω_4	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	0	0	0	1

Table 1.2.: Copy of Table 1.1 extended by four additional effects e_6, e_7, e_8, e_9, converting the non-classical theory into a classical one in higher dimensional space.

Conversely, it is also possible to restrict a classical system in such a way that it seems to acquire non-classical features. Such an example was given by Holevo in 1982 [33]: Take a classical system with four pure states

$$\omega_1 = (1,0,0,0), \quad \omega_2 = (0,1,0,0), \quad \omega_3 = (0,0,1,0), \quad \omega_4 = (0,0,0,1) \tag{1.34}$$

representing four distinguishable values. These extremal states span a three-dimensional tetrahedron of normalized mixed states embedded in four-dimensional space. The corresponding extremal effects are given by the vertices of the four-dimensional hypercube $e = (x_1, x_2, x_3, x_4)$ with $x_i \in \{0, 1\}$, including the zero effect $\emptyset = (0, 0, 0, 0)$ and the unit measure $u = (1, 1, 1, 1)$. By definition, two states $\omega = (y_1, y_2, y_3, y_4)$ and $\omega' = (y'_1, y'_2, y'_3, y'_4)$ are operationally equivalent whenever

$$e(\omega) = e(\omega') \quad \Leftrightarrow \quad \sum_{i=1}^{4} x_i\, y_i = \sum_{i=1}^{4} x_i\, y'_i \tag{1.35}$$

for all available effects e, which in this case means that all components $y_i = y'_i$ coincide.

Now, let us restrict our toolbox of effects to a subset where

$$x_1 + x_2 = x_3 + x_4\,. \tag{1.36}$$

As a result, ω and ω' can be operationally equivalent even if the components y_i and y'_i are different. More specifically, if there is a $t \neq 0$ such that

$$y'_1 = y_1 + t, \quad y'_2 = y_2 + t, \quad y'_3 = y_3 - t, \quad y'_4 = y_4 - t\,, \tag{1.37}$$

then the restricted toolbox of effects does not allow us to distinguish the two states, hence ω and ω' now represent the *same* state in the restricted model.

The extended operational equivalence allows us to choose one of the components, e.g. to set $y_4 = 0$. This means that the four-dimensional parameter space is projected onto a three-dimensional subspace, and the embedded three-dimensional tetrahedron is again projected to a two-dimensional convex object. This projected state space turns out to have the form of a square, as shown schematically in Fig. 1.8. As we have seen before, this is just the state space of a (non-classical) gbit. Therefore, the restriction (1.36) leads effectively to a non-classical behavior. In fact, Holevo showed that such a construction is possible for any probabilistic theory including quantum theory[4].

[4]Note that quantum theory has infinitely many extremal states. The construction therefore requires an infinitely dimensional classical system.

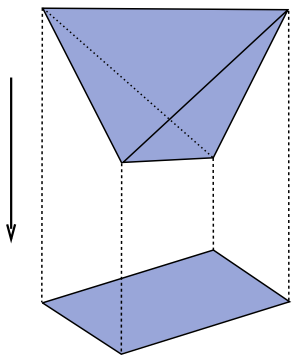


Figure 1.8.: Construction of the two-dimensional gbit state space by projecting a three-dimensional classical state space (adapted from [33]).

To summarize, any non-classical GPT can be extended to a higher-dimensional classical theory by including additional effects. Conversely, non-classical theories can be deduced from a classical one by imposing appropriate restrictions on the available effects. The restrictions allow us to project the classical state space to a non-classical one in lower dimensions, inheriting phenomena like uncertainty relations and non-unique decompositions of mixed states.

Obviously this seems to question the fundamental necessity of non-classical systems. How do we know that all the unusual phenomena in quantum theory do not only result from restrictions in some higher-dimensional space and thus can be explained in classical terms once we extend our theory? However, at this point we should keep in mind that so far we considered only single (non-composite) systems. As we will see in the following section, *multipartite* non-classical systems cannot be described in terms of restricted classical systems. Thus, it would be misleading to conclude that non-classicality only results from restrictions imposed on an underlying higher-dimensional classical system. In fact, the analysis of multipartite systems will allow us to clearly distinguish classical and genuinely non-classical physical systems.

1.5. Multipartite systems

Multipartite systems may be thought of as consisting of several subsystems in which the same type of theory applies. Since such a composed system in itself can be viewed as a single system simply by ignoring its subsystems structure, the consistency conditions discussed in the previous sections obviously apply to multipartite systems as well. However, it turns out that additional consistency conditions arise from the fact that the theory has to be compatible with the partition into given subsystems. In fact, as we will see below, the structure of the subsystems determines a smallest and largest set of joint states and effects that are compatible with the given partition. The actual set of joint elements can be chosen freely within these constraints. This means that a GPT is not yet fully specified by defining states, effects, and the probability table of a single system, instead it is also required to specify how individual systems can be combined to more complex composite systems that embed the joint elements, as the tensor product of unbounded linear spaces is unique. More specifically, one has to define suitable *tensor product* for the sets of states and effects, which is part of the definition of the theory[5].

1.5.1. Separability and the minimal tensor product

In the simplest case, the composite system consists of several independently prepared components. As these subsystems are statistically independent, the joint state describing the overall situation is given by a product state.

As an example let us consider two subsystems A and B which are in the states ω^A and ω^B, respectively. If these systems are completely independent, their joint state is given by a product state, denoted as $\omega^{AB} = \omega^A \otimes \omega^B$. Similarly, the effects of the two subsystems can be combined in product effects $e^{AB} = e^A \otimes e^B$, describing statistically independent measurements on both

[5]The term "tensor product" has been chosen because it formally represents a tensor product of partially ordered vector spaces. Physically, it is just the set of all non-normalized joint elements for given partition into subsystems. This must not be confused with the direct product of the linear spaces V, V^* that embed the joint elements.

sides. In this situation the joint measurement probabilities factorize, i.e.

$$e^{AB}(\omega^{AB}) \;=\; p(e^A \otimes e^B|\omega^A \otimes \omega^B) \;=\; p(e^A|\omega^A)\, p(e^B|\omega^B) \;=\; e^A(\omega^A)\, e^B(\omega^B). \tag{1.38}$$

As a next step, we include classical correlations by randomly choosing preparation procedures and measurement apparatuses in a correlated manner. Formally this can be done by probabilistically mixing the product elements defined above. For example, classically correlated states may be incorporated by including probabilistic linear combinations of the form $\omega^{AB} = \sum_{ij} \lambda_{ij}\, \omega_i^A \otimes \omega_j^B$ with positive coefficients $\lambda_{ij} > 0$. Similarly, one can introduce classically correlated effects.

In the GPT framework the mathematical operation, which yields the set of product elements and their probabilistic mixtures, is the so-called the *minimal tensor product*:

$$\begin{aligned} V_+^A \otimes_{\min} V_+^B &:= \{\omega^{AB} = \sum_{ij} \lambda_{ij}\, \omega_i^A \otimes \omega_j^B \;|\; \omega^A \in V_+^A, \omega^B \in V_+^B, \lambda_{ij} \geq 0\} \\ V_+^{A*} \otimes_{\min} V_+^{B*} &:= \{e^{AB} = \sum_{ij} \mu_{ij}\, e_i^A \otimes e_j^B | e^A \in V_+^{A*}, e^B \in V_+^{B*}, \mu_{ij} \geq 0\}\,. \end{aligned} \tag{1.39}$$

Elements in the minimal tensor product are called *separable* with respect to the partition.

The extremal states in the joint state space $V_+^A \otimes_{\min} V_+^B$ are given by the product of extremal subsystem states. Note that the joint states in this space are not necessarily normalized. Normalized separable joint states can be obtained by forming products of normalized single states or mixtures of them. As a criterion for normalization, the joint unit measure $u^{AB} = u^A \otimes u^B$ is the unique joint effect that gives $u^{AB}(\omega^{AB}) = 1$ on all normalized joint states.

If we apply a joint effect to a joint state, the corresponding measurement

statistics is determined by the weighted sum of factorizing probabilities:

$$\begin{aligned} e^{AB}(\omega^{AB}) = p(e^{AB}|\omega^{AB}) \;&= \Big[\textstyle\sum_{i,j} \mu_{ij}\, e_i^A \otimes e_j^B\Big] \big(\textstyle\sum_{kl} \lambda_{kl}\, \omega_k^A \otimes \omega_l^B\big) \\ &= \textstyle\sum_{ijkl} \lambda_{ij}\, \mu_{kl}\, e_i^A(\omega_k^A)\, e_j^B(\omega_l^B). \end{aligned} \tag{1.40}$$

As the number of combinations and the number of coefficients λ_{ij}, μ_{kl} coincide, it is clear that the measurement statistics obtained from such product effects is sufficient to identify a joint state uniquely. This means that the whole information of classically correlated elements in the minimal tensor product can be extracted by coordinated *local* operations carried out in each of the subsystems.

1.5.2. Entanglement in GPTs

The minimal tensor product defines the sets $V_+^A \otimes_{\min} V_+^B$ and $V_+^{A*} \otimes_{\min} V_+^{B*}$ of classically correlated states and effects which can be seen as subsets of certain vector spaces. For *classical* systems one can show that the minimal tensor product already includes all joint elements that are consistent with the division into classical subsystems [34]. However, in non-classical theories there are generally additional vectors representing elements which are non-separable but nevertheless consistent with the subsystem structure and fully identifiable by local operations and classical communication (LOCC). Such states are called *entangled*. As it is well known, entangled states do exist in standard quantum theory.

In the GPT framework both separable and entangled elements can be represented as vectors in the direct product spaces $V^{AB} = V^A \otimes V^B$ and $V^{AB*} = V^{A*} \otimes V^{B*}$. For separable, classically correlated elements this was directly inherited from classical probability theory. The tensor structure for entangled elements is based on the following additional assumptions[26]: i) *local tomography*, which means that coordinated local operations suffice to identify a joint element, ii) *no-signaling*, stating that local operations in one part of the system have no effect on the local measurement statistics in other parts. As elements of the direct product spaces, we can represent joint elements as $n \times m$ matrices, where $n = \dim V^A = \dim V^{A*}$ and $m = \dim V^B = \dim V^{B*}$ are the dimensions

of the subsystems.

Separable and entangled elements decompose into product elements in a different way. By definition, entangled elements are are not included in the minimal tensor product, meaning that they cannot be written as probabilistic mixtures of product elements. Of course they can still be decomposed into a linear combination of product elements, but such a linear decomposition would inevitably include negative coefficients.

As an example from quantum mechanics let us consider a fully entangled two-qubit Bell state

$$|\psi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)\,. \tag{1.41}$$

Choosing for each qubit the normalized extremal Bloch states

$$\{\omega_1, \omega_2, \omega_3, \omega_4\} = \{\frac{\mathbb{1}}{2}, \frac{\mathbb{1}+\sigma^x}{2}, \frac{\mathbb{1}+\sigma^y}{2}, \frac{\mathbb{1}+\sigma^z}{2}\}\,, \tag{1.42}$$

where $\sigma^{x,y,z}$ are Pauli matrices, a straight-forward calculation shows that the pure Bell state $\omega = |\psi^+\rangle\langle\psi^+|$ can be decomposed into a linear combination

$$\omega = 2\omega_{11} - \omega_{12} + \omega_{13} - \omega_{14} - \omega_{21} + \omega_{22} + \omega_{31} - \omega_{33} - \omega_{41} + \omega_{44} \tag{1.43}$$

of the product states $\omega_{ij} = \omega_i \otimes \omega_j$, which obviously contains negative coefficients.

1.5.3. Entanglement as a genuinely non-classical phenomenon

The previous example of the Bell state illustrates that the phenomenon of entanglement gives rise to additional extremal joint elements in the tensor product which are not part of the minimal tensor product. As we will see in the following, the existence of such non-separable elements makes it impossible to consistently 'simulate' a non-classical system by a classical theory in a higher-dimensional state space.

To see this we first note that the product of extremal elements in the sub-

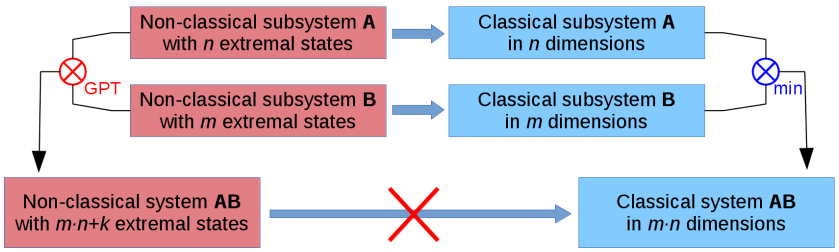


Figure 1.9.: Impossibility to explain entanglement in classical terms. Two non-classical subsystems **A**,**B** containing n and m extremal states are joined to a single non-classical system by means of a nontrivial tensor product $\otimes_{\mathrm{GPT}} > \otimes_{\mathrm{min}}$. The resulting non-classical systems **AB** contains $n \cdot m$ extremal product states plus k additional non-separable extremal states. However, mapping the systems **A**,**B** first to the corresponding higher dimensional classical systems (right side) and combining them by the usual classical tensor product, the resulting $m \cdot n$-dimensional classical system would not be able to account for the additional k entangled states.

systems gives again extremal elements in the composite system. For example, if two non-classical systems A and B have n and m pure states, then the joint system AB possesses at least nm pure product states (see Fig. 1.9). In addition, the joint system also possesses a certain number k of non-separable extremal states, provided that the tensor product is 'larger' than the minimal one.

The existence of non-separable elements is incompatible with the idea of an underlying classical system in a higher-dimensional space with appropriate restriction, as described in Sect. 1.4. The reason is that the combination of two classical systems cannot account for additional non-separable elements. In other words, if we first map the subsystems to the corresponding n- and m-dimensional classical systems and combine them by means of the classical (i.e. minimal) tensor product, the resulting classical system would live in a nm-dimensional space. However, in order to account for the entangled elements, $nm+k$ dimensions would be needed, as illustrated in the figure. In other words, such a construction shows an inconsistent scaling behavior.

The measurement probability $p(e^{AB}|\omega^{AB})$ of an arbitrary joint effect e^{AB} applied to an arbitrary joint state ω^{AB} is still given by (1.40), but in the case

of non-separable elements some of the coefficients λ_{ij} and μ_{kl} will be negative. Since this could lead to negative probabilities $p(e^{AB}|\omega^{AB}) < 0$, further restrictions on the joint elements are needed to ensure positivity.

Entanglement for states and effects is not independent from each other. If we include more and more entangled states, the allowed range of entangled effects becomes smaller, and vice versa. Thus, as in the case of single systems there is a trade-off between entangled states and effects [35]. In particular, if we restrict the range of effects to the minimal tensor product, we can include a certain maximal set of consistent joint states and vice versa. In the following we want to characterize this maximal set of joint states.

1.5.4. Marginal states and conditional states

Before defining the maximal set of joint states, we have to introduce the notion of *marginal states*. To this end let us first consider the measurement statistics of independent local measurements applied to a joint state ω^{AB}, which is given by the joint probability distribution $p(e_i^A, e_j^B|\omega^{AB}) = [e_i^A \otimes e_j^B](\omega^{AB})$. Since the local measurements are independent, we do not have to apply the effects e_i^A und e_j^B at once. In particular, we could observe only the outcome of e^A in part A, ignoring the measurement in part B. The probability of this outcome is given by the marginal probability

$$\begin{aligned} p(e_i^A \,|\, \omega^{AB}) \;&= \textstyle\sum_j p(e_i^A, e_j^B \,|\, \omega^{AB}) = \sum_j [e_i^A \otimes e_j^B](\omega^{AB}) = e_i^A \otimes \left[\sum_j e_j^B\right](\omega^{AB}) \\ &= [e_i^A \otimes u^B](\omega^{AB}) \;=\; e_i^A(\omega^A_{u^B})\,. \end{aligned} \tag{1.44}$$

In the last step of Eq. (1.44) we introduced the so-called *marginal state* $\omega^A_{u^B}$ (analogous to the reduced density matrix from the partial trace in quantum mechanics). This state is the effective subsystem state which predicts the local measurement statistics in part A. Similarly, the marginal state $\omega^B_{u^A}$ determines the measurement statistics in part B.

It is important to note that entangled pure states have mixed marginal states. For example, in standard quantum mechanics the pure state $\rho^{AB} = |\psi^+\rangle\langle\psi^+|$ in Eq. (1.41 has a completely mixed marginal state $\rho^A = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$.

Such a situation, where we have perfect knowledge about the entire system but an imperfect knowledge about its parts, is obviously impossible in classical systems. In addition, the observation leads us to the important insight that the concept of probability in GPTs is not just a matter of incomplete subjective knowledge but rather an important part of the physical laws.

The marginal state $\omega^A_{u^B}$ reflects our knowledge about subsystem A provided that potential measurements on subsystem B are ignored. However, if a particular measurement on B is carried out and the result is communicated to us (via classical communication) our knowledge is of course different. This increased knowledge is accounted for by the conditional probabilities

$$p(e_i^A|e_j^B,\omega^{AB}) = \frac{p(e_i^A, e_j^B \,|\, \omega^{AB})}{p(e_j^B|\omega^{AB})} = \frac{[e_i^A \otimes e_j^B](\omega^{AB})}{e_j^B(\omega^B_{u^A})} = e_i^A\left(\frac{\omega^A_{e_j^B}}{e_j^B(\omega^B_{u^A})}\right) = e_i^A\left(\tilde{\omega}^A_{e_j^B}\right). \tag{1.45}$$

In the last steps we introduced the so-called *conditional state* $\omega^A_{e_j^B}$ and its normalized version $\tilde{\omega}^A_{e_j^B}$. The conditional state $\tilde{\omega}^A_{e_j^B}$ is the effective state in A given that the effect e_j^B was observed in B. The marginal state introduced in Eq. (1.44) is a special conditional state, where the effect e_j^B is just the unit measure in B.

As $\omega^A_{e_j^B}$ depends on the effect e_j^B, it can be interpreted as a linear map from effects in one part onto conditional states in the other.

1.5.5. The maximal tensor product

As outlined above, consistency conditions lead to a trade-off between the sizes of state and effect spaces. Therefore, in order to derive the maximal set of possible joint states, let us assume that the corresponding set of effects is given by the minimal tensor product. Consequently, the joint states have to satisfy two consistency conditions:

1. Applied to product effects they have to give non-negative results.

2. They induce valid conditional states, that is, all conditional states have to be elements of the corresponding subsystem state space.

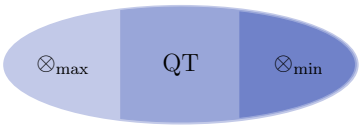


Figure 1.10.: The tensor product of standard quantum theory (QT) lies strictly between the minimal and the maximal tensor product. All tensor products share the extremal states of the minimal tensor product. Quantum theory and the maximal tensor product append additional extremal states.

Note that the second condition always implies the first one, since all factors in (1.45) are non-negative for any valid conditional state. Conversely, in systems obeying the no-restriction hypothesis the first condition also implies the second one. Therefore, it suffices to consider the non-negativity condition alone. With this assumption, the maximal set of non-normalized joint states $V_+^A \otimes_{\max} V_+^B$ for unrestricted systems is just given by the dual cone with respect to product effects:

$$\begin{aligned} V_+^A \otimes_{\max} V_+^B &:= (E_+^A \otimes_{\min} E_+^B)^* \\ &= \left\{ \omega^{AB} \in V^A \otimes V^B \middle| e^A \otimes e^B(\omega^{AB}) \geq 0 \quad \forall e^A \in E^A, e^B \in E^B \right\} . \end{aligned} \tag{1.46}$$

In other words, the maximal tensor product $V_+^A \otimes_{\max} V_+^B$ is simply the set of all joint states which give non-negative results if we apply effects from the minimal tensor product. For restricted systems the situation turns out to get more complicated and requires a generalization of the maximal tensor product, as discussed in Chapter 4.

1.6. Realism versus locality: The meaning of non-local correlations

As shown above entanglement is a strictly non-classical feature of quantum theory. However, a detailed knowledge of the subsystems is required in order to identify a joint element as an entangled one. This issue is addressed in

device-independent set-ups that prove non-classicality without relying on the agreement on a particular theory for subsystems.

The device-independent view abstracts measurement statistics from the specific states and measurements used to produce them. That is, it only considers the probabilities of abstract input-output combinations, with the choice of measurement as input and the outcomes as outputs. Applied to correlations this corresponds to the probability distributions $p(ab|xy)$ to get outcome a and b given that local measurements x and y have been chosen. Note that outputs a, b in this context are no longer referring to particular effects, but are classical labels that simply enumerate outcomes, e.g. the first or the second outcome. Similarly labels x and y for inputs abstract from measurements as specific sets of effects. Specific correlations could in fact result from different combinations of a respective joint state and local measurements. Non-classicality in this context is no longer expressed in non-classical states and effects, but in non-classical probability distributions.

The first and most popular of device-independent non-classicality proofs are the so-called nonlocal correlations that are possible in quantum theory. Already in the early days of quantum mechanics, it was pointed out in the context of the famous EPR gedankenexperiment[36] that the existence of such nonclassical correlations are in conflict with at least one of the following assumptions:

Realism A theory obeys *realism* if measurement outcomes can be interpreted as revealing a property of the system that exists independent of the measurement.

Locality A physical theory obeys *locality* if the measurement on one part of a joint system does not influence measurements on other (spatially separated) parts.

Classical systems satisfy local realism. However, Bell already showed in 1964 that quantum theory can generate correlations that violate at least one of these assumptions [37].

A particularly simple and popular setup that illustrates non-local correlations was introduced by Clauser, Horne, Shimony and Holt (CHSH) in 1969 [38]. The

CHSH setup consists of two parties A and B that share a bipartite quantum state ω^{AB}. Each of the parties chooses between two binary measurements $M_x^A = \{e_{x,a}^A\}_{a=0}^1$ and $M_y^B = \{e_{y,a}^B\}_{a=0}^1$ indexed by $x, y \in \{0, 1\}$. For each choice of x, y we get a probability distribution

$$p(a, b \,|\, x, y) = e_{x,a}^A \otimes e_{y,b}^B(\omega^{AB}) \,. \tag{1.47}$$

The probability distribution generated by a local realistic theory satisfies the *CHSH inequality*

$$S^{\mathrm{LH}} = |C_{0,0} + C_{0,1} + C_{1,0} - C_{1,1}| \leq 2 \tag{1.48}$$

with the correlators

$$C_{x,y} = \sum_{a,b} (-1)^{a \oplus b} \, p(a, b \,|\, x, y). \tag{1.49}$$

In a classical probability theory we have $C_{0,0} = C_{0,1} = C_{1,0} = 1$, implying that the fourth correlator is given by $C_{1,1} = 1$ so that the inequality holds. Quantum theory, however, can violate this inequality as confirmed experimentally in [39]. The theoretical maximum of the CHSH value that can be achieved in quantum theory is given by Tsirelson's bound $S_{\max}^{QT} = 2\sqrt{2}$ [40]. Interestingly, stronger entanglement does not always allow higher violations of a Bell inequality. In fact, it has been shown that in some set-ups even an inverse relationship between entanglement and nonlocality is possible [41]

Also many non-classical GPTs can generate violations of the CHSH inequality. In fact, there exists GPTs that can show violations even exceeding Tsirelson's bound. A frequently studied example is the maximal tensor product of gbits which exhibits so-called PR box correlations [42], violating the CHSH inequality up to its algebraic maximum $S^{PR} = 4$.

Returning to the question of what distinguishes quantum mechanics as the fundamental theory of nature, it is therefore not sufficient to explain the existence of non-classical correlations, one also has to give reasonable arguments why these correlations are not stronger.

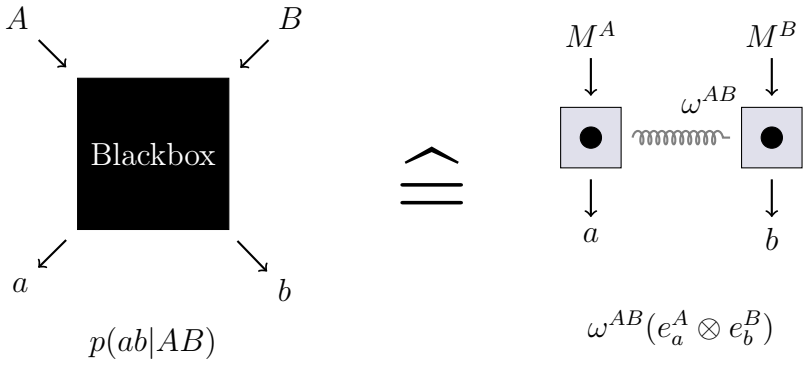


Figure 1.11.: Nonlocal boxes as a device-independent model of nonlocal correlations abstracting from specific measurements and states.

1.7. The no-signaling polytope

Note that up to now we only considered the maximal CHSH violation to characterize device-independent correlations. In order to characterize the exact correlations possible in quantum theory this should be extended to the full device-independent measurement statistics. We therefore consider the full probability distributions for a specific set of inputs and outputs. The abstraction from states and effects is often illustrated by so-called nonlocal boxes as shown in Fig. 1.11. These are black boxes which respond to inputs with certain outputs according to the given probability distribution. This concept allows us to compare general no-signaling correlations to those possible in quantum theory without referring to specific theories.

In general correlations are supposed to occur independently of the separation of the two parts. In particular the parts could be space-like separated, i.e. communication between the parts is regarded impossible. Excluding correlations that would allow superlumial transmission of information, requires the marginal probability distributions to be independent of the choice of the input on the other side of the system. Formally, this refers to the *no-signaling*

principle

$$p(a|x) = \sum_b p(ab|xy) \quad \forall y \tag{1.50}$$
$$p(b|y) = \sum_a p(ab|xy) \quad \forall x.$$

As pointed out before measurement statistics from local measurements on joint GPT systems automatically satisfy this condition.

In the CHSH set-up a, b, x, y are binary. Consequently, the measurement statistics is given by the $2^4 = 16$ probabilities on all possible combinations. But not all these probabilities are independent. First, the probabilities for different outputs a, b sum up to one for any inputs x, y. Apart from the 4 resulting normalization conditions there are 4 conditions from equations (1.50) due to no-signaling. Thus the number of independent probabilities reduces accordingly, which leads to a 8 dimensional space of measurement statistics.

From a geometrical perspective the CHSH term in (1.48) corresponds to a particular direction in this space. The inequality limits the probability distributions in this direction. That is it corresponds to a half-spaces bounded by a hyperplane at $S = 2$. Taking the absolute value implies an equivalent bound in the opposite direction. Further bounds can be achieved by relabelings, that is permutations of the x, y assignments. Together with trivial conditions $p(ab|xy) \in [0, 1]$ on the range of probabilities the intersection of half-spaces yields an eight-dimensional polytope. This so-called *local polytope* $\mathcal{L}$ is spanned by extremal correlations $p(ab|xy) = p(a|x)\, p(b|y)$ with $p(a|x), p(b|y) \in \{0, 1\}$ and includes exactly those correlations that obey local realism.

On the other hand restricting correlations only by the no-signaling principle, allows additional extremal correlations, the so-called PR-boxes of the form

$$p(a, b|x, y) = \begin{cases} \frac{1}{2} & \text{if } a \oplus b = xy \\ 0 & \text{otherwise} \end{cases} \tag{1.51}$$

and equivalent probability distributions with relabeled measurements and out-

comes. These boxes give the algebraic maximum of 4 for the term in (1.48), i.e. violates the CHSH inequality. The PR boxes together with the extremals of the local polytope span the no-signaling polytope $\mathcal{P}$ that includes all correlations that respect the no-signaling principle.

The set of quantum correlations $\mathcal{Q}$ lies between the local and the no-signaling polytope and has infinitely many extremal points, i.e. is not a polytope. The maximum CHSH violation possible for quantum resources is given by Tsirelson's bound of $2\sqrt{2}$ [40]. Although various approximations to $\mathcal{Q}$ have been derived [43, 44, 45, 46] the complete structure of the set of quantum correlations is still unknown. Notably, there are correlations in $\mathcal{P}$ not possible to produce within quantum theory. These post-quantum correlations are regarded to be unphysical. Nevertheless, they have been a subject of extensive research, since they help to understand nonlocal correlations in general. It is an open question whether there is a single physical principle that explains the absence of post-quantum correlations, but at the same time allows the nonlocal correlations that can be explained within quantum theory.

1.8. Discussion: Special multipartite features in quantum theory

As we have seen the phenomena of nonlocality and entanglement are a hallmark of quantum theory but they also exist in many other toy theories. However, in the context of multipartite systems quantum mechanics exhibits various characteristic features which generically do not exist in other GPTs. In this section we are going to review some of these characteristic multipartite quantum features.

1.8.1. Quantum features inherited from subsystems:

Surprisingly, many of these characteristic features are not linked to the structure of the tensor product, rather they are consistently inherited from single systems. In particular the one-to-one correspondence between state spaces and

their information capacity carries over to joint systems. For example, the information capacity of a joint system is simply the product of the single systems information capacities [1].

Since the state spaces are uniquely determined by the information capacity, a joint system consisting of two qubits has the same state space than a single quantum system with information capacity of four. In the case of quantum mechanics the difference between single and joint systems is not reflected in different state space structures but only in a different interpretation of the measurements and states. As composite quantum systems have the same state space structure as their building blocks, multipartite quantum systems inherit all the features from single systems, including e.g. reversible continuous transformations between pure states (transitivity), strong self-duality, non-restricted measurements/states, sharpness and homogeneity. This allows us to interpret the qubit as a fundamental information unit from which any quantum system can be built [47].

In quantum mechanics the equivalence of systems with equal information capacity also manifests itself in the associativity of the tensor product. This means that equal components of a multipartite system can be swapped without changing the state space (compound permutability) [29]. As illustrated in Fig. 1.10, this tensor product lies strictly between the minimal and maximal tensor product, such that potential entanglement in states and effects is perfectly balanced.

The inheritance of such features to joint system in standard quantum theory is quite exceptional, as can be seen when trying to construct something similar for other GPTs. For example, the extremal states of a single gbit can be reversibly transformed into each other and there is an isomorphism between states and effects. However, as shown in [48] the joint states cannot be reversibly transformed when choosing the maximal tensor product. We will show in Chapters 3 that tensor product which inherits an isomorphism between joint states and effects can be constructed, but it treats equal subsystems differently [49].

Given local quantum systems it has been shown that the ordinary quantum

tensor product is the only one that preserves transitivity [50]. Another work explores the opposite direction [51]. The authors assume transitivity on joint states and local systems with state spaces bounded by hyperspheres, which is a generalization of the three-dimensional Bloch ball of qubits known as a *hyperbit* [17]. It was shown that entangled states in such a scenario are only possible for dimension three [51], which was used in [52] to explain why we are living in a three-dimensional world.

1.8.2. Genuine multipartite quantum features:

Beyond the features inherited from single systems there are also genuine multipartite features that are characteristic for quantum theory. Recall that for entangled states the marginal state is mixed even though the joint state might be extremal. Quantum theory allows also the opposite, namely, any mixed state can be regarded as the marginal of a pure extremal state – a process called *purification* [53, 54, 27]. As a consequence any stochastic mapping from one mixed state to another can be realized as a reversible unitary transformation in a higher-dimensional state space without information loss [55].

Since there is a continuum of mixtures, the possibility of purification requires a continuum of pure entangled states. There is also an isomorphism between the transformations of single systems and bipartite joint states [56, 57]. This was recently used to generalize Bayesian inference to quantum states [58]. This isomorphism can be further decomposed into two components. On the one hand, using the self-duality of quantum systems, states are converted to corresponding effects given by the same operator. On the other hand the transformation itself can be realized via *steering*, i.e. the ability to obtain any state as the conditional state of a joint system [59, 60]. Steering is a prerequisite of more advanced multipartite quantum features, like quantum teleportation [61, 30] and entanglement swapping [62].

As pointed out before, nonlocal correlations are a central feature of quantum theory. Several articles have examined the relation between entanglement and non-local correlations in quantum theory. As quantum theory balances

entanglement in states and effects, extending the joint state space to the maximal tensor product would potentially allow for new correlations. While this is not the case for bipartite systems [63], it was found that the maximal tensor product of multipartite systems with more than two subsystems can indeed generate new correlations that are not possible in the standard tensor product [64]. Not only entanglement but also the local structure of the subsystems influences nonlocal correlations.We will discuss a particular class of toy theories in chapter 2. Although each of these theories incorporate joint states that resemble the maximally entangled states in quantum theory the possible nonlocal correlations turn out to strongly depend on the subsystems' structure. A general connection between nonlocality and uncertainty relations of subsystems has been found in Ref. [65].

The device-independent view on nonlocal correlations studies correlations independent of specific joint states and local measurements, simply by considering the probabilities of input-output combinations for a given choice of measurement as input and the outcomes as outputs (see Fig. 1.11). The quantum correlations form a convex subset of the full no-signalling polytope with infinitely many extremal points [66, 67]. For the CHSH set-up this set can be determined by a infinite hierarchy of semi-definite programs, whereas an analytical upper bound known as Q_1 has been derived from the first order [43, 44]. It was shown that any theory that is able to recover classical physics in the macroscopic limit has correlations limited by this bound [68]. Also Q_1 obeys information causality [69, 70]. That is given the nonlocal resource and m bits of classical communication a party on one side can learn at most m bits about the system on the other side. Note that this is a generalization of no-signaling that refers to the situation with $m = 0$. Different to quantum correlations general no-signaling correlations can violate information causality up to the extreme cases of PR boxes that can evaluate any global function that depends on both local inputs from only one classical bit of communication (trivial communication complexity) [71]. Interestingly, for some nonlocal boxes given multiple copies allows to distill PR boxes by using only classical processing at each of the local parts individually [72], whereas the quantum correlations are closed

under such operations [73].

In conclusion quantum theory has a lot of unique characteristic physical features. The framework of Generalized Probabilistic Theories which is used in this work played a crucial role to identify many of those.

2. Local limits on nonlocal correlations

We already mentioned in the last chapter that different theories can produce different sets of nonlocal correlations. Namely, classical theory is limited to correlations obeying local realism, whereas quantum theory allows some, but not all nonlocal correlations. Boxworld, in particular gbits and their maximal tensor product, on the other hand, yields the full set of possible non-signaling correlations.

Remarkably, this limitation of correlations is not a mere consequence of missing entanglement, but can to some extent be traced back to the local subsystems. For example, there are some well known cases where the limitations are unchanged, if all possible joined states are included. That is, the correlations remain the same for joint systems given by the maximal tensor product. For classical systems this is trivial, since the maximal and minimal tensor products coincide [34]. In quantum theory, however, the usual tensor product is strictly smaller, such that a toy theory with joint systems built from the maximal tensor product of quantum systems inhabits new joint states. Nevertheless, in the bipartite case (e.g. the CHSH setting) these additional joint states were shown to only reproduce those correlations that were already available in the standard case [63]. For tripartite systems the maximal tensor product of quantum systems can produce some post-quantum correlations [64]. However, these correlations stay limited, as the set of all non-signaling correlations is strictly bigger.

In this chapter we examine such subsystem-induced limitations of correlations in more detail. We first illustrate the phenomenon. Therefore, we will characterize the nonlocal correlations of toy theories with particular state spaces for subsystems in Section 2.1. We then show in Section 2.2 that a whole class of joint states that we call *inner product states* restricts the measurement statis-

tics to nonlocal correlations. All bipartite correlations in quantum theory and classical probability theory can be reduced to correlations on these states. Finally, the existence of these states is shown to depend on the local state spaces of a theory.

2.1. A family of models

Traditionally the study of nonlocal correlations focuses on three different sets: i) the set of classical correlations that obey local realism, ii) quantum correlations and iii) arbitrary nonlocal correlation that obey the non-signaling condition with the PR boxes as the nonlocal extremal cases for the CHSH set-up. Using bipartite systems constructed by the maximal tensor product, one can choose subsystems with two-dimensional state spaces that yield the respective sets as possible correlations. Classical subsystems given by triangles restricts correlations on bipartite states to classical ones. The maximal tensor product of gbits with a square-shaped state space allows PR boxes and as a consequence all non-signaling CHSH correlations. Whereas the maximal tensor product of disc-shaped subsystems yield exactly the quantum correlations. In this section the dependency between nonlocal correlations and the structure of the local state spaces is further studied by other toy theories.

Our first toy theory is a local modification of boxworld that allows a continuous transition between classical probability theory and the standard version of boxworld by a parameter in section 2.1.1.

A class of theories that include all three of the above cases can be constructed by toy theories with state spaces in the form of regular polygons. As a consequence, this allows us to study transitions from classical probability theory and boxworld to a quantum-like theory by increasing the number of vertices. This quantum-like theory corresponds to the disc-shaped subsystems that emerge in the limit of infinitely many vertices, which can be thought of the equatorial intersection of the Bloch ball of a qubit. These polygon models have been introduced in the master thesis of the author [20]. We explain the structure of these models in section 2.1.2 and extend the approach in this chapter get

further insights on the restrictions of nonlocal correlations.

2.1.1. A local modification of boxworld

Boxworld has single systems given gbits with state spaces having the shape of a square with extremal states given in Equation 1.31. The new model is constructed by exchanging the original extremal state ω_1 of the gbit with

$$\omega_1' = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad x \in [0,1], y \in [x-1, 1-x], \tag{2.1}$$

that can be varied conditioned on the two transition parameters x and y. The effect set is again determined by the no-restriction hypothesis. That is, it is constructed via the dual cone V_+^* according to Equation 1.17. For $x = 0$ one gets the classical case, whereas $x = 1, y = 0$ corresponds to standard boxworld.

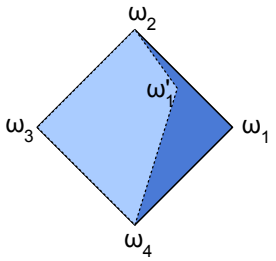


Figure 2.1.: State space of single systems in modified boxworld

Looking at bipartite systems consisting of two of these single systems, it possible to calculate the extremal joint states (again 16 separable and 8 entangled) of the maximal tensor product as a function of x, y. Since the number of different extremal measurements and extremal joint states is rather small, the maximal CHSH-value $S^{\max}$ of the theory can be found by directly comparing the CHSH coefficient of any possible combination.

It is given by:

$$S^{\max} = 2 + \frac{16x^2}{x^2 + (|y| - 1)^2 + 2x(3 + |y|)} \tag{2.2}$$

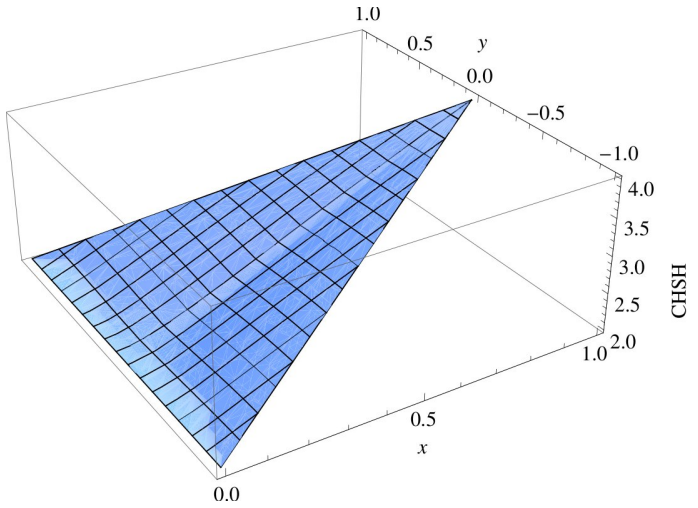


Figure 2.2.: Maximal CHSH-violation as a function of the transition parameters x and y

One can also show that these systems allow correlations not even possible in quantum theory for arbitrary small derivations of the classical state space. This complements a result by Brunner et. al. [72] that there are correlations arbitrary close to classical correlations that are not possible in quantum theory. This section shows that such correlations can be gained by theories with local state spaces including only states that are arbitrary close to those allowed in classical probability theory.

2.1.2. Polygon systems

Let us now discuss a second class of toy systems that are useful to study the limitation of nonlocal correlations that are induced by local state spaces. This class consists of theories with local state spaces given by regular polygons. The different theories are characterized by the respective number n of extremal states. These n vertices that span the subsystems' state spaces can be described

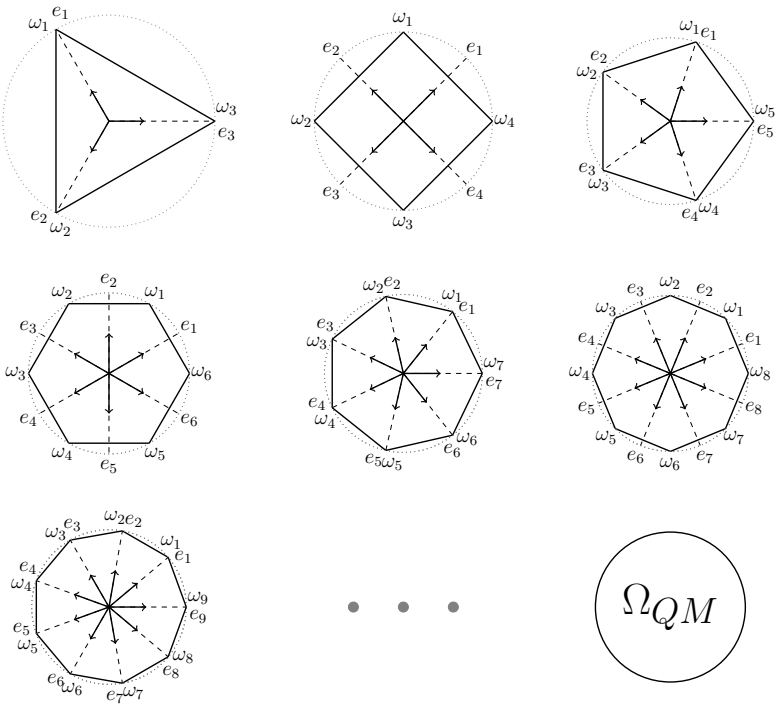


Figure 2.3.: Illustration of the state spaces and ray extremal effects of the polygon models.

by the following equation:

$$\omega_i = \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix} \in \mathbb{R}^3, \tag{2.3}$$

where $r_n = \sqrt{\sec(\pi/n)}$.

The unit effect is

$$u = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{2.4}$$

Different to the extremal states, the description of the extremal effects requires us to distinguish between theories with an even number of vertices and those with odd n. In the case of even n, the set $E(\Omega)$ of all possible measurement

outcomes is the convex hull of the zero effect, the unit effect, and the extremal effects $e_1, \ldots, e_n$, with

$$e_i = \frac{1}{2} \begin{pmatrix} r_n \cos(\frac{(2i-1)\pi}{n}) \\ r_n \sin(\frac{(2i-1)\pi}{n}) \\ 1 \end{pmatrix} . \tag{2.5}$$

A complementary effect can be constructed by $\bar{e_i} = u - e_i$. Hence, a possible dichotomic measurement is $\{e_i, \bar{e_i}\}$. When this measurement is performed on a system in the state ω_j, the probabilities for the two outcomes are given by $e_i \cdot \omega_j$ and $\bar{e_i} \cdot \omega_j$, and satisfy $e_i \cdot \omega_j + \bar{e_i} \cdot \omega_j = 1$. Observe that for even n, $\bar{e_i} = e_{(i+n/2) \bmod n}$.

The case of odd n is slightly different. In this case, define

$$e_i = \frac{1}{1 + {r_n}^2} \begin{pmatrix} r_n \cos(\frac{2\pi i}{n}) \\ r_n \sin(\frac{2\pi i}{n}) \\ 1 \end{pmatrix} \tag{2.6}$$

and again let $\bar{e_i} = u - e_i$, so that a possible dichotomic measurement is $\{e_i, \bar{e_i}\}$. This time, however, $\bar{e_i}$ does not equal e_j for any j. The set $E(\Omega)$ of all possible measurement outcomes is the convex hull of the zero effect, the unit effect, $e_1 \ldots, e_n$, and $\bar{e_1}, \ldots, \bar{e_n}$. As can be seen in Fig. 2.4 in such theories there are effects that are extremal in $E(\Omega)$ (namely the $\bar{e_i}$) but not ray extremal, i.e., they do not lie on an extremal ray of the cone V_+^*. This also happens in quantum mechanics, but only if the dimension of the Hilbert space is larger than two. For example the effect $\mathbb{1} - |\psi\rangle\langle\psi|$ for any rank one projector $|\psi\rangle\langle\psi|$ is then extremal in the set of proper effects, but not ray extremal.

A two-dimensional illustration of the state and effect spaces is given in Fig. 2.3 and a three-dimensional illustration in Fig. 2.4.

As pointed out before the $n = 3$ case corresponds to a classical system with three distinguishable values, whereas $n = 4$ yields the square-shaped state space of a gbit. As $n \to \infty$, the state space tends to a disc of radius one. This makes it similar to a quantum mechanical qubit, whose state space is the Bloch ball.

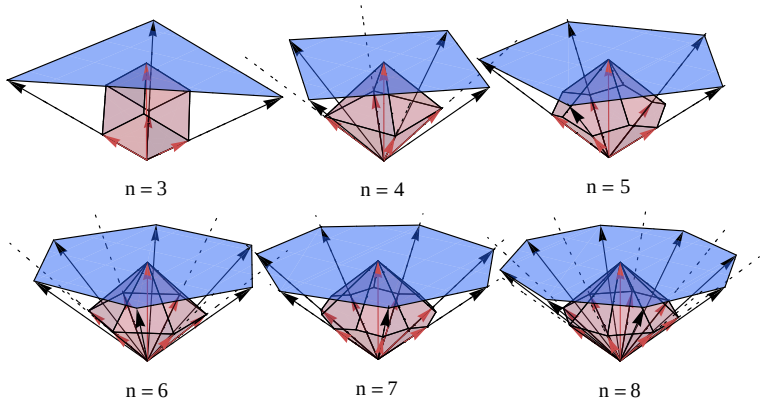


Figure 2.4.: State spaces Ω (blue polygons) and sets of proper effects $E(\Omega)$ (red polytopes) of the polygon toy theories with n vertices. The case $n = 3$ corresponds to a classical system, the $n = 4$ system is capable of generating all non-signaling correlations. In the limit $n \to \infty$ the state space becomes a disc, which can be thought of as the equatorial plane of the Bloch ball.

The disc can be thought of as the equatorial plane of the Bloch ball. We will refer to this case, somewhat loosely, as the quantum case.

2.1.3. Bipartite states of polygon systems

We shall not attempt a complete characterization of the set of all possible non-signaling states $\Omega^A \otimes_{\max} \Omega^B$ for each value of n. Instead, this section describes a particular joint state of two polygon systems, which is the natural analogue of a maximally entangled state of two qubits. The next section examines the nonlocal correlations that can be obtained from performing measurements on these maximally entangled polygon systems.

Recall that a joint state is an element of $V^A \otimes V^B$, hence in the case of two polygon systems, a joint state is an element of $\mathbb{R}^3 \otimes \mathbb{R}^3 = \mathbb{R}^9$. It is convenient to represent the joint state as a 3×3 matrix such that $(e_i \otimes e_j)(\omega^{AB})$ can be calculated by simply left and right multiplying this matrix with the

representations of the effects e_i and e_j in $\mathbb{R}^3$. Define

$$\text{odd n}: \quad \phi^{AB} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\text{even n}: \quad \phi^{AB} = \begin{pmatrix} \cos(\pi/n) & \sin(\pi/n) & 0 \\ -\sin(\pi/n) & \cos(\pi/n) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.7}$$

The state ϕ^{AB} is the natural analogue of a quantum mechanical maximally entangled state for the following reasons. First, it can be verified (see, e.g., Ref. [60]) that except for $n = 3$, ϕ^{AB} is an entangled pure state, where pure means that it is extremal in the maximal tensor product, hence cannot be written as a mixture of other non-signaling states. The $n = 3$ case corresponds to two classical trits, with ϕ^{AB} the maximally correlated state, i.e., if the trit values are 1, 2, 3, then ϕ^{AB} corresponds to $P(11) = P(22) = P(33) = 1/3$. Second, ϕ^{AB} is constructed so that if a measurement is performed on the A system, and outcome e_i obtained, then the updated (or collapsed) state for the B system is ω_i. The marginal probability for Alice to obtain outcome e_i is the same for all i. Compare this with the case of two spin-$1/2$ particles in the state $1/\sqrt{2}(|00\rangle + |11\rangle)$, where $|0\rangle$ and $|1\rangle$ are the eigenstates of spin-z. If a spin measurement in direction $\vec{m}$ in the xz-plane is performed on system A, then the probability of obtaining the up outcome is $1/2$, and if the up outcome is obtained, then the collapsed state of the B system is spin up in direction $\vec{m}$. These quantum predictions are recovered by both versions of ϕ^{AB} in the limit $n \to \infty$.

The following sections investigate the nonlocal correlations that can be produced by performing measurements on two systems in the state ϕ^{AB}. For this it is useful to have an expression for the joint probability of obtaining outcome e_i^A on system A and e_j^B on system B. This is easy to calculate from (2.7). For even n,

$$(e_i^A \otimes e_j^B)(\phi^{AB}) = \frac{1}{4}\left(1 + r_n^2 \cos(\alpha_i - \beta_j)\right), \tag{2.8}$$

where $\alpha_i = \frac{2\pi i}{n}$ and $\beta_j = \frac{(2j-1)\pi}{n}$, and as before, $r_n = \sqrt{\sec(\pi/n)}$. For odd n

$$(e_i^A \otimes e_j^B)(\phi^{AB}) = \frac{1}{(1+r_n^2)^2}\left(1 + r_n^2 \cos(\alpha_i - \beta_j)\right), \tag{2.9}$$

where $\alpha_i = \frac{2\pi i}{n}$ and $\beta_j = \frac{2\pi j}{n}$. Notice the cosine dependence, which is reminiscent of quantum mechanical correlations.

2.1.4. CHSH violations of polygon systems

One commonly used measure of the degree of nonlocality that a bipartite system exhibits is the maximal violation of the Clauser-Horne-Shimony-Holt (CHSH) inequality introduced in 1.6. Whereas classical correlations obey the CHSH inequalities (1.48), PR box correlations correspond to their maximal violation. As discussed before, PR boxes have been explored in the literature and are known to be particularly powerful for certain kinds of information theoretic problem, especially communication complexity problems [71, 74, 72, 69, 68, 75, 76]. The quantum correlations lie strictly between the classical correlations and those that can be produced by PR boxes.

It is interesting to see how the maximal CHSH value obtainable from polygon systems in the state ϕ^{AB} varies as the number of vertices n of the polygon increases. The $n = 4$ case is particularly simple. The optimal choice of measurements to violate the CHSH inequality is

$$x = 0 : \{e_1^A, e_3^A\}, \quad x = 1 : \{e_2^A, e_4^A\}, \quad y = 0 : \{e_2^B, e_4^B\}, \quad y = 1 : \{e_1^B, e_3^B\}, \tag{2.10}$$

and it can be verified from (2.8) in combination with (1.48) that the correlations obtained give $S = 4$. In other words, the maximally entangled state of two $n = 4$ systems can act as a PR box. It follows that this state has the same information theoretic power that PR boxes are known to have.

For general n, assume that Alice's measurement choices are of the form $\{e_i^A, \bar{e}_i^A\}$ and Bob's of the form $\{e_j^B, \bar{e}_j^B\}$. A lengthy but straightforward calcu-

lation gives the following analytic expressions. For even n,

$$S = r_n^2 \sum_{x,y=0,1} (-1)^{xy} \cos\left(\alpha_x - \beta_y\right), \tag{2.11}$$

where as before, $\alpha_x = \frac{2\pi i_x}{n}$ and $\beta_y = \frac{(2j_y-1)\pi}{n}$. For odd n,

$$S = \frac{2}{(1+r_n^2)^2} \left| (r_n^2-1)^2 + 2\, r_n^2 \sum_{x,y=0,1} (-1)^{xy} \cos(\alpha_x - \beta_y) \right|, \tag{2.12}$$

where $\alpha_x = \frac{2\pi i_x}{n}$ and $\beta_y = \frac{2\pi j_y}{n}$. In [20] these expressions were maximized over all possible choices for the angles α_i and β_j to get the maximal violation achievable by local measurements on the maximally entangled state ϕ^{AB}. A detailed analysis of these expressions can be found in A.2. Fig. 2.5 shows the maximal CHSH value for the maximally entangled state of polygon systems as a function of n.

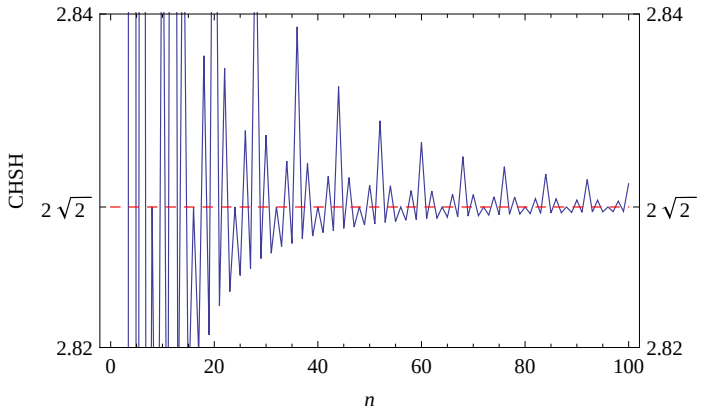


Figure 2.5.: Maximal CHSH value from the maximally entangled state of two polygon systems as a function of the number of vertices n. Tsirelson's bound ($S \leq 2\sqrt{2}$) appears as a natural separation between the case of even n and odd n.

The most important feature of Fig. 2.5 is that the correlations of even n systems can always reach or exceed Tsirelson's bound, while the correlations of odd n systems are always below Tsirelson's bound. Thus Tsirelson's bound appears aass a natural separation between the correlations of these two different

kinds of polygon state spaces. Sections 2.2 and 2.3 show why this is. Section 2.2 shows that for odd n, the maximally entangled state ϕ^{AB} belongs to a broad class of states we call *inner product states*, and that all correlations obtainable from measurements on inner product states satisfy Tsirelson's bound. Section 2.3 goes further, and relates this to a fundamental geometric difference between polygons with even n and odd n. In Fig. 2.3, the difference is seen in the fact that for odd n, the effect cone V_+^* coincides with the state cone V_+, whereas for even n, the effect cone is isomorphic to the state cone but rotated through some angle.

We have only considered correlations obtainable from the maximally entangled state ϕ^{AB}. In principle there could be joint states other than the maximally entangled state which show stronger violations for some Bell inequalities. While this seems unlikely for the CHSH inequality, other Bell inequalities are known to be maximized by non-maximally entangled states in quantum mechanics [77, 41].

2.1.5. The Braunstein-Caves inequalities

The Braunstein-Caves (or *chained*) Bell inequalities [78] are similar to the CHSH inequality, but involve N measurement settings on each system, rather than two. Let Alice's choice of measurement be x, and Bob's y, with $x, y \in \{1, \ldots, N\}$. Let the outcomes be $a, b \in \{0, 1\}$. Local correlations satisfy

$$S_N = \left| \sum_{j=1}^{N-1} (C_{j,j} + C_{j,j+1}) + C_{N,N} - C_{N,1} \right| \leq 2N - 2, \tag{2.13}$$

where as before $C_{x,y} = P(0,0|x,y) + P(1,1|x,y) - P(0,1|x,y) - P(1,0|x,y)$. In the case $N = 2$, this is equivalent to the CHSH inequality, up to relabelling of measurement settings.

The algebraic maximum of S_N is $2N$. This maximum can be attained by performing measurements on the maximally entangled state of even n polygon systems with $n = 2N$. This state is thus tailor made for violating the Braunstein-Caves Bell inequalities. To see this, let Alice's and Bob's measure-

ment choices be given by

$$x = i : \{e_i^A, \bar{e}_i^A\}, \quad i = 1, \ldots, N, \tag{2.14}$$

$$y = j : \{e_j^B, \bar{e}_j^B\}, \quad j = 1, \ldots, N, \tag{2.15}$$

and note that (i) $C_{j,j} = 1$ for $j = 1, ..., N$, (ii) $C_{j,j+1} = 1$ for $j = 1, ..., N-1$ and (iii) $C_{N,1} = -1$. In the case $n \to \infty$, maximal violation of the Braunstein-Caves inequality is achieved in the limit of infinitely many settings. This is also true for a quantum mechanical maximally entangled state, as shown in Ref. [79].

In general, given a set of correlations $P(a, b|x, y)$, they can be written as a mixture

$$P(a, b|x, y) = qP^{\mathrm{NL}}(a, b|x, y) + (1 - q)P^{\mathrm{L}}(a, b|x, y), \tag{2.16}$$

where $0 \leq q \leq 1$, $P^{NL}(a, b|x, y)$ is a set of nonlocal correlations and $P^{L}(a, b|x, y)$ a set of local correlations. Suppose, however, that the correlations $P(a, b|x, y)$ return the maximum value S_N for an appropriate Braunstein-Caves inequality. Then $q(S_N) + (1 - q)(S_N - 2) \geq S_N$, hence $q = 1$. Therefore, the fact that the maximally entangled state of even n polygon systems returns the maximum value for the appropriate Braunstein-Caves inequality indicates that there is no local part in the correlations with $N = n/2$ measurement settings. This was pointed out in the case of quantum systems in Ref. [79, 80]. As a further curiosity, if we did have access to these systems, this feature could be used for secure key distribution, using the protocol of Ref. [81].

2.1.6. Distillation

So far, we have only considered correlations that can be produced by measuring a single copy of a bipartite polygon system. There remains the possibility that stronger correlations could be produced by performing local measurements on multiple bipartite pairs, and locally processing the data (there is a further possibility, involving entangled measurements across multiple copies on each side, which we do not discuss).

Consider the bipartite state ϕ^{AB} of two even n polygon systems, and suppose

that Alice and Bob are choosing from the measurements

$$x = 0 : \{e_1^A, \bar{e}_1^A\}, \quad x = 1 : \{e_2^A, \bar{e}_2^A\}, \quad y = 0 : \{e_1^B, \bar{e}_1^B\}, \quad y = 1 : \{e_2^B, \bar{e}_2^B\}, \tag{2.17}$$

with outcomes $a, b \in \{0, 1\}$ as usual. Recall that $C_{j,j} = 1$ for $j = 0, 1$ and $C_{0,1} = 1$. (2.8) also gives $C_{1,0} = 2\cos(\frac{2\pi}{n}) - 1$. The correlations produced can be written as a probabilistic combination of maximally nonlocal correlations (equivalent up to relabelling to the PR box correlations of (1.51)), and another term which describes local correlations:

$$P_\epsilon(a, b|x, y) = \epsilon P^{\mathrm{PR}}(a, b|x, y) + (1 - \epsilon)P^{\mathrm{L}}(a, b|x, y). \tag{2.18}$$

Here, $0 \leq \epsilon = 1 - \cos(\frac{2\pi}{n}) \leq 1$, P^{PR} is given by

$$P^{\mathrm{PR}}(a, b|x, y) = \begin{cases} \frac{1}{2} & \text{if } a \oplus b = x(y \oplus 1) \\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

and P^{L} is a set of local correlations given by

$$P^{\mathrm{L}}(a, b|x, y) = \begin{cases} \frac{1}{2} & \text{if } a \oplus b = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

In Ref. [72], it is shown that all correlations of the form (2.18) with $0 < \epsilon < 1$ can be distilled into stronger correlations using a protocol that involves two copies of a bipartite system. Importantly, this protocol consists only of local processing and does not involve any communication. In the asymptotic limit of infinitely many copies of a bipartite system, the correlations (2.18) can be distilled to PR box correlations by iterating the protocol. Thus for any finite even n, the polygon systems produce correlations that can be distilled arbitrarily close to PR box correlations (since $\epsilon = 1 - \cos(\frac{2\pi}{n}) > 0$). It is only in the limit $n \to \infty$ (the quantum case), that we get $\epsilon = 0$ and thus lose the ability to distill PR box correlations.

The consequence of the above is that polygon systems with even and finite

n inherit the powerful communication properties of PR boxes as long as there are multiple copies of the maximally entangled state available. For instance, they collapse communication complexity [71], allow for better than classical non-local computation [75], violate information causality [69] and macroscopic locality [68]. Moreover, since the PR box can be considered as a unit of bipartite nonlocality [82, 83, 84], it follows that any bipartite non-signaling probability distribution can be generated from multiple copies of polygon systems with even n. This is particularly surprising as in practice, an individual polygon system with even and very large n would be very difficult to distinguish from one with odd n, and also from the quantum case, i.e. the disc that one gets in the limit $n \to \infty$. These toy theories thus show that practically indistinguishable theories can have fundamentally different limits to the non-local correlations they allow.

For polygon systems with odd and finite n, the situation is dramatically different, as seen in the next section.

2.2. Bounds on correlations

For even n polygon systems, the maximally entangled state can produce arbitrarily strong nonlocal correlations, whereas for odd n polygon systems, the nonlocality is highly constrained. The maximally entangled state of odd n polygon systems cannot, for example, violate Tsirelson's inequality. This section shows that this is a consequence of a much more general result.

We first introduce a class of bipartite states in general theories, which we call *inner product states*. The main theorem establishes a strong constraint on the nonlocal correlations that can be produced from measurements on inner product states. One consequence is that inner product states cannot violate Tsirelson's inequality. The maximally entangled states of odd n polygon systems are inner product states, hence the theorem explains what was only established by direct calculation above — that these states do not violate Tsirelson's inequality. On the other hand, the maximally entangled states of even n polygon systems are not inner product states, which is consistent with them pro-

ducing arbitrary non-signaling correlations. We also show that all classical and quantum states are, in terms of non-local correlations, no stronger than an inner product state.

2.2.1. Inner product states

Recall that a state cone V_+ is the set of unnormalized states of a system, and that these span a vector space V. An effect cone V_+^* is the set of unnormalized measurement outcomes, and these span the vector space V^*. Given two systems A and B, if the state cones V_+^A and V_+^B span vector spaces V^A and V^B respectively, then a joint state is an element of $V^A \otimes V^B$.

Call two distinct systems *similar* if their state spaces are isomorphic. Examples of similar systems are two quantum mechanical qubits, or two classical trits, or two n-vertex polygon systems. For the rest of this section, assume a bipartite system composed of two similar subsystems A and B. In this case, the respective state spaces and effect spaces can be identified, so that $V^A = V^B = V$, $(V^A)^* = (V^B)^* = V^*$, $u^A = u^B = u$, and so on.

Definition 2.1. *A joint state ω^{AB} is* symmetric *if $(e \otimes f)(\omega^{AB}) = (f \otimes e)(\omega^{AB})$ for all measurement outcomes $e, f \in V_+^*$.*

Definition 2.2. *A joint state ω^{AB} is an* inner product state *if ω^{AB} is symmetric, and positive semidefinite, i.e., $(e \otimes e)(\omega^{AB}) \geq 0 \ \forall e \in V^*$.*

Note that by definition of a joint state, it is always true that $(e \otimes e)(\omega^{AB}) \geq 0$ when $e \in V_+^*$, i.e., when e is a valid effect. This is simply a statement of the fact that measurement outcome probabilities have to be greater than or equal to zero. The definition requires something stronger, which is that $(e \otimes e)(\omega^{AB}) \geq 0$ for any e in the whole of the vector space V^*.

Example 1. *Any symmetric product state $\omega^{AB} = \omega \otimes \omega$ is an inner product state.*

Example 2. *Consider two classical systems, each of which is a* nit*, taking values $\{1, \ldots, n\}$. A joint state is simply a joint probability distribution over*

nit values. Write the joint state as a matrix P, where P_{ij} is the joint probability that $A = i$ and $B = j$. This is an inner product state iff the matrix P is symmetric and positive semi-definite. In particular this includes any perfectly correlated state of the form

$$\begin{aligned} P_{ij} &= 0 \quad \text{if} \quad i \neq j \\ P_{ii} &= q_i, \quad q_i \geq 0, \quad \sum_i q_i = 1. \end{aligned}$$

Example 3. *Consider two polygon systems, each corresponding to a state space with n vertices. Section 2.1.3 defined an analogue of a maximally entangled state ϕ^{AB}. In the matrix representation of (2.7), ϕ^{AB} is an inner product state if and only if the matrix is symmetric and positive semi-definite. Hence ϕ^{AB} is an inner product state for odd n, whereas for even n, ϕ^{AB} is not an inner product state.*

Example 4. *The quantum case is slightly subtle. Given two qubits, the maximally entangled state*

$$\Phi^+ = |\Phi^+\rangle\langle\Phi^+|, \qquad |\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \tag{2.21}$$

is symmetric but is not an inner product state, since if σ_y is a Pauli spin matrix, then $(\sigma_y \otimes \sigma_y)(\Phi^+) = -1$. Consider the operator defined by $\tilde{\Phi} = (\mathbb{1} \otimes T)(\Phi^+)$, where T is the linear map that takes an operator in V^B to its transpose with respect to the computational basis. The new operator $\tilde{\Phi}$ is not a valid quantum state. It is locally positive but not globally positive, hence is not a density operator. But it is in the maximal tensor product of two qubits, and it is an inner product state. In fact, $\tilde{\Phi}$ predicts perfect correlation whenever Alice and Bob perform measurements in the same direction. However, the two states are equivalent in terms of the non-local correlations they can produce (as was first shown in Ref. [63]).

Theorem 2.4 below establishes a constraint on the nonlocal correlations that can be obtained from measurements on an inner product state. It may seem as

if the definition of an inner product state is quite restrictive, given that an inner product state must be symmetric, for example, and given that the maximally entangled state Φ_+ of two qubits is not included. This would diminish the interest of the theorem. However, suppose that a bipartite state ω^{AB} can be obtained from an inner product state via a transformation of one of its subsystems. Then any correlations obtained from ω^{AB} could also be obtained from an inner product state. Hence any restriction on the correlations from inner product states also applies to ω^{AB}. Formally,

Theorem 2.1. *Consider a joint state ω^{AB}, which can be written in the form $\omega^{AB} = (\mathbb{1} \otimes \tau)(\sigma^{AB})$, for some $\tau : V_+ \to V_+$ that takes normalized states to normalized states. Any correlations obtained from measurements on ω^{AB} can also be obtained from measurements on σ^{AB}.*

Proof. Define the adjoint map $\tau^\dagger : V_+^* \to V_+^*$ such that for any effect $e \in V_+^*$ and any state $\omega \in V_+$,

$$(\tau^\dagger(e))(\omega) = e(\tau(\omega)). \tag{2.22}$$

Since τ takes normalized states to normalized states, $\tau^\dagger(u) = u$. Given a measurement y on system B, with outcomes $\{f_1, \ldots, f_r\}$, let y' be the measurement with outcomes $\{\tau^\dagger(f_1), \ldots, \tau^\dagger(f_r)\}$. Note that from $f_1 + \cdots + f_r = u$, and $\tau^\dagger(u) = u$, it follows that $\tau^\dagger(f_1) + \cdots + \tau^\dagger(f_r) = u$, as must be the case for y' to be a valid measurement. Then measurements x and y on ω^{AB} have the same joint outcome probabilities as measurements x and y' on σ^{AB}. Hence, if a particular set of correlations can be obtained by performing measurements on ω^{AB}, those same correlations can be obtained by performing different measurements on σ^{AB}. □

Further,

Theorem 2.2. *Given two d-dimensional quantum systems, any pure state $\rho^{AB} = |\psi\rangle\langle\psi|$ can be written in the form $\rho^{AB} = (\mathbb{1} \otimes \tau)(\tilde{\rho}^{AB})$, where $\tau : V_+ \to V_+$ takes normalized states to normalized states, and $\tilde{\rho}^{AB}$ is an inner product state.*

Proof. Using the Schmidt decomposition, every pure quantum state $|\psi\rangle$ can be written in the form:

$$|\psi\rangle = \sum_{i=1}^{r} \lambda_i |a_i\rangle \otimes |b_i\rangle, \tag{2.23}$$

where r is the Schmidt rank, $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ are orthonormal bases and the λ_i are real and positive. A unitary transformation U, on system B, which maps $\{|b_i\rangle\}$ to $\{|a_i\rangle\}$ gives

$$|\psi'\rangle = \sum_{i=1}^{r} \lambda_i |a_i\rangle \otimes |a_i\rangle.$$

Now let

$$\tilde{\rho}^{AB} = (\mathbb{1} \otimes T)(|\psi'\rangle\langle\psi'|),$$

where T is the transpose map, acting on the B system, defined with respect to the basis $\{|a_i\rangle\}$. Note that $\tilde{\rho}^{AB}$ is symmetric since for Hermitian operators E and F,

$$(E \otimes F)(\tilde{\rho}^{AB}) = \mathrm{Tr}[(E \otimes F)\tilde{\rho}^{AB}] = \sum_{ij} \lambda_i \lambda_j E_{ji} F_{ij} = (F \otimes E)(\tilde{\rho}^{AB}).$$

Note also that $\tilde{\rho}^{AB}$ is positive semi-definite since for any Hermitian operator E,

$$(E \otimes E)(\tilde{\rho}^{AB}) = \mathrm{Tr}[(E \otimes E)\tilde{\rho}^{AB}] = \sum_{ij} \lambda_i \lambda_j E_{ji} E_{ij} = \sum_{ij} \lambda_i \lambda_j |E_{ji}|^2 \geq 0.$$

Therefore $\tilde{\rho}^{AB}$ is an inner product state. The quantum state ρ^{AB} can be written $\rho^{AB} = (\mathbb{1} \otimes \tau)(\tilde{\rho}^{AB})$, where τ is the transpose map followed by U^{-1}, which proves the theorem. □

Now any correlations that can be obtained from measurements on a bipartite classical or quantum system, pure or mixed, can also be obtained from measurements on a pure quantum state of two d-dimensional systems for some d. This follows from the fact that mixed quantum states always have a purification on a larger Hilbert space. Combining this observation with theorems 2.1 and 2.2 gives

Theorem 2.3. *Any correlations obtained from measurements on a bipartite, pure or mixed, classical or quantum system could also be obtained from measurements on an inner product state.*

Hence as far as correlations go, the fact that we consider only inner product states is not nearly so restrictive as it looks. By extension, the results apply to all classical and quantum bipartite systems.

2.2.2. The set Q_1

The problem of characterizing those correlations which could in principle be produced by performing measurements on quantum systems, and those that cannot, is an interesting one. Tsirelson's inequality, which limits the possible violation of the CHSH inequality in quantum theory, was the first result in this direction. A great deal of progress is made in Refs. [43, 44, 45], where the problem is reduced to the following form. A hierarchy of sets $Q_1, Q_2, \ldots$ is defined, such that each Q_k is a proper subset of the set of all possible bipartite non-signaling correlations, and each Q_k is strictly contained in its predecessor. For given correlations $P(a, b|x, y)$, and for each k, it is a semi-definite programming problem to determine whether $P(a, b|x, y)$ is contained in Q_k. Furthermore, a given set of correlations $P(a, b|x, y)$ can be obtained from measurements on quantum systems if and only if $P(a, b|x, y)$ is contained in Q_k for some k. Hence the sets Q_k become smaller as k increases, until in the limit $k \to \infty$ they converge towards the set Q of quantum correlations.

The set Q_1, which is the largest in the hierarchy, is of further significance. In Ref. [68] it is shown that correlations in Q_1 satisfy a readily comprehensible physical principle called *macroscopic locality*. For a precise description of what this means, see Ref. [68], but in a nutshell, the principle states that the coarse-grained statistics of correlation experiments involving a large number of particles should admit a description by a local hidden variable model. In other words, the set of microscopic correlations that satisfy the principle of macroscopic locality are those which are compatible with classical physics in a certain limit in which the number of particle pairs being tested is large, and

only coarse-grained statistics, rather than settings and outcomes for every pair, are collected. It is also known that Q_1 is closed under *wiring* [68, 73], in other words it is not possible to distill correlations in Q_1 to correlations outside Q_1 by performing measurements on a number of distinct pairs of systems, and locally manipulating the data. Finally, in the specific case of binary measurement choices and outcomes, all correlations in Q_1 respect Tsirelson's bound of $2\sqrt{2}$ for the CHSH scenario. The main theorem below states that correlations from measurements on inner product states are contained in the set Q_1.

First, we give a formal definition of Q_1. Suppose that Alice and Bob share two systems in a bipartite state, and let Alice choose a measurement x and Bob choose a measurement y. Up to now, when we discussed correlations, Alice's and Bob's outcomes were labelled a and b, and correlations written $P(a, b|x, y)$. For the specific purpose of defining Q_1, however, it is more useful to label the measurement outcomes in such a way that outcomes of distinct measurements have different labels. Hence let the index i range over all possible outcomes of all of Alice's measurement choices. For example, if Alice is choosing from N possible measurements, each of which has k possible outcomes, then i takes values in $\{1, \ldots, kN\}$, with $i = 1, \ldots, k$ the outcomes of the $x = 1$ measurement, $i = k + 1, \ldots, 2k$ the outcomes of the $x = 2$ measurement, and so on. Let the same conventions apply to Bob's outcome, which is denoted j. With a slight abuse of notation, let $x(i)$ denote the unique measurement choice of Alice for which i is a possible outcome. Similarly, $y(j)$. Write $P(i, j)$ for the probability of obtaining outcomes i and j when the measurements $x(i)$ and $y(j)$ are performed. Let $P_A(i)$ denote the marginal probability for Alice to obtain outcome i when she performs measurement $x(i)$, and $P_B(j)$ denote the marginal probability for Bob to obtain outcome j when he performs measurement $y(j)$.

Definition 2.3 ([43, 44, 45, 68])**.** *A set of correlations $P(i, j)$ is in Q_1 iff there exists a positive semi-definite matrix γ of the form*

$$\gamma = \begin{pmatrix} 1 & \vec{P}_A^T & \vec{P}_B^T \\ \vec{P}_A & \tilde{Q} & \tilde{P} \\ \vec{P}_B & \tilde{P}^T & \tilde{R} \end{pmatrix}, \tag{2.24}$$

such that

1. $\vec{P}_A$ *and* $\vec{P}_B$ *are the vectors of probabilities* $P_A(i)$ *and* $P_B(j)$,
2. $\tilde{P}$ *is a matrix with elements* $\tilde{P}_{ij} = P(i,j)$,
3. $\tilde{Q}$ *and* $\tilde{R}$ *are sub-matrices with diagonal elements* $\tilde{Q}_{ii} = P_A(i)$ *and* $\tilde{R}_{jj} = P_B(j)$,
4. $\tilde{Q}_{ii'} = 0$ *if* $i \neq i'$, $x(i) = x(i')$,
5. $\tilde{R}_{jj'} = 0$ *if* $j \neq j'$, $y(j) = y(j')$.

In words, the last two conditions state that elements of $\tilde{Q}$ *and* $\tilde{R}$ *corresponding to different outcomes of the* same *measurement must be zero. The remaining off-diagonal elements of* $\tilde{Q}$ *and* $\tilde{R}$ *can be chosen freely.*

2.2.3. The main theorem

Theorem 2.4. *Consider two similar systems, whose joint state is an inner product state. All correlations that can be obtained from local measurements lie in* Q_1.

Proof. It is sufficient to show that for any set of correlations generated by measurements on an inner product state, there exists a matrix γ of the form (2.24), which is symmetric, positive semi-definite, and has the feature that entries in the blocks $\tilde{Q}$ and $\tilde{R}$ corresponding to different outcomes of the same measurement are zero.

Consider correlations generated by measurements on an inner product state ω^{AB}. Using the notation introduced in section 2.2.2, let e_i be the effect corresponding to Alice's measurement outcome i, and f_j the effect corresponding to Bob's measurement outcome j. Suppose that i ranges from $1, \ldots, n^A$ and j from $1, \ldots, n^B$. Define a vector of effects $g = (u, e_1, \ldots, e_{n^A}, f_1, \ldots, f_{n^B})$, and denote the entries $g_1 = u, g_2 = e_1, \ldots, g_{1+n^A+n^B} = f_{n^B}$. Define the $(1 + n^A + n^B) \times (1 + n^A + n^B)$ matrix $\tilde{\gamma}$ such that $\tilde{\gamma}_{kl} = (g_k \otimes g_l)(\omega^{AB})$.

From the fact that ω^{AB} is an inner product state, it follows directly that $\tilde{\gamma}$ is a symmetric and positive semi-definite matrix [85].

Now define a matrix γ of the form (2.24), with $\gamma_{kl} = \tilde{\gamma}_{kl}$ for all k, l except for the following elements of the sub-matrices $\tilde{Q}$ and $\tilde{R}$:

1. $\tilde{Q}_{ii} = P_A(i)$, and $\tilde{R}_{jj} = P_B(j)$.
2. $\tilde{Q}_{ii'} = 0$ if $i \neq i'$, $x(i) = x(i')$,
3. $\tilde{R}_{jj'} = 0$ if $j \neq j'$, $y(j) = y(j')$.

By construction, γ satisfies conditions (i)-(v) of Definition 2.3, and symmetry of γ follows from symmetry of $\tilde{\gamma}$. It remains to show that γ is positive semi-definite.

To this end, let $\delta = \gamma - \tilde{\gamma}$ and note that δ is of the form

$$\delta = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \delta_Q & \tilde{0} \\ 0 & \tilde{0}^T & \delta_R \end{pmatrix}, \tag{2.25}$$

where δ_Q is an $n_A \times n_A$ sub-matrix, δ_R is an $n_B \times n_B$ sub-matrix, and $\tilde{0}$ is the $n_A \times n_B$ matrix with all entries 0. Since both γ and $\tilde{\gamma}$ are symmetric, δ is also symmetric. We will show that δ_Q and δ_R are positive semi-definite. It follows that δ is positive semi-definite. Since $\gamma = \delta + \tilde{\gamma}$, it follows that γ is also positive semi-definite.

Note that $(\delta_Q)_{ii'} = 0$ for $x(i) \neq x(i')$. It follows that δ_Q is block diagonal, with each block corresponding to a particular measurement choice of Alice. Consider a particular block, corresponding to a measurement with, say, r outcomes. It is of the form

$$M = \begin{pmatrix} e_1 \otimes u - e_1 \otimes e_1 & -e_1 \otimes e_2 & \cdots & -e_1 \otimes e_r \\ -e_2 \otimes e_1 & e_2 \otimes u - e_2 \otimes e_2 & \cdots & -e_2 \otimes e_r \\ & & \vdots & \\ -e_r \otimes e_1 & -e_r \otimes e_2 & \cdots & e_r \otimes u - e_r \otimes e_r \end{pmatrix} (\omega^{AB}). \tag{2.26}$$

Using $e_1 + \cdots + e_r = u$, this matrix can be decomposed into a sum of $(r^2 - r)/2$ matrices

$$M = \sum_{n=2}^{r} \sum_{m=1}^{n-1} M^{mn}, \tag{2.27}$$

where all entries of the matrices M^{mn} are 0, except for

$$(M^{mn})_{mm} = (M^{mn})_{nn} = (e_m \otimes e_n)(\omega^{AB}) \tag{2.28}$$

$$(M^{mn})_{mn} = (M^{mn})_{nm} = -(e_m \otimes e_n)(\omega^{AB}). \tag{2.29}$$

Each M^{mn} is manifestly positive semi-definite, hence M is positive semi-definite. Since each block of δ_Q is positive semi-definite, δ_Q is also positive semi-definite. A similar argument shows that δ_R is also positive semi-definite. Therefore δ and γ are positive semi-definite. This concludes the proof. □

Corollary 2.1. *Consider two systems, whose joint state is of the form* $\omega^{AB} = (\mathbb{1} \otimes \tau)(\sigma^{AB})$*, where* $\tau : V_+ \to V_+$ *takes normalized states to normalized states and* σ^{AB} *is an inner product state. All correlations obtainable from measurements on* ω^{AB} *lie in* Q_1*.*

Proof. This is immediate from theorem 2.4 and theorem 2.1. □

Theorem 2.3 then implies that all correlations from bipartite classical and quantum states lie in Q_1. This was known already of course from Refs. [43, 44, 45]. One could view the theorem and corollary as an independent proof of this fact.

2.3. Polygons revisited

It has already been observed that given two n-vertex polygon systems, the maximally entangled state ϕ^{AB}, defined in section 2.1.3, is an inner product state if and only if n is odd. Theorem 2.4 states that correlations obtained from measurements on an inner product state lie in the set Q_1, which means in particular that they respect Tsirelson's bound for the CHSH inequailty. This

explains why Tsirelson's bound is satisfied by the odd n polygon systems, and is consistent with violation of Tsirelson's bound by the even n polygon systems.

This section relates these observations to simple geometrical properties of the state spaces of polygon systems. A quick glance at figures 2.3 and 2.4 reveals an obvious difference between the odd n and even n cases. For odd n, the effect cone V_+^* coincides with the state cone V_+. For even n on the other hand, the effect cone is isomorphic to the state cone, but is rotated by some non-zero angle. This simple observation lies at the heart of why it is only the maximally entangled states of odd n polygon systems that are inner product states, and hence why it is only these that must satisfy Tsirelson's bound.

The fundamental difference between the odd n and even n state spaces can be stated more formally as follows. First

Definition 2.4 (weakly self-dual)**.** *A system is* weakly self-dual *iff the state and effect cones are isomorphic.*

All of the polygon state spaces are weakly self-dual. The isomorphisms are simply the rotations and improper rotations around the z axis by $(1 + 2k)\pi/n,\ k \in \{0, \ldots, n-1\}$ if n is even and by $2k\pi/n,\ k \in \{0, \ldots, n-1\}$ if n is odd.

The odd n polygon state spaces, on the other hand, satisfy a stronger condition, whereby there are additional restrictions on the isomorphism connecting V_+^* and V_+.

Definition 2.5 (strongly self-dual)**.** *A system is* strongly self-dual *iff there exists an isomorphism $T : V_+^* \to V_+$ which is symmetric and positive semi-definite, i.e., $f[T(e)] = e[T(f)]$ for all $e, f \in V^*$, and $e[T(e)] \geq 0$ for all $e \in V^*$.*

Given the representation of sections 2.1.2 and 2.1.3, the identity map is an example of such an isomorphism. The odd n polygon state spaces are strongly self-dual, but the even n are not.

The concepts of strong and weak self-duality have appeared earlier in the literature, for example in Ref. [30]. Weak self-duality is intimately related to

the operational tasks of probabilistic remote state preparation (steering) and teleportation [60, 30].

Now we can relate these properties of individual systems to the bipartite maximally entangled state ϕ^{AB}. Notice that given two similar systems, any isomorphism $T : V_+^* \to V_+$ corresponds to a bipartite state ω_T^{AB} via

$$(e \otimes f)(\omega_T^{AB}) = \frac{f[T(e)]}{u[T(u)]}. \tag{2.30}$$

The state defined is normalized by construction and is locally positive since $0 \leq f[T(e)]/u[T(u)] \leq 1$ for all $e, f \in E(\Omega)$. Intuitively, ω_T^{AB} is defined so that if Alice performs a measurement and obtains outcome e, then Bob's unnormalized collapsed state, conditioned on that outcome, is $T(e)$.

In the special case that the individual systems are strongly self-dual and the isomorphism T has the additional properties required by definition 2.5, then the induced state ω_T^{AB} is symmetric and positive semi-definite, hence it is an inner product state. This is the case for the maximally entangled state ϕ^{AB} of odd n polygon systems, defined in (2.7), where ϕ^{AB} corresponds to a map T which is simply the identity map. It follows that for odd n, correlations from ϕ^{AB} lie in Q_1.

In the case that individual systems are weakly but not strongly self-dual, the maximally entangled state corresponds to an isomorphism T, but there is no such T with the additional properties of symmetry and positive semi-definiteness, hence the maximally entangled state is not an inner product state. This is the case for the maximally entangled state ϕ^{AB} of the even n polygon systems, defined in (2.7), where ϕ^{AB} corresponds to a map T which is a rotation in $\mathbb{R}^3$ by π/n. This is why for even n, correlations from ϕ^{AB} need not lie in Q_1.

2.4. Correlations outside of Q_1

Correlations obtained from the maximally entangled state of two odd n polygon systems must be contained in Q_1, and this has been seen to be related to the fact that the individual systems are strongly self-dual. It is natural to ask

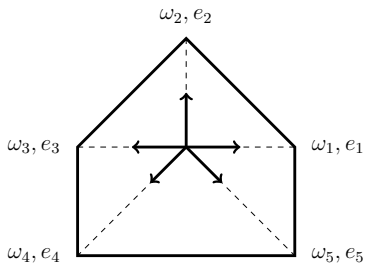


Figure 2.6.: The house-shaped state space is strongly self-dual.

whether the correlations obtained from *any* joint state of strongly self-dual subsystems must also lie in Q_1. An explicit counterexample shows that this is not the case.

Consider a strongly self-dual system with normalized extremal states

$$\begin{aligned} &\omega_1 = (1,0,1)^T && \omega_2 = (0,1,1)^T && \omega_3 = (-1,0,1)^T \\ &\omega_4 = (-1,-1,1)^T && \omega_5 = (1,-1,1)^T, \end{aligned}$$

and normalized ray extremal effects

$$\begin{aligned} &e_1 = \frac{1}{2}(1,0,1)^T && e_2 = \frac{1}{2}(0,1,1)^T && e_3 = \frac{1}{2}(-1,0,1)^T \\ &e_4 = \frac{1}{3}(-1,-1,1)^T && e_5 = \frac{1}{3}(1,-1,1)^T && u = (0,0,1)^T. \end{aligned}$$

The state space for this system looks something like a house and is depicted in Fig. 2.6.

We have explicitly calculated all extremal states in the maximal tensor product of two such systems. One of these joint states can be written as

$$\begin{pmatrix} -1 & -\frac{1}{4} & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{4} & 1 \end{pmatrix}, \tag{2.31}$$

where we have used the same representation as a 3×3 matrix that was introduced in section 2.1.3. This state is extremal in the maximal tensor product,

but is not an inner product state. With a suitable choice of measurements, correlations can be produced which violate Uffink's quadratic inequality [46]

$$(C_{0,0} + C_{1,0})^2 + (C_{0,1} - C_{1,1})^2 \leq 4. \tag{2.32}$$

In particular the measurement choices

$$x = 0 :\{e_5, u - e_5\}, \quad x = 1 :\{e_3, u - e_3\}, \quad y = 0 :\{e_2, u - e_2\}, \quad y = 1 :\{e_3, u - e_3\} \tag{2.33}$$

give

$$(C_{0,0} + C_{1,0})^2 + (C_{0,1} - C_{1,1})^2 = \frac{17}{4} > 4. \tag{2.34}$$

However, satisfaction of Uffink's inequality is known to be a necessary condition for membership of Q_1 [70]; hence these correlations cannot lie in Q_1.

Although these correlations violate Uffink's inequality and lie outside of Q_1, they do not violate Tsirelson's bound for the CHSH inequality. In fact, we have not been able to find a joint state of two strongly self-dual subsystems that violates the CHSH inequality beyond Tsirelson's bound. This leads us to conjecture that Tsirelson's bound holds for every theory with strongly self-dual subsystems.

2.5. Discussion

This work considers a very general setting in which a whole range of probabilistic models can be defined, with the classical and quantum theories as special cases. There is little constraint on the state space, except that it is assumed to be convex, and joint systems are assumed to satisfy a no-signaling principle and a principle of local tomography. The aim of this chapter was to investigate the nonlocal correlations that can be produced by measurements on entangled systems, and to compare and contrast this with the classical and quantum cases.

We revealed an intimate and intricate relationship between the shape of the

state space for an individual system, and the strength of the nonlocal correlations that can be obtained from two systems in an entangled state. This is illustrated by a family of toy models that provide transitions between the usual cases. First, we examine a local modification of boxworld that allows us to characterize the maximal CHSH violation of theories that form a continuum between standard boxworld and classical probability theory. We then considered a second class of models, in which the state space for a single system is a regular polygon with n vertices. Given two such systems, there is an analogue of a maximally entangled state. It turns out that the strength of nonlocal correlations generated by this state depends dramatically on the parity of the number of vertices n of the local polygon. If n is even, maximally nonlocal correlations can be generated, including those that violate macroscopic locality. If n is odd, however, the maximally entangled state respects macroscopic locality. This is in turn explained by the fact that odd n polygons have a geometric property known as strong self-duality, while even n polygons do not. The main theorem, with its corollary, states that correlations from a broad class of bipartite states in probabilistic theories cannot be arbitrarily nonlocal — they are constrained to obey the principle of *macroscopic locality*, or equivalently to lie within the set Q_1, which means in particular that they satisfy Tsirelson's bound for violation of the CHSH inequality. This theorem extends to all bipartite quantum states, which explains why quantum mechanics cannot violate macroscopic locality or Tsirelson's bound.

It would be natural to think that *all* bipartite states of strongly self-dual subsystems would respect macroscopic locality, but the house-shaped counterexample shows that this is not the case. An interesting open question, therefore, is the following: What additional property of local state spaces would ensure that all bipartite states give correlations which respect macroscopic locality? One suggestion is that sharpness might be able to ensure this. Sharpness is the constraint that for any ray extremal effect, there is a unique state on which this effect will occur with certainty. This property is very attractive from a physical point of view. It allows a natural definition of the post-measurement states of these effects, such that repeating a measurement reproduces the same

outcome. This extra constraint is indeed not satisfied by the house model, since the effect e_1 occurs with certainty for both states ω_1 and ω_5, but it is satisfied by odd n polygon models. Another possibility that seems to be plausible is that strong self-duality together with the property that all extremal states of the local systems can be transformed into one another reversibly might limit the set of possible correlations to the ones compatible with macroscopic locality.

Finally, it is worth emphasizing that two theories which have almost identical local state spaces can lead to dramatically different nonlocal correlations. In particular, given any finite level of accuracy, it is always possible to find a polygon model with an even and sufficiently large number of vertices n, which is locally indistinguishable from the quantum-like case, where the state space is a disc. Nevertheless, while quantum correlations are restricted, any non-signaling correlations can be distilled in the former model by using multiple copies of the maximally entangled state.

3. Entanglement swapping in GPTs

Recent developments in the study of quantum foundations concentrate on information theoretical properties. Related to this, is the question where the computational power of quantum computing stems from. Traditional answers to the question suggest "quantum parallelism" associated with quantum superpositions, as well as nonlocal dynamics induced by entanglement to be crucial elements for the speed up in quantum algorithms compared to the best known classical ones. In this chapter, we examine the consequences of these properties for low dimensional GPTs.

Superpositions in GPTs can be defined for non-classical transitive state spaces. That is, superpositions exist in systems with pure states that are not all linearly independent, but can be reversibly transformed into each other. A subset of perfectly distinguishable pure states correspond to the orthogonal basis in quantum theory. The other pure states that result from reversible transformations on these basis states are interpreted as superpositions.

Nonlocal dynamics, on the other hand, we associate with maximal entangled elements and primitives like steering, teleportation and entanglement swapping. In this chapter, we will focus on entanglement swapping, as it requires the other primitives. This means, the other primitives are possible, if entanglement swapping is possible. We will first explain how entanglement swapping works in quantum theory in Section 3.1. We then try to establish entanglement swapping of maximally entangled states for a modified version of boxworld in Section 3.2, but show that inconsistent arise for multipartite systems. These inconsistencies turn out to necessarily occur for any GPT with two-dimensional state spaces and local transitive dynamics (Section 3.3).

A seemingly conflicting theory that allows consistent entanglement swapping, although it has two-dimensional transitive local state spaces, is quantum theory

on a real Hilbert space. We show in Section 3.4 that this is only possible, because the joint systems violate the local tomography assumption of the GPT framework. A similar construction is illustrated to resolve the inconsistency issues for any theory with two-dimensional transitive local state spaces.

3.1. Entanglement swapping in quantum theory

As pointed out in Section 1.5.2 entanglement cannot be generated by mere local operations and classical communication (LOCC). However, it is possible to transfer entanglement. A well-known protocol in quantum information theory that achieves this is *entanglement swapping*. We will illustrate the protocol in the Dirac representation. The equivalent formulation with density matrices, i.e. with states in the sense of the GPT framework, is straight-forward.

The main objects we consider in this discussion are the maximally entangled elements of the Bell basis:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle\right) \tag{3.1}$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle - |11\rangle\right) \tag{3.2}$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle + |10\rangle\right) \tag{3.3}$$

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle - |10\rangle\right) \tag{3.4}$$

The setting consists of two initially separated parties. Let us call them Alice and Charlie. Alice and Charlie do not share entanglement with each other initially. But each of them shares a maximally entangled state $|\Phi^+\rangle$ with a third party that we call Bob. In summary, we have a four-partite system in the state

$$|\varphi\rangle^{AB_1CB_2} = |\Phi^+\rangle^{AB_1} \otimes |\Phi^+\rangle^{CB_2} \tag{3.5}$$

$$= \frac{1}{2} \left(|0000\rangle + |0011\rangle + |1100\rangle + |1111\rangle\right). \tag{3.6}$$

Entanglement swapping allows Alice and Charlie to get entangled with the help of Bob. To achieve this, Bob does a measurement in the Bell basis on his parts of the system. Let us now examine the post measurement state. According to Born's rule it can be determined by the projector on the basis element which corresponds to the outcome. For example suppose Bob gets an outcome corresponding to the basis element $|\Phi^+\rangle^{B_1B_2}$. This results in the post measurement state

$$|\varphi'\rangle = |\Phi^+\rangle^{B_1B_2}\langle\Phi^+|^{B_1B_2}|\varphi\rangle^{AB_1CB_2} = |\Phi^+\rangle^{B_1B_2} \otimes \langle\Phi^+|\varphi\rangle^{AC} \tag{3.7}$$

$$= \frac{1}{2}|\Phi^+\rangle^{B_1B_2} \otimes |\Phi^+\rangle^{AC} \tag{3.8}$$

where we used

$$\begin{aligned}\langle\Phi^+|\varphi\rangle^{AC} &= \frac{1}{2\sqrt{2}}\left(\langle 00| + \langle 11|\right)^{B_1B_2}\left(|0000\rangle + |0011\rangle + |1100\rangle + |1111\rangle\right)^{AB_1CB_2} \\ &= \frac{1}{2\sqrt{2}}\left(\underbrace{\langle 00|00\rangle}_{=1}|00\rangle + \underbrace{\langle 11|00\rangle}_{=0}|00\rangle + \langle 00|01\rangle|01\rangle + \cdots + \langle 11|11\rangle|11\rangle\right)^{AC} \\ &= \frac{1}{2\sqrt{2}}\left(|00\rangle + |11\rangle\right)^{AC} = \frac{1}{2}|\Phi^+\rangle^{AC}. \end{aligned} \tag{3.9}$$

which is the non-normalized conditional state that occurs in AC with probability 1/4. One can easily check that projections from the other basis elements also results in a conditional state in AC that reproduces the basis element up to normalization.

If Bob does not communicate this result to Alice and Bob, they just know that their joint state is given by one of those conditional states with equal probability. That is given by a mixture. More precisely it is given by the maximally mixed state that was also the effective (marginal) state before Bob's measurement. This is no coincidence, but a direct consequence of no-signaling that gets reaffirmed in this example. Communicating the result tells Alice and Charlie the particular conditional state, such that the state in AC reduces to the normalized version of equation (3.9).

In summary, we transformed a given pair of maximally entangled states

shared with Bob to a pair of random but equal Bell states. One of these pairs is completely held by Bob, whereas the other pair is now shared between Alice and Charlie.

3.2. Entanglement swapping in boxworld

There have been various attempts to generalize entanglement swapping, known from quantum mechanics [86], to a swapping protocol for PR boxes, i.e. a LOCC protocol that generates a PR box between two parties Alice and Charlie with the help of a third party (Bob) who shared an PR box with Alice and Charlie individually. A popular approach is known as *nonlocality swapping*. In contrast to entanglement swapping in quantum mechanics nonlocality swapping does not transform entangled states directly but introduces a so-called coupler that acts on probability distributions on Bob's side. It was shown in [87] that being able to apply the same coupler consistently to all PR boxes does not allow any nontrivial coupler for two-input-two-output boxes, i.e. nonlocality swapping is impossible for that case. By exclusion of all PR boxes but one, however, it was shown in [4] that nonlocality swapping can be done. The other PR boxes can be generated by a simple relabelling of measurements and outcomes such that they can be generated from the single box by mere classical local wirings. That the coupler is not sensitive to wirings is explained by introducing a genuine part of the box that corresponds to the (not further specified) state that gives rise to the correlations represented by a nonlocal box. The details on the underlying entanglement swapping on physical states that produces PR box correlations remain unclear.

In [35] it was shown that entanglement swapping in boxworld is not possible. This results from the fact that the standard version of boxworld does not allow any entangled measurements. We show in section 3.2.1 that one can modify joint systems in boxworld such that entangled states as well as entangled measurements are included. We study whether this construction can be extended to multipartite systems and show that this requires a different treatment of bipartite partitions consisting of otherwise equal elementary systems.

It is demonstrated that this modified theory shows potential for entanglement swapping, but treats subsystems differently.

3.2.1. A weakly self-dual version of boxworld

Systems with a one-to-one correspondence between the positive cone of unnormalized states and the dual cone, are referred to as *weakly self-dual* systems.

In boxworld single systems (e.g. a gbit) are weakly self-dual. However, for joint systems, in the standard formulation of boxworld with the maximum tensor product, the systems loose this property. Joint systems include all possible joint states, but no entangled effects. We would like to study a situation similar to quantum theory where entanglement in states and measurements is balanced with a one-to-one correspondence between effects and unnormalized states for single systems and for joint systems.

In order to find such a weakly self-dual tensor product we take a closer look on the tensor product of two qubits and try to mimic it. Exchanging the usual tensor product of two qubits by the maximal tensor product, one gets a joint state space that still includes all pure states available in standard quantum theory. However, it inhabits also some additional pure states that each can be generated by a partial transpose applied to a corresponding entangled pure quantum state. It is well known that such states are non-positive and therefore normally excluded [88]. As these additional unphysical states can be constructed by applying a partial transpose to the normal entangled states, each entangled state in a standard bipartite qubit system has a counterpart in the maximal tensor product that is regarded unphysical. At the same time each new state leads to new constraints on the entangled effects possible. From this point of view the standard tensor product of two qubits can be derived from the maximal tensor product by excluding half of the entangled pure states. This removes constraints on corresponding entangled effects.

We construct a weakly self-dual tensor product for local boxworld systems in a similar manner. We include separable pure states and only four (ω_{17}^{AB}, ω_{18}^{AB}, ω_{19}^{AB}, ω_{20}^{AB}) of the nonlocal extremals as pure joint states. The reduction of the

extremal entangled states, allows for new extremal entangled effects, such that joint states and effects are connected by a one-to-one mapping:

$$e_{17}^{AB} = \frac{2}{3}\left(e_1^A \otimes e_2^B - e_2^A \otimes e_2^B + e_2^A \otimes e_3^b + e_3^A \otimes e_1^B\right) \tag{3.10}$$

$$e_{18}^{AB} = \left(e_2^A \otimes e_2^B - e_3^A \otimes e_3^B + e_3^A \otimes e_4^B + e_4^A \otimes e_3^B\right) \tag{3.11}$$

$$e_{19}^{AB} = \left(e_1^A \otimes e_1^B - e_2^A \otimes e_2^B + e_2^A \otimes e_3^B + e_3^A \otimes e_2^B\right) \tag{3.12}$$

$$e_{20}^{AB} = \frac{2}{3}\left(e_2^A \otimes e_2^B - e_3^A \otimes e_2^B + e_3^A \otimes e_3^B + e_4^A \otimes e_1^B\right) \tag{3.13}$$

Nevertheless, since we keep some of the original entangled pure states, we are still able to produce PR boxes with all the powerful features that come with them.

Dynamics are connected to automorphisms of the state space. Gross et al. have shown that dynamics in standard boxworld are very limited, since invertible maps of the joint systems include only permutations of the local pure states [48]. This means that there is no reversible way to convert pure separable states to pure entangled states, e.g. as it is possible in quantum theory with the CNOT operation. If one want to include at least one of the original entangled pure states of boxworld, one can get the others by the local symmetries. Thus, excluding some of the entangled extremals leads to a situation where these local symmetries would map joint states allowed in the theory to those that have been excluded. This gives rise to a notion of complete positivity as opposed to positivity, similar to the situation in quantum theory. As a consequence the construction above restricts also local dynamics to be consistent with the joint elements.

3.2.2. Multipartite extension and entanglement swapping

Given the new tensor product, we want to extend it to multipartite systems as well. Using the example of a four-partite system $ABCD$, we will show in this section that this, if possible at all, can only be done under conditions that could be regarded unphysical, namely the symmetry under swapping of the identical elementary systems has to be broken.

No matter what the exact structure of the four-partite tensor product might be, it must at least contain any product state consisting of two bipartite states in AB and CD of the bipartite tensor product chosen before. Assuming that the systems BD allow for the same joint measurements, it can be shown that applying the entangled effects in (3.10) to the product $\omega^{AB} \otimes \omega^{CD}$ of two nonlocal extremal states would result in a collapse to an entangled pure state in AC, i.e. the realization of entanglement swapping. The resulting state ω^{AC} in AC however is not guaranteed to be in the bipartite tensor product chosen before. One can find combinations of states and effects such that ω^{AC} is one of the states of the maximal tensor product that we abolished for the construction of our new tensor product. For example if one gets outcome e_{17}^{BD} when measuring the state $\omega_{17}^{AB} \otimes \omega_{18}^{CD}$ the resulting collapsed state in AC is ω_{22}^{AC} that is not included in the weakly self-dual tensor product chosen before. This means that the bipartite tensor product for subsystem BD and AC must be different to the tensor product chosen for AB and CD. Hence, the initial construction can only be extended in a way that does not treat each pairs of particles in the same way.

Since elementary systems of the same form should be indistinguishable, it is questionable if this is consistent with a reasonable notion of a multipartite system. More formally, this shows that the construction of the weakly self-dual tensor product above cannot be extended to an associative tensor product.

3.3. Entanglement swapping in low dimensional GPTs

The inconsistencies shown in the last section are not a peculiarity of boxworld. In fact we will show in the following that the limitations are generic for two-dimensional single systems in the GPT framework.

3.3.1. Transitive two-dimensional state spaces

The pure states in quantum theory are connected by unitary transformations. That is, each pure state can be transformed into each other reversibly. Although most of the toy theories discussed so far also have this property, this is not the case for all GPTs. Counter-examples are the house-shaped state space and the kite-shaped state spaces in the transition between the classical and the gbit case in Chapter 2. We will call state spaces *transitive*, when all extremal states can be reversibly transformed into each other[1].

It is clear that transitivity implies a symmetry that is reflected in a group structure for the reversible transformations. As Dakic et al. point out in [23], the Shur-Auerbach lemma guarantees that every compact group has an orthogonal representation [89], whereas compactness is given for GPT transformations due to the fact that the range $[0, 1]$ of probabilities need to be preserved [1]. Consequently, using a suitable representation of the state space according to (1.22) it is always possible to arrange the extremal states of a transitive state space in such a way, that each state lies on a circumscribed sphere. The group of reversible transformations is a subgroup of the orthogonal group $O(n)$ on the state space dimension n.

Considering two-dimensional systems, we have to focus on the the subgroups of $O(2)$. These are $O(2)$ itself, $SO(2)$ as well as the dihedral groups D_n and the cyclic groups C_n. The elements of all these groups are two-dimensional rotations $R^+(\alpha)$ and reflection-rotations $R^-(\alpha)$ given by

$$R^+(\alpha) = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.14}$$

$$R^-(\alpha) = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ \sin(\alpha) & -\cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.15}$$

[1] More precisely, such geometrical bodies are called *vertex-transitive* or *isogonal* in mathematics.

where the finite groups D_n and C_n allow only discrete rotation angles α. The state space corresponding to these symmetries are the regular polygon systems and the disc-shaped state space from chapter 2, as well as truncated versions of the regular polygons.

3.3.2. Incompatible maximally entangled elements

We want to restrict our discussion on the entanglement swapping of maximally entangled states. The definition of maximally entangled elements requires an isomorphism between extremal effects and extremal states, which excludes the truncated polygons. So we only need to consider the polygon systems and the disc system from Chapter 2.

As introduced in this chapter the maximally entangled elements are given by the isomorphisms between cones and dual cones that are unbiased, i.e. have maximally mixed marginal states. The set of all maximally entangled states is given by all unbiased reversible maps of effects that acts on one side of the bipartite system to conditional states at the other part. Transitivity allows us to get all maximally entangled elements from a single generic one combined with the automorphism on local states and effects. In particular, we chose the generic state Φ as defined in equation (2.7). The other potential maximal entangled states Φ' are generated by matrix multiplication with the local symmetry transformations from Equation (3.14):

$$R' = \Phi' = R \cdot \Phi. \tag{3.16}$$

That Φ' is again a rotation/reflection matrix, results from the algebraic structure of the rotations/reflections. In particular, combinations via the usual

matrix product and their transpose yield the following relations:

$$\begin{aligned} R^+(\alpha)\cdot R^+(\beta) &= R^-(\alpha)\cdot R^-(\beta) = R^+(\alpha+\beta) \\ R^+(\alpha)\cdot R^-(\beta) &= R^-(\alpha+\beta) \\ R^-(\alpha)\cdot R^+(\beta) &= R^-(\alpha-\beta) \\ R^{+T}(\alpha) &= R^+(-\alpha) \\ R^{-T}(\alpha) &= R^-(\alpha). \end{aligned} \tag{3.17}$$

Potential maximally entangled effects corresponds basically to the same expressions as the potential maximally entangled states, since they correspond to the isomorphisms in the opposite direction. The difference is that they are weighted with some scalar normalization factor that limits the maximal measurement probabilities to one. For the discussion in this chapter it suffices to examine whether combinations of effects and states result in non-negative values. We will therefore omit normalization factors and replace both maximally entangled states and effects by the unscaled rotation/reflection matrices defined in (3.14).

The application of a joint effect on a joint state is given by the sum of the products of corresponding matrix components of both matrices, which can conveniently be written as a Hilbert-Schmidt inner product:

$$e(\omega) \propto \mathrm{tr}[R(\alpha)^T \cdot R(\beta)] = \sum_{ij} r_{ij}(\alpha)\, r_{ij}(\beta) \tag{3.18}$$

Using the algebraic relations in (3.17) we can reduce this to the trace of a single transformation matrix with some angle α':

$$e(\omega) \propto \begin{cases} \mathrm{tr}[R^+(\alpha')] = 1 + 2\,\cos(\alpha') \\ \mathrm{tr}[R^-(\alpha')] = 1 \end{cases} \tag{3.19}$$

In order to get positive values the combinations of the transformations corresponding to joint states and effects, have to result in either a reflection or in a rotation with rotation angles in the range $\alpha' \in [-2\,\pi/3, 2\,\pi/3]$. On the other

hand, combinations of joint states and effects that result in a rotation with α' in the range of $]2\,\pi/3, 4\,\pi/3[$ would result in unphysical negative values.

In the classical case, the local state space is given by a triangle with a D_3 symmetry, that can be generated by $R^+(2\,\pi/3)$ and $R^-(2\,\pi/3)$ respectively. Notably, the corresponding transformation matrices have a non-negative trace and are closed under matrix multiplication and transposition. A classical theory does of course not have entanglement. However, there are separable joint elements that reproduce the isomorphic mapping behavior that we defined as "maximally entangled". These elements are given by the same rotation/reflection matrices as the local symmetries. As these are closed under compositions we get non-negative values in (3.19), even if we include all "maximally entangled" elements.

For general regular polygons the assumption of local transitivity requires us to at least include all elements of the cyclic group C_n as local transformations, as the smallest transitive subgroup. Likewise, for the disc we have to at least include the local transformations of $SO(2)$. Consequently, if we have one maximally entangled state that is a rotation we get all the rotations with rotation angles that respect the polygon symmetry. Also the local transformations produce all reflections from a single reflection as a maximally entangled state. However, these minimal local symmetry transformations would not be able to convert between rotations and reflections. The same is true for maximally entangled effects. If a combination of maximally entangled effects and states result in a rotation, this necessarily allows negative values in (3.19) for some angles α' except for the triangle representing the classical case. On the other hand, if combinations result in a reflection we get non-negative values for any angles and therefore for any polygon. Consequently, the only possibilities to chose maximally entangled elements consistently are i) to have entanglement only for joint states or for effects, ii) to have maximally entangled states given by rotations and maximally entangled effects given by reflections, or iii) the other way round.

As we want to generate theories that allow nonlocal primitives like teleportation and entanglement swapping, we require entanglement for states and effects.

This excludes option i). The other options give valid results for any combinations of a single pair of a bipartite state and effect. However, we will show in the next section that inconsistencies arise in bigger multipartite systems.

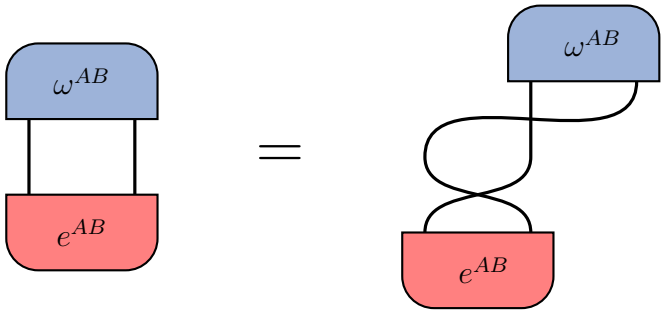
3.3.3. Composing nonlocal primitives

In the last section we used matrix products and the transpose operation exclusively in combination with the trace to get a convenient representation of inner products. There are also operational interpretations for these operations themselves. We will now discuss the meaning of the mathematical operations. In addition we introduce a graphical representation to illustrate more complex constructions.

Therefore, we map the experimental setup to a circuit of boxes with a logical flow from left to right or top to bottom. We depict states as blue boxes with one or more open "outputs" that can serve as inputs to measurements. Effects are the corresponding counterparts with open "inputs" where we chose a red color. A circuit that connects all outputs from states with all inputs from effects represent the probability of the corresponding state/effect combination. Similar graphical formalisms have been introduced by several authors [27, 90, 29].

Given a bipartite state in matrix representation, applying a matrix on the left side corresponds to a transformation of the first part, whereas applying something on the right acts on the second part of the system. The transpose swaps both parts. Bipartite states act also as maps from local effects on one part to the conditional state on the other part and likewise applying a local state to a bipartite effect yields an effective local effect on the other part. The trace connects the first part of a system with the last one. Using our graphical notation, we illustrate how these operations are used in the Hilbert-Schmidt inner product in Figure 3.1.

We can apply these rules to chain bipartite states and effects together, which allows us to build nonlocal primitives like teleportation and entanglement swapping.



$e^{AB}(\omega^{AB})$ $\qquad$ $\mathrm{tr}[e^{AB,T} \cdot \omega^{AB}]$

Figure 3.1.: Illustration of the mathematical operations involved in the Hilbert-Schmidt inner product: The transpose exchanges the inputs of the effect e^{AB}. Matrix multiplication then connects the second input of e^{AB} with the first output of the state ω^{AB}. The trace finally connects the remaining loose ends.

Teleportation We can combine an entangled effect in parts AB with an entangled state in parts BC to produce nonlocal transformations. Using maximally entangled elements, the construction transforms any state in A to a state in C that differs from the original state only by a reversible local transformation. Dependent on the measurement outcome in AB, given by different effects, we also get different transformations. Communicating the measurement outcome and reverting the transformation finally gives us the usual teleportation protocol as known from quantum information theory.

A particular combination of maximally entangled effects and states as shown in Figure 3.2 can itself be regarded as probabilistic teleportation [30], as it still depends on the corresponding outcome to occur.

Entanglement swapping As illustrated in Figure 3.3 entanglement swapping can be described by composing a maximally entangled state on AB with a maximally entangled effect in BC and finally another maximally entangled state in CD. The result is a reversible transformation from effects in A to states in D, which corresponds to a maximally entangled state in AD.

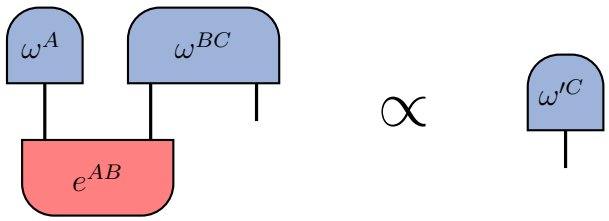


Figure 3.2.: Probabilistic teleportation is done by chaining a maximally entangled effect and a maximally entangled state: The maximally entangled effect e^{AB} maps any state ω^A to a effect in B that is again transformed to a state ω'^C in C via ω^{BC}. As maximally entangled elements are defined by isomorphisms, the whole construction forms an isomorphism of states in A to states in B.

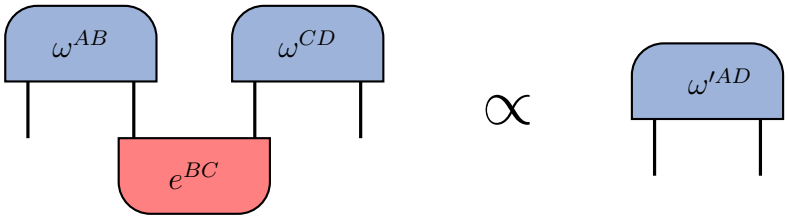


Figure 3.3.: Graphical illustration of entanglement swapping.

It is reasonable to assume that bipartite systems of the same single systems should always be described by the same composition rules. That is, we suspect the joint state in AD is ensured to be one of those allowed in AB and CD respectively. If this is the case, we call this consistent entanglement swapping. In the following, we show that consistent entanglement swapping is impossible for transitive two-dimensional systems.

3.3.4. Inconsistencies in multipartite systems

As pointed out the only ways to get compatible maximally entangled effects and states is to strictly define one of them to be proper rotations, whereas the others are given by reflections. The entanglement swapping protocol turns out to break this classification.

Suppose we choose maximum entangled joint state to be given by rotations,

implying reflections for maximally entangled effects. Entanglement swapping yields a reflection as a state in AD as it combines two rotations and one reflection:

$$R^{-,AD}(\alpha_1 + \alpha_2 - \alpha_3) = R^{+,AB}(\alpha_1) \cdot R^{-,BC}(\alpha_2) \cdot R^{+,CD}(\alpha_3). \tag{3.20}$$

Similarly, the opposite classification of states and effects transforms reflections to rotations by:

$$R^{+,AD}(\alpha_1 - \alpha_2 + \alpha_3) = R^{-,AB}(\alpha_1) \cdot R^{+,BC}(\alpha_2) \cdot R^{-,CD}(\alpha_3). \tag{3.21}$$

Consequently, entanglement swapping introduces inconsistencies for systems with two-dimensional transitive state spaces in the GPT framework.

3.4. Entanglement swapping enabled by violating local tomography

We have illustrated that the GPT framework forbids consistent entanglement swapping for systems with two-dimensional transitive state spaces. However, we can circumvent these issues, when we weaken the local tomography assumption. We first show that the violation of local tomography enables entanglement swapping in quantum theory on a real Hilbert space and then show that the same construction can be exploited to allow consistent entanglement swapping for the previously conflicting systems.

3.4.1. Quantum theory on a real Hilbert space

Quantum theory on real valued Hilbert space (*real quantum theory* for short) is known to allow consistent entanglement swapping, even though having subsystems with transitive two-dimensional state spaces. However, this theory does not allow to identify joint states by local measurements [91]. That is, it violates the local tomography assumption of the GPT framework. In fact, the density

matrices in real quantum theory are not elements of a tensor space anymore, but includes purely global degrees of freedom.

Before illustrating this, let us consider the pure states of real quantum theory in the Dirac formalism. As in the usual case, we have a Hilbert space with a orthonormal basis of Ket-vectors $|i\rangle$ with general pure states $|\Psi\rangle$ given by superpositions:

$$|\Psi\rangle = \sum_i c_i |i\rangle \qquad \sum_i |c_i|^2 = 1$$
$$\langle i|j\rangle = \delta_{ij} \tag{3.22}$$

The only difference to standard quantum theory is that coefficients $c_i \in \mathbb{R}$ are not complex, but restricted to real numbers. As in the usual case joint systems have states given by the same rules, with the basis of the joint Hilbert space, given by products:

$$|\Psi\rangle^{AB} = \sum_{ij} c_{ij} |i\rangle \otimes |j\rangle \qquad \sum_{ij} |c_{ij}|^2 = 1$$
$$\langle i|^A \otimes \langle j|^B (|k\rangle^A \otimes |l\rangle^B) = \langle i|k\rangle \langle j|l\rangle = \delta_{ik}\, \delta_{jl}, \tag{3.23}$$

where c_{ij} are again real coefficients. So, considering only Ket-vectors, it seems there is not much difference to the standard case and the tensor product structure is preserved. This changes when we look at the actual measurement statistics and the associated density matrices.

Density matrices in the standard case are given by positive Hermitian matrices. Hermitian matrices acting on a n-dimensional Hilbert space have n^2 independent degrees of freedom. The reason that both ket-vectors and density matrices scale multiplicatively results from $(m \cdot n)^2 = n^2 \cdot m^2$. We can decompose any Hermitian matrix ρ into a symmetric matrix S and an antisymmetric matrix A by

$$\rho = S + i\, A \tag{3.24}$$

Switching to real valued Hilbert spaces reduces the density matrices to positive symmetrical matrices with $d(n) = n\,(n+1)/2$ independent degrees of freedom. Therefore, the dimension of the Hilbert space dimensions scales differently than the dimension of operator spaces as

$$d(n \cdot m) \neq d(n) \cdot d(m). \tag{3.25}$$

The reason for this different scaling behavior is that Kronecker products of two antisymmetric matrices yields a symmetric matrix. Hence, there are valid global components that result from the product of components that are invalid for the local systems. This leads to a super-multiplicative scaling.

For example consider the density matrix of a standard qubit from a complex Hilbert space. It can be composed from a basis of the identity matrix $\mathbb{1}$ and the three Pauli matrices σ_x, σ_y, σ_z. The matrices σ_x and σ_z are symmetric, whereas σ_y is antisymmetric and imaginary. The density matrix for a system from a two-dimensional real Hilbert space, a so-called rebit, has no σ_y components. The local state space is given by a disc. A two-rebit system on the other hand, can have a $\sigma_y \otimes \sigma_y$ component, as the product is a real symmetric matrix. An example that inhibits such a component is the density matrix Φ^+ of the following real valued maximally entangled state

$$\begin{aligned} \Phi^+ &= |\Phi^+\rangle\langle\Phi^+| \qquad |\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right) \\ &= \frac{1}{4}\left(\mathbb{1} + \sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z\right). \end{aligned} \tag{3.26}$$

This purely global $\sigma_y \otimes \sigma_y$ component cannot be detected by local measurements. It turns out that entanglement in two-rebit systems is completely characterized by this non-local component [91]. This violates local tomography that is assumed in the GPT framework.

3.4.2. Entanglement swapping in real quantum theory

As entangled elements in real quantum theory have an additional $\sigma_y \otimes \sigma_y$-component, we can not represent them as 3×3-matrices anymore. Thus, we use the original basis of the standard qubit with all Pauli matrices. This allows us to embed real quantum theory in this larger space, where we set all components zero that can only result from states in a complex Hilbert space. In particular, we describe local elements by a vector with four components, where one component corresponding to σ_y is always zero. The joint elements are represented by 4×4-matrices, where the components have a fixed value of zero that correspond to the product of σ_y with another different Pauli matrix or the identity. This way we can still use matrix arithmetrics, even though the tensor product structure is broken in real quantum theory.

The set of all maximally entangled states are given by the $O(2)$ symmetries of the disc. In addition the $\sigma_y \otimes \sigma_y$-component is either 1 or -1 dependent whether the symmetry on the disc is a reflection or a rotation, such that the overall matrix is a reflection on $\mathbb{R}^4$. Consequently, the combination of a maximally entangled effect in AB and a maximally entangled state in BC yield a nonlocal rotation. As pointed out before, this can be interpreted as a probabilistic teleportation protocol. As the combination of rotations gives again a rotation, we can chain several of such constructions together and still get rotations. This way, real quantum theory guarantees consistent teleportation.

On the other hand, we can connect both ends of the teleportation construction by the trace and reinterpret it as the usual application of joint effects on joint states. It is easy to see that non-negativity of the trace behaves opposite to the previous case with 3×3 matrices that obey local tomography. Rotations give always non-negative values, such that applying maximally entangled effects on states in real quantum theory yields valid probability valued outcomes, although both are chosen as reflections. The trace of reflections themselves, however, can give negative values. With the given choice of joint elements this is not a problem as the trace of a single entangled state without an effect does not have any operational meaning.

Entanglement swapping can be build upon the teleportation construction. We only need to apply one side of a maximally entangled state to the teleportation apparatus. Recall that maximally entangled elements are defined as unbiased isomorphisms. A composition of isomorphisms obviously results in another isomorphism. Consequently, the result of the entanglement swapping set-up build from maximally entangled elements always yields a maximally entangled state. However, this state might not be valid, as it might conflict with the maximally entangled effects that have been defined. That is, some maximally entangled effects could give negative results on the generated maximally entangled states. We showed that this happens for two-dimensional transitive local state spaces and maximally entangled elements that obey local tomography. In real quantum theory, on the other hand, entanglement swapping turns out to be consistent. Teleportation corresponds to a rotation, whereas every entangled element is a reflection. This implies that the whole entanglement swapping set-up results in a reflection. As maximally entangled elements were defined as reflections in the beginning, there is no conflict.

3.4.3. Nonlocal dynamics for non-local-tomographic polygon systems

We can construct consistent entanglement swapping for systems with regular polygons as local state spaces in a similar way as provided in real quantum theory. That is we chose the both types of transformations, rotations and reflections for maximally entangled states and effects, but add an additional strictly global component that is set 1 or -1, such that the overall transformation is a reflection on a bigger linear space. As shown for all continuous rotation angles for the disc case that corresponds to real quantum theory, this prevents the conversion between rotations and reflections for joint elements and guarantees that bipartite joint states and effects are compatible, such that measurement probabilities are non-negative. Obviously, this is also guaranteed for discrete rotation angles. In addition, it is impossible to get invalid angles due to the group structure of the discrete rotation/reflections.

4. GPTs without the no-restriction hypothesis

The standard GPT framework assumes the no-restriction hypothesis, in which the state space of a physical theory determines the set of measurements. However, this assumption seems not physically motivated. In this chapter we generalize the framework to account for systems that do not obey the no-restriction hypothesis. We then show how our framework can be used to describe new classes of probabilistic theories, for example those which include intrinsic noise. Relaxing the restriction hypothesis also allows us to introduce a 'self-dualization' procedure, which yields a new class of theories that share many features of quantum theory, such as obeying Tsirelson's bound for the maximally entangled state. We then characterize joint states, generalizing the maximal tensor product. We show how this new tensor product can be used to describe the convex closure of the Spekkens toy theory, and in doing so we obtain an analysis of why it is local in terms of the geometry of its state space. We show that the unrestricted version of the Spekkens toy theory corresponds to a variant of boxworld and therefore would allow maximal nonlocal correlations.

The idea of removing the no-restriction hypothesis (or replacing it with other assumptions) has appeared sporadically in other works [27, 18]. However, until now a systematic analysis of the consequences of doing so has been lacking. In this chapter we provide a well-defined framework with the no-restriction hypothesis omitted, whilst keeping the other assumptions of the GPT framework. Our work then proceeds in two parts.

In the first part we show that this new framework encompasses more theories than before. For example, we show that theories with intrinsic noise can be described in our framework, but not in the existing GPT framework. We also

provide a procedure for constructing a *self-dual* theory from a theory which is not self-dual. The importance of this is that self-duality has been shown to imply 'quantum-like' (for example, limiting bipartite nonlocality to Tsirelson's bound for the maximally entangled state [3]). Hence this allows us to introduce a new class of probabilistic theories with 'quantum-like' behaviour, and crucially, this is a class of theories which does not satisfy the no-restriction hypothesis.

In the second part, we develop the treatment of composite systems. In particular, we show that our extension requires a new (and more general) definition of the tensor product for describing composite systems. This significantly extends the GPT framework, since it allows us to analyze the relationship between nonlocality and the geometry of the state space of a theory, building on previous work in this direction. For example, we show how the Spekkens toy theory [21] (for which the connection to GPTs had not been previously established) can be viewed as a GPT, but only in our more general framework. Moreover, this allows us to give an analysis of why the Spekkens theory is local, using the geometry of its state space.

4.1. The no-restriction hypothesis

We now consider in detail the no-restriction hypothesis, and the consequences of relaxing it.

4.1.1. Defining the set of effects

Let us recall the interdependence of state spaces and effect sets in the traditional framework. In general effects are restricted to give values in the range of $[0, 1]$ when applied to normalized states. As pointed out in the introduction of the traditional GPT framework in Section 1.2.9, the set of effects E is not restricted any further. That is, the set of effects is exactly the set of all probability-valued linear functionals on the given states. We will call this relationship between states and effects the *no-restriction hypothesis*, in accordance with [27]. It is

satisfied for classical probability theory and quantum theory.

As shown in Section 1.2.9 the set of effects under the no-restriction hypothesis is given by

$$E^{\max} := V_+^* \cap (u - V_+^*) \tag{4.1}$$

with the so-called dual cone

$$V_+^* := \{e \in V^* \,|\, e(\omega) \geq 0 \quad \forall \omega \in \Omega\}\,. \tag{4.2}$$

If the no-restriction hypothesis holds, then a theory is completely determined by the state space, since the effect set can be derived from the state space via the dual cone. On the other hand, the unnormalized states can be recovered as the dual of the dual cone, which is the *primal cone* V_+:

$$V_+ := \{\lambda\,\omega \,|\, \omega \in \Omega, \lambda \geq 0\} = (V_+^*)^*. \tag{4.3}$$

In the following we want develop the framework of GPTs without the no-restriction hypothesis. There are two main reasons for doing so:

1. The necessity of the no-restriction hypothesis is questionable from an operational perspective. Indeed, considering the physical meaning of states and effects there is no reason to believe that the possible preparation procedures determine possible measurements.

2. The possibility to cope with restricted systems generalizes the GPT framework to cover new scenarios that have not been accessible within the old framework.

4.1.2. Relaxing the no-restriction hypothesis

Let us note the constraints that still apply when the no-restriction hypothesis is removed. Clearly, effects still need to give probabilities when applied to any state. That is, when allowing violations of the no-restriction hypothesis, the set of probability-valued linear functionals on states in (4.1) remains an upper

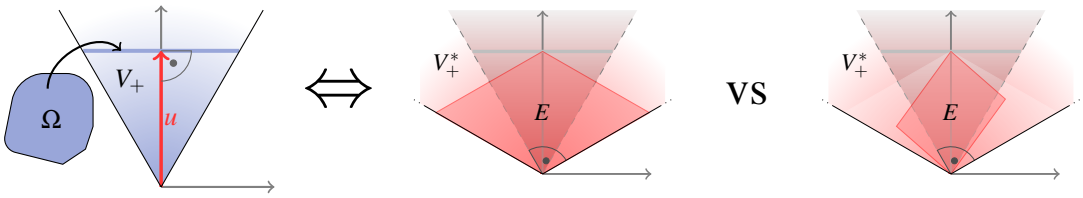


Figure 4.1.: The construction of the effect set E in the traditional GPT framework with no-restriction hypothesis is shown in the middle. Without the no-restriction hypothesis the definition of the effect set gets a independent part of the theory specification (right picture).

bound for possible effects. However, in general not all elements in this set need to represent a valid measurement outcome. Consequently, the set of effects E may actually be given by a *subset* of (4.1):

$$E \subseteq E^{\max}. \tag{4.4}$$

This is the crucial new ingredient in the GPT framework that we shall use in subsequent sections.

For operational reasons, we shall assume that E is a convex set. We have also identified the following four consistency conditions that also have to be met:

i) The unit measure u needs to be included in the restricted set as it is crucial for the definition of measurements.

ii) For every effect e included in E, the complement effect $\bar{e} = u - e$ needs to be included as well. We will show in Section 4.4 that including an effect, but not the complement can yield inconsistencies for joint states.

iii) For any effects e_i, e_j that are included in a common measurement M: $e_i + e_j \in E$. Operationally this corresponds to a 'coarse graining' where the discrimination between outcomes e_i and e_j is waived.

iv) Transformations map valid states to valid states. However, for any transformation T on states, there is also an adjoint transformation $T^\dagger$ on ef-

fects defined by $e[T(\omega)] = [T^\dagger(e)](\omega)$ for all states and effects. The effect set must be consistent with transformations, and hence we must have $\forall e \in E : T^\dagger(e) \in E$.

Apart from these consistency restrictions, the definition of the effect set E is now an independent part of the specification of the theory. In other words, the effect set E does not depend on the state space now, and the dual cone V_+^* is irrelevant for single systems. However, we will see in Section 4.5.2 that we still need it to classify consistent joint states.

Let us now consider how removing the no-restriction hypothesis will be useful. As shown above, the no-restriction hypothesis connects a set of states and effects via the respective dual-cone. Taking a closer look at the dual cone construction in (4.2), it can easily be seen that each extremal point of the primal cone describes a facet of the dual cone and the other way round. Therefore, arbitrary small changes in the primal cone, can have an enormous impact on the form of the dual cone. Consequently, the no-restriction hypothesis makes it extremely difficult to alter a theory in a controlled way. However, it has always been a central motivation for the framework of generalized probabilistic theories to find alternatives to quantum theory.

We shall now show in Section 4.2 and Section 4.3 that new models with interesting features can indeed be constructed when accepting violations of the no-restriction hypothesis. Furthermore, for joint systems, we will see in Section 4.4 and Section 4.5 how consistency conditions are affected.

4.2. Theories with intrinsic noise

The no-restriction hypothesis guarantees that for any pure state ω, there is an effect e, with $e \neq u$, such that $e(\omega) = 1$. In contrast, removing the no-restriction hypothesis allows for the modeling of systems with intrinsic noise, i.e. systems for which the unit measure is the *only* certain outcome for any state. For example, an isotropic unbiased implementation of noise can be achieved by restricting the effects to a set where the original extremal effects are replaced

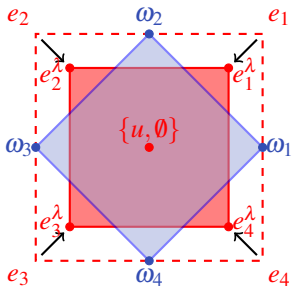


Figure 4.2.: Inclusion of noise into boxworld: State space and effects are both embedded into $\mathbb{R}^3$ and shown from above for illustration. The state space (blue) is given by a square. The effect set is the octahedron spanned by the extremal effects e_i, u and $\emptyset$. The noisy theory has a restricted effect set with extremal effects e_i^λ.

by mixtures with $u/2$ (except for $\emptyset$ and u itself). In order to combine noise and bias one can mix the extremal with another effect instead of $u/2$.

The inclusion of intrinsic noise by a modification of boxworld is illustrated in Fig. 4.2. The state space of a single system is given by a square. In the traditional model the effect set is determined by the no-restriction hypothesis. A noisy version of boxworld is given by mixing the extremal effects e_i with $u/2$:

$$e_i \mapsto e_i^\lambda = \lambda\, e_i + (1-\lambda)\,\frac{u}{2}. \tag{4.5}$$

This model is particularly interesting with respect to its potential non-local correlations in joint systems. This will be examined in more detail after introducing joint system of restricted theories in Section 4.4.

4.3. Self-dualization procedure

A particular class of systems that has gained a lot of interest recently are so-called *(strongly) self-dual* systems [30, 3, 31]. These are systems with a particular geometrical structure, shared by both classical probability theory and quantum theory. For strongly self-dual systems states and effects can be identified with each other and thus be represented by the same mathematical

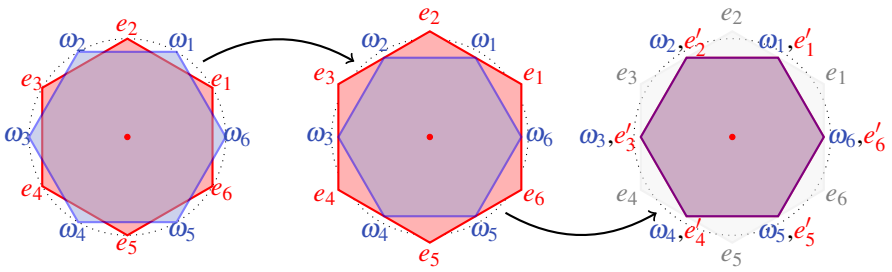


Figure 4.3.: Self-dualization of a hexagon system: The pictures show the state-space (blue) and the intersection of the effect cone (red) that lies in the same plane. In the first step the state cone will be embedded into the effect cone by an equivalence transformation (1.22). In the second step the effects not included in the state cone are abandoned.

objects. E.g. in quantum theory both states and effects are represented by positive hermitian operators.

Formally, strong self-duality is given by the following definition.

Definition 4.1. *A system is* strongly self-dual *iff there exists an isomorphism* $\Phi : V_+^* \mapsto V_+$ *giving rise to a corresponding symmetric bilinear form* T *with* $T(e, f) = e[\Phi(f)] = T(f, e)$ *and* $T(e, e) \geq 0$ *for all* $e, f \in V^*$.

That is, T provides a semi inner product on effects. In a similar way for strongly self-dual systems the inverse map Φ^{-1} leads to a semi inner product on states.

Strong self-duality greatly restricts the class of possible systems. As we describe below, the property of 'bit-symmetry' implies that a system is strongly self-dual [31], and there is evidence that non-local correlations of self-dual systems are limited [3]. In this section we provide a general construction rule to modify any system, such that it resembles the behaviour of strongly self-dual systems.

Theorem 4.1. *Any theory in the GPT framework can be modified to resemble strongly self-dual systems respecting Definition 4.1 with the dual cone* V_+^* *replaced by a truncated cone* $\mathcal{V}_+^*$.

Proof. Using our representation, we assume an embedding of effects and states in a common vector space with a scalar product mediating the application of effects on states. We start from an arbitrary theory for which the no-restriction hypothesis holds. The freedom of linearly transformations L^T from (1.22) allows us to strictly enlarge the effect cone V_+^*, while the corresponding inverse L^{-1} constricts the cone of unnormalized states V_+ to be strictly smaller. Hence, one can always represent the same physical theory, with V_+ embedded in V_+^*. We can then define a truncated the effect cone from $\mathcal{V}_+^* \subseteq V_+^*$, such that $\mathcal{V}_+^*$ coincides with the state cone V_+. Hence we can describe unnormalized effects and states with the same set of vectors. Consequently, the restriction of effects yields the vector space's scalar product to act as an inner product between states. This satisfies the definition of strong self-duality with the dual cone V_+^* exchanged for the truncated effect cone $\mathcal{V}_+^*$. The set of effects is then constructed from $\mathcal{V}_+^*$ by $E = \mathcal{V}_+^* \cap u - \mathcal{V}_+^*$. □

The connection between self-dualized systems and actual strongly self-dual systems is not only limited to a mere formal resemblance. In fact, the following example shows that self-dualized systems have features that strongly self-dual systems have when the no-restriction hypothesis is assumed.

4.3.1. Example: self-dualized polygons

Let us illustrate the self-dualization procedure on polygon systems introduced in chapter 2. As before the regular polygon-shaped state spaces of these systems are spanned by the extremal states from equation (2.3).

In the previous discussion of the theory we set $E(\Omega)$ of all possible measurement outcomes to be determined by the no-restriction hypothesis which results in the extremal effects given by equations (2.6) and (2.5) respectively.

While the cases with an odd number of vertices n are strongly self-dual, the even cases are not. The extremal rays of the dual cone of polygon systems with odd number of vertices, coincide with the scaled extremal states. However, for polygon system with an even number of vertices the primal and dual cones are only isomorphic and can be matched by a rotation of $\frac{\pi}{n}$. That is, the even

polygons are not strongly self-dual in the original models. We will now self-dualize these even-polygon systems using the procedure described in Theorem 4.1.

As discussed in Section 1.2.12 there is always the freedom to apply arbitrary bijective linear maps to all effects and the corresponding inverse map on all states. We use this to shrink the state space by $r_n \mapsto 1$ to fit in a circumscribed circle of radius one. Applying the inverse map to effects results in a effect cone with $r_n \mapsto r_n^2$. This new effect cone is strictly bigger than the cone of unnormalized states. By truncating this effect cone, such that the new extremal effect e_i' are given by

$$e_i' = \frac{1}{2}\left(e_i + e_{(i+1)\mathrm{mod}\ n}\right) = \frac{\omega_i}{2}, \tag{4.6}$$

the primal cone coincide with the new effect cone generated by the restricted effect set.

Let us demonstrate the self-dualization procedure explicitly, by using the polygon with $n = 4$ (this is the boxworld model). In the first step the pure states and effects are transformed to the equivalent representation given in (1.31) and (1.32). In this representation the effect cone is completely embedded in the cone of unnormed states. The actual self-dualization is then done by exchanging e_i for $e_i' = \omega_i/2$, shrugging off the effects not included in the primal cone.

For all self-dualized polygon models, another interesting feature emerges for the restricted case. Namely, there exists a specific pure state $\bar{\omega}$ for each pure state ω, such that they can be perfectly distinguished by an effect e with $e(\omega) = 1$ and $e(\bar{\omega}) = 0$. Furthermore, each pair of perfectly distinguishable states can be mapped reversibly to any other pair of perfectly distinguishable states. This feature is known as *bit symmetry,* and was shown to only hold for strongly self-dual systems in the traditional framework [31].

This demonstrates that the self-dualization procedure can actually reproduce properties thought to be specific for actual strongly self-dual systems. Note that the mathematical description of actual strongly self-dual systems can be complex. Using self-dualized systems might be an alternative that helps to

identify new features of strongly self-dual systems, even if one is not interested in the relaxation of the no-restriction hypothesis.

4.3.2. Spekkens's toy theory

In [21] Spekkens introduced a toy theory which replicates many features of quantum theory. For example, it exhibits a no-cloning theorem and a teleportation protocol. The theory is not explicitly probabilistic, since outcomes are not explicitly assigned probabilities. Instead, a graphical calculus is used. Given a state ω, the outcome i is only specified to be 'possible' or 'impossible'. The Spekkens theory in its original form also has no notion of arbitrary *convex mixing*, i.e. it does not have the property for any pair of states ω_1 and ω_2, there exists a state $p\omega_1 + (1-p)\omega_2$ for all probabilities $p \in [0, 1]$.

The ability to form convex mixtures is crucial to GPTs, and in particular to its operational motivation. Fortunately, there is a natural extension of Spekkens theory which is probabilistic and which does allow convex mixing (the probabilistic version of this theory was also introduced previously by Hardy in [92]). The state space Ω of a single system is then the octahedron. In the representation that we have used, the six extremal states (i.e. the pure states) are just given by the co-ordinates of the octahedron in $\mathbb{R}^3$, with an extra component for normalization. For example, the four extremal states that form the square

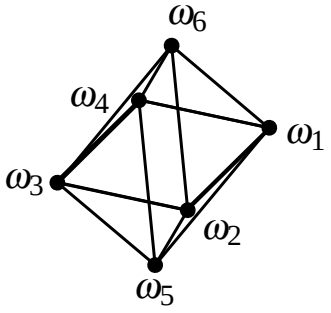


Figure 4.4.: The state space of the Spekkens model, with the six pure states ω_i labelled.

base of each tetrahedron are identical to the states for boxworld (see Fig. 4.4).

That is, for $i = 1, \ldots, 4$ the states are:

$$\omega_i = \begin{pmatrix} \cos(\frac{2\pi i}{4}) \\ \sin(\frac{2\pi i}{4}) \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^4, \tag{4.7}$$

and for $i = 5, 6$ the states are

$$\omega_i = (0, 0, \pm 1, 1)^T \in \mathbb{R}^4, \tag{4.8}$$

Now, the dual space of an octahedron is the cube. However, in the Spekkens theory, the space of effects is *identical* to the state space: it is also the octahedron depicted in Fig. 4.4. Since the octahedron can be obtained by restricting the cube (in the same way that is depicted for the hexagon in Fig. 4.3),we see that the Spekkens theory provides an example of a self-dualized theory. In particular, the convex probabilistic version of it is obtained using the self-dualization procedure defined in Theorem 4.1, and as described above for self-dualized polygons. Indeed, as with boxworld, the restricted effects are given by:

$$e'_i = \frac{\omega_i}{2}$$

Hence we see that, at least for single systems, self-dualized boxworld can be seen as an *extension* of the Spekkens theory: the state and effect space of the Spekkens theory are given by the extremal states and effects of self-dualized boxworld. We develop the analysis of joint systems for the Spekkens theory in Section 4.5.3.

We note that the single-system state space of Spekkens theory is also identical to that of *stabilizer quantum mechanics*, for which the only allowed states are the eigenstates of the Pauli operators, and the allowed transformations are the Clifford operations. As discussed in [21] and further in [93], the Spekkens theory and stabilizer quantum mechanics, however, differ in the group of reversible transformations that each theory specifies.

4.4. Joint systems in the traditional GPT framework

In the preceding sections we have not distinguished between single systems and joint systems. That is, our discussion so far (e.g. of self-dualization) has not involved any potential *subsystem* structure, whereby a system C can be divided into subsystems A and B, with each subsystem having well-defined states and effects. In the next section we shall consider how relaxing the no-restriction hypothesis affects composite systems. Before doing so, let us recall the treatment of joint systems in the traditional framework, i.e. when the no-restriction hypothesis is assumed to hold.

We will restrict the discussion of joint systems to the bipartite case with two subsystems, as the generalization of multipartite systems is straightforward. Bipartite joint states are given by elements of the product space

$$V^{AB} = V^A \otimes V^B \tag{4.9}$$

and joint effects are elements of $V^{AB*} = V^{A*} \otimes V^{B*}$ respectively [1].

We will represent joint states and joint effects by $n \times m$ matrices, with $n = \dim V^A = \dim V^{A*}$, $m = \dim V^B = \dim V^{B*}$. As for single systems, the application of effects on states results in the sum of the entry-wise products. This can be elegantly written as the Hilbert-Schmidt inner product

$$e^{AB}\left(\omega^{AB}\right) = \mathrm{tr}\left(e^T \cdot \omega\right) = \sum_{i,j} \epsilon_{ij}\, w_{ij}, \tag{4.10}$$

where we write $e^T \cdot \omega$ for the matrix product between the transpose of matrix e representing the joint effect e^{AB} and the matrix ω representing the joint state ω^{AB}.

[1] It can be shown that this follows from two conditions on joint states: i) local tomography ii) the no-signaling principle. The no-signaling principle forbids sending information by local operations on a joint state, and will be explained in more detail in Section 4.4. Local tomography is the identification of joint states by combinations of local measurements.

To define a composite system for a particular GPT (with specified state and effect spaces for individual systems), we must define the set of joint states $\Omega^{AB} = \{\omega^{AB}\}$, and the set of joint effects $E^{AB} = \{e^{AB}\}$, such that these are consistent with the individual systems. If the no-restriction hypothesis holds, then, as before, once the set of joint states Ω^{AB} is defined, the set of effects E^{AB} is determined. In this situation we need only consider the definition of Ω^{AB} in order to specify the behaviour of composite systems. There is much freedom in defining Ω^{AB}, but there are two boundary cases which we now discuss.

Lower bound on joint systems

Consider *independently prepared* systems A and B with states $\omega^A \in \Omega^A$, $\omega^B \in \Omega^B$. Treating the systems *jointly* as a composite AB, the overall preparation is represented by the product state $\omega^{AB} = \omega^A \otimes \omega^B$, with $\omega^{AB} \in V^{AB}$. However, just as classical mixtures are allowed for single systems, for joint systems mixtures between product states give valid joint states again. This corresponds to the ability of experimenters to classically correlate the preparations and measurements of the individual systems, e.g. two experimenters can agree on specific settings.

The set of unnormalized states only containing product states and their mixtures is known as the *minimal tensor product* $A_+ \otimes_{\min} B_+$.

Definition 4.2. *The* minimal tensor product *is given by*

$$A_+ \otimes_{min} B_+ := \left\{ \omega^{AB} \in V^{AB} \,\middle|\, \omega^{AB} = \sum_i \lambda_i \, \omega_i^A \otimes \omega_i^B, \quad \omega_i^A \in A_+, \omega_i^B \in B_+, \lambda_i \geq 0 \right\}. \tag{4.11}$$

It is the smallest possible set of unnormalized joint states ω^{AB} that is compatible with given state cones $A_+ \equiv V_+^A$, $B_+ \equiv V_+^B$ of subsystems A,B.

Similar reasoning applies to measurements, and so the set of joint effects is lower-bounded by the convex hull of product effects. Importantly, this includes the joint unit measure $u^{AB} = u^A \otimes u^B$, which is uniquely defined due to the

equivalence principle. Hence, normalization of joint states ω^{AB} is represented by the condition $u^{AB}(\omega^{AB}) = 1$. This allows us to define the bipartite state space $\Omega^{AB}_{\min}$ corresponding to the minimal tensor product:

$$\Omega^{AB}_{\min} := \left\{\omega^{AB} \in A_+ \otimes_{\min} B_+ \,\middle|\, u^{AB}\left(\omega^{AB}\right) = 1\right\} \tag{4.12}$$

$$= \left\{\omega^{AB} \in V^{AB} \,\middle|\, \omega^{AB} = \sum_i p_i\, \omega_i^A \otimes \omega_i^B, \right. \tag{4.13}$$

$$\left. \omega_i^A \in \Omega^A, \omega_i^B \in \Omega^B, p_i \geq 0, \sum_i p_i = 1 \right\}.$$

For classical subsystems (i.e. a simplex), the joint states and effects defined by the minimal tensor product is sufficient to describe joint classical systems. Theories with non-classical subsystems, however, allow joint states that cannot be interpreted as a mixture of product states, i.e. entangled states. The other extreme to the minimal tensor product allows all possible entangled states, as we now show.

Upper bound on joint systems

Everything introduced so far is valid independent of the no-restriction hypothesis. This changes now, as we ask for the maximal sets of joint states and effects consistent with the structure of the single systems.

First, let us focus on the traditional GPT framework with single systems obeying the no-restriction hypothesis. Given a specific state space the no-restriction hypothesis determines the effects for the single systems. As argued above, the joint system should at least incorporate product effects and their mixtures. Applying such joint effects to any potential joint state ω^{AB} should give probabilities. In particular this implies that the joint states form a subset of the following set of linear elements.

Definition 4.3. *The* maximal tensor product *is defined as*

$$A_+ \otimes_{\max} B_+ := \{\omega^{AB} \in V^{AB} | (e^A \otimes e^B)[\omega^{AB}] > 0, \quad \forall e^A \in E^A, e^B \in E^B\} \tag{4.14}$$

$$= \left(A_+^* \otimes_{min} B_+^*\right)^* . \tag{4.15}$$

It is the largest possible set of unnormalized joint states ω^{AB} that is compatible with given state cones A_+, B_+ of subsystems A, B that respect the no-restriction hypothesis.

Note that the second equality arises just by definition of the dual cone (4.2). Hence, we see that the maximal tensor product for states is given by the maximal set of joint states consistent with the minimal tensor product for effects. Similarly the maximal tensor product for effects is defined as the maximal set of joint effects consistent with the minimal tensor product for states. Elements in the maximal tensor product, but not in the minimal tensor product are called *entangled*.

To summarise the standard constructions of joint systems: the definition of a GPT includes the tensor product, which specifies the composition of subsystems. The minimal and maximal tensor product are only the extreme cases where the joint state space Ω^{AB} is chosen as smallest or the biggest set compatible with the state spaces Ω^A, Ω^B of single systems. In general, a GPT can be defined to include any set of joint states between those extremes.

For example, the joint state space in quantum theory lies strictly between the minimal and maximal tensor product. E.g. the partial transposed of density matrices representing entangled states of two qubits or a qubit and a qutrit are known to give invalid states for the quantum tensor product, because they are not positive on all entangled effects [88]. However, these states give positive results for separable measurements, i.e. they are in the maximal tensor product. Note that these states should not be misunderstood as part of quantum theory, but form a separate toy theory that omits any entangled measurements. Nevertheless, the additional states in the maximal tensor product of local quantum systems are useful for the study of entanglement in standard quantum theory,

as they correspond exactly to the set of entanglement witnesses.

Joint states as linear maps

For our generalization of the maximal tensor product, we shall use the following conception of joint states. Joint states can linearly map effects from one part of the joint system to unnormalized states of the other subsystem. This can be conveniently shown in the representation of joint states as matrices, since

$$\left(e^A \otimes e^B\right)\left[\omega^{AB}\right] = \mathrm{tr}\left[\left(e^A \otimes e^B\right)^T \cdot \omega^{AB}\right] = \left(e^A\right)^T \cdot \omega^{AB} \cdot e^B.$$

Using associativity of the matrix product, we can interpret parts of this expression as 'effective' states of the subsystems A and B. We define these *conditional states* as

$$\omega^A_{e^B} := \omega^{AB} \cdot e^B \tag{4.16}$$

$$\omega^B_{e^A} := \left(e^A\right)^T \cdot \omega^{AB} \tag{4.17}$$

The process of remotely preparing a state by a measurement outcome on the other part of a joint state is usually referred to as 'steering' [65]. It demonstrates that, when measuring only part of a joint system, the joint state acts as a linear map from effects of one side of the system to unnormalized states of the other part. The maximal tensor product coincides exactly with all possible linear maps of this form, i.e. it corresponds to all potential joint states that have valid conditional states for non-restricted systems [30]. This property will be central for the generalization of the maximal tensor product in the next section.

Conditional states at A are unnormalized: they are weighted with the probability of obtaining the corresponding measurement outcome at B. That is, the probability accounts for the potential ignorance of the outcome for observers at B. Consequently, if one knows the measurement outcome in B the effective description of the state in A is given by the normalized conditional state:

$$\tilde{\omega}^A_{e^B} = \frac{\omega^A_{e^B}}{p(e^B|\omega^{AB})} = \frac{\omega^A_{e^B}}{u(\omega^A_{e^B})}.$$

The *marginal state* or *reduced state* $\omega^A_{u^B}$ gives the description of the effective state on part A of a joint state ω^{AB}. This is a conditional state with $e^B = u^B$, and is already normalized i.e. $\tilde{\omega}^A_{u^B} = \omega^A_{u^B}$.

Note that this formalism still applies if the parts of the system are space-like separated, i.e. if there is no causal relationship between the measurement on the system B and the system A. However, the no-signaling principle states that steering cannot be used to transmit information, i.e. it does not allow for communication faster than the speed of light. The relationship between steering and the no-signaling principle is shown by the following theorem. First, we call a set of effects $\{e^A_i\}_i$, for any system A, a *complete measurement* if

$$\sum_i e^A_i = u^A.$$

An *incomplete measurement* is a set of effects $\{e^A_i\}_i$ that sum up to less than the unit measure.

Theorem 4.2. *Assuming the no-signaling principle, steering implies that all measurements are complete measurements.*

Proof. Consider two observers in part A and B respectively sharing a joint state ω^{AB}. The observer in B performs a measurement on his part and gets some measurement outcome e^B_j. Knowing the outcome the description of the system in A from his point of view is given by the normalized conditional state $\tilde{\omega}^A_{e^B_j}$. The other observer 'knows' only the coarse graining of the different measurement outcomes. I.e. from his point of view the state in A is an ensemble of possible 'post-measurement states' $\{\tilde{\omega}^A_{e^B_i}\}$.

Remember that the equivalence principle gives a one-to-one correspondence of states and specific measurement statistics. Consequently, no-signaling requires the state in A after the measurement on B to be identical to the original marginal state $\omega^A_{u^B}$ in order to prevent information transfer, i.e.

$$\sum_i p_i \, \tilde{\omega}^A_{e^B_i} = \sum_i \omega^A_{e^B_i} = \omega^A_{\sum_i e^B_i} = \omega^A_{u^B},$$

where we used the definition of the normalized conditional state and the linearity of effects.

Since the coarse grained conditional state needs to be equal to the marginal state for any joint state

$$\sum_i e_i^B = u^B.$$

□

As a consequence even an imperfect detector that might fail to 'click' requires to be modelled by a complete measurement which includes an additional 'no click'-effect for consistency with steering and no-signaling.

We will use the interpretation of the maximal tensor product as the set of all positive linear maps to generalize it for systems violating the no-restriction hypothesis.

4.5. The generalized maximal tensor product

As we have discussed, by removing the no-restriction hypothesis, the definition of a physical system now needs a specification of *both* the state space and the effect set. That is, the set of allowed states and the set of allowed effects can be chosen independently—except for the constraints discussed in Section 4.1. Let us now consider the specification of joint systems when the no-restriction hypothesis is removed.

The definition of the minimal tensor product $A_+ \otimes_{\min} B_+$ makes no reference to the effect sets E^A and E^B. I.e. it is constructed by products and their convex combinations. Therefore the minimal tensor product can be defined without assuming the no-restriction hypothesis, and hence carries over to our more general situation. Indeed, everything that we have introduced for joint systems so far is valid independently of the no-restriction hypothesis — with one exception.

The exception is the maximal tensor product. As before, we expect the

maximal tensor product to comprise all joint states that are compatible with the given subsystems. Compatibility can be broken down to two requirements: i) non-negative results on local effects ii) valid conditional states. For non-restricted systems both requirements are equivalent, as the no-restriction hypothesis implies consistent mappings (i.e. valid conditional states) if and only if local effects give non-negative results on joint states. Now, for the general case (i.e. without the no-restriction hypothesis), valid conditional states still guarantees non-negativity on local effects. However, the implication in the other direction that is used in the construction of the maximal tensor product is no longer secured.

4.5.1. Failure of the traditional maximal tensor product

Before generalizing the maximal tensor product we will show that the traditional construction rules fail for restricted systems.

The traditional maximal tensor product $A_+ \otimes_{\max} B_+$ is given by the dual of the set of separable effects. For restricted systems this yields two different variants. Eq. (4.14) seems to suggest a construction based on the restricted effects, whereas Eq. (4.15) uses the subsystems' dual cones, which are generated by the set of unrestricted effects. We show that neither choice gives the set of all joint states consistent with restricted subsystems.

The first variant is constructed as follows. Consider the restricted effects E^A of a subsystem A with an effect cone $E^A_+ := \{\lambda\, e^A \,|\, e^A \in E^A, \lambda \geq 0\}$. Following Eq. (4.3) we can construct a virtual, non-restricted system $\mathcal{A}$ with the state cone given by

$$\mathcal{A}_+ := \left\{\omega^A \in V^A \,\middle|\, e^A(\omega^A) \geq 0 \quad \forall e^A \in E^A_+\right\} \supseteq A_+ \tag{4.18}$$

$$\Rightarrow \mathcal{A}^*_+ = E^A_+. \tag{4.19}$$

I.e. the virtual system extends the unnormalized states, such that the no-restriction hypothesis is satisfied. The potential joint states from Eq. (4.14), correspond actually to the standard maximal tensor product $\mathcal{A}_+ \otimes_{\max} \mathcal{B}_+ = (E^A_+ \otimes_{\min} E^B_+)^*$ of the virtual systems $\mathcal{A}$, $\mathcal{B}$. Recall that according to the

interpretation of joint states as positive linear maps, $\mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$ is exactly the set of all maps from the restricted effect cones E_+^A (E_+^B) to the unnormalized virtual states $\mathcal{B}_+$ ($\mathcal{A}_+$) on the other side of the bipartite system. In other words, this construction includes joint states that allow the preparation of conditional states in the subsystems not limited to the initial definition of the state spaces Ω^A, Ω^B, but to those of the virtual systems instead.

For example in a bipartite system of self-dualized boxworld with extremal states according to (1.31) and restricted extremal effects $e_i' = \omega_i/2$ the potential joint state

$$\omega^{AB} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \Omega_{\max}^{\mathcal{AB}} \tag{4.20}$$

gives positive values on any pair of restricted effects. However, some conditional states are not valid for the actual system A, e.g. $\tilde{\omega}_{e_1'}^A = (-1, 1, 1)^T \notin \Omega^A$.

The second variant of the traditional maximal tensor product $(A_+^* \otimes_{\min} B_+^*)^*$ is based on the dual cones A_+^*, B_+^* according to Eq. (4.15). The resulting joint states are also consistent with the restricted effects, since the latter is included in the set of all of effects. However, this construction omits joint states which are consistent *only* with the restricted effects. For example, for self-dualized boxworld the tensor product defined in such a way would not include the identity matrix. As the identity matrix gives valid conditional states and returns positive values for any pair of effects, there is no reason to regard it as invalid, though.

4.5.2. Construction of the generalized maximal tensor product

As shown above, the traditional construction rules for the maximal tensor product lead to inconsistencies when applied to theories not obeying the no-restriction hypothesis. In this section we shall construct a *generalized* maximal tensor product $A_+ \overline{\otimes}_{\max} B_+$:

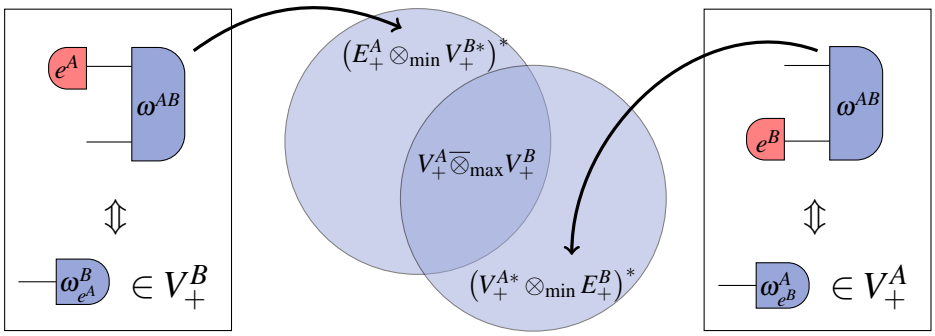


Figure 4.5.: Illustration of the construction of the generalized maximal tensor product.

Definition 4.4. *The* generalized maximal tensor product *of systems A, B with primal cones A_+, B_+, dual cones A^*_+, B^*_+ and effect cones E^A_+, E^B_+ is given by*

$$\begin{aligned} A_+ \overline{\otimes}_{\max} B_+ &:= \left(E^A_+ \otimes_{min} B^*_+\right)^* \cap \left(A^*_+ \otimes_{min} E^B_+\right)^* \\ &= \left(E^A_+ \otimes_{min} B^*_+ \cup A^*_+ \otimes_{min} E^B_+\right)^*. \end{aligned} \tag{4.21}$$

For the saturated case, dual cones and effect cones are identical, and we recover the usual maximal tensor product. Hence our construction is indeed a generalization of the existing definition of the maximal tensor product. For restricted systems it strictly lies between both traditional variants introduced in Section 4.5.1. That is,

$$(A^*_+ \otimes_{\min} B^*_+)^* \subseteq A_+ \overline{\otimes}_{\max} B_+ \subseteq (E^A_+ \otimes_{\min} E^B_+)^*. \tag{4.22}$$

The generalized maximal tensor product determines all joint states consistent with general subsystems regardless whether the no-restriction hypothesis holds or not. I.e. all joint states with valid conditional states are included, as shown in the following theorem.

Theorem 4.3. *Let $\omega^{AB} \in V^{AB}$. Then $\omega^{AB} \in A_+ \overline{\otimes}_{\max} B_+$ iff ω^{AB} has well-defined conditional states:*

$$\tilde{\omega}^A_{e^B} \in \Omega^A \text{ and } \tilde{\omega}^B_{e^A} \in \Omega^B$$

for all $e^A \in E^A$ *and* $e^B \in E^B$.

Proof. We shall show that

$$\omega^B_{e^A} \in B_+ \text{ iff } \omega^{AB} \in \left(E^A_+ \otimes_{\min} B^*_+\right)^* \tag{4.23}$$

and that

$$\omega^A_{e^B} \in A_+ \text{ iff } \omega^{AB} \in \left(A^*_+ \otimes_{\min} E^B_+\right)^*. \tag{4.24}$$

Since $A_+ \overline{\otimes}_{\max} B_+$ is defined as the intersection of sets of linear maps $\left(E^A_+ \otimes_{\min} B^*_+\right)^*$ and $\left(A^*_+ \otimes_{\min} E^B_+\right)^*$, this will establish the thesis.

First we show the $A \to B$ direction, i.e. (4.23), which is the statement that $\left(E^A_+ \otimes_{\min} B^*_+\right)^*$ is the set of all and only those joint states ω^{AB} such that each ω^{AB} defines a map from effects $e^A \in E^A_+$ on system A to valid unnormalized states $\omega^B \in B_+$.

Recall that for non-restricted systems the traditional maximal tensor product is already known to give all positive linear maps from the effect cone of one part of the system to the state cone of the other part for both directions. We now show that $\left(E^A_+ \otimes_{\min} B^*_+\right)^*$ can be interpreted as the traditional maximal tensor product $\mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$ of two virtual systems $\mathcal{A}$ and $\mathcal{B}$ that obey the no-restriction hypothesis. $\mathcal{A}$ is the virtual system that has already been introduced in (4.18). It has an extended virtual state cone $\mathcal{A}_+$, since $(E^A_+)^* \subset \mathcal{A}_+$. However, the actual effect cone $E^A_+ = \mathcal{A}^*_+$ is kept. The opposite situation applies to $\mathcal{B}$. Here, the effect set E^B_+ is extended to the dual cone B^*_+, so that $E^B_+ \subset B^*_+$, where B^*_+ representing the full set of potential unnormalized effects. However, the original state cone $B_+ = \mathcal{B}_+$ is kept. With these conventions

$$\left(E^A_+ \otimes_{\min} B^*_+\right)^* = \mathcal{A}_+ \otimes_{\max} \mathcal{B}_+$$

follows directly from the definition of the traditional tensor product in (4.15). Hence for the $A \to B$ direction, this means that $\left(E^A_+ \otimes_{\min} B^*_+\right)^*$ contains all positive linear maps from the restricted effects in A to allowed states in B. That is, $\omega^{AB} \in \left(E^A_+ \otimes_{\min} B^*_+\right)^*$ is a sufficient and necessary condition for $\omega^B_{e^A} \in B_+$

which proves (4.23). However, for the traditional maximal tensor product the same joint states also coincide with the positive maps in the other direction $\mathcal{B} \to \mathcal{A}$, potentially including invalid mappings.

By swapping the roles of A and B in the above argument, we similarly obtain that the set $\left(A_+^* \otimes_{\min} E_+^B\right)^*$ includes all linear maps that are consistent for the $B \to A$ direction, but also those which lead to inconsistencies in $A \to B$ opposite direction. Hence we obtain (4.24) as the intersection of the sets $\left(E_+^A \otimes_{\min} B_+^*\right)^*$ and $\left(A_+^* \otimes_{\min} E_+^B\right)^*$, which includes only those joint states that are consistent in both directions. □

If ω^{AB} has well-defined conditional states, then in particular it is locally positive:

$$\left(e^A \otimes e^B\right)\left[\omega^{AB}\right] \geq 0$$

which provides a useful necessary condition that joint states must satisfy. Note that the pure product states stay extremal in $A_+ \overline{\otimes}_{\max} B_+$, as they are the unique joint states with pure marginals.

As mentioned above the generalized maximal tensor product reduces to the traditional maximal tensor product if both systems obey the no-restriction hypothesis. In the following we show that the tensor products are equivalent even if only one of the systems is unrestricted.

Theorem 4.4. *The generalized maximal tensor product $A_+ \overline{\otimes}_{\max} B_+$ of two systems reduces to the traditional maximal tensor product $(A_+^* \otimes_{min} B_+^*)^*$ with respect to the dual cones A_+^*, B_+^*, if one of the systems obeys the no-restriction hypothesis.*

Proof. Without loss of generality, assume that system A is the unrestricted system. Since system A is unrestricted, we have $A_+^* = E_+^A$ and consequently

$$A_+ \overline{\otimes}_{\max} B_+ = (A_+^* \otimes_{\min} B_+^*)^* \cap (A_+^* \otimes_{\min} E_+^B)^*. \tag{4.25}$$

Thus, the generalized maximal tensor product is now given by the intersection of the standard maximal tensor product from the dual cones and the standard maximal tensor product with respect to the dual cone in A, but the original

restricted effects in B. As discussed above, the construction of the traditional maximal tensor product is defined by positivity with respect to the product effects. In general, the restriction of a set of functionals enlarges the set of elements for which each functional yields a positive real. Hence the restriction of the local effect cone from B_+^* to E_+^B yields a strictly larger tensor product, and so $(A_+^* \otimes_{\min} B_+^*)^* \subset (A_+^* \otimes_{\min} E_+^B)^*$. The intersection of a set with a strictly larger set obviously gives the original set, which in this case is the standard maximal tensor product $(A_+^* \otimes_{\min} B_+^*)^*$ constructed from the local dual cones. □

In the traditional GPT framework the choices of tensor products for states and effects are not independent, as the no-restriction hypothesis does not only apply to single systems, but to the joint system as well. Having the minimal tensor product for joint states (effects) does in fact constitute the maximal tensor product for the set of joint effects (states). This restriction seems inappropriate given that arbitrary single systems can actually be emulated by classical systems with constrained measurements [33], whereas entanglement is a strictly non-classical feature.

In our modified framework that is also valid for systems violating the no-restriction hypothesis, this is no longer the case. We have seen that we can generalize the maximal tensor product, but nevertheless we are not forced to use this for states when we choose the minimum tensor product for effects and the other way round.

4.5.3. Examples of joint systems

To give some specific examples for the generalized maximal tensor product, we have calculated it for the toy theories introduced in Section 4.2 and Section 4.3. For systems with extremals having strictly rational coordinates this is done using the program lrs [94], whereas for the other cases the double description method [95] is used for an analytic calculation.

Noisy boxworld

In the original unrestricted version of boxworld joint systems are given by the maximal tensor product, including the 16 extremal product states and 8 pure entangled joint states $\Phi^{AB} = \frac{1}{2}(\omega_1 \otimes \omega_2 - \omega_2 \otimes \omega_2 + \omega_2 \otimes \omega_3 + \omega_3 \otimes \omega_1)$ and respectively the states transformed by local symmetries.

These entangled extremals can be interpreted as a maximally entangled state of two such systems, as they form a isomorphic map and have totally mixed reduced states. They correspond to a rotation of $\frac{\pi}{4}$ and the local symmetries of the state spaces.

This theory has become very popular as it shows nonlocal correlations beyond those possible in quantum theory, when choosing between two possible binary measurements at each side of the bipartite systems. Let us denote the two measurements $\{M_x^A\}$ and $\{M_y^B\}$ for each of the systems A and B respectively: we index the measurements at each system with $x, y \in \{0, 1\}$. Each measurement has binary outcomes, labelled with $a, b \in \{0, 1\}$ for systems A and B respectively. For example, the $x = 0$ measurement on system A consists of a pair of effects $M_x^A = \{e_0^x, e_1^x\}$ satisfying $e_0^x + e_1^x = u$; similarly for the $x = 1$ measurement on system A, and $y \in \{0, 1\}$ measurements on system B. This leads to a bipartite conditional probability distribution

$$P(a, b|x, y) := (e_a^x \otimes e_b^y)[\omega^{AB}] \tag{4.26}$$

We define the correlation

$$C_{xy} := P(a = b|x, y) - P(a \neq b|x, y).$$

To introduce the Clauser-Horne-Shimony-Holt (CHSH) inequality for demonstrating nonlocality, we introduce the parameter

$$S := |C_{00} + C_{01} + C_{10} - C_{11}|,$$

For classical systems it is upper bounded by the CHSH inequality [38]

$$S^{\mathrm{C}} \leq 2,$$

whereas for quantum theory it must satisfy $S^{\mathrm{Q}} \leq 2\sqrt{2}$ [40]. However, local measurements on the maximally entangled state Φ in boxworld can produce correlations which reach the algebraic maximum $S^{\max} = 4$, i.e. the theory allows the post-quantum correlations known as PR boxes [42].

For the noisy version of boxworld introduced in Section 4.2 there is still a notion of a maximally entangled state in the generalized maximal tensor product, namely

$$\Phi^{\lambda} = \omega^{AB}_{\mathrm{ent},1} = \mathrm{diag}(\frac{1}{\lambda}, \frac{1}{\lambda}, 1) \cdot \Phi^{1}, \tag{4.27}$$

i.e. the original maximally entangled state Φ^1 combined with a mapping of the effects on one side of the system to the original unrestricted set. Note that this map does not undo the restriction of effects completely. The reversion only happens to occur in this particular case when mapping to states of the other part. On the other part, however, only restricted effects can be applied to. Consequently, the correlations possible with restricted systems will be different to those possible in unrestricted systems.

Furthermore, constructing the generalized maximal tensor product it turns out there are 4 different classes of new pure joint states that are entangled but not maximally entangled. These are representatives of each class

$$\begin{aligned}
\omega^{AB}_{\mathrm{ent},2} &= -\alpha\,\omega_2 \otimes \omega_2 + \beta\,\omega_2 \otimes \omega_4 + \beta\,\omega_4 \otimes \omega_2 - \alpha\,\omega_4 \otimes \omega_4 \\
\omega^{AB}_{\mathrm{ent},3} &= -\alpha\,\omega_2 \otimes \omega_2 + \beta\,\omega_2 \otimes \omega_3 + \beta\,\omega_4 \otimes \omega_2 - \alpha\,\omega_4 \otimes \omega_3 \\
\omega^{AB}_{\mathrm{ent},4} &= -\alpha\,\omega_2 \otimes \omega_2 + \beta\,\omega_2 \otimes \omega_4 + \beta\,\omega_3 \otimes \omega_2 - \alpha\,\omega_3 \otimes \omega_4 \\
\omega^{AB}_{\mathrm{ent},5} &= -\alpha\,\omega_3 \otimes \omega_2 + \beta\,\omega_4 \otimes \omega_1 + \beta\,\omega_4 \otimes \omega_2 - \alpha\,\omega_4 \otimes \omega_3 \\
\text{with } \alpha &= \frac{1-\lambda}{4\,\lambda}, \beta = \frac{1+\lambda}{4\,\lambda},
\end{aligned} \tag{4.28}$$

where the other elements of the class only differ by the local symmetries.

In conclusion the generalized maximal tensor product is spanned by 96 pure states. Namely, it consists of 16 local pure states, 8 pure entangled states of class $\omega^{AB}_{\text{ent},1}$, 8 of class $\omega^{AB}_{\text{ent},2}$, 16 of class $\omega^{AB}_{\text{ent},3}$, 16 of class $\omega^{AB}_{\text{ent},4}$ and 32 states of class $\omega^{AB}_{\text{ent},5}$.

Considering local measurements on one instance of any of the nonlocal extremal states the maximal CHSH violation S^λ as a function of the parameter λ of the restricted model can be shown to be $4\,\lambda^2$. Note that this bound is only guaranteed for the correlations that occur from direct measurements. However, it is known that wiring the measurements on multiple joint states via classical post-processing, might give rise to a distillation of correlations beyond for some values of λ [73].

Self-dualized polygons

Interestingly, not only boxworld but all bipartite polygon systems allow a joint state with features known from the maximally entangled state of ordinary quantum theory. Namely, the linear maps corresponding to these states are given by isomorphisms of the dual and primal cones with maximally mixed reduced states. The $2\,n$ different maximally entangled states correspond to the elements of the dihedral group. For even n, the maximally entangled states include an additional rotation of π/n mapping the dual cone of one part to the primal cone of the other part. It was shown that non-local correlations based on two binary local measurements on the maximally entangled states at each side show correlations strictly weaker than quantum correlations for the odd case, whereas the unrestricted even case shows correlations as strong as those of quantum theory or stronger [3].

Replacing the original polygon systems with even n by their self-dualized versions, the maximally entangled states lose the additional rotation as the new effect cone and the state cone coincide. Note, that the self-dualized single systems become subtheories of the theory given in the limit $n \to \infty$, i.e. the quantum case, as both states and effects form strict subsets. Thus, the correlations on the maximally entangled state form a strict subset of those in quantum theory, in contrast to the unrestricted case which allows post-quantum correla-

tions. In fact, the maximally entangled states of general self-dualized theories in general are inner product states, as defined in Chapter 2, where we also proved that these have correlations limited by Tsirelson's bound, which cannot be distilled beyond.

For self-dualized boxworld the generalized maximal tensor product is given by the 16 local pure states, the 8 states $\omega^{AB}_{\text{ent},1}$ representing the identity and symmetry mappings as well as a class of 64 pure entangled states $\omega^{AB}_{\text{ent},2} = 1/4(-\omega_1 \otimes \omega_1 + \omega_1 \otimes \omega_3 + 2\,\omega_2 \otimes \omega_4 + \omega_3 \otimes \omega_1 - \omega_3 \otimes \omega_3 + 2\,\omega_4 \otimes \omega_2)$. The generalized maximal tensor product for the pentagon system is equal to the traditional one. For hexagon systems one gets already 7672 extremals with 475 classes of entangled pure states.

Systems with a higher number of vertices are neither computational feasible with our methods nor do we expect further insights from their characterization.

Spekkens's toy theory

The Spekkens theory that we introduced earlier is a local theory, meaning that (in the probabilistic version) it cannot violate any Bell inequalities. However, as discussed in [21], the Spekkens theory has entangled states. This raises the question of why the Spekkens theory does not exhibit bipartite nonlocality. In contrast, a classical theory, i.e. a simplex, is local but does not have entangled states. One could then ask, given that the Spekkens theory has entangled states, but is local, what must be *added* to the definition of the theory to make it nonlocal?

In our framework, the answer to this question can be clearly understood in terms of the geometry of the state space. First, recall that the state space Ω^A of a single system in the Spekkens theory is an octahedron, and the effect space E^A is identically the same, i.e. E^A not the full dual space. Consider a pair of single systems A and B in the Spekkens theory. Since the effect space E^A is not the full dual space A^*_+, we must use the generalized tensor product $\Omega^{AB} = A_+ \overline{\otimes}_{\max} B_+$ to define the bipartite states. Note that Ω^{AB} contains more states than the original Spekkens theory, since the latter has 24 extremal bipartite entangled states, whereas the former is of the order of 10^6 (calculated

using lrs). However, consider the following bipartite state:

$$\omega^{AB} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.29}$$

It is straightforward to verify that ω^{AB} leads to well-defined conditional states for system B for all effects $e^A \in E^A$, i.e.:

$$\tilde{\omega}^B_{e^A} \in \Omega^B$$

and correspondingly for conditional states for system A when using effects on system B. In particular, it is also easily checked that $(e^A \otimes e^B)[\omega^{AB}] \geq 0$ for any pair of effects e^A and e^B. Hence by Theorem 4.3, this shows that ω^{AB} is in the generalized tensor product $A_+ \overline{\otimes}_{\max} B_+$. It is also a state in the convex closure of the original Spekkens theory, since it decomposes as $\omega^{AB} = \frac{1}{2}\sigma^{AB} + \frac{1}{2}\tau^{AB}$, where σ^{AB} and τ^{AB} are both in the set of 24 extremal bipartite entangled states defined in the original Spekkens theory.

Now, since the Spekkens theory is local, the CHSH inequality (4.5.3) is satisfied for any choice of measurements M_x and M_y on the state ω^{AB}, or any other bipartite state. However, let us consider the unrestricted effect space A_+^* from which the restricted space E^A for the Spekkens theory was derived. The unrestricted effect space of the octahedron is the cube. We can represent the normalised extremal effects as the vertices of a cube:

$$e_i = \frac{1}{2}\,(\pm 1, \pm 1, \pm 1, 1)^T$$

Now, suppose that we use the cube to be the effect space for the octahedron, i.e. we use the full dual space. It is easily shown that the state ω^{AB} defined in Eq. 4.29 is again in the generalized maximal tensor product $A_+ \overline{\otimes}_{\max} B_+$ (which coincides with the usual maximal tensor product in this case). However, we can now provide measurements which violate the CHSH inequality. In

particular, consider two measurements for Alice given by $M_0^A = \{e_0, u - e_0\}$ and $M_1^A = \{e_1, u - e_1\}$ where:

$$e_0 = \frac{1}{2}(1, 1, -1, 1)^T, \; e_1 = \frac{1}{2}(-1, -1, -1, 1)^T$$

and two measurements for Bob given by $M_0^B = \{e_0, u - e_0\}$ and $M_1^B = \{e_1, u - e_1\}$. By using these choices of measurements in Eq. 4.26 and the following equations, we obtain the value of the CHSH parameter: this is $S = 4$. This is the value attained by PR boxes, and hence using the full effect space essentially yields the same nonlocality as boxworld.

We therefore see that the Spekkens theory can be embedded into a nonlocal theory by embedding the effect space of single system into the full dual cone. Moreover, we see completing the Spekkens theory in this way yields boxworld. This provides a new understanding of why the Spekkens theory is local: the measurements are too restricted.

Nonlocality in linearly restricted theories

Now, while restriction of an effect space yields fewer local measurements, there are at the same time also additional joint states in the generalized maximal tensor product. An open question is then whether restriction of local effects always limits the possible correlations to those in the unrestricted systems. In the following we provide a positive answer to the question for the special case where the restriction results from a linear bijection.

Theorem 4.5. *Let A and B be systems with restricted effect sets $E^A = L^A \cdot E^{A,\max}$ that can be generated by a linear bijective map L^A from the full unrestricted set of probability valued functionals $E^{A,\max}$. Then correlations on the generalized maximal tensor product $A_+ \overline{\otimes}_{\max} B_+$ of the restricted systems A, B can be at most as strong as on the maximal tensor product $(A_+^* \otimes_{min} B_+^*)^*$ of the systems with unrestricted effect sets.*

Proof. First let us rewrite the correlations possible in the unrestricted systems. Note that for any theory there is some freedom in choosing representations of

effects and states. In particular, one can transform the effect set by a linear bijection L, and counter it with a corresponding transformation $(L^{-1})^T$ on states; this will not affect measurement probabilities, since:

$$(L \cdot e)[(L^{-1})^T \omega] = (L \cdot e)^T \cdot (L^{-1})^T \cdot \omega = e^T \cdot (L^{-1} \cdot L)^T \cdot \omega = e^T \cdot \omega = e(\omega). \tag{4.30}$$

The set of bipartite correlations in a GPT results from applying each of the possible combinations of local effects in $E^{A,\max}$ and $E^{B,\max}$ to each of the joint states in the traditional maximal tensor product $(A_+^* \otimes_{\min} B_+^*)^*$. Accordingly, we denote the set of all correlations as $E^{A,\max} \otimes E^{B,\max}[(A_+^* \otimes_{\min} B_+^*)^*]$. Using Eq. (4.30) this can be translated into

$$\begin{aligned} &E^{A,\max} \otimes E^{B,\max} \left[\left(A_+^* \otimes_{\min} B_+^*\right)^* \right] \\ &\quad = \left(L^A \cdot E^{A,\max} \otimes L^B \cdot E^{B,\max}\right) \left[\left(L^{A,-1} \otimes L^{B,-1}\right)^T \cdot \left(A_+^* \otimes_{\min} B_+^*\right)^* \right] \\ &\quad = E^A \otimes E^B \left[\left(E_+^A \otimes_{\min} E_+^B\right)^* \right]. \end{aligned} \tag{4.31}$$

Consequently, the correlations of the unrestricted effects on the standard maximal tensor product of the unrestricted systems is equal to the correlations by the restricted effect set on the standard maximal tensor product of the restricted systems.

We now compare this to the correlations that are actually possible in the restricted systems given by $E^A \otimes E^B(A_+ \overline{\otimes}_{\max} B_+)$, which uses the generalized maximal tensor product. According to Eq. (4.25) we have $A_+ \overline{\otimes}_{\max} B_+ \subseteq (E_+^A \otimes_{\min} E_+^B)^*$, and together with Eq. 4.31 this implies:

$$\begin{aligned} E^A \otimes E^B \left(A_+ \overline{\otimes}_{\max} B_+\right) &\subseteq E^A \otimes E^B \left[\left(E_+^A \otimes_{\min} E_+^B\right)^* \right] \\ &= E^{A,\max} \otimes E^{B,\max} \left[\left(A_+^* \otimes_{\min} B_+^*\right)^* \right]. \end{aligned} \tag{4.32}$$

□

Note that this result only applies to restrictions by a linear bijection. The general question if non-linear restrictions of effects can produce correlations

exceeding the unrestricted case is still open.

4.6. Discussion

In this chapter, we have extended the framework of generalized probabilistic theories. Given an arbitrary state space the traditional framework determines the possible measurement outcomes as corresponding to the complete set of probability valued linear functionals on states. In contrast to the traditional framework, our generalization allows the set of states and and the set of effects to be defined separately. As a result the upper bound for the set of joint states, known as the maximal tensor product, is no longer valid in its traditional form, but has to be replaced by a generalized version.

As an application for restricted models, we provided a self-dualization procedure that alters any theory by restricting the set of effects, such that states and the restricted effects are similarly related as states and unrestricted effects in strongly self-dual systems. We introduce specific examples for which the self-dualization does not only give a formal resemblance but reproduces a phenomenon called *bit symmetry* shown to only hold for strongly self-dual systems in the traditional framework [31]. Furthermore, these self-dualized models show quantum correlations, whereas the original models have correlations that are stronger than quantum correlations. In particular, the correlations of boxworld—a theory known to allow correlations only restricted by the no-signaling principle—has classical correlations if self-dualized, even though the generalized maximal tensor product includes maximally entangled states. We showed how the Spekkens theory is related to this model, since it is also self-dual and violates the no-restriction hypothesis: but were it to satisfy this principle, by taking the full dual cone, it would produce nonlocal correlations.

As another application for restricted models, we show that restrictions can be used to alter theories, such that their measurements are inherently noisy. This is different to the unrestricted theories, since in our noisy theories it holds that for pure states there is no non-trivial extremal effect occurring with certainty. We derive the maximal CHSH violation [38] of a noisy version of boxworld as

a function of a noise parameter.

The modified framework is therefore suitable for examining new situations that could not be addressed using the traditional framework. In particular the self-dualization procedure might be useful for the study of strong self-duality that has recently received much interest [30, 3, 31].

5. Conclusions

The framework of generalized probabilistic theories (GPTs) is a well-established tool used to study the physical foundations of quantum theory. The basic idea of the GPT framework is to introduce a general notion of a physical theory combined with the tools to formally describe theories. The mathematical formalism of the framework connects arbitrary measurement statistics with states and 1-bit measurements in a similar way as the density matrix formalism in quantum theory. This is used to construct toy theories and study the relations between their mathematical and physical properties. As quantum theory is a special case, we can get insights on its particular physical properties by contrasting them to properties seen in other theories. This thesis uses the GPT framework to get a better understanding of the roles of nonlocality and entanglement from a foundational point of view. In addition, we also extend the framework itself.

An introduction of the GPT framework is given in Chapter 1. We put particular emphasis on a some aspects that are usually not addressed in the literature: i) the GPT framework is able to incorporate arbitrary measurement statistics into a physical theory ii) if only single systems are considered any seemingly non-classical properties could potentially be explained by a higher-dimensional classical theory with restricted measurements iii) entanglement and nonlocal correlations in joint systems are genuine non-classical properties that forbid such explanations.

Conversely, we pointed out that the properties of joint systems are not completely independent of the structure of the single systems. In particular, we show in Chapter 2 that the subsystems indirectly affect the nonlocal correlations possible in a theory. The specific correlations of quantum theory are shown to be connected to those on joint states that act like an inner product

on the measurements. The existence of such joint states and equivalent states depend on a geometrical property of the subsystems called strong self-duality. Strong self-duality is notably present in both quantum and classical systems, but very special among toy theories [3, 30, 31].

In order to extend the range of physical theories and properties that can be studied with the GPT framework, we furthermore provided modifications of the framework itself. Chapter 3 shows how we can use similar representations of states and effects, when allowing purely global degrees of freedom (i.e. how to describe theories that violate the so-called local tomography assumption). Chapter 4 generalizes the framework to allow restricted measurements, whereas the so-called no-restriction hypothesis in the traditional framework completely determines measurements from possible states.

These generalizations not only allow to describe completely new situations, but also provide technical tools that simplifies the study of phenomena that were already addressed in the traditional framework and in quantum information theory. They therefore build the ground for possible future works in several directions. The modification in Chapter 3 was illustrated to enable entanglement swapping to be examined with structurally simple lower dimensional systems. This might also be useful for the study of other phenomena connected to nonlocal dynamics. Chapter 4 shows how intrinsic noise can be manifested in the definition of a effective theory. One could use this to examine the noise sensitivity of quantum information theoretical protocols by applying them to a manifestly noisy version of quantum theory. The self-dualization procedure introduced in Section 4.3 simplifies the study of strong self-duality that has recently gathered lots of attention. Another possible avenue for further work is the study of the relationship between our constructions of more general joint systems in Section 4.5 and the formalism of categorical quantum theory [96] in which the definition of composite systems is the primary formal device.

A. Appendix

A.1. Partial order of effects

Given two effects e_i, e_j one is dominated by the other iff it occurs with lower probability for any state:

$$e_i \leq e_j \Leftrightarrow e_i(\omega) \leq e_j(\omega) \forall \omega \in \Omega \tag{A.1}$$

Note that there are effects that cannot be compared in such a way. There might be states that give higher probabilities for e_i, while other states give higher probabilities on e_j. Therefore this is called a partial order.

A partial order on the elements of a vector space can be induced by a convex cone. The partial order of effects is based on the dual cone V_+^* with

$$e_i \leq e_j \Leftrightarrow e_i \in e_j - V_+^*. \tag{A.2}$$

To see that this is equivalent to (A.1) recall that the dual cone V_+^* is given by all elements in V^* with non-negative results on the state space elements $\omega \in \Omega$. Consequently, subtracting one of these elements from an effect cannot result in a bigger value for any state.

Note that the dual cone depends on the state space Ω. Extension of a model with new states might therefore also affect the partial order.

A.2. Maximal CHSH value of the polygon systems

In the main text, we gave expressions for the maximal CHSH value returned by measurements on a maximally entangled state of two n-vertex polygon systems. The expression for even n is given in (2.11), and for odd n, in (2.12). The choice of angles that maximize these quantities is not unique. We will see below that we have to take two different sets of optimal angles into account.

Table A.1.: Optimal angles

	α_0^*	α_1^*	β_0^*	β_1^*
Set 1	0	$\frac{\pi}{2}$	$\frac{\pi}{4}$	$-\frac{\pi}{4}$
Set 2	0	$\frac{\pi}{2}$	$-\frac{3\pi}{4}$	$\frac{3\pi}{4}$

Note that the optimization has been performed without any restriction on the values of the angles α_x^* and β_y^*. However, due to the polygon structure of our model, only specific angles, corresponding to extremal effects, are admissible. Thus the optimal CHSH values are obtained by taking the extremal effects which are closest to the optimal angles. The deviation from the optimal angles will be called $\Delta\alpha_0, \Delta\alpha_1, \Delta\beta_0, \Delta\beta_1$. Without loss of generality we set $\Delta\alpha_0$ to 0. A detailed analysis reveals a total of eight classes of deviation angles characterized by the remainder $x = n \mod 8$ of the division of n by 8. For a free choice of angles both sets in Table A.1 lead to the same maximum value of the CHSH-coefficient. Whether the available extremal effects are closer to the angles of set 1 or set 2, however, depends on the number of vertices. It turns out that for even n as well as for $x \in \{1, 7\}$ this is the case for set 1, whereas for $x \in \{3, 5\}$ the smallest derivation can be achieved to set 2. The maximal CHSH value for each polygon system is given by the following parameters for (2.11) and (2.12):

$$\begin{aligned} \beta_y &= \beta_y^* + \Delta\beta_y \\ \alpha_x &= \alpha_x^* + \Delta\alpha_x \end{aligned}$$

Table A.2.: Analytical expression for the maximal CHSH-violation of polygon boxes

x	$\Delta\alpha_1$	$\Delta\beta_0$	$\Delta\beta_1$	S
0	0	$\frac{\pi}{n}$	$\frac{\pi}{n}$	$2\sqrt{2}$
1	$\frac{-\pi}{2n}$	$\frac{-\pi}{4n}$	$\frac{\pi}{4n}$	$\frac{2}{\left(1+\sec\left(\frac{\pi}{n}\right)\right)^2}\left[1+\sec\left(\frac{\pi}{n}\right)\left(2\cos\left(\frac{n+3}{4n}\,\pi\right)+6\sin\left(\frac{n+1}{4n}\,\pi\right)+\sec\left(\frac{\pi}{n}\right)-2\right)\right]$
2	$\frac{\pi}{n}$	$\frac{\pi}{2n}$	$\frac{-\pi}{2n}$	$\sec\left(\frac{\pi}{n}\right)\left[3\cos(\frac{n+2}{4n}\,\pi)+\sin\left(\frac{n+6}{4n}\,\pi\right)\right]$
3	$\frac{\pi}{2n}$	$\frac{\pi}{4n}$	$\frac{-\pi}{4n}$	$\frac{-2}{\left(1+\sec(\frac{\pi}{n})\right)^2}\left[1-\sec\left(\frac{\pi}{n}\right)\left(6\cos\left(\frac{n+1}{4n}\,\pi\right)+2\sin\left(\frac{n+3}{4n}\,\pi\right)-\sec\left(\frac{\pi}{n}\right)\right)\right]$
4	0	0	0	$2\sqrt{2}\,\sec(\frac{\pi}{n})$
5	$\frac{-\pi}{2n}$	$\frac{-\pi}{4n}$	$\frac{\pi}{4n}$	$\frac{-2}{\left(1+\sec(\frac{\pi}{n})\right)^2}\left[1-\sec\left(\frac{\pi}{n}\right)\left(6\sin\left(\frac{n+1}{4n}\,\pi\right)+2\cos\left(\frac{n+3}{4n}\,\pi\right)-\sec\left(\frac{\pi}{n}\right)\right)\right]$
6	$\frac{\pi}{n}$	$\frac{-\pi}{2n}$	$\frac{\pi}{2n}$	$\sec(\frac{\pi}{n})\left[\cos\left(\frac{n+6}{4n}\,\pi\right)+3\sin\left(\frac{n+2}{4n}\,\pi\right)\right]$
7	$\frac{\pi}{2n}$	$\frac{\pi}{4n}$	$\frac{-\pi}{4n}$	$\frac{2}{\left(1+\sec\left(\frac{\pi}{n}\right)\right)^2}\left[1+\sec\left(\frac{\pi}{n}\right)\left(2\sin\left(\frac{n+3}{4n}\,\pi\right)+6\cos\left(\frac{n+1}{4n}\,\pi\right)+\sec\left(\frac{\pi}{n}\right)-2\right)\right]$

The eight classes can clearly be seen in Fig. 2.5. The analytic expressions for the maximal CHSH value as a function of the number of vertices n and the remainder x are given in Table A.2.

Bibliography

[1] L. Hardy, "Quantum theory from five reasonable axioms," *preprint* (2001) , arXiv:quant-ph/0101012.

[2] P. Ball, "Physics: Quantum quest," *Nature* **501** (2013) no. 7466, 154–156.

[3] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, "Limits on nonlocal correlations from the structure of the local state space," *New J. Phys.* **13** (2011) no. 6, 063024.

[4] P. Skrzypczyk, N. Brunner, and S. Popescu, "Emergence of quantum correlations from nonlocality swapping," *Phys. Rev. Lett.* **102** (2009) 110402.

[5] P. Janotta and R. Lal, "Generalized probabilistic theories without the no-restriction hypothesis," *Phys. Rev. A* **87** (2013) 052131.

[6] P. Janotta and R. Lal, "Non-locality in theories without the no-restriction hypothesis," in Proceedings 10th International Workshop on *Quantum Physics and Logic,* Barcelona, Spain, July 17-19, 2013, vol. XX of *to be published in EPTCS*, p. XXX. Open Publishing Association, 2013.

[7] L. Hardy and R. Spekkens, "Why physics needs quantum foundations," *Phys. Can.* **66** (2010) no. 2, 73–76. http://www.cap.ca/en/article/why-physics-needs-quantum-foundations.

[8] P. Janotta and H. Hinrichsen, "Generalized probability theories: What determines the structure of quantum theory?." accepted for publication in J. Phys. A, 2014.

[9] G. Mackey, *Mathematical Foundations of Quantum Mechanics.* Mathematical physics monograph series. Benjamin, 1963.

[10] G. Ludwig, *Foundations of quantum mechanics.* No. Bd. 1 in Texts and monographs in physics. Springer, 1983.

[11] G. Ludwig, *An Axiomatic Basis for Quantum Mechanics: Volume 1: Derivation of Hilbert Space Structure.* An Axiomatic Basis for Quantum Mechanics: Derivation of Hilbert Space Structure. Springer, 1985.

[12] K. Kraus, *States, Effects, and Operations Fundamental Notions of Quantum Theory*, vol. 190 of *Lecture Notes in Physics*. Springer Berlin / Heidelberg, 1983.

[13] T. Heinosaari and M. M. Wolf, "Nondisturbing quantum measurements," *J. Math. Phys.* **51** (2010) 092201.

[14] D. Reeb, D. Reitzner, and M. M. Wolf, "Coexistence does not imply joint measurability," *Journal of Physics A: Mathematical and Theoretical* **46** (2013) 462002.

[15] C. Ududec, *Perspectives on the Formalism of Quantum Theory*. PhD thesis, University of Waterloo, Canada, 2012.

[16] M. P. Müller, O. C. O. Dahlsten, and V. Vedral, "Unifying typical entanglement and coin tossing: on randomization in probabilistic theories," *Communications in Mathematical Physics* **316** (2012) no. 2, 441–487.

[17] M. Pawłowski and A. Winter, "Hyperbits: The information quasiparticles," *Phys. Rev. A* **85** (2012) 022331.

[18] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, "Entropy and information causality in general probabilistic theories," *New J. Phys.* **12** (2010) no. 3, 033024.

[19] C. Pfister and S. Wehner, "An information-theoretic principle implies that any discrete physical theory is classical," *Nature Comm.* **4** (2013) 1851.

[20] P. Janotta, "Nonlocal correlations in generalized probabilistic theories," Master's thesis, University of Würzburg, Germany, 2010.

[21] R. W. Spekkens, "Evidence for the epistemic view of quantum states: A toy theory," *Phys. Rev. A* **75** (2007) 032110.

[22] A. J. P. Garner, O. C. O. Dahlsten, Y. Nakata, M. Murao, and V. Vedral, "A framework for phase and interference in generalized probabilistic theories," *New J. Phys.* **15** (2013) no. 9, 093044.

[23] B. Dakic and C. Brukner, *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, ch. Quantum theory and beyond: is entanglement special?, pp. 365–392. Cambridge University Press, 2011.

[24] L. Masanes and M. P. Müller, "A derivation of quantum theory from physical requirements," *New J. Phys.* **13** (2011) no. 6, 063001.

[25] A. Wilce, "Conjugates, correlation and quantum mechanics," *preprint* (2012) , arXiv:1206.2897.

[26] J. Barrett, "Information processing in generalized probabilistic theories," *Phys. Rev. A* **75** (2007) 032304.

[27] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Informational derivation of quantum theory," *Phys. Rev. A* **84** (2011) 012311.

[28] I. Bengtsson, S. Weis, and K. Zyczkowski, "Geometry of the set of mixed quantum states: An apophatic approach," in *Geometric Methods in Physics. XXX Workshop 2011*, Trends in Mathematics, pp. 175–197. Springer, 2013.

[29] L. Hardy, "Reformulating and reconstructing quantum theory," *preprint* (2011) , arXiv:1104.2066v3.

[30] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Teleportation in general probabilistic theories," in *Mathematical Foundations of Information Flow (Clifford Lectures Information 2008)*, Proceedings of symposia in applied mathematics, pp. 25–47. American Mathematical Society, 2012. arXiv:0805.3553.

[31] M. P. Müller and C. Ududec, "Structure of reversible computation determines the self-duality of quantum theory," *Phys. Rev. Lett.* **108** (2012) 130401.

[32] L. Masanes, A. Acin, and N. Gisin, "General properties of nonsignaling theories," *Phys. Rev. A* **73** (2006) 012112.

[33] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*. Publications of the Scuola Normale Superiore. Springer, 2011.

[34] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, "Generalized no-broadcasting theorem," *Phys. Rev. Lett.* **99** (2007) 240501.

[35] A. J. Short and J. Barrett, "Strong nonlocality: a trade-off between states and measurements," *New J. Phys.* **12** (2010) no. 3, 033034.

[36] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?," *Phys. Rev.* **47** (1935) 777–780.

[37] J. S. Bell, "On the einstein podolsky rosen paradox," *Physics* **1** (1964) 195–200.

[38] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.* **23** (1969) 880–884.

[39] A. Aspect, "Bell's inequality test: more ideal than ever," *nature* **398** (1999) 189–190.

[40] B. Cirelson, "Quantum generalizations of Bell's inequality," *Lett. Math. Phys.* **4** (1980) no. 2, 93–100.

[41] G. Vallone, G. Lima, E. S. Gomez, G. Canas, J.-A. Larsson, P. Mataloni, and A. Cabello, "Bell scenarios in which nonlocality and entanglement are inversely related," *Phys. Rev. A* **89** (2014) 012102.

[42] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Found. Phys.* **24** (1994) no. 3, 379.

[43] M. Navascues, S. Pironio, and A. Acin, "Bounding the set of quantum correlations," *Phys. Rev. Lett.* **98** (2007) 010401.

[44] M. Navascues, S. Pironio, and A. Acin, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," *New J. Phys.* **10** (2008) no. 7, 073013.

[45] A. Doherty, Y.-C. Liang, B. Toner, and S. Wehner, "The quantum moment problem and bounds on entangled multi-prover games," in *Computational Complexity, 2008. CCC '08. 23rd Annual IEEE Conference on*, pp. 199–210. 2008.

[46] J. Uffink, "Quadratic bell inequalities as tests for multipartite entanglement," *Phys. Rev. Lett.* **88** (2002) 230406.

[47] L. Masanes, M. P. Mueller, R. Augusiak, and D. Perez-Garcia, "Existence of an information unit as a postulate of quantum theory," *PNAS* **110** (2013) no. 41, 16373.

[48] D. Gross, M. P. Müller, R. Colbeck, and O. C. O. Dahlsten, "All reversible dynamics in maximally nonlocal theories are trivial," *Phys. Rev. Lett.* **104** (2010) 080402.

[49] P. Janotta, "Generalizations of boxworld," in Proceedings 8th International Workshop on *Quantum Physics and Logic,* Nijmegen, Netherlands, October 27-29, 2011, vol. 95 of *EPTCS*, pp. 183–192. Open Publishing Association, 2012.

[50] G. de la Torre, L. Masanes, A. J. Short, and M. P. Müller, "Deriving quantum theory from its local structure and reversibility," *Phys. Rev. Lett.* **109** (2012) 090403.

[51] L. Masanes, M. P. Müller, D. Perez-Garcia, and R. Augusiak, "Entanglement and the three-dimensionality of the bloch ball," *preprint* (2013) , `arXiv:1111.4060v2`.

[52] M. P. Müller and L. Masanes, "Three-dimensionality of space and the quantum bit: an information-theoretic approach," *New J. Phys.* **15** (2013) no. 5, 053040.

[53] I. Gelfand and M. Neumark, "Mat. sb.," *Math. Sb.* **12(54)** (1943) 197–217. `http://mi.mathnet.ru/eng/msb6155`.

[54] I. E. Segal, "The group algebra of a locally compact group," *Trans. Amer. Math. Soc.* **61** (1947) 69–105.

[55] W. F. Stinespring, "Positive functions on c^*-algebras," *Proc. Am. Math. Soc.* **6** (1955) 211.

[56] M.-D. Choi, "Completely positive linear maps on complex matrices," *Linear Algebra and its Applications* **10** (1975) no. 3, 285–290.

[57] A. Jamiolkowski, "Linear transformations which preserve trace and positive semidefiniteness of operators," *Rep. Math. Phys.* **3** (1972) no. 4, 275–278.

[58] M. S. Leifer and R. W. Spekkens, "Towards a formulation of quantum theory as a causally neutral theory of bayesian inference," *Phys. Rev. A* **88** (2013) 052130.

[59] E. Schrödinger, "Discussion of probability relations between separated systems," *Proc. Cambridge Phil. Soc.* **31** (1935) no. 4, 553.

[60] H. Barnum, C. P. Gaebler, and A. Wilce, "Ensemble steering, weak self-duality, and the structure of probabilistic theories," *Found. Phys.* **43** (2013) no. 12, 1411–1427.

[61] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels," *Phys. Rev. Lett.* **70** (1993) 1895–1899.

[62] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, "Quantum repeaters: The role of imperfect local operations in quantum communication," *Phys. Rev. Lett.* **81** (1998) 5932–5935.

[63] H. Barnum, S. Beigi, S. Boixo, M. B. Elliott, and S. Wehner, "Local quantum measurement and no-signaling imply quantum correlations," *Phys. Rev. Lett.* **104** (2010) 140401.

[64] A. Acin, R. Augusiak, D. Cavalcanti, C. Hadley, J. K. Korbicz, M. Lewenstein, L. Masanes, and M. Piani, "Unified framework for correlations in terms of local quantum observables," *Phys. Rev. Lett.* **104** (2010) 140404.

[65] J. Oppenheim and S. Wehner, "The uncertainty principle determines the nonlocality of quantum mechanics," *Science* **330** (2010) no. 6007, 1072–1074.

[66] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," *Phys. Rev. A* **71** (2005) 022101.

[67] V. Scarani, *Ultracold Gases and Quantum Information*, ch. Quantum Information: Primitive notions and quantum correlation, pp. 105–177. Oxford University Press, 2009.

[68] M. Navascues and H. Wunderlich, "A glance beyond the quantum model," *Proc. Roy. Soc. Lond. A* **466** (2010) 881–890.

[69] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, "Information causality as a physical principle," *Nature* **461** (2009) 1101–1104.

[70] J. Allcock, N. Brunner, M. Pawlowski, and V. Scarani, "Recovering part of the boundary between quantum and nonquantum correlations from information causality," *Phys. Rev. A* **80** (2009) 040103.

[71] W. van Dam, "Implausible consequences of superstrong nonlocality," *preprint* (2005) , arXiv:quant-ph/0501159.

[72] N. Brunner and P. Skrzypczyk, "Nonlocality distillation and postquantum theories with trivial communication complexity," *Phys. Rev. Lett.* **102** (2009) 160403.

[73] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vertesi, "Closed sets of nonlocal correlations," *Phys. Rev. A* **80** (2009) 062107.

[74] G. Brassard, H. Buhrman, N. Linden, A. A. Methot, A. Tapp, and F. Unger, "Limit on nonlocality in any world in which communication complexity is not trivial," *Phys. Rev. Lett.* **96** (2006) 250401.

[75] N. Linden, S. Popescu, A. J. Short, and A. Winter, "Quantum nonlocality and beyond: Limits from nonlocal computation," *Phys. Rev. Lett.* **99** (2007) 180502.

[76] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acin, N. Gisin, and S. Pironio, "Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage," *Phys. Rev. Lett.* **104** (2010) 230404.

[77] A. A. Methot and V. Scarani, "An anomaly of non-locality," *Quant. Inf. Comp.* **7** (2007) 157–170. http://www.rintonpress.com/xqic7/qic-7-12/157-170.pdf.

[78] S. L. Braunstein and C. M. Caves, "Wringing out better bell inequalities," *Ann. Phys.* **202** (1990) no. 1, 22 – 56.

[79] J. Barrett, A. Kent, and S. Pironio, "Maximally nonlocal and monogamous quantum correlations," *Phys. Rev. Lett.* **97** (2006) 170409.

[80] A. C. Elitzur, S. Popescu, and D. Rohrlich, "Quantum nonlocality for each pair in an ensemble," *Physics Letters A* **162** (1992) no. 1, 25 – 28.

[81] J. Barrett, L. Hardy, and A. Kent, "No signaling and quantum key distribution," *Phys. Rev. Lett.* **95** (2005) 010503.

[82] J. Barrett and S. Pironio, "Popescu-rohrlich correlations as a unit of nonlocality," *Phys. Rev. Lett.* **95** (2005) 140401.

[83] N. S. Jones and L. Masanes, "Interconversion of nonlocal correlations," *Phys. Rev. A* **72** (2005) 052312.

[84] M. Forster and S. Wolf, "The universality of non-local boxes," in *9th Int. Conf. on Quantum Communication, Measurement and Computing, Calgari, 2008*, pp. 117–120. AIP, 2009.

[85] R. Bhatia, *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

[86] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, ""event-ready-detectors" bell experiment via entanglement swapping," *Phys. Rev. Lett.* **71** (1993) 4287–4290.

[87] A. J. Short, S. Popescu, and N. Gisin, "Entanglement swapping for generalized nonlocal correlations," *Phys. Rev. A* **73** (2006) 012101.

[88] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of mixed states: necessary and sufficient conditions," *Phys. Lett. A* **223** (1996) no. 1 & 2, 1–8.

[89] H. Boerner, *Representations of Groups*. North-Holland Publishing, 1963.

[90] B. Coecke, "Quantum picturalism," *Contemporary Physics* **51** (2010) no. 1, 59–83.

[91] C. M. Caves, C. A. Fuchs, and P. Rungta, "Entanglement of formation of an arbitrary state of two rebits," *Foundations of Physics Letters* **14** (2001) no. 3, 199–212.

[92] L. Hardy, "Disentangling nonlocality and teleportation," *preprint* (1999) , `arXiv:quant-ph/9906123`.

[93] B. Coecke, B. Edwards, and R. Spekkens, "Phase groups and the origin of non-locality for qubits," *Electronic Notes in Theoretical Computer Science* **270** (2011) no. 2, 15–36.

[94] D. Avis, "lrs program documentation.".

[95] K. Fukuda and A. Prodon, "Double description method revisited," in *Combinatorics and Computer Science*, M. Deza, R. Euler, and I. Manoussakis, eds., vol. 1120 of *Lecture Notes in Computer Science*, pp. 91–111. Springer, 1996.

[96] S. Abramsky and B. Coecke, "A categorical semantics of quantum protocols," in *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, pp. 415–425. July, 2004.