Universidade de São Paulo
Instituto de Física

Aspectos computacionais de contextualidade e não-classicalidade

Giulio Halisson Camillo da Silva

Orientadora: Profa. Dra. Bárbara Lopes Amaral

Tese de doutorado apresentada ao Instituto de Física da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutor em Ciências.

Banca Examinadora:
Profa. Dra. Bárbara Lopes Amaral - Orientadora (Universidade de São Paulo)
Prof. Dr. Marcelo Martinelli (Universidade de São Paulo)
Prof. Dr. Roberto Dobal Baldijão (Uniwersytet Gdański)
Prof. Dr. Cristhiano A. G. Duarte Carneiro Silva (Universidade Federal da Bahia)
Profa. Dra. Jara Juana Bermejo-Vega (Universidad de Granada)

São Paulo
2023

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Silva, Giulio Halisson Camillo da

Aspectos computacionais de contextualidade e não-classicalidade. São Paulo, 2023.

Tese (Doutorado) - Universidade de São Paulo. Instituto de Física. Depto. de Física Matemática

Orientador: Profa. Dra. Bárbara Lopes Amaral.

Área de Concentração: Informação Quântica.

Unitermos: 1. Informação quântica; 2. Mecânica quântica; 3. Física matemática.

USP/IF/SBI-105/2023

University of São Paulo
Physics Institute

Computational aspects of contextuality and non-classicality

Giulio Halisson Camillo da Silva

Supervisor: Prof. Dr. Bárbara Lopes Amaral

Thesis submitted to the Physics Institute of the University of São Paulo in partial fulfillment of the requirements for the degree of Doctor of Science.

Examining Committee:
Prof. Dr. Bárbara Lopes Amaral - Supervisor (Universidade de São Paulo)
Prof. Dr. Marcelo Martinelli (Universidade de São Paulo)
Prof. Dr. Roberto Dobal Baldijão (Uniwersytet Gdański)
Prof. Dr. Cristhiano A. G. Duarte Carneiro Silva (Universidade Federal da Bahia)
Prof. Dr. Jara Juana Bermejo-Vega (Universidad de Granada)

São Paulo
2023

Agradecimentos

Acredito que é de conhecimento comum as dificuldades que a formação acadêmica, tal como é no presente momento, impõe a quem decide se submeter a ela, ainda mais após um governo fracassado e revoltante em tantos níveis no Brasil deste tempo. Penso que é importante acrescentar também o fato de que pelo menos metade deste trabalho foi realizado em período pandêmico, sob isolamento domiciliar e restrição de atividades presenciais. No entanto, longe de querer atenção por conta disso, essas palavras iniciais, que transcrevem minimamente a mensagem de algumas lágrimas enquanto escrevo, são na verdade para contextualizar, e portanto proporcionar, um ponto de vista adequado que faça justiça à importância que dou e reflita a gratidão que tenho a todes que menciono a seguir.

Eu gostaria de começar agradecendo à professora Bárbara Amaral por me aceitar para este programa de doutorado e por ter sido tão solidária ao longo desses quatro anos. Além de me ajudar a superar as dificuldades de me introduzir em uma área diferente daquela do meu mestrado, ela esteve sempre me apresentando oportunidades e outras linhas de pesquisa. Somo a isso tudo sua empatia. O resultado foi um bom trabalho e uma amiga.

Certamente eu devo parte dessa enriquecedora jornada ao meus colegas da USP também. Gostaria de agradecer aos meus amigos e pares de doutorado, Ana Camila e Ivan Romualdo, por toda ajuda e conversa, seja pessoalmente, enquanto dividíamos a mesma sala 3106, seja virtualmente. Incluindo nessa a Brenda Malabarba, com quem também aprendi muito sobre café, e também o Alisson Tezzin, por todo esforço em revisões e opiniões do meu trabalho, e o Ari da Costa, pelas colaborações e discussões. Agradeço também a todos os meus irmãos de orientação, em especial Gustavo, Amanda, Naim, Baldi, Rafa, Leo, Danilo e Luis, por todas as discussões, reuniões esclarecedoras, companhia em viagens, e também pela inabalável disposição para religar meu computador. E, com certeza, agradeço também a todo o time do instituto e do DFMA, em especial aos queridos da secretaria pela competência excepcional.

Também gostaria de agradecer a amigos importantes fora da universidade. Aos Rafael Chaves, Davide Poderini, Jacinto, Leandro e Vítor, pela excelente companhia durante minha estadia em Natal, por tudo que aprendi e por me apresentarem ao surf. Ao Víctor Cervantes, pela enriquecedora colaboração, pela paciência e simpatia. E aos queridos da minha cidade, em especial Rebeca, Rafael, Ana, Júlio, Dani, Rafa, Salman e Karine, pela consideração, preocupação e festas que me mantiveram com forças.

Sem dúvidas a minha família merece um agradecimento especial. Principalmente meu irmão, Gian, por sempre encontrar tempo para me ajudar com meus problemas e por me inspirar a dar o meu melhor. E também meus pais, Carlos e Maristela, pelo amor inalterável, por suportar minhas escolhas e por todo o conforto nesse tempo. E à Luciana, Joênio, Augusto e Maria Rita, pelo carinho, pelas viagens, pela companhia, e por me fazerem sentir parte de uma segunda família.

Por último, mas não menos importante, agradeço a minha namorada, Amanda. Por

todo esforço para ajudar a esfriar minha cabeça e superar os obstáculos do caminho. Pelo amor, preocupação e todo o carinho. Amo você.

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

Resumo

No que se segue, o leitor encontrará uma dissertação de doutorado no uso de métodos e abordagens computacionais para o estudo de contextualidade e outros fenômenos não clássicos. O capítulo 1 introduz o tópico de foco maior, isto é, contextualidade, com a linguagem matemática necessária para seu entendimento, enquanto o capítulo 2 dá a estrutura necessária para qualquer investigação sistemática a ser feita ao longo deste manuscrito, incluindo apresentação de problemas de otimização e usos típicos em problemas relacionados e a construção de quantificadores de contextualidade, além de uma pequena contribuição do autor nesse sentido. Esses dois capítulos resumem o conteúdo que o autor precisou aprender para se inserir na área. Os capítulos seguintes compõem outros dois resultados encontrados em colaboração com colegas durante o programa. O primeiro deles está relacionado com a construção de quantificadores de não-localidade para estados quânticos a partir de quantificadores similares para comportamentos em cenários de medição; nós apresentamos uma nova medida daquele recurso, iluminando debates conhecidos na literatura. Já no segundo deles, investigamos conjuntos de correlações típicos em redes causais de interesse, a fim de entender melhor a estrutura desses conjuntos dentro do universo de correlações possíveis para tais redes; em particular mostramos vantagens em utilizar programação quadrática ao invés de outra estratégia mais comum na literatura e também o impacto de intervenções para a detecção de não classicalidade. Conhecimento muito avançado em matemática não é presumido para os primeiros dois capítulos. Entretanto, um certo conhecimento em física quântica é desejável para apreciar melhor todo o trabalho.

Palavras-chaves: Contextualidade; Não-localidade; Não-classicalidade; Quantificadores; Causalidade

Abstract

In what follows, the reader is going to find a doctoral dissertation on the use of computational methods and approaches to the study of contextuality and other non-classical phenomena. Chapter 1 introduces the topic of main attention, that is, contextuality, with the appropriate required mathematical language to its understanding. Chapter 2 gives the necessary structure to any systematic investigation to appear along this manuscript, including the presentation of optimization problems and typical uses in cases related to the main topic, and also the construction of contextuality quantifiers and a small contribution of the author to this. These first two chapters summarize the content that the author had to learn in order to insert himself into the area of research. Succeeding chapters compose other two results found in collaboration with colleagues along the program. The first of them is related to the construction of non-locality quantifiers for quantum states from similar quantifiers for behaviors in measurement scenarios; we present a new measure of the resource, shedding some light in some debates in the literature. In the second of them, we investigate typical correlation sets in causal networks of interest, with the end of improving our understanding the structure of such sets within the universe of possible correlations in such networks; in particular, we show advantages of quadratic programming over another common strategy in the literature as well as the impact of interventions on the detection of non-classicality. Advanced mathematical knowledge is not presumed for the first two chapters. However, some knowledge in quantum physics is desirable to appreciate the work completely.

Keywords: Contextuality; Non-locality; Non-classicality; Quantifiers; Causality

List of Figures

List of Tables

Contents

Introduction

Quantum physics is a subject intriguing physicists since its discovery at the beginning of the past century. Throughout the years of scientific research on it, revolutionary knowledge has been developed, however several aspects of its impact, be on science or on our lives, are yet to be discovered. The crucial lesson being that it is powerful and worth our effort towards fully understanding it.

In its youngest face, we have quantum computing. This topic, that has already reached regular newspapers, breaking the usually hard barriers of the academic bubble, sees on growing attention day after day, under the promise/hope to provide technological changes able to solve problems and to promote developments like never seen before.

Although, from the eyes of this author, we are still at the embryonic stage of a future like this, it is undeniable that recent advances make it clearer more and more. At the heart of such progress, however, lies not only practical questions as those aiming the final goal of having a workable quantum computer, but also problems that drive us into a well established theory, with comprehensible phenomena and complete control over them.

Of particular importance to the present manuscript, *quantum contextuality* is one of these phenomena, which concerns to the possibility of accessing properties of a quantum system simultaneously and the dependency of such access on the context of realization, meaning which ones are being measured together.

Despite the simple informal statement of such phenomenon, it is remarkable to notice that its study has allowed scientists to uncover various facets of long standing debates in Physics, including for example the *completeness* of quantum mechanics and the challenging *non-local* nature of it. This last one, in particular, has led (so far) to one recent (2022) Nobel Prize, overcoming a *loophole free* experimental violation of a Bell inequality for local theories.

As it might not seem evident at a first sight, it turns out there is a lot one can do using computational approaches and techniques to investigate, learn and contribute to that end. And this is precisely the viewpoint we are going to adopt in this manuscript and discuss about. By the end of this reading we want the reader to feel comfortable with usual formulations of contextuality mainly, but hopefully of related non-classical phenomena too.

However, and to clarify a bit more the paragraph just above, this thesis is not intended to be a first reading on the mentioned topics. It assumes the presence of an already interested reader, previously exposed to some ideas, and was planned to work as a roadmap in the study of contextuality for them. As such, one will find a main line of reasoning around the traditional approach to the phenomenon and several paths for exploration from it, and relations between them.

In that sense, it is our intention to provide some basic understanding on the typical structures arising on such aspects of the research. As well as making the best we can to review and therefore to make of this thesis the shortcut for whichever path of interest one

may find on the subject. Because of this, the details will be more on the surveying of ideas and less on a full discussion/analysis of them.

Here, therefore, the reader is going to encounter every thought that the author found it could be useful to share when learning about the theme. Especially from someone coming from an almost totally unrelated area of specialization for his Master's degree. And as novelty in the connected literature, some very small contributions from this author elaborated along his graduate program.

The first chapter clarifies most of the concepts, math, and frameworks fundamental to the topic. Particular attention being given to the traditional *Kochen-Specker* approach to contextuality and possible extensions and alternatives to it. It is supposed the least technical as possible, while capable of aiming the reader with the required material to understand and criticize succeeding chapters and to introduce oneself into any connected discussion.

Following it, in contrast, one will find the most technical chapter, containing most of the computational tools underlying the contributions to be presented in this dissertation. "Most" because, as it is going to be clear later, the author has managed to use the knowledge acquired during the program to collaborate with a group focused on a different line of research of non-classicality.

His first work appears in the third chapter, a primarily mathematical result, but with interesting consequences to the resource theory of contextuality. This important side of the phenomenon is in fact what motivates the focus on quantifiers of the resource throughout the chapter, which constitute a if not the typical use depending on computers.

Next chapters are then just working examples of the content preceding them. Each one within the respective context, the problems are specified along their reading and any additional concept and/or mathematical requirement is fulfilled when necessary. Our choice on this is to not distract the reader from the narrative constructed in this work. Namely, that contextuality has guided and motivated the author to pursue challenges even outside its immediate surrounding.

In the first of them, one will find a work on the idea of quantifying non-locality (a special instance of contextuality) for quantum states, which is mainly motivated by the idea of extracting a resource from such systems for practical purposes. Now, the second employs a different, though related and comprehensible, framework, that goes by the name of *causal networks*, to explore the idea of multiple sources of common cause among the variables of interest in a physical scenario, particularly focused on tripartite networks. Networks beyond those present in the paradigmatic Bell's theorem can lead to new kinds and applications of non-classical behavior.

For ease if necessary, sections and subsections have been marked with an asterisk in the table of contents in the case of containing contributed work.

Chapter 1

Contextuality

This chapter focuses on introducing the key topic within which the present contribution is situated. In addition to providing the necessary foundational knowledge for a better understanding of the issues and main aspects involved, we aim to keep the information on each problem up to date.

While a historical introduction may have been a fitting choice for this purpose, we have opted for a different yet closely related perspective. In the following sections, one will discover what can be regarded as a recap of previously explored areas. Our aim is to present well-established concepts and notable alternatives from the literature while offering a glimpse of their potential future implications, with [Bud+21] proving instrumental in this regard.

As mentioned in the introduction, this dissertation has not been planned to work as a first reading in the topic, but just as a general roadmap containing specific contributions of this author around the implied main line of presentation of the area that we follow. Having this in mind, we assume the reader already has a motivation for studying contextuality. However, to record a bit about its importance, firstly we notice its fundamental consequence that statistical predictions of quantum theory can not be explained by models in which measurements reveal properties determined prior to the knowledge of what other compatible measurements are being made together [BEL66; KS75]; an intrinsically non-classical feature, fruit of incompatibility between measurements. Secondly, we refer the reader to [Naw+13; Cab13; CSW14; ACC14] for its role in the understanding of more essential aspects of quantum theory. And lastly, we mention the fact that contextuality has been recognized as a potential *resource* for quantum computing, particularly in the scheme of *Measurement-based quantum computing* [Rau13; How+14; Del+15], and has been employed for random number generation [Um+13] and other information processing tasks considering space-like separated systems [Bru+14].

Our preferred approach to introducing the phenomenon (and, to some extent, the mathematical concept) of interest is through an illustrative example, particularly one drawn from the realm of quantum physics – the birthplace and central theme of this thesis. This choice allows us to acquaint readers with fundamental terms that will be indispensable in the chapters to come.

1.1 The Peres-Mermin magic square

Asher Peres and David Mermin provided in some of their works a beautiful and simple example to illustrate the non-intuitive nature of quantum predictions ([Mer90; Mer93;

Per90; Per91; Per92; Per02]). It consists in nine dichotomic[1] measurements arranged in a square prescribing compatibility among them:

A	a	α
B	b	β
C	c	γ

Measurements in the same row or column are said to be compatible (or commuting). In more tangible terms, this implies that their results can be accessed simultaneously. Intuitively, this leads us to believe that a measurement procedure unveils preexisting properties, a perspective we embrace in the upcoming discussion. However, in contrast to this intuition, one could argue that the measurement procedure is, in fact, attributing outcomes when it occurs. In other words, prior to the measurement, it may not be possible to make any definitive claims about whether the system possesses a specific property. This is a longstanding debate, which, while beyond the scope of this manuscript, is well-documented in references such as [EPR35], [PBR12], and [Spe07].

In this scenario, let us consider assigning values to these properties before any measurements occur. Each of them must attain one out of two possibilities. We will use +1 if the property exhibits a given form and -1 otherwise. This approach not only provides definite values for the expectation values of individual measurements but also yields definite values for the expectations of products involving those measurements . This notion holds significance for properties that share a row or a column, such as $\langle Aa\alpha\rangle$, $\langle Bb\beta\rangle$, $\langle Cc\gamma\rangle$, $\langle ABC\rangle$, $\langle abc\rangle$, and $\langle\alpha\beta\gamma\rangle$. It is worth noting that while it is possible to assign a value to expectation values like $\langle Ab\rangle$, this value exists only *counterfactually*, meaning it cannot be experimentally accessible. This is because, *a priori*, A and b are not compatible.

The magic of this square is that each measurement takes place in both a specific row and column. So, when you alter the value assigned to one of these measurements, it affects the sign of the product involving that particular row and column. For instance, imagine that initially, all properties are assumed to reveal +1, and then you reset one of them to -1. In such cases, one can always verify the following relationship holds true

$$\langle Aa\alpha\rangle + \langle Bb\beta\rangle + \langle ABC\rangle + \langle abc\rangle + \langle\alpha\beta\gamma\rangle - \langle Cc\gamma\rangle \leq 4\ . \tag{1.1}$$

With that settled, let us see a quantum realization of that system. As in the original works ([Mer90; Mer93; Per90; Per91; Per92; Per02]), that can be achieved through a quantum set up of two spin 1/2 particles and the following measurements:

$\sigma_z \otimes \mathbb{1}$	$\mathbb{1} \otimes \sigma_z$	$\sigma_z \otimes \sigma_z$
$\mathbb{1} \otimes \sigma_x$	$\sigma_x \otimes \mathbb{1}$	$\sigma_x \otimes \sigma_x$
$\sigma_z \otimes \sigma_x$	$\sigma_x \otimes \sigma_z$	$\sigma_y \otimes \sigma_y$

.

The sigmas stand for the usual Pauli matrices and $\mathbb{1}$ for the identity:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.2}$$

[1] With only two possible outcomes.

and their eigenvalues (+1 or -1) are the property values revealed by the measurement associated, i.e. the spin orientation: up or down in the given direction.

One can confirm that, in this scenario, the expectation value of the product of measurements within the same row or column equals 1, except for the last column, where it amounts to -1, just notice that:

$$\sigma_j\sigma_j = \mathbb{1},\ \sigma_z\sigma_x = i\sigma_y,\ \text{and}\ \sigma_j\sigma_k = -\sigma_k\sigma_j,\ \text{for}\ j \neq k \in \{x, y, z\}. \tag{1.3}$$

As a result, regardless of the expectation value assigned to each measurement, the quantum prediction for the left-hand side of inequality (1.1) is 6, clearly violating it.

Now, we are faced with two possible interpretations of this observation. It could be that the entire notion of assigning pre-existing values to properties of physical systems is flawed, or perhaps the way we are going about it is not quite right. The first situation brings to the forefront a discussion about the nature of the state of a (quantum) physical system, where questions about whether their properties are truly being unveiled through experiments arise. The second introduces an alternative perspective: the value assigned to a measurement may depend on the other measurements being performed alongside it, which we will refer to as a *context*[2]. With that in mind, look again at the magic square, and imagine that, for example, the value of γ is equal to -1 when in $\langle Cc\gamma\rangle$, but 1 when in $\langle\alpha\beta\gamma\rangle$; the dependence on the context leads us to the violation of the inequality.

In light of this second possibility, we can then describe the phenomenon we have outlined as *quantum contextuality*. The reason we have dropped the word "quantum" from the chapter title is that, as we will soon see, contextuality is a phenomenon that also manifests in probabilistic systems outside the realm of (quantum) physics.

Given the fundamental nature of this subject, there is a vast body of literature that we can't cover comprehensively here. However, in the upcoming sections, we will present common ideas, different formulations, and various approaches as a guide. Along the way, we will aim to highlight the advantages and disadvantages of each approach and their practical applicability.

1.2 Generalities

As demonstrated in the previous section, regardless of the approach chosen to describe contextuality, certain fundamental aspects of it inherently lend themselves to the language of probabilistic theories. These aspects encompass the use of concepts like expectation values and the assignment of values to properties, among others. Therefore, a crucial prerequisite for what follows is a belief in the description of physical systems through the lens of random variables. While advocating for this perspective lies outside the scope of this work, it is worth noting that the countless successful experimental results that rely on this treatment speak for themselves.

In this section, our aim is to provide an introduction to the foundational framework of probabilistic theories, drawing upon [Apo69], and its essential components, which will be pivotal in the discussions to come. Furthermore, we will formalize and imbue mathematical rigor into certain aspects highlighted in the previous example. Despite the significance of these concepts, we will maintain the mathematical level necessary for comprehension.

[2]Although this debate is out of the scope of this dissertation, notice that these alternatives are not completely exclusive, after all it seems plausible to have both the reality of state and the dependence of our knowledge about them on the context questioned at the same time

Our starting point is *random variables*. These are among the elementary objects of a probabilistic theory. To avoid unnecessary complications, here we are going to consider discrete real random variables only. However, before going into its formal definition, we need some brief words on the mathematical structures behind it. This will give an intuition on the relations among random variables later.

Let S be a non-empty set with finite number of elements (our *sample space*) and let Σ be the power set of S. The pair (S, Σ) is a measurable space, because:

$$\begin{aligned} &0)\ \emptyset \in \Sigma \\ &1)\ \text{if } A \in \Sigma \text{ then } A^c \in \Sigma, \text{ and} \\ &2)\ \text{if } A_{\mathrm{i}} \in \Sigma \text{ for each i in a countable index set } I, \text{ then } \bigcup_{\mathrm{i}\in I} A_{\mathrm{i}} \in \Sigma. \end{aligned} \tag{1.4}$$

Each element of Σ is called an *event* or a *measurable set*.

Take then p some function from Σ to the interval $[0, 1]$, called a *probability measure*. As such, it satisfies

$$\begin{aligned} &p(S) = 1 \\ &p\left(\bullet\bigcup\nolimits_{j=1,2,\ldots} S_j\right) = \sum_{j=1,2,\ldots} p(S_j), \end{aligned} \tag{1.5}$$

where $\{S_j\}$ is a countable collection of pairwise disjoint sets in Σ.

A discrete real random variable X on S can then be defined as follows.

Definition 1.2.1. A real random variable X is a function[3] $X : S \to \mathbb{R}$ from a sample space (S, Σ) like above into the reals, such that S belongs to the probability space (S, Σ, p).

The probability of each outcome in S determines the function p, which in turn can be used to find $p_X(x)$ (or just $p(x)$), the probability distribution of X (sometimes also called its *mass function*), by pushing-forward the measure p into a probability measure on the image of X, that is,

$$p_X(x) \equiv p(X^{-1}(x)) \equiv p(X = x). \tag{1.6}$$

Clearly then, the probability distribution $p_X(x)$ satisfies the following properties:

$$\begin{aligned} &(I)\ p_X(x) \geq 0\ \forall x, \\ &(II)\ \sum_x p_X(x) = 1, \end{aligned} \tag{1.7}$$

that is to say, it is (I) non-negative and (II) normalized. However, as we will see throughout this work, it can be useful sometimes to relax such conditions, in order to consider (if not I) *quasi*, (if not II) *sub* or *over* -normalized probabilities.

A robust foundation for probabilistic physical theories is provided by the so- called "operational" formulation of *generalized probabilistic theories* (for an example, see [DCP17]). In this framework, every component of the *operational process*, namely *preparation*, *transformation* and *measurement*, is endowed with physical significance, ultimately leading to the establishment of a meaningful association with a probability distribution for each operation that can be applied to the system. Nevertheless, for our purposes here, we will focus on only two fundamental concepts introduced within this framework: *observables*

[3]For those familiar with the mathematical formalism, we record that any such function will me measurable when the σ-algebra is the power set.

and *effects*. In the terms mentioned before, the latter concept falls under the scope of an event, while the former should be viewed as an assembly of mutually exclusive events (read effects), the union of which results in X. In this context, the collection of output values in a measurement aligns with our representation of X, while the effects constitute the elements of Σ. The probabilities associated with these effects can be derived from the probability distribution prescribed by the measurement process. In this sense, somewhat informally, we could consider measurements themselves as random variables.

Before delving into the topic of contextuality, we must introduce two crucial concepts related to random variables: *expectation value* and *compatibility*. Given a real random variable A, its expectation value $\langle A \rangle$ is defined as

$$\langle A \rangle = \sum_{a \in \mathcal{O}} a \; p(a), \tag{1.8}$$

where a denotes an output from the set $\mathcal{O} \subset \mathbb{R}$ of possible outputs and $p(a)$ the probability it occurs in a run, i.e. experimental realization, of A.

Two random variables A and B can be said to be compatible if they are defined on the same probability space. In such case we can define their joint probability distribution, corresponding to the probability distribution on all possible pair (a, b) of output values: $p(a, b) \equiv p(A = a \; and \; B = b)$. This allows us also to define the expectation value of the product of those random variables AB as

$$\langle AB \rangle = \sum_{a \in \mathcal{O}_A} \sum_{b \in \mathcal{O}_B} a \; b \; p(a, b), \tag{1.9}$$

where $\mathcal{O}_A$ and $\mathcal{O}_B$ are the respective sets of possible outputs for A and B. It also encodes the marginal distributions, i.e., the distribution of each individual random variable, that can be obtained by summing up the joint probabilities for a fixed output of the desired marginal over all outputs of the other, e.g:

$$p(a) = \sum_{b \in \mathcal{O}_B} p(a, b).$$

Finally, the idea can be extended to more compatible variables in the same fashion.

From a physics perspective, compatibility is typically expressed in terms of what is known as *joint measurability* of measurements (as discussed in [BLW14], [BLM96], and [Bus+16]). In this context, two measurements, denoted as A and B with corresponding effects A_j and B_k, are considered to be *jointly measurable* if there exists a third measurement, represented as C , with effects C_{jk} such that

$$A_j = \sum_k C_{jk} \text{ and } B_k = \sum_j C_{jk}. \tag{1.10}$$

Connecting to previously mentioned operational terms just to fix ideas, for a quantum physicist, that approximately perform projective measurements (PVM's for short) in the lab, A_j and B_k will be projectors, while A and B the observables themselves. Equivalently, one may also see compatibility being stated in terms of the commutator between A and B, namely

$$[A, B] \equiv AB - BA, \tag{1.11}$$

when the product is defined, for example if these elements can be just matrices - which is the case of quantum theory in textbooks -, resulting in zero for compatibility.

To relate all these notions, in a very non-rigorous manner, notice that (a) If two observables A and B can be written as functions of a common observable X, then their commutator vanishes, because $[X, X] = 0$; (b) If their commutator vanishes, then there exists a common basis that diagonalizes them, which, with the appropriate treatment, relates to the common probability space; (c) If they are defined in the same probability space, therefore existing a joint probability distribution, by treating the pair of outputs as referring to a single random variable, A and B as functions of it could be found by inspecting their marginal probabilities.

With this clarification in place and considering the example discussed in the previous section, it is evident that our focus lies in the examination of systems of random variables that exhibit non-trivial relationships of compatibility among them as we delve into the properties of expectation values associated with these variables.

In this sense, the central definition to what follows is that of a *context*:

Definition 1.2.2. A context is a maximal set of compatible random variables.

The inclusion of the *maximal* condition is typically warranted to prevent redundancies. For example, if A, B, and C are compatible, there exists a joint probability distribution for them, from which the marginals for AB, BC, and AC can be derived. These marginals would represent the joint distribution for the possible contexts formed by those pairs, eliminating the need to consider them separately.

In the subsequent sections, this will be the definition we adopt. However, it is important to mention that alternative definitions exist. In fact, as we will explore in the next section, extensions to the concept of contextuality aim to accommodate features such as imperfect or non-ideal measurements within the description of a physical system. This can be addressed through an alternative definition in which a context is constructed by a single observable, with the primary focus placed on the effects and their relationships as components of the same observable (as discussed in [Spe05] and [Spe14]).

To comprehensively address the manifested phenomenon we observed in the magic square with the quantum realization, it is imperative to introduce the concept of *hidden variable* models, originally known as *local hidden variable models*, due to spatial separation between parties in the analyzed situation at the time (as discussed in [Bel64]), although such separation was not inherently there in the example we have discussed before. These models offer an appropriate framework for the idea of revealing pre-existing property values in physical systems.

The core principle underlying hidden variable models is the recognition that the considered variables within a system may not fully disclose all the information needed. Therefore, properties described by hidden (read inaccessible) variables are introduced, typically through a parameter denoted as λ. These hidden variables introduce a probabilistic dimension into the description, configuring our system to be described by a hypothetical state where all properties have pre-determined values. Therefore, presuming that those values can be accessed simultaneously, i.e. considering the hypothesis that every random variable in the system is compatible.

In more objective terms, notice that, for two compatible random variables in the system, say A and B, with outputs $a \in \mathcal{O}_A$ and $b \in \mathcal{O}_B$, the relationship between their joint probability distribution $p(a, b)$ and the hidden variables is described by

$$p(a, b) = \sum_{\lambda} p(a, b|\lambda)\ p(\lambda), \tag{1.12}$$

where λ represents a state with determinate properties, occurring with a certain probability $p(\lambda)$ among all possible states. The term $p(a, b|\lambda)$ corresponds to the probability distribution of the outputs a and b within a specific hidden state identified by λ. This is the usual notation for conditional probability, in which we have

$$p(x, y) = p(x|y)\ p(y), \tag{1.13}$$

given that $p(y) \neq 0$. The conditioning value y that occurs with probability $p(y)$ restricts and redefines the probability space of interest for obtaining x given that y has been observed, assigning the probability $p(x|y)$ to it.

Since these random variables were assumed to be compatible, we can write

$$p(a, b|\lambda) = p(a|\lambda)\ p(b|\lambda). \tag{1.14}$$

Now, the decision of assigning values to random variables within a system without taking into account the other random variables and the specific contexts in which they appears corresponds to imposing the above relation between any two A, B. Or equivalently, to apply the split to all the variables in the system at once. This amounts to a reality in which only the states λ's are important to describe what happens to each random variable independently, whatever is the involved context. Or yet, that for determining the probabilities of individual random variables, factors beyond the hidden variables become irrelevant. This leads us to the following definition.

Definition 1.2.3. For any set of probabilities and corresponding random variables, or just system, if there can be found a set of hidden variables grouped into the random variable Λ with probabilities $p(\lambda)$ such that the probability distribution of every pair of random variables A and B in that set satisfies

$$p(a, b) = \sum_{\lambda} p(a|\lambda)\ p(b|\lambda)\ p(\lambda), \tag{1.15}$$

then the system is said to be *non-contextual*, and such description a *non-contextual hidden variable* (NCHV) model for the system.

This motivates the qualitative definition of our phenomenon.

Definition 1.2.4. Any probabilistic system incompatible with definition 1.2.3 is said to be *contextual*.

It is also common to make those assertions in terms of *behaviours*, which just stands for the set of joint probabilities prescribed by a system within each context. A given behaviour is thus said to be contextual or non-contextual.

It is worth emphasizing that, despite the definitions provided in the preceding paragraphs, contextuality (or non-contextuality) was assessed through a distinct approach in our example. Rather than relying on the probabilities resulting from a proper quantum realization, we employed the algebra of Pauli matrices to calculate the expectation values of their products, which appeared in inequality (1.1). The conclusion drawn from this analysis was that the quantum value for the left-hand side of the inequality exceeded the value predicted by non-contextual hidden variable models. To be more precise in future collocations, consider the definition below:

Definition 1.2.5. A non-contextuality inequality is a linear function of joint probabilities that adheres to the resulting bound when applied to non-contextual systems. If the application of this function yields a value beyond the bound, we conclude that the inequality has been violated by that system.

In more contemporary terms, the quantum system we discussed earlier has, in fact, violated the non-contextuality inequality. This is another way of stating that the system cannot be described by a non-contextual hidden variable model. The emphasis here is on the modern perspective. Testing the contextuality of probabilistic systems through inequalities is a relatively recent approach, with notable examples being the CHSH scenario [Cla+69] and the KCBS scenario [Kly+08].

Originally, the first proof of this phenomenon was presented through the lens of logical assignments to propositions and the identification of the so-called *KS*-sets of vectors that satisfy certain functional conditions imposed by the system's structure [KS75]. In essence, a proposition was associated with a rank-1 projector in a d-dimensional Hilbert space, and a set of mutually exclusive propositions constituted a context. Assigning "true" or "false" values to these propositions was done in such a way that only one "true" value could be assigned within each context.

Kochen and Specker provided a physical interpretation for a system with $d = 3$, involving spins, and established a relationship between a context of three orthogonal projectors (essentially, a basis for measurement) and the three directions that defined them. Naturally, the same vector could appear in different sets. The proof revealed that there existed a finite set known as the KS-set, composed of vectors in $\mathbb{R}^3$ for which "true" or "false" values could not be consistently assigned to all of them while adhering to the logical constraints mentioned earlier within all sets of three orthogonal vectors. In the original work, this KS-set consisted of a total of 117 vectors [KS75].

For this reason, the way we have been approaching contextuality and also some extensions to it are generally referred to as *Kochen-Specker contextuality*; the principle behind them is the same (that is, hidden variable models, value assignments to properties and compatibility between them), what changes is the interpretation of the random variables in relation to the definition of contexts, as we will see. It should be pointed also that the perspective of the aforementioned proof should not be understood as outdated though, after all recent works still investigate details on it, for example the possibility of reducing the number of vectors required in such a proof (see[CEG96; XCG20; Lis+14]).

Now, the differences highlighted in the previous paragraph become more pronounced when confronted with actual experiments. The reason is that, within an experiment, there is no predefined recipe for realizing a specific context, as provided by the theoretical framework presented earlier. One key challenge is how to interpret the simultaneous implementation of two measurements.

In systems like those in physics, where a real physical state exists to convey information to be extracted, conducting measurements often requires interaction with the state itself. This suggests that addressing the issue mentioned above must encompass the concept of *non-disturbance* to the system.

A widely employed approach to address this issue begins with the fundamental principle that measurements are conducted in a *sequential* manner. This implies that the outcome of one measurement serves as the input for the subsequent one. Compatibility between measurements is then defined in terms of their mutual non-detectability when performed in sequence. In other words, if one measurement cannot be detected between two realizations of the other on the same system, which means that they produce the

same response in spite of the intermediate realization of the other, they are considered to be compatible. This feature is often referred to as *outcome repeatability*, where compatible measurements do not interfere with each other's outcomes in any sequential realization. For more detailed insights into this concept, consult references such as [HW10] and [UVB19].

Besides, it is important to recognize that, in practice, ideal measurements are essentially non-existent, after all we are always subjected to the imprecision on the position or settings of our detectors and also the unreachable requirement of infinite repetitions for perfect characterization of a given measurement.

Consequently, it becomes necessary to quantify the level of disturbance introduced during an experiment and to understand the adjustments required to ensure a reliable verification process. This aspect is examined within the contemporary perspective we are discussing here in references like [Kir+09] and [Jer+16], which explore how non-contextuality inequalities are adapted to accommodate experimentally accessible probabilities, i.e. those corresponding to compatible observables, along with proper error correction procedures. References such as [Sza15] and [SKG13] delve into the potential impact of time evolution on the hidden states between subsequent measurements instead.

We now grasp the pivotal role that non-contextuality inequalities play in the contemporary study of this subject. As we contemplate the violations of these inequalities, an intriguing idea naturally emerges: the extent to which the bounds are exceeded might carry profound meaning. This opens up an entirely new realm of possibilities, paving the way for substantial advances in our understanding of the subject. In particular towards ways of characterizing and quantifying contextuality properly, which will be the driving topic of the next chapter.

For further motivation, within the realm of non-locality, a form of non-contextuality in which compatibility is enforced by spatial separation among the components conducting measurements in different parts of a multipartite physical system, a remarkable discovery revealed non-locality as a foundational element for secure communication, as elucidated in works such as [Eke91] and [BHK05].

1.3 Extended notions of KS-Contextuality

In the earlier section, we observed that addressing contextuality in practical scenarios requires careful consideration of crucial subtleties, namely possible disturbance on our systems and limitations of experiments. More specifically, there we discussed a method to tackle the challenge of performing simultaneous measurements, or more broadly, realizing random variables concurrently. Simultaneity was replaced by a sequential approach without inducing disturbance.

Nevertheless, this is not the sole available approach. In fact, it is precisely a distinct perspective that paves the way for alternative pathways to a concept of contextuality more aligned with experimental considerations — especially from a physicist's standpoint. However, it should be pointed out, following the discussion addressed on [Tez+22; DK23a; DK23b], that so far we have not accomplished such a construction that can completely handle the principle of *post-processing* (see [Tez+22] for example), which in a few words can be understood as the principle stating that any manipulation of measured data can not bring more contextuality into your system. Beyond physics, the rationale supporting such formulations are more compelling though.

The underlying concept behind these notions hinges on the idea of a context within the realm of real-life experiences. Let us consider a simple experiment as an example: imagine a beam of polarized light, and its polarization and color are intended to be determined by two distinct apparatus, denoted as X and Y respectively, in a sequential manner. One could argue that measuring X before Y is fundamentally different from the reverse order. This is due to the fact that it would necessitate an entirely new configuration of the experimental setup, despite the apparent compatibility of properties. Following this line of reasoning, one might even question whether each apparatus remains the same in the two situations.

In this sense, different setups give rise to distinct contexts in which compatible measurements are conducted. Measurements designed to unveil the same property in various contexts are then regarded not as inherently equal but rather as interconnected in some manner, taking into account the restrictions imposed by their individual probability distributions.

Up to this point, our discussion has been quite pragmatic. However, in order to delve into a theoretical understanding of the previously mentioned aspects, we need to introduce the necessary mathematical framework. To accomplish this, we will formally present some concepts elucidated in the preceding section.

Irrespective of the nature of a system, the random variables that underlie it can be encapsulated within a structure termed a *compatibility scenario*, or simply a scenario.

Definition 1.3.1. A scenario is a triple $(\mathcal{R}, \mathcal{C}, \mathcal{O})$: a set $\mathcal{R}$ comprising random variables, a family $\mathcal{C}$ consisting of subsets of $\mathcal{R}$ that delineate our contexts, and a set $\mathcal{O}$ encompassing possible outputs for our random variables[4].

The joint probability distribution of random variables $(R_1, ..., R_{|C|})$ belonging to a given context $C \in \mathcal{C}$ with respective outputs $(a_1, ..., a_{|C|})$ is represented by

$$p(a_1, ..., a_{|C|}|R_1, ..., R_{|C|}) = p(R_1 = a_1 \text{ and } R_2 = a_2 \text{ ...and } R_{|C|} = a_{|C|}), \tag{1.16}$$

in harmony with our previous notation by having in mind the following: the choice of random variables to be realized is itself subject to some process that is usually random, therefore the expression should be seen as conditioning the outputs on the input random variables forming that context. The set of all such joint probabilities is called a behaviour B for the scenario.

Recalling our previous definition 1.2.3 of non-contextuality, it can be rephrased now in terms of a *global* probability distribution

$$p(a_1, ..., a_{|\mathcal{R}|}|R_1, ..., R_{|\mathcal{R}|}) \tag{1.17}$$

for all random variables in $\mathcal{R}$[5].

Definition 1.3.2. A behaviour B of a scenario $(\mathcal{R}, \mathcal{C}, \mathcal{O})$ is non-contextual if and only if there exists a global probability distribution (1.17) producing the correct marginal distributions (1.16) for every context $C \in \mathcal{C}$.

That is to say there is a description in which all the random variables could be compatible and simultaneously realizable while generating the observed behaviour.

[4] As before, all these sets are presumed to be finite.

[5] By taking the outputs of Λ to be every possible combination (i.e. $|\mathcal{R}|$-tuple) of outputs for the $|\mathcal{R}|$ random variables in the scenario.

Now, aside from some specialized terminology, the rationale outlined above is applicable across all approaches to Kochen-Specker contextuality[6]. What the expanded notions introduce as innovation is the departure from a feature known in the traditional formulation as the *non-disturbance* condition. A behavior satisfies this condition if any random variable appearing in two different contexts maintains the same marginal distribution, independent of the joint probability distribution from which it is marginalized. For instance, let A, B, and C be three random variables with outputs a, b, and c, respectively, and suppose $\{A, B\}$ and $\{A, C\}$ are contexts. In this case, the condition implies:

$$\sum_b p(a, b|A, B) = \sum_c p(a, c|A, C) \equiv p(a|A) \ . \tag{1.18}$$

Discarding this condition allows a random variable to exhibit different marginals depending on the context. If two random variables do not have the same marginal probability distribution, it follows that they should not be treated as equal in the first place. This prompts the introduction of a new label for each random variable, indicating the context to which it belongs. In other words, each original random variable is split into m new ones, with one for each of the m contexts in which it appears.

However, instead of adhering to the "split" terminology and reasoning, we opt to treat the modified scenario as an independent scenario. A random variable R_q^c is then associated with the property (or general content) q it informs us about in the context c where it is observed. While variables sharing a q are not considered identical, we anticipate there to be some relation between them. Therefore, we designate them as a special set of random variables in our scenario, termed a *connection*. Different contexts from which the random variables originate render them mutually exclusive; technically, they are referred to as *stochastic unrelated* (refer to [Dzh16]), indicating that they lack a joint distribution.

These notations were introduced in the development of approaches known as *Contextuality by Default* (refer to[DK16a] and [DK16b] for an updated version) or *Extended Contextuality* ([AD19a]); contexts are also referred to as *bunches* in [DK16a].

In these formulations, assertions about contextuality (non-contextuality) gravitate towards the non-existence (existence) of the so-called *couplings*, which exhibit special features. To comprehend this, let us first define these objects.

Definition 1.3.3. A coupling for a set $\{X_j : j \in I\}$ of random variables is a set $\{T_j : j \in I\}$ of jointly distributed random variables where each X_j has the same distribution as T_j.

Definition 1.3.4. The set $\{T_q^c : \forall q, \forall c, \text{ s.t. } q \text{ is measured in } c\}$ is a coupling for a behaviour B if for every context c the set $\{T_q^c : \forall q, \text{ s.t. } q \text{ is measured in } c\}$ is a coupling for it.

The distinctiveness arises from having different random variables responding to the same question q. The idea is that, if they share a content q, they should be as equal as possible to approach a non-contextual description. In the literature, two paths are typically followed: the use of either *maximal* or *multi-maximal* couplings.

Definition 1.3.5. A maximal coupling for a given behaviour B is a coupling $\{T_q^c\}$ for which the probability that all random variables answering the same question q are equal is the maximum allowed by the probabilities for each R_q^c in the behaviour.

[6]The reader might explore an alternative mathematical construction using *Sheaf-theory* ([AB11]), for instance.

Maximal coupling always exists [ADO18], but are not necessarily unique.

Similarly, a multi-maximal coupling satisfies the maximum condition for any pair of random variables sharing a content q. In other words, the coupling is maximal for any pair of such random variables.

Definition 1.3.6. A multi-maximal coupling for a given behaviour B is a coupling $\{T_q^c\}$ for which the probability that any two random variables R_q^c and $R_q^{c'}$ answering the same question q are equal is the maximum allowed by the individual probabilities.

Then we can define non-contextuality in the extended formulation.

Definition 1.3.7. A behaviour, in an extended formulation, is said to be *multi-maximally* or *maximally non-contextual* if there is a multi-maximal or a maximal coupling, respectively, yielding all the correct marginals in the behaviour.

Otherwise, it is considered contextual.

The first path was characterized in [ADO18] and further exploration can be found on [AD19a; AD19b; AC18] for example. The second path, has seen significant progress in recent works. It is known (see [KD22], for example) that for a non-contextual system of dichotomic (or *binary*) random variables, there exists a unique multi-maximal coupling for it. This is a robust result, particularly when combined with the dichotomization procedure for a system of random variables (also [KD22]). Broadly speaking, dichotomization involves exchanging each random variable in the system for an equivalent set of dichotomic random variables.

As a generalization of the original formulation of KS-contextuality, the extended version must be reducible to that one in some way. The way of achieving this is through a specific instance of coupling called the *identity* coupling, where the random variables are identical. A behavior is non-contextual in that sense if there exists a multi-maximal (or maximal) coupling for it such that it is an identity coupling for all variables sharing a content, for any content. In the other direction, by considering systems in which the probability of random variables sharing a content being equal is 1, we are essentially in the traditional construction. In fact, such systems are termed *consistently connected* in the terms of these extended formulations. In other words, the extended approaches incorporate instances that are *inconsistently connected* into the analysis.

It turns out that this incorporation also expands the fields to which contextuality analysis can be significant. Indeed, it has been shown that probabilistic systems of human choices, for example, exhibit contextuality ([CD18]). Moreover, works like [Bru+15] have inspired the search for contextuality in the subject of natural language as well as [Wan+21; Wan21]. These are the reasons why the title of this chapter does not carry the adjective "quantum" in it.

To illustrate this, and also to emphasize the multidisciplinary aspect of such approaches, we expose now an example given in [Dzh22], which has personal value for this manuscript, fruit of studies in a pandemic epoch. Suppose we ask to a random person two questions:

q_1- Would you like to take an overseas vacation this summer?

q_2- Are you wary of contracting COVID-19?

The person can answer YES or NO. And suppose also that we ask in two different orders:

c_1- q_1 then q_2;

c_2- q_2 then q_1.

We cannot determine the answer in advance because the person was randomly chosen, so the answers are assumed to be random variables. The following scheme synthesizes the scenario:

R_1^1	R_2^1	c_1
R_1^2	R_2^2	c_2
q_1	q_2	$\mathbf{C}_2$

Table 1.1: Cyclic scenario of rank 2.

.

Answers to the same question in different contexts are associated to different random variables in accordance with our previous construction. $\mathbf{C}_2$ is just a special label to our scenario: the $\mathbf{C}$ stands for the family of *cyclic* scenarios that we will come across again at some point and, therefore, it will be properly explained later; the number 2 denotes its rank, the amount of contexts and/or connections in the scenario.

The temptation is to assume that the same question has equal distribution independent on the context it is. Considering that $+1$ stands for YES and -1 for NO, we would have then in this case, for $a > 0$ and $b > 0$ and any context:

$$p(R_1 = +1) = a,\ p(R_1 = -1) = 1 - a \text{ and } p(R_2 = +1) = b,\ p(R_2 = -1) = 1 - b. \quad (1.19)$$

Now, denoting the following joint probabilities:

$$p(R_1^1 = +1 \text{ and } R_2^1 = +1) = r_1 \text{ and } p(R_1^2 = +1 \text{ and } R_2^2 = +1) = r_2, \quad (1.20)$$

the assumption of the existence of a maximal (or equivalently in this case, multi-maximal) coupling would amount to say that the following tables are equal. Noticing that there

c_1	$R_2^1 = +1$	$R_2^1 = -1$	
$R_1^1 = +1$	r_1	$a - r_1$	a
$R_1^1 = -1$	$b - r_1$	$1 - a - b + r_1$	$1 - a$
	b	$1 - b$	

,

c_2	$R_2^2 = +1$	$R_2^2 = -1$	
$R_1^2 = +1$	r_2	$a - r_2$	a
$R_1^2 = -1$	$b - r_2$	$1 - a - b + r_2$	$1 - a$
	b	$1 - b$	

.

is only one degree of freedom in each table, their equality would boil down to requiring that $r_1 = r_2$. Therefore, if $r_1 \neq r_2$ we clearly have a context dependence on the joint distributions. This is the same as stating that the order in which questions are answered is important. It turns out this was already shown to be true in [Moo02]. Therefore the system is contextual. In fact this is the simplest scenario able to exhibit contextuality.

As a final comment on this section, it is worth noting that more recently, Contextuality by Default (or just CbD) has been linked to causal approaches to contextuality (see [Jon19]). At this point, it is clear that in this dissertation, we are exclusively dealing with random variables with non-trivial compatibility relations among them, but different

perspectives do exist, especially physics-inspired ones. The so-called *M-Contextuality*, short for *Model-based Contextuality*, describes systems through probabilistic causal models that require stronger (hidden) direct influences. The equivalence between them, as shown in the aforementioned reference, provides a new interpretation of CbD in terms of such models, i.e., as the impossibility of modeling a system without hidden direct influences among the random variables; in contrast to the interpretation we have been using, in which the phenomenon is related to different distributions in different contexts.

1.4 Further comments

To conclude this chapter, we present approaches to contextuality that differ somewhat from the reasoning we have been developing in the previous sections. After doing so, we believe that one should have a first roadmap of current methods to explore the impossibility of describing data from a (quantum) experiment using a single classical probability space. This section offers alternative perspectives on the same phenomenon with fewer details, allowing the reader to contemplate and critique the main line of reasoning in this work through different well-established frameworks. Questions regarding numerical results will be addressed in the next chapter with the necessary background.

There are approaches to be mentioned yet. One of them is distinct in a more fundamental sense: compatibility among random variables is abandoned, and the entire construction is based on equivalence relations among procedures. For this reason, we reserve it for the end. The other one shifts the focus, while maintaining the compatibility intuition we developed behind the scenes.

To comprehend this, let us first translate what we know into the language of graph theory. A graph is a pair $G = (V, E)$, where V is a set of vertices and E is a set of edges linking two vertices, i.e., a set of unordered pairs $\{j, k\}$, with j and k in V; two vertices not connected are said to be independent. A hypergraph H is defined similarly, with the difference being that edges are allowed to connect more than two vertices, forming subsets of the power set of V.

As detailed in [AC18], considering vertices as measurements (i.e., random variables) and representing compatibility between them using edges, scenarios naturally gain a description through the use of graphs. Contexts become hyperedges. We refer to them as *compatibility hypergraphs*[7]. For instance, in the traditional notion of KS-contextuality, for simplicity, the (hyper) graph representation of the famous CHSH scenario is shown in figure 1.1 (see also the next section for more).

Also, notice that the measurements were separated into two different sets: A or B with subscripts, or equivalently red and blue. This is to highlight the physical motivation behind the scenario, where two distant parts possess two measurements each, and the compatibility between them is implied by the long distance[8], which prohibits disturbance in one side by the performance of a measurement on the other.

Now, this framework is already proving useful as it stands. Indeed, investigations into

[7] According to the principle in [Cab12], the story is simpler when involving sharp measurements: graphs can be used, and we can stick with *compatibility graphs* instead; roughly speaking, the principle states that for such kinds of measurements, i.e. ideal or error free, pairwise compatibility is equivalent to global compatibility.

[8] In the relativistic sense that if some action is performed in one part of the shared system, the information takes at least the time required for the light to travel all the distance to reach the other part. Acting on the other side during this interval prevents the perturbation.

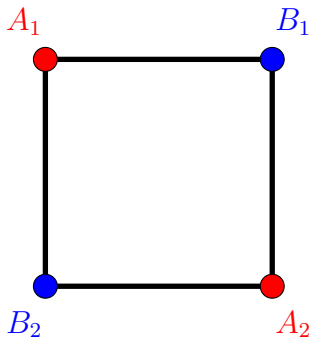


Figure 1.1: Compatibility (hyper-)graph of the CHSH scenario. Four measurements compose it: A_1, A_2, B_1 and B_2. Contexts are represented by edges.

the connections between graph-theoretic properties and the existence of NCHV have been explored in several works (such as [BM10], [KRK12], and [Ram+12]). However, it is not drastically different from what we were doing before; essentially, we have redefined the concepts using a different set of objects, equipped with new tools that might come in handy. Nonetheless, what truly captures our interest is a divergence in the approach.

This innovative idea, initially introduced in [CSW10] and [CSW14], shifts our focus to the compatibility relation between events rather than measurements. Given n events produced by a set of measurements and an initial state, the exclusivity relation among these events can be graphically represented by an n-vertex graph, where edges connect mutually exclusive events. Two outputs in a experimental realization are considered exclusive if there exists a measurement that produces them following respective two orthogonal effects.

Investigations then shifted their focus to *exclusivity graphs* rather than compatibility ones, opening the door to studies that do not assume specific scenarios from the start, driving conclusions about them as a consequence when the compatibility structure contains such events as possible outputs in their description. This change is particularly crucial in the pursuit of principles that can elucidate the correlations observed in nature ([Fri+13] and [ACC14] on *local orthogonality*, and [Paw+09] on *information causality*, to name a few). This endeavor has also led to the development of refined concepts of classicality, as well as an enhanced understanding of entanglement (as seen in [DEP20a] and [DEP20b]).

The essence of this shift lies in the fact that graph properties are extensively understood in the existing literature. Once established, this formulation automatically brought forth a wealth of information. Notably, graph-theoretic quantities like the *independence number* and the *Lovász number* promptly became relevant to the discussion [CSW14]. The former provides non-contextual bounds for defined inequalities, while the latter sets a limit on the violation of quantum correlations subjected to the linear functions that define such inequalities (recall definition 1.2.5). Figure 1.2 illustrates the case for the KCBS scenario. In this scenario, we have five dichotomic measurements $\{A_0, ..., A_4\}$ with outputs either $+1$ or -1. Contexts are created by pairs of measurements with labels differing by 1[9]. Non-contextual systems adhere to probabilities for these events that satisfy

[9]Modulo 5, i.e., 4+1=0, so $\{A_0, A_4\}$ is included as a context.

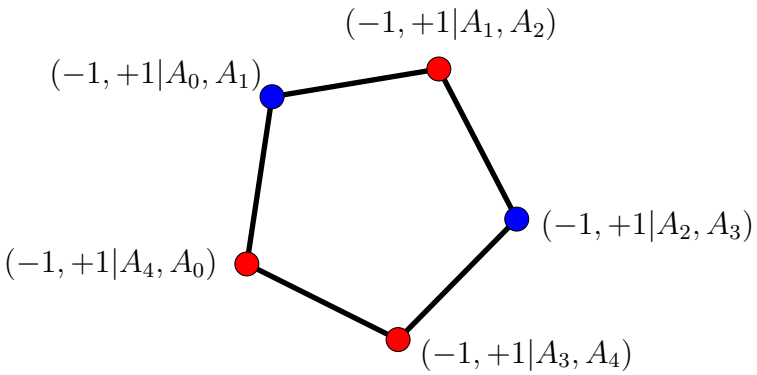


Figure 1.2: Exclusivity graph associated to the inequality S_{KCBS} defined by five events in the KCBS scenario with associated probabilities $p(a_j, a_k|A_jA_k)$, i.e. the probability of having the pair of outputs (a_j, a_k) given that (A_j, A_k) have been decided as inputs. The maximum number of independent vertices is 2, exemplified through the blue ones.

the inequality:

$$S_{KCBS} = \sum_{j=0}^{4} p(-1,+1|A_j, A_{j+1}) \leq 2, \tag{1.21}$$

for which we already have reports in the literature for experimental realization of quantum violations [Lap+11; Ahr+13].

Towards the conclusion of this chapter, we introduce the last remaining approach to contextuality. While different lines of research deviate from the one we intend to follow in the next chapters, it is precisely this diversity that makes it worth mentioning, offering readers a taste of the various flavors within the field. It is known in the literature by *Generalized Contextuality* or *Contextuality by Spekkens*, and it reduces to the Kochen-Specker's traditional notion under some conditions [KS15].

The fundamental components here are *preparations* and *measurements* (and also *transformations*, but we won't delve into them for simplicity). These are specialized instances of the effects we mentioned earlier. In simpler terms, the former requires no income system; it *prepares* outgoing states, while the latter lacks an outcome system; it *measures* incoming states. Additionally, we need rules for calculating probabilities $p(k|P, M)$, indicating the likelihood of obtaining output k given a preparation P and a measurement M setup.

In this operational perspective, non-contextuality is constructed based on the indistinguishability of these procedures, mathematically expressed as two procedures being equivalent (and thus represented by the same elements of the theory, forming an *equivalence class*) if they yield the same statistics.

Explorations within this approach delve far beyond the reasoning we have developed so far, so we conclude our discussion here. However, for those keenly interested, we recommend [Spe05] for an in-depth presentation. it is worth noting that experimental tests have been conducted ([Spe+09] and [Maz+16], for instance), underscoring the central role of non-contextuality inequalities from this viewpoint too.

1.5 A nobel winner example

This section is in certain way optional and has been included in the work to highlight the importance of the field by contemplating the awarded works of Alain Aspect, John F. Clauser and Anton Zeilinger, "for experiments with entangled photons, establishing the violation of Bell inequalities and pioneering quantum information science". It is based on the scientific report on the Nobel Prize in Physics of 2022 provided by the Nobel Committee for Physics [Phy22].

The path leading to the results of these scientists refers to the seminal work by Albert Einstein, Boris Podolski and Nathan Rosen (EPR) in 1935 [EPR35] and the puzzle presented by a thought experiment on it. Consider two distant parts, traditionally Alice and Bob, sharing a pair of spin one-half entangled particles. Each party can measure their piece with spin operators that do not commute. If both parties happen to perform a measurement of the same direction of the spin, their results will be completely correlated or anti-correlated depending on the entangled state. However, even when they do not use the same direction, they seem to obtain *sharp* values for the component of the spin measured. And the problem relies in the fact that quantum mechanics does not allow us to assign sharp values simultaneously for quantum numbers referring to non-commuting observables. Either then, the description provided by the theory is incomplete or non-commuting operators make the simultaneous reality of the physical properties associated to them impossible.

While the negation of the second possibility has led to the appearance of different interpretations of quantum mechanics [Eve57; Boh52] with the passing of the years, its acceptance originated John Bell's work "On the einstein-podolsky-rosen paradox" [Bel64]. By making use of a special form of the previous thought experiment, he showed the impossibility of reproducing the results prescribed by quantum mechanics by means of local hidden variable models. To do that, Bell derived an inequality as the ones we described before, and then showed that for some experimental conditions quantum mechanics violates it. However, some unjustifiable assumptions on the detectors for real experiments made it not suitable for tests.

John Clauser, Michael Horne, Abner Shimony and Richard Holt (CHSH) then removed this obstacle in 1969 [Cla+69] by proposing a modified scenario, illustrated in figure 1.1, in which each party could now perform two different experiments. It can now be realized many times, and in each instance, Alice and Bob *choose* the setup to be used in their side. Figure 1.3 below illustrates the experimental setup required. Local hidden variable models should then obey the following inequality:

$$S_{CHSH} = |\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle| \leq 2, \tag{1.22}$$

which can be violated by quantum mechanics using particular orientations for the detectors, producing the optimal value of $S = 2\sqrt{2}$. Recalling that local refers to the idea of locality implied by special relativity and the existence of a common past.

Then Clauser, together with Stuart Freedman, conducted the first test in 1972 by leveraging previously available setup of the research of Carl Kocher on the time correlation of photons originated from the same source [FC72]. The apparatus employed a hydrogen arc lamp to excite calcium atoms out of their ground states to produce entangled photons from subsequent decays. However the polarizers available initially hadbeen identified by them as inefficient for the purposes they had, specially because of angle controlling, so they rebuilt the experiment and after two years of preparation finally recorded the data.

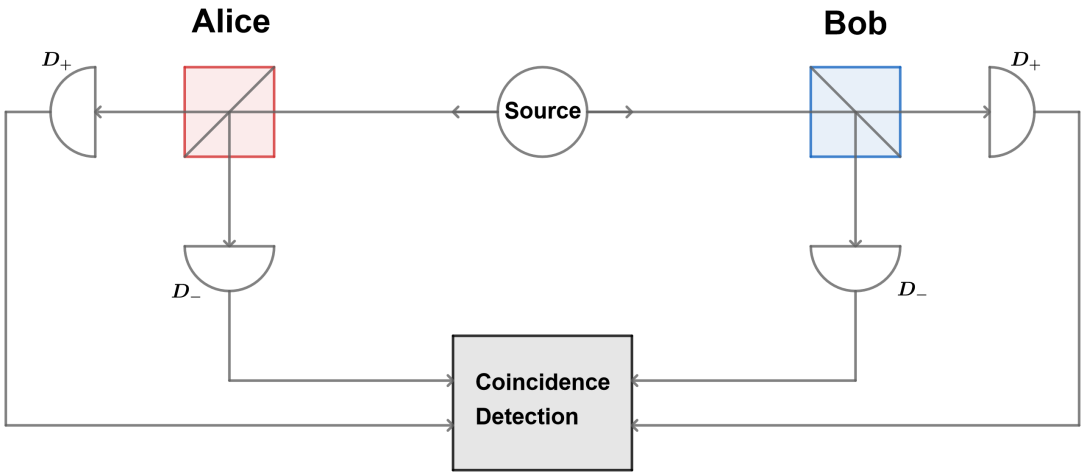


Figure 1.3: Generic experimental apparatus for realizing a CHSH test. A source emits entangled photons in opposite directions. Each of them encounters an adjustable polarizer, and the signals of the respective channels are detected by single photon detectors (D_+ and D_- in the picture). Their coincidence is counted by a separated unit. The expectation values of interest can then be computed by recording the numbers of coincidence events corresponding to simultaneous detection by Alice and Bob.

Using a slightly modified version of the inequality above, they have found a violation of it within six standard deviations, the value $\delta = 0.050 \pm 0.008$ for the inequality:

$$\delta_{CHSH} = \left| \frac{R(22.5°)}{R_0} - \frac{R(67.5°)}{R_0} \right| - \frac{1}{4} \leq 0, \tag{1.23}$$

where $R(\phi)$ is the coincidence rate at an angle ϕ between the polarizers, and R_0 the value when they are not present.

Nevertheless, there was still room for criticism about it. An important step to obtain (1.22) is the assumption of independent choice of measurements for Alice and Bob in each run, which is doubtful to be the case in this realization. Alain Aspect was the first one to find a way around this problem (also called the *locality loophole*), by using polarization settings that randomly changed while the photons were still travelling [AGR81; AGR82; ADR82]. In a series of three experiments, he and collaborators first improved the method of exciting calcium atoms, then employed two-channel polarizers that allowed to obtain the largest violation at the time, and finally made use of optical switchers to get rid of the aforementioned challenge by allowing photons to switch between two paths in shorter timescales than those required to travel to the detectors.

Almost two decades after this, Anton Zeilinger's group repeated the test now under much more strict (locality) conditions and with other technical improvements [Wei+98], including 400m of distance between the parties. The addressed issues were mainly related to the challenging task of performing *loophole-free* tests, which require among other things a separation of 300m at least [Bru+14]. Zeilinger's experiments managed to close the locality and the *detection* loopholes simultaneously. This last one has to do with the fact that no detector is completely efficient, which may cause unfair violation of such inequalities.

Chapter 2

Computational tools

The potential of contexuality in the realm of quantum computation has garnered recent attention, both in terms of practical applications (as explored in [How+14], [Rau13], and [Del+15]), as well as its role in the pursuit of computational advantages (refer to [Kle+11]; additionally, consider [Gup+22] with respect to communication complexity advantages). The utilization of contextuality in communication and computational tasks as valuable resources ([Ber+17; Rau+17]) has also led to the development of a dedicated *resource theory* to effectively address this aspect, as detailed in [Ama19].

An integral element within resource theories is, without a doubt, the concept of *quantifiers*. Quantifiers play a crucial role in measuring the quantity of the resource of interest within a given system. Hence, it becomes evident that special attention must be devoted to assessing the degree of contextuality a system may possess.

This chapter is reserved for presenting the systematic line of investigation regarding contextuality quantifiers while placing it within the computational aspects related to the topic, including an overview of optimization problems and important tasks that appear in (quantum) foundational research. That is because, since the way to characterize it derives from a more general approach, it makes sense and seems natural to look for the usability of these "general" problems in the characterization of quantum or non-classical phenomena themselves.

Each quantifier has its own interpretation and specific applications, but there exists shared reasoning underlying the definition of all of them. As mentioned earlier, the goal is to establish criteria that rank our behaviours as "more" or "less" contextual, attributing to them a classification through a numerical value whenever possible.

As we progress through this chapter, the comments and references not only offer detailed sources for interested readers but also demonstrate that quantifiers themselves constitute an important and active area of research. Indeed, the final sections are dedicated to exemplifying this with recent advancements.

2.1 Linear programming

The preceding words are quite suggestive. A computational procedure is unavoidable to our purposes, specially when one notices how large the space of parameters defining our behaviours can be; for example the simplest Bell scenario with two inputs and two outputs, leading to sixteen degrees of freedom. Therefore, we start with the exposition of the structure that our computational tasks will share under the translation of the classification and quantification problems into the language of computers, i.e. algorithms.

Such structure is known as *Linear Programming* and it is as simple as the reason for its name: all its defining pieces carry the form of a mathematical linear (or *affine*) functions. It is an instance of a family of computational problems that we call *optimization problems*. A member of it has an *objective function* (sometimes called *cost function*), say $f(\mathbf{x})$, whose values we want to explore in order to maximize (or minimize) it over a given domain of allowed values for our variables organized in a vector $\mathbf{x}$. Details beyond what is about to be presented here can be found in [GO11].

As we have just discussed, a linear programming (LP from now on) task will aim to optimize an affine function. Labelling $\{x_j\}_{j\in I}$, with I set of indices, the set of variables of the program, any such function can be written as $f(\mathbf{x}) = \sum_j a_j x_j + k$ for a set of given coefficients $\{a_j\}$ and a constant k, all real.

The allowed region of the variables are the real vectors delimited by linear *constraints*. A constraint is an equality or inequality involving the variables of the program. By a linear constraint we will mean an inequality (or equality, when it is the case) of the form $\sum_j m_j x_j \leq b_j$, $j \in I$. And since it is common to have more than one such restriction, the coefficients are grouped into a matrix M and a vector $\mathbf{b}$, so that the constraints can be stated in the form $M\mathbf{x} \leq \mathbf{b}$. In what follows we consider finite regions only.

Therefore, the LP task we will cross throughout this work can be summarized in the following way:

$$\begin{aligned} &\text{Maximize } f(\mathbf{x}) \\ &\text{such that } M\mathbf{x} \leq \mathbf{b} \ . \end{aligned} \tag{2.1}$$

Any $\mathbf{x}$ satisfying the constraints is called a *feasible* solution to the LP task. An *optimal* solution accomplishes the demanded maximization.

The reason for looking to linear formulations of computational problems, besides their simplicity, has to do with the speed of the computations and also the well-known properties they have. To mention the ones of special significance in what follows, we note the fact that graphs of linear functions are just hyper-planes within the real vector space that contains $\mathbf{x}$, therefore the intersection of their implied hyper-spaces delimit a convex region (when finite) of allowed vectors, called a *polytope*[1]; unbounded regions can give raise to unsolvable problems. Moreover, the existence of solutions is well established in the literature, the reference to which we again indicate [GO11]. The most important feature is that optimal solutions are located at one (or some) of the vertices of the defining *feasibility polytope*, i.e. the polytope of solutions to (2.1).

The implementation of this kind of task in a computer can be easily achieved using any programming language, provided the appropriate packages are utilized. In this work, any result regarding LP's will come from simulations written in Julia ([Bez+17]), applying JuMP as model builder for optimization problems ([DHL17]) and the CLP package ([For+22]) to provide the body ((2.1) in computer terms) of the LP's, which we fulfill with a proper soul (the defining coefficients) then make it run after an optimal solution. For other optimization problems, as the ones to be introduced in the next section, we mostly make use of Python and a series of packages supported by it, to be mentioned when convenient later on.

2.1.1 A simple non-locality test

To illustrate this section and give a working example on the use of this tool, we are going to revisit the CHSH scenario of the first chapter (see figure 1.3) to provide a way of testing

[1] These objects are generalizations of 3d-polyhedra in higher dimensions.

whether a given behaviour or (set of) correlations is in agreement or not with a local (i.e. classical) description. That is to say, the LP task to be constructed here is a test for non-locality of the joint probability distributions $p^*(a,b|A,B)$ that Alice and Bob may have in hands after their experiment.

As a particular instance of contextuality, the exhibition of non-locality by their correlations amounts to the non-existence of a probability distribution $p(\lambda)$, the hidden variable model, determining their joint distributions, while also in accordance with the no-signalling (or no-disturbing) condition if necessary.

Explicitly, since we know that λ encodes all possible deterministic strategies, indexing them by the corresponding outputs values a_1, a_2 for measurements A_1, A_2 of Alice and b_1, b_2 of Bob, giving the correct marginal amounts to have

$$\begin{aligned}
\sum_{a_2,b_2} p(a_1,a_2,b_1,b_2) &= p^*(a,b|A_1,B_1),\\
\sum_{a_2,b_1} p(a_1,a_2,b_1,b_2) &= p^*(a,b|A_1,B_2),\\
\sum_{a_1,b_2} p(a_1,a_2,b_1,b_2) &= p^*(a,b|A_2,B_1),\\
\sum_{a_1,b_1} p(a_1,a_2,b_1,b_2) &= p^*(a,b|A_2,B_2).
\end{aligned} \tag{2.2}$$

Therefore, the task of testing non-locality of a given $p^*(a,b|A,B)$ can be stated as the following LP.

$$\begin{aligned}
&\max_{\mathbf{p}} \ 1\\
&\text{such that } \sum_{a_2,b_2} p(a_1,a_2,b_1,b_2) = p^*(a,b|A_1,B_1),\\
&\sum_{a_2,b_1} p(a_1,a_2,b_1,b_2) = p^*(a,b|A_1,B_2),\\
&\sum_{a_1,b_2} p(a_1,a_2,b_1,b_2) = p^*(a,b|A_2,B_1),\\
&\sum_{a_1,b_1} p(a_1,a_2,b_1,b_2) = p^*(a,b|A_2,B_2).\\
&p_j \geq 0 \ \forall j,\\
&\sum_j p_j = 1.
\end{aligned} \tag{2.3}$$

That is, it is equivalent to find a feasible point for the linear programming task above, which is also referred to as a *feasibility* problem, after all any feasible point would provide the same value for the objective function. And the same idea can be extended to more complicated scenarios of non-locality. Adding details, or a *degree* of non-locality, to the objective function is exactly the role of quantifiers, as we are going to see later on. And to finish, if for some reason signalling needs to be tested or taken into account for further exploration too (for example include tolerated errors and so on), this can done by adding

the extra conditions

$$\begin{aligned}\sum_b p^*(a,b|A,B) &= \sum_b p^*(a,b|A,B'),\ \forall a,A,B,B',\\ \sum_a p^*(a,b|A,B) &= \sum_a p^*(a,b|A',B),\ \forall b,B,A,A'\end{aligned} \tag{2.4}$$

which, of course, could be verified previous to the LP test, after all it depends on $\mathbf{p}^*$ only.

2.2 Semi-definite programming

With this section we want to generalize the set of optimization problems that are known to be efficiently solvable in a computer, in the sense of the time efficiency of the algorithms involved. Besides, we want to show a natural path to approach further problems in quantum physics research once one is aware of linear programming. More specifically, they are a particular case of the broader family described as *semi-definite programs* (SDP). As such, one would expect to find useful tools on it that could fit well on tackling optimization tasks raising in quantum foundations and applications . The reason for their efficiency relies on interior points methods [Kar84; KK92; KK93; NN94], providing polynomial time implementation. In what follows, after making the connection just mentioned, we point out typical uses of it that one can find in research on quantum (or more generally non-classical) physics, highlighting those that are going to be useful for us in succeeding chapters. In order to do so we follow [Tav+23]. See also the recent book [SC23] for a focus in quantum applications.

While keeping a linear objective function, SDP's aim to optimize it over a domain for the variables now restricted to be part of a set of positive semi-definite (PSD) matrices. That is to say, we want to extremize a linear function over a set of matrices lying in the intersection of the PSD cone with hyperplanes and/or hyperspaces. Generically, this can be phrased as:

$$\begin{aligned}&\text{Maximize } \mathrm{Tr}\,(CX)\\ &\text{such that } \mathrm{Tr}\,(A_j X) = \mathbf{b}_j,\ \forall j,\\ &\qquad\qquad X \geq 0,\end{aligned} \tag{2.5}$$

where C (the coefficients for the objective function), X and A_j are hermitian matrices and $\mathbf{b}_j$ a real vector. Inequalities can be included as before in addition to or as substitutes for the equality ones above, performing the logical operation component-wisely.

Notice now that by demanding X to be diagonal with each element non-negative, we arrive at the equation (2.1). More important is to note that to every SDP (and therefore LP) we can associate another SDP of the form[2]:

$$\begin{aligned}&\text{Minimize } \mathbf{b}\cdot\mathbf{y}\\ &\text{such that } \sum_j A_j\ y_j \geq C,\end{aligned} \tag{2.6}$$

where the minimization is over real vectors $\mathbf{y}$ and the dot is to denote the inner product between vectors. This is called the *dual problem* associated to equation (2.5), which would then be the *primal problem*. And the importance of it has to do with two known results in their theory, called *weak* and *strong dualities*, which we just enunciate in the form of a non-rigorous theorem here as follows.

[2]For the interested reader, it is obtained via Lagrange multipliers. See [SC23] for example.

Theorem 2.2.1. *(weak and strong dualities) A feasible solution to the dual problem provides an upper bound on the optimal value of the primal, while a feasible solution to the primal provides a lower bound on the optimal value of the dual. The strong duality holds when we have the equality of these bounds, i.e. when the optimal values coincide.*

The equality of the bounds are known to always be the case for LP's. In the case of SDP's, strict feasibility ($X^* > 0$ feasible or equivalently $\mathbf{y}^*$ feasible with $\sum_j A_j \; y_j > C$) is sufficient [Sla13].

2.2.1 Quantum applications

The form of the objective function in equation (2.5) is then quite suggestive on the applications for quantum mechanics when one recall the Born rule that assigns the probability of obtaining the outcome j associated to the element E_j of a generic (i.e. not necessarily projective) measurement, also known as *POVM*[3] , when measured on a state ρ. That means, for $\{E_j\}$ with $j \in O$ finite, such that

$$\begin{aligned} &\sum_j E_j = \mathbb{1}, \\ &E_j \geq 0 \; \forall j \in O, \end{aligned} \tag{2.7}$$

i.e. with $\{E_j\}$ a set of positive semi-definite matrices summing up to the identity, quantum mechanics postulates that the outcome j occurs with probability $p_j(\rho)$ given by

$$p_j(\rho) \equiv \mathrm{Tr}\left(\rho E_j\right), \tag{2.8}$$

where ρ is the quantum state being measured.

Immediately we see then that one could use this with the following purpose, for example. For a fixed finite set of incoming quantum states $\{\rho_j\}$, with characteristics predetermined by the application, one can optimize for the best POVM $\{E_j^*\}$ choice that maximizes the chance p_{guess} of guessing correctly the incoming state in successive runs:

$$p_{guess} \equiv \sum_j \mathrm{Tr}\left(\rho_j E_j\right), \tag{2.9}$$

that is to say, p_{guess} would be our objective function. This is of particular importance in protocols of *quantum Bit Commitment* [BC91], in which one wants to check potential risks of cheating by the receiving part. The protocol is to make a party commit to a bit that will then be sent secretly ("within" a safe) to a second party, with subsequent disclosure of its value by mediation of a key.

In a similar fashion, one can employ a simple heuristic that goes by the name of *seesaw* [PV10; WW01b] in which now one optimizes for states and measurements, each one in a separated step while keeping the others fixed, in order for example to look for the largest quantum violation of a Bell inequality. The candidate would be the one to which the procedure converges to. There is a caveat, however, that one could fall into a local and not a global extreme.

In the examples above there are at least other two other important features playing important roles behind the scenes: *steering* and *entanglement*. And because of the central

[3] It is short for *positive-operator-valued measurement*.

role they can have among other applications, it becomes crucial then to be able to certify their presence (for entanglement) and absence (for the steering). This is where semi-definite programming comes in again.

To understand how we provide such certificates, let us briefly review what is the meaning of those terms, by considering the typical scenario (recall figure 1.3) with Alice and Bob sharing a quantum system ρ. Starting with the simplest one, entanglement, which is just a mathematical feature at the end of the day.

Definition 2.2.1. (*entanglement*) An entangled state ρ in a Hilbert space $\mathcal{H}$ resulting of the product between two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ stands for a state that cannot be written as a probabilistic mixture $p(\lambda)$ of some states ρ_λ^1 and ρ_λ^2 respectively prepared by each party separately, that means

$$\rho \neq \sum_\lambda p(\lambda)\ \rho_\lambda^1 \otimes \rho_\lambda^2, \tag{2.10}$$

where $\otimes$ is the usual tensor product.

The crucial aspect of entanglement is that the state correlates the subsystems. And because of this, by making use of the same well characterized measurement in each party individually, one party can infer (however, not change) the result of the other on the realization of that measurement[4].

Now, steering has to do with a different aspect of correlation between parties. More specifically, it is understood under the scope of only one of Alice and Bob. Say Bob, trusting only on the measurements he possess and the subsystem he holds, wants to answer the question whether his state can be correlated or not to the realized measurements and revealed outputs by Alice. In other words, he wishes to rule out or not the possibility that the information of inputs and outputs from the other side can change the description of his state (not his actual state) and therefore his results, in the sense of not being able to describe his state by local pieces only.

With the partial information that Alice is supposed to select among possible inputs X, corresponding to realizing measurements A_X, providing an output a to it, Bob knows that in his hands lies what we call an *assemblage* of states ("prepared" by Alice), namely

$$\rho_{a|X} = \mathrm{Tr}_A\left(\left(A_X^a \otimes \mathbb{1}\right)\rho\right), \tag{2.11}$$

where Tr_A denotes the *partial trace* over part A. In a quantum system composed of parts 1 and 2, with respective basis $\{|\phi_j^1\rangle\}$ and $\{|\phi_k^2\rangle\}$, *tracing out* 1 or taking the partial trace over 1 of an operator with matrix elements T^{jk}, for example, i.e

$$T = \sum_{j,k} T^{jk} |\phi_j^1\rangle\langle\phi_k^2|, \tag{2.12}$$

means to define an operator on party 2, with elements say T_2^k, such that

$$T_2^k \equiv \mathrm{Tr}_1\left(T\right)^k = \sum_j T^{jk}, \tag{2.13}$$

which is usually referred to, in the case of quantum states, as the reduced state to the party.

The point is that, if there is no correlation between Alice's action at her side and the description that Bob gives to the subsystem in his side, then he should be able to model it by using local hidden states only [WJD07], which leads us to the following definition.

[4] This inference is perfect for maximally entangled states only.

Definition 2.2.2. (*steering*) An assemblage $\rho_{a|X}$ is local or *unsteerable* if

$$\rho_{a|X} = \sum_{\lambda} p(\lambda)\ p(a|X,\lambda)\ \sigma^*_\lambda, \tag{2.14}$$

where $\{\sigma^*_\lambda\}$ is some set of local states at Bob's side, and $p(a|X,\lambda)$ is some probability of having the output a as conditioned by hidden a variable λ and the input X only. Otherwise we say it is *steerable*.

Checking agreement with this definition amounts to test the strict feasibility of the following SDP task, which demands non-negativity of t [Tav+23].

$$\begin{aligned} \max_{\{\sigma_\lambda\},t} \quad & t \\ \text{such that} \quad & \rho_{a|X} = \sum_{\lambda} D(a|X,\lambda)\ \sigma_\lambda, \text{ and} \\ & \sigma_\lambda \geq t\ \mathbb{1}\ \forall\lambda, \end{aligned} \tag{2.15}$$

where $D(a|X,\lambda) = \delta_{r_\lambda(X),a}$ stands for the finite set of deterministic functions $r_\lambda(X)$ indexed by λ mapping the inputs X into fixed output values; the last condition should be seen as between the eigenvalues of σ_λ and t. In the case of reaching optimal solution with the desired condition on t, the dual SDP provides what we call a *certificate* of no steering. To see this, let us explicit the dual form of the task above.

$$\begin{aligned} \min_{\{W_{a|X}\}} \quad & \sum_{a,X} \mathrm{Tr}\left(W_{a|X}\ \rho_{a|X}\right) \\ \text{such that} \quad & \sum_{a,X,\lambda} \mathrm{Tr}\left(W_{a|X}\right)\ D(a|X,\lambda) = 1, \text{ and} \\ & \sum_{a,X} W_{a|X}\ D(a|X,\lambda) \geq 0\ \forall\lambda. \end{aligned} \tag{2.16}$$

The dual then looks for objects $\{W_{a|X}\}$ that add up to form positive semi-definite matrices, as some kind of (pseudo-)element of a (sub-normalized) POVM, because individually there is no need for being PSD and they not necessarily add up to the identity. Strict feasibility in this case, however, means that the objective function needs to be strictly greater than zero, which means that a feasible $\{W^*_{a|X}\}$ provides an inequality that is violated by any steerable assemblage, namely:

$$\sum_{a,X} \mathrm{Tr}\left(W^*_{a|X}\ \rho_{a|X}\right) \geq 0, \tag{2.17}$$

which we call a *witness* of steering.

Certifying entanglement using this approach follows a similar reasoning. In this case we want to certify that a given state ρ is in accordance with equation (2.10). Let Alice and Bob be our experimenters again and call their shared state ρ_{AB}. Then assume they have known quantum measurements $\{A_X\}$ and $\{B_Y\}$, respectively, with corresponding inputs $\{X\}$ and $\{Y\}$. Now, by determining the correlations

$$p(a,b|X,Y) = \mathrm{Tr}\left(\left(A^a_X \otimes B^b_Y\right)\ \rho_{AB}\right), \tag{2.18}$$

they can reconstruct the density matrix ρ_{AB} and decide separability by means of some analytical criterion. However, the problem is that, a necessary and sufficient criterion only

exists for qubit-qubit or qubit-qutrit systems, known as *PPT* criterion [HHH96; Per96], enunciated but not proved below, which is only necessary condition in higher dimensions. Moreover, this decision problem is known to be NP-hard [Gha10; Gur03].

The criterion is stated in terms of the *partial transposition* operation, defined in the following equation for Alice for example.

$$\begin{aligned}\rho_{AB}^{T_A} \equiv (T \otimes \mathbb{1})\ \rho_{AB} &= \sum_{j,k,l,m} \rho_{AB}^{jk,lm}\ T\left(|j\rangle\langle k|\right) \otimes |l\rangle\langle m| \\ &= \sum_{j,k,l,m} \rho_{AB}^{jk,lm}\ |k\rangle\langle j| \otimes |l\rangle\langle m| \\ &= \sum_{j,k,l,m} \rho_{AB}^{kj,lm}\ |j\rangle\langle k| \otimes |l\rangle\langle m|,\end{aligned} \tag{2.19}$$

where T stands for the transposition operator (exchange of row and column indices).

Theorem 2.2.2. *(PPT) If ρ_{AB} is separable then all the eigenvalues of $\rho_{AB}^{T_A}$ are non-negative.*

A typical use of such criterion is not only to detect, but also to quantify the *robustness* of the entanglement, by computing how much of the maximally mixed state (read the identity) needs to be mixed with ρ_{AB} to make it PPT, that is to say, with *positive* partial transposition. It can be computed by means of the following SDP [Tav+23].

$$\begin{aligned}\max_{t} \quad & t \\ \text{such that} \quad & \rho_{AB}^{T_A} \geq\ t\ \mathbb{1},\end{aligned} \tag{2.20}$$

providing a lower bound to that quantity since we are taking in consideration only a necessary condition. This is sometimes referred to as taking a *relaxed* version of the actual problem.

And as before, the dual SDP associated to this problem provides a witness of entanglement, namely

$$\begin{aligned}\min_{W} \quad & \mathrm{Tr}\left(W\rho_{AB}\right) \\ \text{such that} \quad & \mathrm{Tr}\left(W\right) = 1,\ \text{and} \\ & W^{T_A} \geq 0,\end{aligned} \tag{2.21}$$

which is very useful considering the difficulty of realizing full tomography of a given state. By violating an inequality defined by a solution of the task above we already certify the presence of the property in the target state.

Wrapping all about this two components up then, we see that different aspects of correlations, more specifically non-classical ones, can be assessed and leveraged by the same computational approach. And following the reasoning we have used here, naturally the reader could wonder about the situation in which parties are interested in investigating the case in which only information of inputs and outputs to each one is available. However, this aligns precisely with the challenge addressed by non-locality that we have already seen, therefore a "third" form of correlation between Alice and Bob, irrespective of the internal mechanisms of the apparatus on each side.

This does not exhaust all the interesting lines within quantum physics though. The final section of this chapter will deal with another (but not the only remaining) crucial

question on the foundations of correlations that can be also addressed by a similar approach. The reason for having a separate part for it is because it incorporates some more ideas that could make its presentation cumbersome to follow if we not put aside the last paragraphs. Besides, at some point later in this thesis the reader might find useful to have this topic in quick access.

2.3 NPA hierarchy

For the last example of how to apply optimisation tasks to quantum information, which is though the most important for a later chapter, we will also work on a modified version of an actual problem, in the sense of considering formulations of it that can be addressed computationally in a efficient manner. The approach now, instead of considering only relaxed constraints as we have just seen, explores also the idea of approximating the desired result in some limit; convergence and its speed becoming the most interesting aspects of it then. More specifically, we address the challenge of characterizing quantum correlations.

Such challenge is a particular instance of a family of problems that can be approached in a similar fashion. The idea behind is to approximate an optimization task that is not SDP by a series of SDP with increasing degree of complexity, or level. That is, a *hierarchy* of SDP, each level in the hierarchy providing tighter bounds to the solution than the previous. And the class of problems in question is called *non-commutative polynomial optimization*, which is known to be NP-hard [Nes00]. It can be stated as below.

$$\begin{aligned} \max_{\rho,\{X_j\}} \quad & \mathrm{Tr}\left(\rho\ f\left(X_1,...,X_n\right)\right) \\ \text{such that} \quad & \mathrm{Tr}\left(\rho\ h_l\left(X_1,...,X_n\right)\right) \geq 0\ \forall l, \\ & g_m\left(X_1,...,X_n\right) \geq 0\ \forall m, \\ & \mathrm{Tr}\left(\rho\right) = 1,\ \text{and} \\ & \rho \geq 0, \end{aligned} \tag{2.22}$$

where the maximization is over all states ρ and bounded operators $\{X_j\}_{j\in\{1,...,n\}}$ on a Hilbert space $\mathcal{H}$, and the polynomials f, h_l and g_m are all hermitian[5].

The approach we are going to consider for the hierarchy procedure of the task above makes use of what is called *moment-matrices*. To understand it, we first make an useful definition.

Definition 2.3.1. A *monomial* is any composition of the operators $\{X_j\}$ and its length is the number of elements in the composition. The identity operator has length 0.

Let us denote as S_k the set of monomials with length smaller or equal to $k \in \mathbb{N}$. If X_i is not hermitian, for some i, then we also include its adjoint $X_\mathrm{i}^\dagger$ in the generating set of S_k. Then, for any feasible point $(\mathcal{H}, \rho, X_1, ..., X_n)$, that is any state ρ and set of operators $\{X_1, ..., X_n\}$ in $\mathcal{H}$ satisfying the constraints in (2.22), we can define a moment matrix Γ^k, of level k, with elements given by

$$\Gamma^k(M,N) = \mathrm{Tr}\left(\rho\ M^\dagger N\right), \tag{2.23}$$

[5]Since they are functions of operators, they are operators themselves. Therefore the "hermitian" adjective here has the same meaning as that for operators, which are just matrices in finite dimensions. That is to say, an hermitian polynomial (operator) is a polynomial which is equal to its (i.e. the resulting operator's) conjugate transpose.

with $M, N \in S_k$. These matrices are positive semi-definite, as follows from

$$\begin{aligned}\langle\phi|\Gamma^k|\phi\rangle &= \sum_{M,N} \mathrm{Tr}\left(\rho\ M^\dagger N\right)\langle\phi|M\rangle\langle N|\phi\rangle \\ &= \mathrm{Tr}\left(\sum_{M,N} \rho\ M^\dagger N\langle\phi|M\rangle\langle N|\phi\rangle\right) \\ &= \mathrm{Tr}\left(\rho\ R^\dagger R\right),\ \text{with}\ R = \sum_N \langle N|\phi\rangle N \\ &\geq 0,\end{aligned} \tag{2.24}$$

for any state $|\phi\rangle$.

The fundamental step that prescribes the hierarchy appears when one notice that for any polynomial $p\left(X_1, ..., X_n\right)$ of degree smaller or equal than $2k \in \mathbb{N}$, the quantity $\mathrm{Tr}\left(\rho\ p\left(X_1, ..., X_n\right)\right)$ becomes a linear combination of the elements of Γ^k. Notice that the products $M^\dagger N$ in (2.23) include any term that could appear in $p\left(X_1, ..., X_n\right)$, therefore forming a basis for those polynomials; the linear combination in question follows then from linearity of the trace.

And keeping a similar reasoning, we can also introduce *localizing moment matrices* $\Gamma^{k_m}_{g_m}$ as follows

$$\Gamma^{k_m}_{g_m}(M, N) = \mathrm{Tr}\left(\rho\ M^\dagger g_m\left(X_1, ..., X_n\right) N\right), \tag{2.25}$$

which helps with the relaxation of the polynomial constraints g_m, where now M and N are monomials in S_{k_m}, k_m being the level of the matrix. The choice of level k_m is tied to the level k of the moment matrix Γ^k, in order to be able to express the elements above in terms of the previous ones. Naturally, choosing $k_m = k - (\text{degree of } g_m)/2$ guarantees that is the case. Moreover, once g_m is required to be positive semi-definite, it can be shown that so will be $\Gamma^{k_m}_{g_m}$.

With that in hands, we can now define the level k relaxation of the task (2.22) to be the SDP

$$\begin{aligned}\max_{\Gamma^k, \Gamma^{k_m}_{g_m}} \quad & \sum_{M,N \in S_k} f_{M,N}\ \Gamma^k(M, N) \\ \text{such that} \quad & \sum_{M,N \in S_k} h^l_{M,N}\ \Gamma^k(M, N) \geq 0\ \forall l, \\ & \Gamma^{k_m}_{g_m} \geq 0\ \forall m,\ \text{and} \\ & \Gamma^k \geq 0,\end{aligned} \tag{2.26}$$

provided the degrees of f and g_m are not larger then $2k$; $f_{M,N}$ and $h^l_{M,N}$ being just the coefficients of the expansions of the polynomials mentioned before. The optimization above looks for PSD matrices of level k as above that, while working as basis for expanding the functions on it, also respects constraints from the original problem. And to recall, this is a relaxation because the quantities above are formulated as dependent on a feasible point of (2.22), therefore it provides an upper bound to that maximization only.

The problem of testing whether a given behaviour is quantum or not can then be cast in these terms as following. For simplicity, and to contemplate the origin of the method, we assume the system of interest is bipartite, as in the scenario with Alice and Bob in section 1.5. Just to recall, for a given set of inputs $\{X\}$ and $\{Y\}$, respectively for Alice and Bob, a fully characterization of their correlations is given by the set of distributions $p(a, b|X, Y)$, for every pair of inputs (X, Y). And if it accepts a quantum description, we

should be able to find a state ρ and quantum measurements $\{A_X\}$ and $\{B_Y\}$ such that equation (2.18) holds.

While membership of p on the associated quantum set Q is known to be an undecidable task [Ji+21], it is also known that Q may be approximated by a sequence of sets $\{Q_k\}_{k\in\mathbb{N}}$ such that

$$Q_1 \supseteq Q_2 \supseteq ... \supseteq Q_\infty \supseteq Q, \tag{2.27}$$

that it to say, it is a sequence of tighter outer approximations of Q. And testing membership of p on Q_k is a SDP. This hierarchy of SDP carries the name of its creators: Navascués-Pironio-Acín (NPA) [NPA07]. The convergence to Q is true for finite dimensions [NPA07].

To make it more explicit, first define the set of operators that needs to be taken into consideration for the scenario. In this case they are the measurements $\{A_X\}$, $\{B_Y\}$ and the identity $\mathbb{1}$. Let S_k be the set of monomials with length not greater than k. For an unknown state ρ we associate then the moment matrix

$$\Gamma^k(M,N) = \mathrm{Tr}\left(\rho\, M^\dagger N\right), \tag{2.28}$$

which encodes constraints from quantum theory, namely $\Gamma^k(\mathbb{1},\mathbb{1}) = 1$ (normalization), $\Gamma^k(M,N) = \Gamma^k(M',N')$ if $M^\dagger N = M'^\dagger N'$ and $\Gamma^k(M,N) = 0$ whenever $M^\dagger N = 0$ (orthogonal projective measurements, for example). In addition, we want it to produce the correct marginal probabilities under test, which means $\Gamma^k(A_X, B_Y) = p(a,b|X,Y)$, $\Gamma^k(A_X,\mathbb{1}) = p(a|X)$ and $\Gamma^k(\mathbb{1},B_Y) = p(b|Y)$. The other variables just need to be such that $\Gamma^k \geq 0$ (PSD). In summary then, it amounts to optimize the SDP

$$\begin{aligned} \max_{\Gamma^k} \quad & \mathbb{1} \\ \text{such that} \quad & \Gamma^k(A_X,B_Y) = p(a,b|X,Y), \\ & \Gamma^k(A_X,\mathbb{1}) = p(a|X), \\ & \Gamma^k(\mathbb{1},B_Y) = p(b|Y), \text{ and} \\ & \Gamma^k \geq 0, \end{aligned} \tag{2.29}$$

the objective function $\mathbb{1}$ is because it is a feasibility test, that is, we only want to find a valid Γ^k. Notice also that Γ^k can be taken to be real, after all if Γ^k is a solution, so is Γ^{k*} and therefore $(\Gamma^k + \Gamma^{k*})/2$.

As one may have noticed, extension to more parties is straightforward and it will be made explicit in future chapters if it turns out to be necessary. Besides, it should also be pointed out that the dual of such SDP is as important as before: by formulating in terms of optimization of a function defining non-contextuality inequality (in this case Bell inequality), it provides an optimal bound on the quantum violation.

Finally, by considering only commuting operators in the formulation of (2.22), we arrive in a SDP hierarchy with a finite sequence of outer approximations $\{\mathcal{L}_k\}$ that can be used instead to test compatibility with a classical description $\mathcal{L}$, i.e with local hidden variable models. The finiteness of the sequence occurring because increasing the number of elements in a monomial after some point becomes redundant in view of the commutation relations.

Chapter 3

Contextuality quantifiers

We are now well-prepared to delve into the process of characterizing and quantifying behaviours with respect to contextuality. The upcoming sections should be regarded as working examples, stemming from the initial one. Quantifiers will be presented mostly within the framework of CbD. However, as mentioned earlier, instances within the traditional approach can be derived by imposing additional constraints. Consequently, the quantifiers should maintain their form in that perspective as well. The rationale behind this choice will become more apparent as we progress through the section. The key idea is to encompass a broader set of systems. Furthermore, our goal here is not to establish their validity as quantifiers within a resource theory ([Ama19] covers this aspect), primarily because it is not the main focus we want to have on the subject. Secondly, these quantifiers are well-established entities in the literature, and our references will guide the reader in that direction if desired.

We are going to warm up in the topic without initially making use of the structure from the last section. The intention is to fix the qualitative part of the process in our heads before going into computations that can become rather abstract and non-sense once not performed with pen and paper.

We call our first example of contextuality quantifier CNT_0. It is the straightforward one that the reader would have in mind at this point after the first chapter. With a non-contextuality inequality $S \leq B_{NC}$ in hands, where B_{NC} stands for the non-contextual bound, the criterion (known as *Bell* criterion [Bel64]) for having a contextual behaviour is, of course, $S > B_{NC}$.

Surpassed the bound, the amount exceeding it is then a natural manner to quantify our resource. We have then:

$$CNT_0 = \max(S - B_{NC}, 0) \ , \tag{3.1}$$

where the maximum is used to give a proper characterization of non-contextual systems too, i.e. zero degree of contextuality in such cases. One could even change the last step and normalize the final quantity, for example by dividing it by the greatest algebraic value S could reach, to get a normalized result; in [ABM17] some properties of CNT_0 are studied under this standpoint.

Let us see an explicit example, while it is simple enough so we can compute using paper and pen. Consider the CHSH scenario illustrated in figure 1.1. It is in the traditional approach to contextuality, but there is no need to complicate things reformulating it in the CbD view; we would only have a different non-contextuality inequality and a greater set of joint probabilities to deal with, which would simplify to this one under the added

exigence of consistent connectedness. We have four dichotomic random variables A_1, A_2, B_1 and B_2, with outcomes ± 1, and four contexts (A_1B_1), (A_1B_2), (A_2B_1) and (A_2B_2). Our analysis of a given behavior consists in taking the joint probabilities and inserting them into the following inequality:

$$S_{CHSH} \equiv |\langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle| \leq 2 \ . \tag{3.2}$$

The modulus is there to deal with the dance of signs underlying the choice of labels to the random variables.

Consider then the following behaviour $\mathbf{p}$, obtained from a simulation of the statistics of a quantum measurement process:

Context	p(-1,-1)	p(-1,1)	p(1,-1)	p(1,1)
(A_1B_1)	0.42	0.08	0.08	0.42
(A_1B_2)	0.38	0.12	0.12	0.38
(A_2B_1)	0.28	0.22	0.22	0.28
(A_2B_2)	0.04	0.46	0.46	0.04

Table 3.1: CHSH contextual example.

For this behavior, each expectation value (recall (1.8)) appearing in S_{CHSH} can be computed by summing up the probabilities at the beginning and end of the corresponding line above, and then subtracting the probabilities in the middle of the same line. We have then:

$$S_{CHSH}(\mathbf{p}) = |0.68 + 0.52 + 0.12 + 0.84| = 2.16. \tag{3.3}$$

Hence $CNT_0(\mathbf{p}) = 0.16$.

This quantifier is rather instructive. However, it can be very problematic depending on the scenario. That is because, in general, there are non-equivalent non-contextuality inequalities associated to the same scenario[1]. It can be the case of having a behaviour that violates one inequality but not another. One could argue then that a manner to bypass this problem is by considering all those inequalities. Nevertheless, determining all of them is known to be a hard problem (see [Pit08] for example).

The next quantifiers are options to avoid such struggle. The cost we pay for it, nonetheless, is having to appeal to computers to do the job. With the current technology, this is, on the contrary to what may sound, "happy news". We follow the nomenclature adopted in [KD19] for them.

Before delving into it, however, we need to establish a useful vectorial representation of behaviours. This is a common procedure found in the literature ([KD19] and [AB11], for example). Recall that in CbD, a system $\mathcal{R}$ is represented by a set of random variables labeled by the bunch (context) $c \in \mathcal{C}$ and the connection $q \in Q$ they appear in, i.e. R_q^c.

A complete description of the system is given by a vector $\mathbf{p}$ made of two other vectors concatenated:

$$\mathbf{p}_{(.)} = \begin{bmatrix} \mathbf{p_{(b)}} \\ \mathbf{p_{(c)}} \end{bmatrix} \ , \tag{3.4}$$

[1]By equivalence in this case we mean having quantities that can be obtained from the other by permuting the random variables appearing on it or exchanging their subscripts.

where $\mathbf{p_{(b)}}$ and $\mathbf{p_{(c)}}$ stand for all the joint probabilities for each context and any coupling for the connections, respectively. For each bunch c, we can construct a vector $\mathbf{p}^{(c)}$ by stacking the joint probabilities; see it as a column vector constructed by piling/stacking its entries. If we now stack all such $\mathbf{p}^{(c)}$'s we get $\mathbf{p_{(b)}}$. In the same fashion, for each connection q, we write an analogous $\mathbf{p}^{(q)}$ from a coupling and stack them to form $\mathbf{p_{(c)}}$. Notice that, for a consistently connected system, the information about connections could be dropped, and the description is just as simple as the example we have seen above in table 3.1.

For example, for the simplest scenario in table 1.1, in which we have only four binary variables $R_1^1, R_1^2, R_2^1, R_2^2$, with two contexts $c_1 = \{R_1^1, R_2^1\}, c_2 = \{R_1^2, R_2^2\}$ and two connections $q_1 = \{R_1^1, R_1^2\}, q_2 = \{R_2^1, R_2^2\}$, we would have

$$\begin{aligned}
\mathbf{p}^{(c_1)^\intercal} &= \left[p(0,0|R_1^1,R_2^1), p(0,1|R_1^1,R_2^1), p(1,0|R_1^1,R_2^1), p(1,1|R_1^1,R_2^1)\right], \\
\mathbf{p}^{(c_2)^\intercal} &= \left[p(0,0|R_1^2,R_2^2), p(0,1|R_1^2,R_2^2), p(1,0|R_1^2,R_2^2), p(1,1|R_1^2,R_2^2)\right], \\
\mathbf{p}^{(q_1)^\intercal} &= \left[p(0,0|R_1^1,R_1^2), p(0,1|R_1^1,R_1^2), p(1,0|R_1^1,R_1^2), p(1,1|R_1^1,R_1^2)\right], \\
\mathbf{p}^{(q_2)^\intercal} &= \left[p(0,0|R_2^1,R_2^2), p(0,1|R_2^1,R_2^2), p(1,0|R_2^1,R_2^2), p(1,1|R_2^1,R_2^2)\right],
\end{aligned} \tag{3.5}$$

where the probabilities appearing in $\mathbf{p}^{(q_1)}$ and $\mathbf{p}^{(q_2)}$ are determined from those in $\mathbf{p}^{(c_1)}$ and $\mathbf{p}^{(c_2)}$ depending on the adopted notion of coupling. The vector $\mathbf{p_{(b)}}$ in this case would then be a concatenation of $\mathbf{p}^{(c_1)}$ and $\mathbf{p}^{(c_2)}$, and $\mathbf{p_{(c)}}$ a concatenation of $\mathbf{p}^{(q_1)}$ and $\mathbf{p}^{(q_2)}$. And therefore, $\mathbf{p}_{(.)}$ in this example would be a concatenation of the vectors in the expressions above, from top to bottom.

Nonetheless, this is not the most convenient representation to work with yet, because it does not allow to change one component without affecting others. For this reason, we are going to reduce the description excluding redundancies ([DK16a]). In view of the uniqueness of multi-maximal couplings for systems of binary random variables and the dichotomization procedure, we assume from now on that all R_q^c are binary, with outcomes say 0 and 1.

Within each bunch c, replace $\mathbf{p}^{(c)}$ with a stack of

$$\Pr(R_q^c = 1 \text{ and } R_{q'}^c = 1) = \left\langle R_q^c R_{q'}^c \right\rangle \ , \tag{3.6}$$

for all q and q' appearing in c, with $q \neq q'$. And do the analogous within each connection. Notice we are grouping expectation values of marginals of pairs of random variables up to now. Let's call those new vectors $\mathbf{p_b}$, the bunches marginals, and $\mathbf{p_c}$, the connections marginals.

In the example just above, those vectors amount to

$$\begin{aligned}
\mathbf{p_b}^\intercal &= \left[p(1,1|R_1^1,R_2^1), p(1,1|R_1^2,R_2^2)\right], \\
\mathbf{p_c}^\intercal &= \left[p(1,1|R_1^1,R_1^2), p(1,1|R_2^1,R_2^2)\right].
\end{aligned} \tag{3.7}$$

Now, a sufficient description for the system in this reduced perspective will also require the consideration of *lower* marginals, i.e. expectation values containing a smaller number of random variables within the brackets. The missing parts are the normalization of the probabilities[2], under the convention to it always be the first entry (i.e. topmost) in the reduced representation, and the expectation value of each variable alone, $\langle R_q^c \rangle$. We stack

[2] In the literature usually represented by the symbol $\langle\rangle$.

them into a vector that we call $\mathbf{p_l}$. The resulting vector for the whole system will then be:

$$\mathbf{p} = \begin{bmatrix} \mathbf{p_l} \\ \mathbf{p_b} \\ \mathbf{p_c} \end{bmatrix} . \tag{3.8}$$

In our working example, we would have

$$\mathbf{p_l}^\intercal = \left[1, \langle R_1^1\rangle, \langle R_1^2\rangle, \langle R_2^1\rangle, \langle R_2^2\rangle\right] , \tag{3.9}$$

and therefore

$$\mathbf{p}^\intercal = \left[1, \langle R_1^1\rangle, \langle R_1^2\rangle, \langle R_2^1\rangle, \langle R_2^2\rangle, \langle R_1^1, R_2^1\rangle, \langle R_1^2, R_2^2\rangle, \langle R_1^1, R_1^2\rangle, \langle R_2^1, R_2^2\rangle\right] , \tag{3.10}$$

where the 1 is for an assumption of normalized probabilities, and the expectation values of single variables are obtained from previous $\mathbf{p}^{(c)}$'s, for instance $\langle R_1^1\rangle = p(1,0|R_1^1, R_2^1) + p(1,1|R_1^1, R_2^1)$.

We can go a step further. Notice that, for a scenario with N binary random variables, there is a total of 2^N possible deterministic behaviours, corresponding each to fixed outputs for all of them, to which we can assign probabilities. Indeed, in the previous chapter we called the vector $\mathbf{x}$ of assigned probabilities (summing up to 1) to each of these events a coupling for a given behaviour.

For each expectation value appearing in $\mathbf{p}$ there is a conjunction of events in which the random variables within it assume the outcome 1. A natural manner to express $\mathbf{p}$ in terms of couplings then emerges: consider a matrix $\mathbf{M}$, in which the entry M_{ij} assumes the value 1 if the event j has outcome 1 for the random variables appearing in the component i of $\mathbf{p}$ and 0 otherwise; the first line of $\mathbf{M}$ is filled with 1's, in accordance with its normalization. Therefore, we also have:

$$\mathbf{p} = \mathbf{M}\mathbf{x} .$$

We may also use $\mathbf{p_l} = \mathbf{M_l}\mathbf{x}$, $\mathbf{p_b} = \mathbf{M_b}\mathbf{x}$ and $\mathbf{p_c} = \mathbf{M_c}\mathbf{x}$ for the lines of $\mathbf{M}$ restricted to the corresponding vectors in $\mathbf{p}$. This construction can be related to equations (1.12) and (1.15) directly as well.

To specify a system $\mathbf{p}$ starting from this reduced representation directly, we need to provide $\mathbf{p_l}$ and $\mathbf{p_b}$ either estimated by an experiment or predicted by a model. With that in hands, $\mathbf{p_c}$ is filled with the connections probabilities for multi-maximal couplings; for binary random variables, each entry in $\mathbf{p_c}$ is obtained by taking the smallest expectation value in $\mathbf{p_l}$ among the ones of the random variables involved on it (see [AD19a] and [DK16b] for example).

The reader should be guessing by now what is the point of this short digression.

Definition 3.0.1. A system $\mathbf{p}$ is non-contextual if there is a non-negative $\mathbf{x}$, component wise, such that

$$\mathbf{p} = \mathbf{M}\mathbf{x}, \text{ with } |\mathbf{x}|_1 = 1. \tag{3.11}$$

Otherwise the system is contextual.

This definition is in accordance with chapter 1 and $|.|_1$ denotes the L_1 norm of a vector (the sum of the absolute value of its entries).

We have found a criterion for characterizing a system as contextual or not through this representation. In particular, it should be noted that it is given by a linear restriction.

And this is not surprising, because our next measures of contextuality explores it to construct a manner of quantifying the resource through linear programming.

In on hand, if there is no multi-maximal coupling for a given data (i.e. $\mathbf{p_b}$ and $\mathbf{p_l}$), we could find a coupling $\mathbf{x}^*$ that produces a $\mathbf{p}^*$ that is the closest (with respect to some norm, we choose L_1) to $\mathbf{p_c}$, while giving the correct lower and bunches marginals. That is to say, for a given system $\mathbf{p}$ as before, define the following polytope

$$\mathbb{P}_\mathbf{c} = \{\bar{\mathbf{p}}_\mathbf{c} : \mathbf{M_c x} = \bar{\mathbf{p}}_\mathbf{c}, \text{ for some } \mathbf{x} \geq 0, \ \mathbf{M_l x} = \mathbf{p_l}, \ \mathbf{M_b x} = \mathbf{p_b}\}, \tag{3.12}$$

where the non-negativity is component wise as before. It corresponds to the set of all possible couplings for our system $\mathbf{p}$. Then find $\mathbf{p}^*_\mathbf{c}$ in $\mathbb{P}_\mathbf{c}$ that minimizes the L_1-distance between $\mathbf{p}$ and the polytope (first appeared in [KDL15] and [KD16]):

$$CNT_1(\mathbf{p}) = \min_{\bar{\mathbf{p}}_\mathbf{c} \in \mathbb{P}_\mathbf{c}} \left|\mathbf{p_c} - \bar{\mathbf{p}}_\mathbf{c}\right|_1 = |\mathbf{p_c}|_1 - \max_{\bar{\mathbf{p}}_\mathbf{c} \in \mathbb{P}_\mathbf{c}} |\bar{\mathbf{p}}_\mathbf{c}|_1. \tag{3.13}$$

Noticing that we have:

$$\max_{\bar{\mathbf{p}}_\mathbf{c} \in \mathbb{P}_\mathbf{c}} |\bar{\mathbf{p}}_\mathbf{c}|_1 = \max_{\mathbf{x} \geq 0, \mathbf{M_l x} = \mathbf{p_l}, \mathbf{M_b x} = \mathbf{p_b}} \left|\mathbf{M_c x}\right|_1, \tag{3.14}$$

a solution to the task can be found via the LP program in table 3.2.

Find	$\mathbf{x}$
maximizing	$\left\|\mathbf{M_c x}\right\|_1$
such that	$\mathbf{x} \geq 0$, $\mathbf{M_l x} = \mathbf{p_l}$, $\mathbf{M_b x} = \mathbf{p_b}$

Table 3.2: CNT_1 LP task.

Optimal condition is reached for $\mathbf{x}^*$. Therefore, according to this measure, our system $\mathbf{p}$ has the following amount of contextuality:

$$CNT_1(\mathbf{p}) = \left|\mathbf{p} - \mathbf{M x}^*\right|_1. \tag{3.15}$$

One the other hand, the same reasoning can be reproduced in terms of bunch marginals instead. In this case, we look for a multi-maximal coupling $\mathbf{x}^{**}$ that yields a $\mathbf{p}^{**}$ the closest to $\mathbf{p_b}$ as possible, in L_1 norm, while $\mathbf{p}^{**}_\mathbf{l} = \mathbf{p_l}$ and $\mathbf{p}^{**}_\mathbf{c} = \mathbf{p_c}$. Or, repeating the previous idea, define the polytope

$$\mathbb{P}_\mathbf{b} = \{\bar{\mathbf{p}}_\mathbf{b} : \mathbf{M_b x} = \bar{\mathbf{p}}_\mathbf{b}, \text{ for some } \mathbf{x} \geq 0, \ \mathbf{M_l x} = \mathbf{p_l}, \ \mathbf{M_c x} = \mathbf{p_c}\}. \tag{3.16}$$

This polytope corresponds to the set of all non-contextual systems that align with the individual expectation values for the random variables encoded in $\mathbf{p_l}$ and with the distributions for each context $\mathbf{p_c}$. In simpler terms, the elements of this set are reduced vectors containing information about joint probabilities within contexts that could have generated the expectation values for each random variable observed in $\mathbf{p_l}$. This is done while adhering to the multi-maximal condition required in constructing $\mathbf{p_c}$. Our measure reads (as first introduced in [KD19])

$$CNT_2(\mathbf{p}) = \min_{\bar{\mathbf{p}}_\mathbf{b} \in \mathbb{P}_\mathbf{b}} \left|\mathbf{p_b} - \bar{\mathbf{p}}_\mathbf{b}\right|_1 = \min_{\mathbf{x} \geq 0, \mathbf{M_l x} = \mathbf{p_l}, \mathbf{M_c x} = \mathbf{p_c}} \left|\mathbf{p_b} - \mathbf{M_b x}\right|_1. \tag{3.17}$$

Find	$\mathbf{x}$
minimizing	$\lvert\mathbf{d}\rvert_1$
such that	$\mathbf{x} \geq 0$, $\mathbf{d} \geq 0$,
	$\mathbf{M_l x} = \mathbf{p_l}$, $\mathbf{M_c x} = \mathbf{p_c}$,
	$-\mathbf{d} \leq \mathbf{p_b} - \mathbf{M_b x} \leq \mathbf{d}$

Table 3.3: CNT_2 LP task.

The minimum is attained by the optimal solution $\mathbf{x}^{**}$ of the LP in table 3.3. The additional vector $\mathbf{d}$ is there to deal with the L_1 norm appearing in equation (3.17); it should be interpreted as a vector of differences (component wise) between $\mathbf{p_b}$ and $\mathbf{M_b x}$, which we want to minimize.

After this measure, options seem to be exhausted. However, as we will see along the rest of this section, there are alternative ways of thinking that still allow for the construction of valid quantifiers. These possibilities show up when one decides to explore subtle mathematical aspects of equation (3.11).

Notice, to start, that $\mathbf{x}$ must be non-negative on it. It could be argued then that a contextual system $\mathbf{p}$, inconsistent by construction with equation (3.11), could satisfy it if we abandoned the non-negativity condition. Indeed, this idea has been already explored ([DK16a]) and has given birth to a measure that we will call CNT3. In the spirit of the previous paragraphs, start considering the polytope

$$\mathbb{Y} = \left\{ \mathbf{y} : \mathbf{My} = \mathbf{p}, \sum_{\mathrm{i}} y_{\mathrm{i}} = 1 \right\}, \tag{3.18}$$

which was shown in [DK16a] to be non-empty.

Since there is no $\mathbf{x} \geq 0$ included in the set, any element $\mathbf{y}$ on it satisfy

$$|\mathbf{y}|_1 > 1. \tag{3.19}$$

The natural measure elapsing from it is then:

$$CNT_3(\mathbf{p}) = \min_{\mathbf{y} \in \mathbb{Y}} (|\mathbf{y}|_1 - 1). \tag{3.20}$$

It quantifies how much the sum of the absolute values of the entries in $\mathbf{y}$ can be made close to 1, while generating the desired $\mathbf{p}$; 1 because proper couplings sum up to this value as we have seen. In fact, an appropriate manner to refer to a $\mathbf{y} \in \mathbb{Y}$ at this point is as a quasi-coupling to the vector $\mathbf{p}$. This quantifier can be seen as connected to the debate involving negativity and contextuality as equivalent notions of non-classicality in Physics (for a synthesis see [Spe08]).

An optimal quasi-coupling for equation (3.20) can be found through the optimization of the LP task in table 3.4. We write $\mathbf{y} = \mathbf{y}^+ - \mathbf{y}^-$, a difference between two non-negative vectors, and minimize the norm of the one carrying the negative entries in $\mathbf{y}$.

The last measure we are going to present here follows the same steps above in its formalization. The only difference relies in the need of using "complete" vectors as in (3.4) in replacement of $\mathbf{p}$ in (3.11); notice that this also implies the need of substituting

Find	$\mathbf{y}^+$ and $\mathbf{y}^-$
minimizing	$\lvert\mathbf{y}^-\rvert_1$
such that	$\mathbf{M}\left(\mathbf{y}^+ - \mathbf{y}^-\right) = \mathbf{p}$,
	$\mathbf{y}^+ \geq 0$, $\mathbf{y}^- \geq 0$

Table 3.4: CNT_3 LP task.

$\mathbf{M}$ by a similar $\mathbf{M}_{(.)}$ acting in a coupling, say $\mathbf{z}$, for the system[3]. That is because it requires a look at probabilities themselves instead of expectation values, as we will see. Consider then the following polytope

$$\mathbb{Z} = \left\{\mathbf{z} : \mathbf{M}_{(.)}\mathbf{z} \leq \mathbf{p}_{(.)}, \mathbf{z} \geq 0, \lvert\mathbf{z}\rvert_1 \leq 1\right\}, \tag{3.21}$$

where the inequalities are imposed component wise as always. We see that this set stands for all couplings and sub-normalized couplings (i.e. couplings with $\lvert\mathbf{z}\rvert_1 < 1$) yielding marginal probabilities below the respective component in $\mathbf{p}_{(.)}$, for all of them. It is always a non-empty set: $\mathbf{z} = \mathbf{0}$ (zero for all entries) is always there. Nonetheless, the important situation occurs when there is a $\mathbf{z}$ for which $\lvert\mathbf{z}\rvert_1 = 1$. This implies that $\mathbf{M}_{(.)}\mathbf{z} = \mathbf{p}_{(.)}$, because within every $\mathbf{p}^{(\mathbf{c})}$ and $\mathbf{p}^{(\mathbf{b})}$ the probabilities sum to 1. In other words, it means $\mathbf{p}_{(.)}$ is non-contextual.

Contextual instances do not manifest this last feature. More specifically, any $\mathbf{z}$ in the polytope defined for it is sub-normalized. It is natural then to define a new quantifier (see [AB11], [ABM17] and again [KD19]) of contextuality as below

$$CNTF(\mathbf{p}_{(.)}) = 1 - \max_{\mathbf{z}\in\mathbb{Z}}(\lvert\mathbf{z}\rvert_1), \tag{3.22}$$

where the maximum is achieved by an optimal solution to the LP task in table 3.5. There is no reason for it to be unique *a priori*.

Find	$\mathbf{z}$
maximizing	$\lvert\mathbf{z}\rvert_1$
such that	$\mathbf{M}_{(.)}\mathbf{z} \leq \mathbf{p}_{(.)}$,
	$\mathbf{z} \geq 0$, $\lvert\mathbf{z}\rvert_1 \leq 1$

Table 3.5: $CNTF$ LP task.

Observe that the idea of this measure is to find a behavior $\mathbf{M}_{(.)}\mathbf{z}$ that approximates $\mathbf{p}_{(.)}$ from below the best it can for all joint probabilities at the same time by moving $\mathbf{z}$ as close as possible to a proper coupling. In fact, as presented in the original work [AB11], this optimizer could be seen as the maximal *non-contextual fraction* in a convex combination of behaviors summing up to $\mathbf{p}_{(.)}$; $CNTF$ would then represent the minimum *contextual*

[3]It is constructed in the same way, just abandon the brackets and look at the joint probabilities defined by the deterministic outcomes for the random variables involved. The matrix $\mathbf{M}_{(.)}$ at the end of the day performs the marginalization of the coupling $\mathbf{z}$ onto all the joint probabilities within bunches and connections.

fraction required in such decomposition. That is to say, our contextual behavior could be written as

$$\mathbf{p}_{(.)} = CNTF\,\mathbf{p}^{\mathrm{c}}_{(.)} + (1 - CNTF)\,\mathbf{p}^{\mathrm{nc}}_{(.)}, \tag{3.23}$$

for two behaviors $\mathbf{p}^{\mathrm{c}}_{(.)}$, contextual, and $\mathbf{p}^{\mathrm{nc}}_{(.)}$, non-contextual. Recall that a convex combination of two vectors is a linear combination of them with the two non-negative coefficients summing up to 1.

We conclude this section with a comment on the utilization of the L_1-norm in our definitions. The primary reason lies in the fact that for non-negative vectors (component wise), it represents the simple sum of its coordinates. Secondly, for CNT_2, it is known to have an operational meaning of how well the two distributions can be distinguished considering an optimal event [BAC18]. And a third reason stems from the ease with which it lends itself to linear programming (LP) constructions. In fact, the non-linearity of geometric distances poses more challenging optimization problems, particularly quadratic ones where the objective functions involve second powers of the variables. These quadratic problems are generally impractical due to the complexity reflected in the time required for optimization.

Lastly, while normalization constants can be introduced before the linear objective functions employed in our programs, their introduction, despite any potential physical meaning (as exemplified in [BAC18] where CNT_2 assumes the significance of the *trace distance* between a behavior and the set of non-contextual polytope), serves as mere artifacts for practical purposes.

3.1 Relations between measures of contextuality

We have seen throughout the last section different manners to locate contextual examples within the whole universe of existing systems for a scenario by means of notions of closeness to convex sets determined by non-contextual features. In view of that, at this point the reader must be wondering why the existence and the consideration of so many quantifiers.

The main reason behind it is that this subarea is rather recent, all the aforementioned works (except those related to CNT_0) date less than ten years ago; that is, it is still under development. And also, *a priori* different quantifiers are expected to produce non-equivalent catalogs of the resource for the same scenario. In the sense that, provided two quantifiers are proportional to each other, the knowledge of the catalog produced by one would be sufficient to produce the other, as rulers using distinct marks.

Therefore, the rest of this chapter is dedicated to uncover what we know about the relations among them, a natural line of progress one would expect, and about a deeper feature carried by some of them: extendability to measures of non-contextuality.

Nevertheless, up to the knowledge of this author, relations between quantifiers of contextuality have been explored (in parts) only for a specific, but important family of scenarios. Known as *cyclic scenarios*, the already mentioned CHSH ([Cla+69]) and KCBS ([Kly+08]) are members of particular importance as they are recurrent figures in the literature; we add to the pile of references we already have on this the works [DKC20] and [Abr13], which shows that systems without cyclic subsystems (i.e. obtained by taking a subset of the set of random variables) are necessarily non-contextual. For this reason, we reserve some lines to a proper presentation before going into the point of interest.

As exhibited in [Cer23], a system $\mathcal{R}$ is cyclic if:

1) Each of its contexts are composed of two jointly distributed binary random variables,

2) Each content q appears in two contexts, and

3) There is no proper subsystem of $\mathcal{R}$ satisfying 1 and 2.

A cyclic system of *rank* n refers to a cyclic system with n contexts. Table 1.1 shows a cyclic scenario of rank 2, the simplest one. Contexts and connections can be arranged and labelled so that

$$\mathcal{R}_n = \left\{\left\{R^{\mathrm{i}}_{\mathrm{i}}, R^{\mathrm{i}}_{\mathrm{i}\oplus 1}\right\} : \mathrm{i} = 1, ..., n\right\}, \tag{3.24}$$

where $\oplus$ denotes cyclic shift: $1 \to 2, ..., n-1 \to n, n \to 1$. These arrangement allows then a graphical disposition, similar to figure 1.1, with vertices and edges. The vertices representing random variables and the edges bunches and connections; two edges are linked to each vertex, one for the context the random variable appears in and another for the connection. Figure 3.1 shows the same rank 2 scenario through this layout.

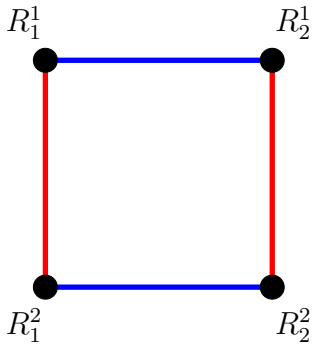


Figure 3.1: Rank 2 cyclic scenario: $c \in \{1, 2\}$ and $q \in \{1, 2\}$. In this case, horizontal (blue) edges are contexts, vertical (red) ones are connections.

Going back to the topic, in [DKC20] it has been shown that for such systems we have

$$CNT_1(\mathbf{p}) = CNT_2(\mathbf{p}). \tag{3.25}$$

The proof is a consequence of properties of the polytopes defined by (3.12) and (3.16). It can be shown that the L_1-distance of $\mathbf{p}$ from them is single-coordinate (i.e. can be computed along a single coordinate), the same in any of the coordinates of $\mathbf{p_b}$ or $\mathbf{p_c}$. The equality is then immediate from this fact. In parallel to the proof, a measure like CNT_0 is shown to be four times the value of them.

Succeeding this result, in [KD19] has been conjectured that for these scenarios we may have

$$CNT_3(\mathbf{p}) = \frac{2CNT_1(\mathbf{p})}{n-1} = \frac{2CNT_2(\mathbf{p})}{n-1}, \tag{3.26}$$

with n denoting the rank of the cyclic system $\mathbf{p}$.

A recent work on the subject ([Cer23]) managed to include $CNTF$ into the picture. It has been proved that the following equality also holds for those scenarios, namely that

$$CNTF(\mathbf{p}_{(.)}) = 2CNT_2(\mathbf{p}). \tag{3.27}$$

It would be of great importance to put an end in this conjecture by proving or disproving it. In particular to Physics, in which the possible veracity of the conjecture would lead to the unavoidable conclusion that all these measures provide the same information about the resource, despite their rather different formulations. At least within this class of scenarios. And the next subsection is dedicated exactly to this end.

3.1.1 Closing the conjecture

The content of this section is the first presentation of an academic contribution of this author, fruit of an international collaboration (with my colleague Víctor H. Cervantes). For this reason, in order to contemplate the original work, it is going to be presented in a form very close to the pre-print version [CC23], therefore potentially overlapping with what we have seen up to this point in this manuscript.

Our main result relates the measures CNT_3 and $CNTF$, proving their proportionality and thereby closing the aforementioned conjecture. To relate them in a cyclic system $\mathcal{R}_n$, we consider the set of its *defective* quasi-couplings, those that fail to be a coupling, but still obey some conditions, explicit in the expression below. Let

$$\mathcal{Q}_n = \left\{ \mathbf{x} \in \mathbb{R}^{2^{2n}} : \mathbf{M}_{(.)}\mathbf{x} \leq \mathbf{p}^*_{(.)} \text{ and } \mathbf{1}^\intercal \mathbf{x} \leq 1 \right\}, \tag{3.28}$$

the convex pyramid obtained by the intersection of the convex polyhedral cone[4] defined by the half-spaces $\mathbf{M}_{(.)}\mathbf{x} \leq \mathbf{p}^*_{(.)}$ and the half-space $\mathbf{1}^\intercal \mathbf{x} \leq 1$. Figure 3.2 schematically illustrates the set $\mathcal{Q}_n$. We see that the intersection of hyperplane $\mathbf{1}^\intercal \mathbf{y} = 1$ and $\mathcal{Q}_n$ defines the face of the pyramid on which all solutions to task 3.4 used to compute CNT_3 lie. Similarly, the intersection of hyperplane $\mathbf{1}^\intercal \mathbf{z} = 1 - CNTF$, $\mathcal{Q}_n$, and the nonnegative orthant of $\mathbb{R}^{2^{2n}}$, defines a slice on whose surface lie all solutions to task 3.5 used to compute CNTF.

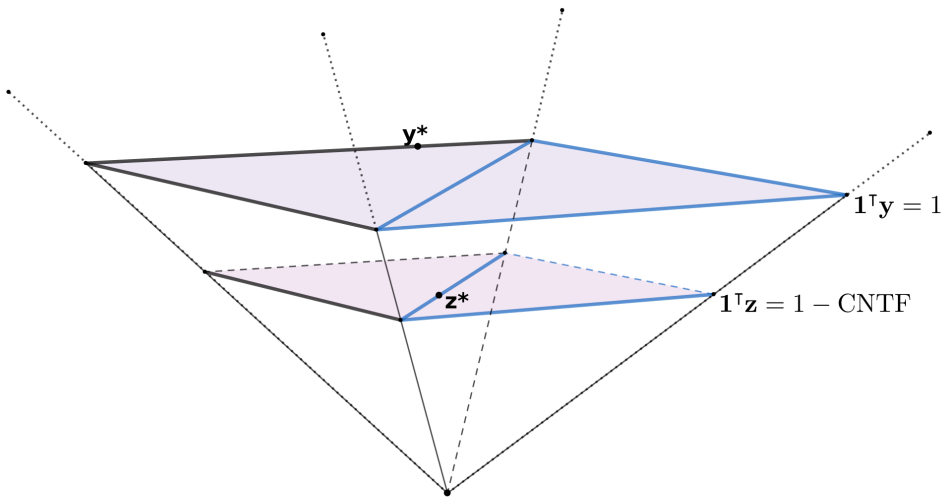


Figure 3.2: Scheme of the pyramid of defective quasi-couplings $\mathcal{Q}_n$. The intersection of $\mathcal{Q}_n$ and the nonnegative orthant of $\mathbb{R}^{2^{2n}}$ is illustrated via the blue lines on the two depicted slices cutting through $\mathcal{Q}_n$. Quasi-couplings $\mathbf{y}^*$ lie on the slice $\mathbf{1}^\intercal \mathbf{y} = 1$ and defective couplings $\mathbf{z}^*$ that are solutions to task (3.5) lie within the closed region delimited by blue edges on the slice $\mathbf{1}^\intercal \mathbf{z} = 1 - CNTF$.

Lemma 3.1.1. If a cyclic system $\mathcal{R}_n$ is contextual, there exists some solution $\mathbf{y}^*$ of task 3.4 with a single negative component.

[4] that is, a space closed under addition and multiplication by non-negative scalars generated by the intersection of a finite number of half- spaces which have $\mathbf{0}$ on their boundary [LP86; Wey52]

Proof. Fix $i \in \{1,\dots,n\}$, and choose an event $S = \left\{S_i^i = 1, S_i^{i\ominus 1} = 0\right\}$ such that a multi-maximal coupling of $\mathcal{R}_i$ has, without loss of generality, $\Pr\left(T_i^i = 1, T_i^{i\ominus 1} = 0\right) = 0$.[5] Look at the row u of $\mathbf{M}_{(.)}$ corresponding to $\Pr\left(T_i^i = 1, T_i^{i\ominus 1} = 0\right) = 0$ and let V be the set of indices $j \in \{1,\dots,2^{2n}\}$ such that $\mathbf{M}_{(.),u,j} = 1$. Choose any $v \in V$, and let s_v be the vth component of S. Define $\mathbf{q}^*_{(.)}$ component-wise by taking $\mathbf{q}^*_{(.),i} = \mathbf{p}^*_{(.),i} + \frac{1}{2}CNT_3$ if the event s' whose probability is the ith component of $\mathbf{p}^*_{(.)}$ is contained in s_v, and $\mathbf{q}^*_{(.),i} = \mathbf{p}^*_{(.),i}$, otherwise. Lastly, let

$$\mathcal{H}_v = \left\{\mathbf{x} \in \mathbb{R}^{2^{2n}} : \mathbf{1}^\intercal(\mathbf{x} - \mathbf{e}_v) = 1 + \frac{1}{2}CNT_3 \text{ and } \mathbf{M}_{(.)}(\mathbf{x} - \mathbf{e}_v) = \mathbf{q}^*_{(.)}\right\}, \tag{3.29}$$

where $\mathbf{e}_v$ is the unit vector with a 1 on its vth component, and choose a point $\mathbf{w}^*$ with zero vth component in the intersection of $\mathcal{H}$ and the nonnegative orhtant of $\mathbb{R}^{2^{2n}}$. Clearly, the point $\mathbf{y}^* = \mathbf{w}^* - \frac{1}{2}CNT_3\mathbf{e}_v$ is a solution of task 3.4 with $\mathbf{y}^*_v = -\frac{1}{2}CNT_3$ its sole negative component. □

Lemma 3.1.2. Let $\mathcal{R}_n$ be a contextual cyclic system. Given a solution $\mathbf{y}^*$ of task 3.4 as in Lemma 3.1.1, a solution $\mathbf{z}^*$ of task 3.5 can be constructed such that $|\mathbf{y}^*_i| \geq \mathbf{z}^*_i$, $i = 1,\dots,2^{2n}$, and $||\mathbf{y}^* - \mathbf{z}^*||_1 = nCNT_3$.

Proof. Choose a solution $\mathbf{y}^*$ in accordance to Lemma 3.1.1. Let $\hat{\mathbf{x}}_1 = \mathbf{y}^*_v\mathbf{e}_v$ where v is the index of the only negative component of $\mathbf{y}^*$. Using this v, let s_v and $\mathbf{q}^*_{(.)}$ be defined as in Lemma 3.1.1, and let U be the set of indices $u \in \{1,\dots,12n\}$ such that $\mathbf{M}_{(.),u,v} = 1$. Note that $|U| = 4n$, where there are n indices such that $\mathbf{p}_{(.),u}$ corresponds to $\Pr(R_i^i = r_i^i, R^i_{i\oplus i} = r^i_{i\oplus 1})$, one for each of the n contexts of $\mathcal{R}_n$; another n correspond to $\Pr(T_i^i = r_i^i, T_i^{i\ominus 1} = r_i^{i\ominus 1})$, one per content; and $2n$ correspond to one probability $\Pr(R_j^i = r_j^i)$ for each random variable in the system.

Let $\mathbf{M}_U$ be the submatrix of $\mathbf{M}_{(.)}$ whose rows are indexed by U, and $\mathbf{M}_{U'}$ the matrix with the remaining rows of $\mathbf{M}_{(.)}$. (Note that matrix $\mathbf{M}_U$ is a reduction of matrix $\mathbf{M}_{(.)}$ in the same manner as $\mathbf{M}$, with the event s_v taking the place of the event $\left\{S_i^i = 1, S^i_{i\oplus 1} = 1\right\}_{i=1,\dots,n}$ for its construction, see Ref. [DK16a].) Define $\mathbf{p}^*_U$ and $\mathbf{p}^*_{U'}$ analogously. We can then rewrite $\mathcal{Q}_n$ as the intersection of

$$\mathcal{Q}_U = \left\{\mathbf{x} \in \mathbb{R}^{2^{2n}} : \mathbf{M}_U\mathbf{x} \leq \mathbf{p}^*_U \text{ and } \mathbf{1}^\intercal\mathbf{x} \leq 1\right\}, \tag{3.30}$$

and

$$\mathcal{Q}_{U'} = \left\{\mathbf{x} \in \mathbb{R}^{2^{2n}} : \mathbf{M}_{U'}\mathbf{x} \leq \mathbf{p}^*_{U'} \text{ and } \mathbf{1}^\intercal\mathbf{x} \leq 1\right\}. \tag{3.31}$$

From the definition of $\mathbf{M}$ (see Ref.[DK16a]), we have that the dimension of $\mathcal{Q}_n$ is $4n+1$. Similarly, the dimension of $\mathcal{Q}_U$ is $4n+1$ because it is constructed by a minimal subset of defining inequalities of $\mathcal{Q}_n$.

Define $\mathbf{w}^* = \mathbf{y}^* - \hat{\mathbf{x}}_1$. Since $\mathbf{M}_{(.)}\mathbf{w}^* = \mathbf{q}^*_{(.)}$, $\mathbf{w}^* \notin \mathcal{Q}_n$. Clearly, $\mathbf{w}^* \notin \mathcal{Q}_U$ and $\mathbf{w}^* \in \mathcal{Q}_{U'}$.

[5] If for no $\mathcal{R}_i$, $\Pr\left(T_i^i = 1, T_i^{i\ominus 1} = 0\right) = 0$, replace R_i^c in the system with $1 - R_i^c$ for some i.

Let us next consider the task

find	minimizing	subject to
$\mathbf{x}$	$\mathbf{1}^\intercal\mathbf{x}$	$\mathbf{M}_U(\mathbf{w}^*-\mathbf{x})\leq\mathbf{p}^*_U$
		$\mathbf{M}_{U'}(\mathbf{w}^*-\mathbf{x})\leq\mathbf{p}^*_{U'}$
		$\mathbf{x}\geq 0$
		$\mathbf{1}^\intercal(\mathbf{w}^*-\mathbf{x})\leq 1$
		$\mathbf{e}_v^\intercal\mathbf{x}=0$

. (3.32)

This task must have a solution, since $\mathbf{x}^*=\mathbf{w}^*$ satisfies all its constraints. Additionally, it is evident that the second set of restrictions (those associated with $\mathbf{p}^*_{U'}$) place no restriction to finding the solution because, by construction, $\mathbf{M}_{U'}\mathbf{w}^*=\mathbf{p}^*_{U'}$; hence, any vector $\mathbf{x}\geq 0$ will satisfy that set of inequalities. Further examination of the constraints shows immediately that for any solution $\mathbf{x}^*$, $\mathbf{1}^\intercal\mathbf{x}^*\geq -2\mathbf{y}^*_v$. Similarly, inspecting the constraints associated with $\mathbf{p}^*_U$ reveals that whenever a vector $\mathbf{x}'$ satisfies $\mathbf{M}^\intercal_{U,u}(\mathbf{w}^*-\mathbf{x}')\leq\mathbf{p}^*_{Uu}$, where $\mathbf{p}^*_{Uu}$ is a probability $\Pr(R^i_i=r^i_i,R^i_{i\oplus i}=r^i_{i\oplus 1})$, then $\mathbf{M}^\intercal_{U,t}(\mathbf{w}^*-\mathbf{x}')\leq\mathbf{p}^*_{Ut}$, where $\mathbf{p}^*_{Ut}$ is a probability $\Pr(R^i_i=r^i_i)$ or $\Pr(R^i_{i\oplus 1}=r^i_{i\oplus 1})$ —for the same i in the event corresponding to $\Pr(R^i_i=r^i_i,R^i_{i\oplus i}=r^i_{i\oplus 1})$—, will also be satisfied. An analogous observation can be made when $\mathbf{p}^*_{Uu}$ is a probability $\Pr(T^i_i=r^i_i,T^{i\ominus 1}_i=r^{i\ominus 1}_i)$. Therefore, at most $2n$ of the constraints imposed via matrix $\mathbf{M}_U$ are active in determining the solution space of task (3.32).

Let $\mathbf{M}_w$ and $\mathbf{p}^*_w$ contain the rows and probabilities of $\mathbf{M}_U$ and $\mathbf{p}^*_U$, respectively, corresponding to bunch and connection probabilities. Since the rows of $\mathbf{M}_U$ are linearly independent, so are the rows of $\mathbf{M}_w$, and the latter has full row rank $2n$. Given the considerations in the above paragraph, task (3.32) is equivalent to task

find	minimizing	subject to
$\mathbf{x}$	$\mathbf{1}^\intercal\mathbf{x}$	$\mathbf{M}_w(\mathbf{w}^*-\mathbf{x})\leq\mathbf{p}^*_w$
		$\mathbf{x}\geq 0$
		$\mathbf{e}_v^\intercal\mathbf{x}=0$

. (3.33)

Now, the constraint $\mathbf{e}_v^\intercal\mathbf{x}=0$ can be replaced by a modification of column v of matrix $\mathbf{M}_w$ in which the column is replaced by a vector of zeros. This effectively reduces its rank to $2n-1$. We further note that, in standard form, the constraints for task (3.33) are $\mathbf{M}_w\mathbf{x}\geq\mathbf{M}_w\mathbf{w}^*-\mathbf{p}^*_w$, and the deficiency in rank just introduced implies that there is some row of $\mathbf{M}_w$ that may be safely removed for purposes of finding a solution $\mathbf{x}^*$. Since (assuming the modified matrix) $\mathbf{M}_w\mathbf{x}\geq\mathbf{M}_w\mathbf{w}^*-\mathbf{p}^*_w$ is an underdetermined system with $2n-1$ inequalities, there exists a solution $\mathbf{x}^*$ such that all components of $\mathbf{x}^*$ but $2n-1$ are zero. Therefore, we see that for a solution $\mathbf{x}^*$, $\mathbf{M}_w\mathbf{x}^*=\mathbf{M}_w\mathbf{w}^*-\mathbf{p}^*_w$, and that $\mathbf{1}^\intercal\mathbf{x}^*=\mathbf{1}^\intercal(\mathbf{M}_w\mathbf{w}^*-\mathbf{p}^*_w)=-(2n-1)\mathbf{y}^*_v$. The statement is obtained by noting that task (3.32) is equivalent to maximizing $\mathbf{1}^\intercal(\mathbf{w}^*-\mathbf{x})$ under the same constraints, which is essentially task 3.5. In other words, $\mathbf{z}^*\equiv\mathbf{y}^*-\hat{\mathbf{x}}_1-\mathbf{x}^*$ is an optimal solution to task 3.5. □

Lemma 3.1.3. $\|\mathbf{y}^*-\mathbf{z}^*\|_1=CNTF+CNT_3$

Proof. Choose solutions $\mathbf{y}^*$ and $\mathbf{z}^*$ in accordance to Lemmas 3.1.1 and 3.1.2. Then

$$\begin{aligned}\|\mathbf{y}^*\|_1 &= \|\mathbf{y}^* - \mathbf{z}^* + \mathbf{z}^*\|_1 \\ &= \|\mathbf{y}^* - \mathbf{z}^*\|_1 + \|\mathbf{z}^*\|_1 \\ &= \|\mathbf{y}^* - \mathbf{z}^*\|_1 + 1 - CNTF\end{aligned}$$

where the second line follows by the choice of $\mathbf{y}^*$ and $\mathbf{z}^*$. The statement follows immediately by noting that $\|\mathbf{y}^*\|_1 = 1 + CNT_3$. □

Theorem 3.1.1. *If $\mathcal{R}_n$ is a cyclic system of rank n, then $CNTF(\mathcal{R}_n) = (n-1)CNT_3(\mathcal{R}_n)$*

Proof. The relation in the statement is trivially true for any noncontextual system; hence, assume that $\mathcal{R}_n$ is a contextual cyclic system of rank n. Choose solutions $\mathbf{y}^*$ and $\mathbf{z}^*$ in accordance to Lemmas 3.1.1 and 3.1.2. By Lemma 3.1.2,

$$\|\mathbf{y}^* - \mathbf{z}^*\|_1 = nCNT_3.$$

And from Lemma 3.1.3, it follows that

$$CNTF = (n-1)CNT_3. \tag{3.34}$$

□

Below we illustrate the results of Theorem 3.1.1 and the application of Lemmas 3.1.1 and 3.1.2 in two example cyclic systems of rank 3, illustrated in table 3.6.

R_1^1	R_2^1		c_1
	R_2^2	R_3^2	c_2
R_1^3		R_3^3	c_3
q_1	q_2	q_3	$\mathcal{R}_3$

Table 3.6: Cyclic system of rank 3.

Example 3.1.1 (Consistently connected system)**.** Consider a cyclic system $\mathcal{R}_3$ with joint distributions of bunches $i \in \{1,2,3\}$ given by

	$R_i^i = 0$	$R_i^i = 1$
$R_{i\oplus 1}^i = 0$	$1/8$	$3/8$
$R_{i\oplus 1}^i = 1$	$3/8$	$1/8$

. (3.35)

The system $\mathcal{R}_3$ is consistently connected and can be represented by the vector:

$$\mathbf{p}^{*\mathsf{T}} = (1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/8, 1/8, 1/8, 1/2, 1/2, 1/2).$$

Let $\{\mathbf{e}_j\}_{j=1,\ldots,64}$ be the standard basis of $\mathbb{R}^{64}$, then we can write a solution to task 3.4 with a single negative mass (as in Lemma 3.1.1) in its 14th component:

$$\mathbf{y}^* = \frac{1}{16}\left(3\mathbf{e}_7 - \mathbf{e}_{14} + 2\mathbf{e}_{25} + \mathbf{e}_{26} + 3\mathbf{e}_{31} + 2\mathbf{e}_{34} + \mathbf{e}_{38} + 2\mathbf{e}_{40} + \mathbf{e}_{42} + 2\mathbf{e}_{58}\right).$$

To highlight the dimension of the solution space of task (3.32), this solution can be further re- expressed as a linear combination of the following L_1-orthonormal vectors $\{\hat{\mathbf{x}}_j\}_{j=1,\dots,6}$:

$$\begin{aligned} \hat{\mathbf{x}}_1 &= -\mathbf{e}_{14}, & \hat{\mathbf{x}}_4 &= (\mathbf{e}_{26} + \mathbf{e}_{38} + \mathbf{e}_{42})/3, \\ \hat{\mathbf{x}}_2 &= \mathbf{e}_7, & \hat{\mathbf{x}}_5 &= \mathbf{e}_{31}, \\ \hat{\mathbf{x}}_3 &= (\mathbf{e}_{25} + \mathbf{e}_{40})/2, & \hat{\mathbf{x}}_6 &= (\mathbf{e}_{34} + \mathbf{e}_{58})/2. \end{aligned}$$

In terms of these vectors, we have

$$\mathbf{y}^* = \frac{1}{16}\left(\hat{\mathbf{x}}_1 + 3\hat{\mathbf{x}}_2 + 4\hat{\mathbf{x}}_3 + 3\hat{\mathbf{x}}_4 + 3\hat{\mathbf{x}}_5 + 4\hat{\mathbf{x}}_6\right).$$

Now, we can use the construction in Lemma 3.1.2 to find the point

$$\mathbf{z}^* = \frac{1}{16}\left(0\hat{\mathbf{x}}_1 + 2\hat{\mathbf{x}}_2 + 3\hat{\mathbf{x}}_3 + 2\hat{\mathbf{x}}_4 + 2\hat{\mathbf{x}}_5 + 3\hat{\mathbf{x}}_6\right),$$

which is a solution to task 3.5 to compute $CNTF$. For this system $CNT_3 = 1/8$ and

$$CNTF = \frac{1}{4} = (3-1)CNT_3.$$

Example 3.1.2 (Inconsistently connected system)**.** Consider the system $\mathcal{R}'_3$ in which the distribution of the third bunch of system $\mathcal{R}_3$ from Example 3.1.1 is replaced by

	$R_3^3 = 0$	$R_3^3 = 1$
$R_1^3 = 0$	$1/8$	$7/16$
$R_1^3 = 1$	$3/8$	$1/16$

. (3.36)

The system $\mathcal{R}'_3$ is inconsistently connected and the following vector can represent it—with the entries that differ from the corresponding vector representing system $\mathcal{R}_3$ from Example 3.1.1 in bold:

$$\mathbf{p}^{*\mathsf{T}} = (1/2, 1/2, 1/2, 1/2, \mathbf{7/16}, 1/2, 1/8, 1/8, \mathbf{1/16}, 1/2, 1/2, \mathbf{7/16}).$$

One possible solution $\mathbf{y}^*$ of task 3.4 for system $\mathcal{R}'_3$ can be written as a linear combination of the following L_1-orthonormal vectors $\{\hat{\mathbf{x}}_j\}_{j=1,\dots,6}$:

$$\begin{aligned} \hat{\mathbf{x}}_1 &= -\mathbf{e}_{49}, & \hat{\mathbf{x}}_4 &= (\mathbf{e}_{23} + \mathbf{e}_{39})/2, \\ \hat{\mathbf{x}}_2 &= (\mathbf{e}_7 + \mathbf{e}_{31} + \mathbf{e}_{40})/3, & \hat{\mathbf{x}}_5 &= \mathbf{e}_{25}, \\ \hat{\mathbf{x}}_3 &= \mathbf{e}_{27}, & \hat{\mathbf{x}}_6 &= (\mathbf{e}_{34} + \mathbf{e}_{58})/2, \end{aligned}$$

with

$$\mathbf{y}^* = \frac{1}{16}\left(\hat{\mathbf{x}}_1 + 6\hat{\mathbf{x}}_2 + \hat{\mathbf{x}}_3 + 2\hat{\mathbf{x}}_4 + 2\hat{\mathbf{x}}_5 + 6\hat{\mathbf{x}}_6\right).$$

Similarly to the previous example, use the construction in Lemma 3.1.2 to find the point

$$\mathbf{z}^* = \frac{1}{16}\left(0\hat{\mathbf{x}}_1 + 6\hat{\mathbf{x}}_2 + \hat{\mathbf{x}}_3 + 0\hat{\mathbf{x}}_4 + \hat{\mathbf{x}}_5 + 4\hat{\mathbf{x}}_6\right),$$

which is a solution to task 3.5 to compute $CNTF$. For this system $CNT_3 = 1/8$ and

$$CNTF = \frac{1}{4} = (3-1)CNT_3.$$

3.1.2 Discussion

We can now claim that all the fundamentally different approaches to quantify contextuality applicable to general cyclic systems of random variables currently found in the literature are proportional to each other within the class of cyclic systems. The proportionality relations among the measures are:

$$2CNT_0 = 2CNT_1 = 2CNT_2 = CNTF = (n-1)CNT_3. \tag{3.37}$$

The equality of the first three measures was shown in Ref. [DKC20], the third equality was proved in Ref. [Cer23], and the last equality, in the paper behind the last subsection. It should also be noticed that the hierarchical measure of contextuality proposed in Ref. [CD20] (see also next section) reduces to CNT_2 for cyclic systems; thus, it also satisfies the proportionality to the other measures.

However, as noted in Refs.[DKC20; CD23; Cer23], the relations among these measures are not as simple in other classes of systems of random variables. In Ref. [DKC20], one can find examples of non-cyclic systems for which CNT_1 and CNT_2 are not functions of each other, in the sense that pairs of systems can be found such that one measure obtains different values while another measure remains constant. A class of examples is considered in Ref. [Cer23; CD20] to show the same lack of functional relation between CNT_2 and CNTF outside of cyclic systems. Lastly, in Ref. [CD23], some examples of hypercyclic systems of order higher than 2 — cyclic systems are a special case of this class where order equals 2 — show that in general there is no functional relation among any of the measures here considered.

Note also that the chain of equalities in expression (3.37) involves only measures of degree of contextuality proposed so far in the literature which can be applied to general systems of random variables and that can be described as relaxing at least one of the three constraints required for a system to be noncontextual in CbD: i) That a joint probability distribution can be found; ii) that such joint distribution agrees with the distributions observed within each context; iii) that such joint distribution agrees with the equality across contexts (recalls multi-maximality). Our result completes the theoretical description of the interrelations of those measures in cyclic systems. However, this set of measures does not necessarily exhaust the possible ways to quantify the degree of contextuality; additional measures can always be proposed.

We conclude with some considerations regarding other possible measures, as well as a few immediate — not necessarily complete — observations on how they could relate to the ones discussed previously. We look into two possible avenues. On one hand, we may try to find new measures of contextuality by considering additional aspects of the couplings that could be imposed on the system; on the other hand we may approach the difference between such couplings using a variety of distance functions.

Regarding the first approach, let us recall that CNT_3 deals with quasi-couplings, and that $CNTF$ is associated with sub-normalized couplings. Hence, a possibility that may appear yet to be explored lies in using over-normalized couplings of the system, which means that we deal with a finite mass function with entries summing up to more than 1. Clearly, it should equal 1 for a non-contextual system, and one should aim therefore for the minimization of the excess above 1. However, such a quantity is already found as part of solving task 3.4; it is $\mathbf{y}^*_+$ and a measure of degree of contextuality defined as its excess above 1 equals $\frac{1}{2}CNT_3$.

CNT_2 is instead computed from the deviations of the bunches to the corresponding marginals of couplings S, subject to matching the multi-maximal couplings. CNT_1, on

the other hand, is computed from the deviation of the multi-maximal couplings of the connections to the bunches. CNT_0 is computed from the deviations of both bunches and couplings of connections to the corresponding marginals of couplings S, subject to matching both as much as possible. On this front, as shown in Ref. [DK16a], multi-maximal couplings are entirely characterized by pairwise joint distributions of the binary random variables in the system, whereas bunches admit a hierarchical description of joint distributions of k-tuples of random variables, as explained in Ref. [CD20]. This property is exploited to define the *hierarchical* version of CNT_2. It appears to us that hierarchical analogues of CNT_3 and CNTF could readily be defined by modifying *mutatis mutandis* the sets of tasks presented in Ref. [CD20][6]. Naturally, such new hierarchical measures coincide with the regular measures when applied to cyclic systems; hence satisfying the same relations already outlined above.

Another possibility is to consider additional measures that capitalize on features of contextuality different to those afforded within the Contextuality-by-Default framework. For instance, Contextual Fraction was originally proposed in the Sheaf-Theoretic approach to contextuality. Its translation into the framework of CbD is straightforward (see Ref. [Dzh23; AD19a]) and produces a generally applicable measure. Other measures within alternative frameworks may be available or could be proposed such that their translation is different to any of the measures discussed here.

As for the second approach, note that the distance used in the definition of all the discussed measures is defined using the L_1 norm. As discussed in Ref. [KD19], L_1 may be preferable given the additivity of probabilities; however, different distances can always be chosen. If a different p-norm is used to define a distance, the measures analogous to CNT_0, CNT_1, and CNT_2, are also equal to each other and proportional to the L_1 based measures with constants $n^{\frac{p-1}{p}}$ as presented in Ref. [DKC20]. Similarly, families of measures associated with CNT_3 and $CNTF$ could be derived by using other p-norms. By properties of p-norms, the same constants will bound the respective measures. It remains to be seen, however, if equality would also be attained for these measures.

In addition to using p-norm derived distances, there are several alternative distances, as well as several functions satisfying fewer requirements than that of a metric, that have been proposed to compare probability distributions and that could be used to define further analogous measures by relaxing some of the three defining constraints. For instance, in Ref. [Gru+14], the Kullback–Leibler divergence is used to define measures wherein the requirement to agree with the distributions within each context is relaxed. Future work should address in what manner these relative entropy measures could be translated to the CbD framework and generalized to systems with disturbance, and whether they are functionally related to the measures here considered for some class of systems, including cyclic systems.

Recently, contextuality has been regarded as a quantum resource (see e.g., Ref. [Ama19]). Moreover, within the framework of resource theory, some measures of degree of contextuality have been identified as quantifiers of such resource; that is, the measures satisfy certain properties of interest, such as monotonicity, with respect to some transformations of the systems of random variables. The fact that measures of contextuality are not generally functions of each other, as exemplified by the measures considered in our result, highlights that the relations between different systems with respect to contextuality is complex and that each measure may only reveal partial information about it which gets complemented by other quantifiers. On the other hand, the existence of functional

[6] Without number on page 4 between expressions (17) and (18).

relations among these quantifiers within the class of cyclic scenarios suggests that the structure of (non)contextual resources may be simple for those scenarios. This reinforces and motivates the need for better understanding the patterns of contextuality within different scenarios to shed light into their structure and how they partition into contextual and non- contextual systems.

3.2 Non-contextuality quantifiers

The intention behind this final section of the chapter is to address the question of also being able to quantify the degree of non-contextuality of a given system. We have been focusing on characterizing the protagonist phenomenon in this writing, but it should be fair and natural to consider a similar treatment of non- contextual systems as well. Not with the goal of exploring their peculiarities, of course; after all, they constitute the standards that contextual instances seem to challenge. Instead, the aim is to extend known quantifiers to refine them or even filter them by checking their agreement and consistency with the standards.

In this sense, still in [KD19] it has been suggested that CNT_2 can achieve this. We will call its non-contextuality counterpart as $NCNT_2$, following the same reference. It is defined as:

$$NCNT_2(\mathbf{p}) = \inf_{\bar{\mathbf{p}}_\mathbf{b} \notin \mathbb{P}_\mathbf{b}} \left|\mathbf{p}_\mathbf{b} - \bar{\mathbf{p}}_\mathbf{b}\right|_1 = \min_{\bar{\mathbf{p}}_\mathbf{b} \in \partial\mathbb{P}_\mathbf{b}} \left|\mathbf{p}_\mathbf{b} - \bar{\mathbf{p}}_\mathbf{b}\right|_1, \tag{3.38}$$

where inf stands for the infimum[7] of the set of values accessed by the expression and $\partial\mathbb{P}_\mathbf{b}$ stands for the boundary of the convex polytope $\mathbb{P}_\mathbf{b}$ (recall (3.16)).

As stated in [KD19], the measure is well defined, in the sense of satisfying the conditions to be a proper quantifier. Noticing that, for $\mathbf{p}$ non-contextual, it lies within the polytope $\mathbb{P}_\mathbf{b}$ by construction, the quantifier can be interpreted as how far the bunches of this behavior are from those not compatible with the multi-maximal coupling (s, maybe, if for non binary random variables) of its connections. To compute it, a similar result (see [DKC20]) to the one leading the authors to the relation (3.25) is applied. That is, for a $\mathbf{p}_\mathbf{b}$ interior to $\mathbb{P}_\mathbf{b}$, its distance to the boundary of the polytope is a single coordinate L_1-distance too. So we just need to increase or decrease the probabilities in each $\mathbf{p}^{(c)}$ one by one as far as we can without leaving the polytope, and to choose the smallest change at the end.

So, enumerating each element of $\mathbf{p}_\mathbf{b}$ from i =1, to say, K, for each i one computes the LP tasks shown in table 3.7 and saves the results of d_i^+ (the positive change required for i to reach the boundary) and d_i^- (the analogous negative change). The vector $\mathbf{e}_\mathrm{i}$ is the unitary one with entry 1 in the ith component[8]. The measure is then expressed through the equation below:

$$NCNT_2(\mathbf{p}) = \min_{\mathrm{i}=1,\ldots,K} \left\{\min\left(d_\mathrm{i}^+, d_\mathrm{i}^-\right)\right\} \ . \tag{3.39}$$

[7] The infimum a of a subset S of a partially ordered set P is the greatest lower bound in P to all the elements of S, i.e. $a \leq s \ \forall s \in S$ with $a \in P$. A partially ordered set P is a set aimed with a binary operation (i.e. with two inputs) $\leq$ that tells us that, for some pairs of elements in P, one element comes before the other in the order. In this case P can be seen as the non-negative real numbers and S as the set of values of the expression determined by all $\bar{\mathbf{p}}_\mathbf{b}$ outside the polytope $\mathbb{P}_\mathbf{b}$.

[8] It could be formulated with vectors $\mathbf{d}_\mathrm{i}^+$ and $\mathbf{d}_\mathrm{i-}$ instead, but this would mean more variables within the program, which for larger systems can make the computation harder.

Find	$\mathbf{x}$, d_{i}^{+}
maximizing	d_{i}^{+}
such that	$\mathbf{p_b} + d_{\mathrm{i}}^{+}\mathbf{e}_{\mathrm{i}} = \mathbf{M}\mathbf{x}$,
	$\mathbf{x} \geq 0$, $d_{\mathrm{i}}^{+} \geq 0$,
	$\mathbf{p_l} = \mathbf{M_l}\mathbf{x}$, $\mathbf{p_c} = \mathbf{M_c}\mathbf{x}$

,

Find	$\mathbf{x}$, d_{i}^{-}
maximizing	d_{i}^{-}
such that	$\mathbf{p_b} - d_{\mathrm{i}}^{-}\mathbf{e}_{\mathrm{i}} = \mathbf{M}\mathbf{x}$,
	$\mathbf{x} \geq 0$, $d_{\mathrm{i}}^{-} \geq 0$,
	$\mathbf{p_l} = \mathbf{M_l}\mathbf{x}$, $\mathbf{p_c} = \mathbf{M_c}\mathbf{x}$

.

Table 3.7: $NCNT_2$ LP tasks

In spite of the intuitive connection between those quantities, we see that equations (3.17) and (3.38) do not have the same structure. However, both CNT_2 and its counterpart $NCNT_2$ have been glued in a more recent work through the elaboration of a more sophisticated quantifier; the term "glued" is insightful and very suitable, since the distances will share a common surface as reference. The new measure that goes by the name of *Hierarchic Measure of Contextuality*[9] (first appeared in [CD20]) not only link the previous quantities, but also provides a hierarchy to catalog probabilistic systems beyond cyclic ones. For details we refer the reader to the previous citation in this paragraph, but a general idea can be given here. The reduced vector representation we have seen some sections behind can be modified to include marginals of higher orders for systems containing contexts with more than two random variables on them. And in the same reasoning as before, vectors $\mathbf{p_s}$ and matrices $\mathbf{M_s}$ can then be defined for marginals of order $s > 2$ for bunches and connections besides the older ones. Non-contextual systems will keep satisfying equation (3.11), but with different degrees of non-contextuality depending on the order of the bunch marginals considered. Conversely, contextual systems will violate that equation for some marginals; more specifically for all $s \leq s^*$ for some s^*. This allow us to establish levels s of degree of contextuality and non-contextuality for all systems, or their hierarchic positions.

To conclude this enriching section and the chapter, we address the unsettling question in the mind of the reader: What about the other measures? The exposition we have just seen was planned in this way because, in fact, none of the other candidates we have discussed before can be explored using similar ideas.

Starting with CNT_1, an analogous construction as in equation (3.38) is possible, but useless. This is the case because, if $\mathbf{p_c}$ was in the interior of $\mathbb{P}$, it would mean that the collection of $\mathbf{p}^{(q)}$'s was not a probability distribution of a multi-maximal coupling for the connections in the first place, after all they could be increased without leaving the polytope delimited by the non-contextual relation (3.11) (for fixed $\mathbf{p_l}$ and $\mathbf{p_b}$).

For CNT_3 the reason is simpler. A system is non-contextual under the existence of a coupling $\mathbf{x}$ such that $\mathbf{M}\mathbf{x} = \mathbf{p}$. It follows then that $CNT_3(\mathbf{p}) = |\mathbf{x}|_1 - 1 = 0$, for all non-contextual systems; as pointed in [KD19], it seems impossible to extend it without altering its logic. This is also the case for $CNTF$. Notice that it does not provide a way for distinguishing between non-contextual instances, since all those systems have zero for it. In comparison to the previous measures, the problem with these ones relies in the fact that the distances involve couplings and quasi-couplings alone, without comparing them with similar objects. Or yet, they look for objects always within a polytope, regardless they position in relation to that set.

[9] The hierarchy here has nothing to do with SDP hierarchies previously presented.

Chapter 4

Non-locality quantifiers for states

In this chapter we apply some of the ideas that we have seen before to the field of quantum Physics and its applications. More precisely, the work to be presented deals with the extension of quantifiers for the phenomenon we have been talking about within probabilistic systems to measures of the same resource in a different object in that theory: states. Most of what one is going to see below was taken from the article that contains results obtained by the author of this thesis in collaboration with his advisor (Bárbara) and colleagues (Ari and Fernando); it can be accessed online through [Pat+22].

Our starting point is the clarification of a strange word in the title: *non-locality*. We have come across this term in the first chapter, but now we are going to give the proper attention it deserves. For all intents and purposes, it is just a special name for special cases of contextuality used by physicists. It applies to physical situations in which signalling between parties is not allowed or can not influence the other party on time, because for example the distance between them[1], therefore becoming the reason behind the compatibility of measurements realized in each of their partial systems. Scenarios carrying such interpretation go by the generic name of *Bell scenarios*; it is common to label them by the number m of parties, the amount of measurements l each party has and also how many outcomes d each measurement has: (m, l, d).

Going a little bit further in this qualitative perspective, *quantum non-locality*, from now on, is a manifestation of the eccentric nature of quantum theory as seen by a classical viewer, which means an observer used to the dynamic of the macroscopic world it lives in. Also known as *Bell non-locality* (after [Bel64]), this phenomenon is seen as a consequence of entanglement. In plain language, it means that the statistics provided by some distant measurements performed in sides of a shared entangled quantum state can not be explained by local hidden variable models.

Although entanglement is a necessary ingredient for having non-locality, they refer to different resources[2]([Bru+14]). That is because there are entangled states that can only produce local correlations ([Wer89]), i.e. joint probabilities in accordance with such local models. This motivates an independent study of non-locality as a resource.

We will consider a bipartite Bell scenario where two distant parts, Alice and Bob, share a possibly correlated pair of physical systems, in which they perform measurements, obtaining measurement outcomes. Alice and Bob are not allowed to communicate and the choices of inputs are assumed to be independent. Moreover, measurements will

[1] Following a relativistic point of view and the limitation of the light velocity, as we have discussed in section 1.4.

[2] In the sense of resource theory we have already commented about in the first chapter.

not receive different labels for different contexts they are measured in, i.e. we assume consistent connectedness. The reader should then expect to see the traditional groundwork of contextuality in what follows.

Preceding chapters taught us how to look at the behaviors generated by systems like these. However, it is also interesting to focus on the state shared by Alice and Bob. This is related to the idea of how much non-locality can be "extracted" from that state. In this sense, in what follows we want to investigate how non-locality quantifiers for behaviors can be used to build non-locality quantifiers for states, ending with the proposition of a new one. We also give attention to the phenomenon known as *anomaly of non-locality* ([Aci+02]), in special the *weak anomaly* observed in [Ros+17].

In order to do this, we first give a geometric overview of the sets of correlations of interest. Besides giving intuition, this also motivates the quantifier defined in our work. We show how such object can be constructed and what are the properties it has. Then we apply it to the study of the aforementioned anomalies.

4.1 Geometry of the set of Bell correlations

We consider a bipartite Bell scenario in which Alice and Bob perform measurements labeled by variables A_x and B_y, obtaining measurement outcomes described by variables a and b, respectively, as shown in fig. 4.1. The description of Alice and Bob's outcomes is given by a set of probability distributions $p(a, b|x, y)$, or *behavior*, that gives the probability of outcomes a and b given inputs x and y. The set of behaviors under consideration depends on the physical assumptions we make about the system Alice and Bob share.

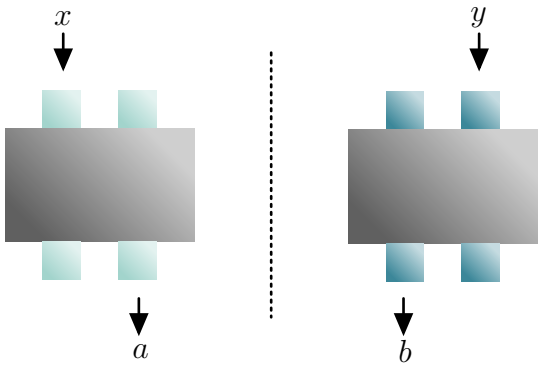


Figure 4.1: Alice and Bob have each a measurement device with inputs x and y and outputs a and b. The device is described by the set of probability distributions $p\,(a, b|x, y)$, that gives the probability of outcomes a and b given inputs x and y.

The fact that Alice and Bob are spatially separated and cannot communicate with each other implies that the statistics of a measurement on one part is independent of the measurement choice of the other. We recognize these assumptions as the non-disturbance conditions described in the first chapter (see (1.18)), but here we will give it a more suitable name due to the circumstances: *non-signalling conditions*. It implies a set of linear constraints as below:

$$\begin{aligned} p(a|x) &= \sum_b p(a, b|x, y) = \sum_b p(a, b|x, y') \\ p(b|y) &= \sum_a p(a, b|x, y) = \sum_a p(a, b|x', y) \end{aligned} . \tag{4.1}$$

A stronger constraint on the description of the experiment is that the statistics of Alice and Bob be consistent with the assumption of local causality. This can be understood in two steps. First we assume that there is a common past of configurations of the system for the events occurring at the distant parties, encoded in the form of a probability distribution for deterministic states parameterized by λ. This means we should write:

$$p(a,b|x,y) = \sum_{\lambda} p(a,b|x,y,\lambda) \ .$$

Now, the second step is a consequence of the distance between the parties. For the same reason that motivates the introduction of a common cause in the past, i.e. a relativistic point of view of the physical situation, local causality here means that outcomes from one side cannot cause any change in the data of the other. This implies that the quantities in the r.h.s of the equation above can be decomposed:

$$p(a,b|x,y,\lambda) = p(\lambda)p(a|x,\lambda)p(b|y,\lambda) \ .$$

Such that the joint probabilities we are interested in read:

$$p(a,b|x,y) = \sum_{\lambda} p(\lambda)p(a|x,\lambda)p(b|y,\lambda) \ . \tag{4.2}$$

It should be noted the resemblance between this equation and (1.15) we have seen before.

For this type of behavior, correlations between Alice and Bob are mediated by the variable λ that thus suffices to compute the probabilities of each of the outcomes, that is, $p(a|x,y,b,\lambda) = p(a|x,\lambda)$, and similarly for b. The behaviors that can be decomposed in this way are called *local behaviors*. And, in accordance with our first chapter, models describing systems like these are also called *local hidden variable models*.

The result known as Bell's theorem (see [Bel64], also source of the nomenclature just above) states that Alice and Bob can perform measurements in a entangled quantum state to generate behaviors that cannot be decomposed in the form of equation (4.2). These can be obtained by local measurements M_a^x and M_b^y on distant parts of a bipartite state ρ that, according to quantum theory, can be described by

$$p(a,b|x,y) = \mathrm{Tr}\left[(M_a^x \otimes M_b^y)\,\rho\right] \ , \tag{4.3}$$

the famous Bohr's rule for quantum predictions.

In general, the set of local behaviors $\mathcal{L}$ is a strict subset of the set of quantum behaviors $\mathcal{Q}$ that, in turn, is a strict subset of the set of non-signaling behaviors $\mathcal{NS}$, as shown in figure 4.2.

The local set is a polytope, whose vertices are deterministic systems labeled by λ, and hence any local behavior can be written as a convex sum of a finite set of points at the boundary. If we represent our behavior $\mathbf{p}$ as a vector with $|x||y||a||b|$ components, where $|.|$ indicates cardinality of the collection of possible values, so condition (4.2) can be written succinctly as

$$\mathbf{p} = A \cdot \lambda \ ,$$

with λ being a probability vector over the set of variables λ, with components $\lambda_i = p(\lambda = i)$, and A being a matrix indexed by i and the multi-index variable $j = (x,y,a,b)$ with $A_{j,i} = \delta_{a,f_a(x,\lambda=i)}\delta_{b,f_b(y,\lambda=i)}$, where f_a and f_b are deterministic functions that give the values of measurements x and y when $\lambda = i$.

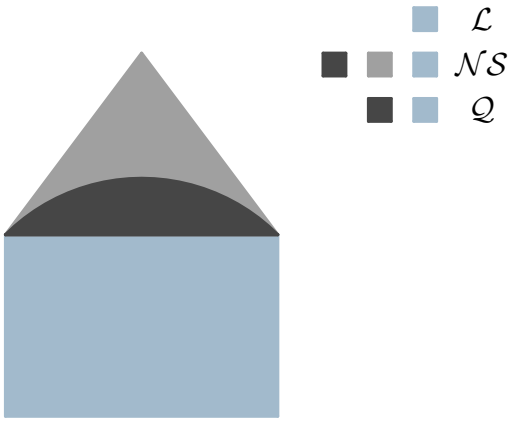


Figure 4.2: Typical structure of sets of behaviors of interest.

Notice that, in the language of our last chapter, $\mathbf{p}$ is the complete description of a probabilistic system (see equation (3.4)) in which information about connections can be ignored, because they are not necessary as pointed before.

It is now clear how the content of chapter 2 will be useful to us here. Checking membership of a given behavior in the local polytope $\mathcal{L}$ is part of the work of the quantifiers we have seen. Of particular importance to what we are going to see in the following sections are the measures we have called CNT0 and CNT2 before.

The first one, see equation (3.1), made use of non-contextuality inequalities (now Bell inequalities) and looked at violations of them as criterion. The great problem with it was the fact that different inequalities in the same scenario provide different quantifiers and a proper characterization should then take into account all of them. Now, CNT2 (equation (3.17)) was one of those that made use of linear programming to find the best L_1-approximating local member of $\mathcal{L}$ to the target behavior.

Instead of using equation (3.17) directly, however, we modify it a little bit into a style more appropriate to the physical situation here. In essence, we are going to put a normalization constant in front of the objective function of the LP task in 3.3 to give it a meaningful interpretation in terms of an existent definition in physics: the *trace distance* (see [BAC18]) $\mathbf{D}(\mathbf{q}, \mathbf{p})$ between two probability distributions $\mathbf{q}$ and $\mathbf{p}$:

$$\mathbf{D}(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \sum_x |q(x) - p(x)| \ , \tag{4.4}$$

where $|.|$ stands for absolute value of the number inside and x labels the entries of the vectors in question.

For the systems we are interested in, the vectors are joint probabilities with entries labeled by the values of inputs x and y and outputs a and b, as described some paragraphs above. Under the consideration of uniform distribution of the inputs, i.e.

$$p(x, y) = \frac{1}{|x||y|} \ ,$$

with $|.|$ the cardinality again, the objective function of the mentioned LP task can be modified in order to the optimal condition be achieved by the local behavior that minimizes the trace distance to the target behavior. That is, we have a non-locality quantifier

for behaviors similar to equation (3.17):

$$\begin{aligned} \mathrm{NL}(\mathbf{q}) &= \frac{1}{|x||y|} \min_{\mathbf{p}\in\mathcal{L}} \mathbf{D}(\mathbf{q},\mathbf{p}) \\ &= \frac{1}{2|x||y|} \min_{\mathbf{p}\in\mathcal{L}} \sum_{x,y,a,b} \left| q(a,b|x,y) - p(a,b|x,y) \right| \end{aligned} . \tag{4.5}$$

This can be implemented either by changing the task 3.3 to have the correct normalization or with the expression above if the packages in your favourite code language offers support for objective functions containing L_1-norms of vectors with possible negative entries. The visual aspect of this procedure is illustrated in figure 4.3

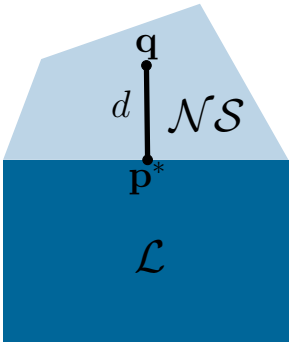


Figure 4.3: Schematic drawing of a distribution $\mathbf{q} \in \mathcal{NS}$ and $d = \mathrm{NL}(\mathbf{q})$, the trace distance from $\mathbf{q}$ to the closest local distribution $\mathbf{p}^* \in \mathcal{L}$.

4.2 Quantifying non-locality of a quantum state

We now address the problem of quantifying non-locality of a quantum state ρ. Since from ρ we can generate many different non-local behaviors by varying the measurements applied to it, this is not a trivial problem.

One way of defining a non-locality quantifier for quantum states is to maximize the degree of violation of a Bell inequality over all possible measurements for Alice and Bob and to associate a greater numerical violation with a greater degree of non-locality. This association has generated some debatable conclusions. For example, using the usual measure, the so called *anomaly of non-locality* appears ([MS06],[Aci+02]). Consider the Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality in the $(2,2,3)$ scenario ([Col+02]):

$$\begin{aligned} S_{\mathrm{CGLMP}} &= p\left(a=b|0,0\right) + p\left(a=b|0,1\right) + p\left(a=b|1,0\right) + p\left(a=b+2|1,1\right) \\ &\quad - p\left(a=b+1|0,0\right) - p\left(a=b+2|0,1\right) - p\left(a=b+2|1,0\right) - p\left(a=b|1,1\right) \\ &\leq 2 \end{aligned} \tag{4.6}$$

and a two qutrit system in state

$$|\psi\left(\gamma\right)\rangle = \frac{1}{\sqrt{2+\gamma^2}} \left(|00\rangle + \gamma|11\rangle + |22\rangle\right) . \tag{4.7}$$

For the maximally entangled state $\gamma = 1$, the best choice of measurements gives $S_{\mathrm{CGLMP}} = \frac{4(2\sqrt{3}+3)}{9} \simeq 2.873$ [Col+02]. However, the authors in [Aci+02] found that, for

the very same choice of settings, another state gives a higher violation. Specifically, the violation $S_{\text{CGLMP}} = 1 + \sqrt{\frac{11}{3}} \simeq 2.915$ is obtained for the non-maximally entangled state with $\gamma = \frac{\sqrt{11}-\sqrt{3}}{2} \simeq 0.792$. This fact is known as *anomaly of non-locality*.

Besides the non-expected feature of anomaly of non-locality, a quantifier like that clearly inherits the problem of choice of an inequality in its definition. Now, although keeping the same problem, in reference [FP15], the authors present an alternative measure to quantify non-locality, called *volume of violation*, that is an important step into the investigation. While the previous measure takes only the settings that lead to the maximum violation, the volume of violation takes into account all the settings that produce violation of a Bell inequality. For the calculation of this new quantity, for a particular state, we make an integration in the region that leads to the violation of a fixed Bell inequality [MS06]. In general, we can write

$$V_I(\rho) = \frac{1}{V_T} \int_{\Gamma} d^n x \ ,$$

where Γ is the set of measurement choices for Alice and Bob that lead to a violation of the inequality I for state ρ and

$$V_T = \int_{\Lambda} d^n x \ ,$$

is the volume of the set Λ of all possible measurement choices for Alice and Bob. Note that $d^n x$ will display the format that gives equal weights to any setting. Thus, state ρ is more non-local than state σ if and only if $V_I(\rho) > V_I(\sigma)$. Also, if for state ρ the volume of violation is $V_I(\rho) = 0$, we say that state ρ is local with respect to the given inequality. Following the same reasoning, $V_I(\rho) = 1$ indicates that ρ is maximally non-local with respect to that inequality.

This measure uses the relative volume of measurement choices that lead to violation of a particular Bell inequality to quantify non-locality. Hence, $V_I(\rho)$ has a direct interpretation: it corresponds to the probability of violating a particular Bell inequality with state ρ when the measurement configuration is chosen randomly.

The volume of violation is a measure of non-locality for quantum states with many good properties, as already reported in reference [FP15]. Nevertheless, $V_I(\rho)$ does not take into account all the measurement configurations that lead to non-local behaviors, but only the ones that lead to non-local behaviors that violate a particular Bell inequality. Except for the simplest scenario $(2, 2, 2)$, there are many non-equivalent families of Bell inequalities and hence $V_I(\rho)$ gives only a lower bound to the relative volume of measurement choices that lead to non-locality.

In reference [Ros+17] the authors consider a modification of the volume of violation, called *non-local volume*, replacing the violation of a Bell inequality by membership of the corresponding behavior in the polytope of local correlations. Hence, the non-local volume takes into account all the settings that produce a non-local behavior, which is in general strictly larger than the set of behaviors violating a single Bell inequality. For the calculation of this new quantity, for a particular state, we make an integration in the region Δ of the set of measurement setups that lead to a non-local behavior. In general, we can write

$$V(\rho) = \frac{1}{V_T} \int_{\Delta} d^n x \ ,$$

where V_T is defined as before. Again we consider an integration that gives equal weights to any setting.

In this contribution we consider another option in which we integrate over the set of measurement choices that give a non-local behavior but using a non-locality quantifier Q defined in the set of behaviors as a weight in the integral:

$$V_Q(\rho) = \frac{1}{V_T} \int_\Delta Q(x) d^n x \ , \tag{4.8}$$

where $Q(x)$ is the quantifier Q for the behavior generated by the state and measurement settings x and Δ and V_T are defined as before. We consider only *faithful* quantifiers: $Q(\mathbf{p}) > 0$ if and only if $\mathbf{p} \notin \mathcal{L}$.

This new quantifier, that we call *Q-weighted non-local volume*, can be interpreted in a similar way. As the non-local volume, it takes into account all the settings that produce a non-local behavior for state ρ, but it sums with a higher weight the behaviors that are more non-local according to the quantifier Q. In particular, we are interested in the non-locality quantifier for states obtained when we choose $Q = \mathrm{NL}$:

$$V_{NL}(\rho) = \frac{1}{V_T} \int_\Delta NL(x) d^n x \ , \tag{4.9}$$

where $NL(x)$ is the trace distance for the behavior generated by the state and measurement settings x and Δ and V_T are defined as before. We call this quantifier *trace-weighted non-local volume*.

4.3 Properties of the trace-weighted non-local volume

In reference [Lip+18], the authors show that the non-local volume is invariant under local unitaries and that it is strictly positive for pure entangled bipartite states. The proofs can be slightly modified to show that these properties are also true for V_Q.

Theorem 4.3.1. *V_Q is invariant under local unitaries.*

Proof. Let $\rho' = U_1 \otimes U_2 \rho U_1^\dagger \otimes U_2^\dagger$ where U_1 and U_2 are local unitaries for subsystems 1 and 2, respectively. The behavior generated with measurements $\{M_\mathrm{i}\}$ and $\{N_\mathrm{i}\}$ and state ρ is equal to the behavior generated with measurements $\left\{U_1^\dagger M_\mathrm{i} U_1\right\}$ and $\left\{U_2^\dagger N_\mathrm{i} U_2\right\}$ and state ρ'. Hence, the sets Δ are the same for ρ and ρ', which implies $V_Q(\rho) = V_Q(\rho')$. □

Theorem 4.3.2. *If Q is a continuous function, for all pure bipartite entangled states, in a scenario with at least two choices of two-outcome measurements, V_Q is strictly positive. That is, $V_Q(|\psi\rangle) = 0$ if and only if $|\psi\rangle$ is a product state.*

Proof. Since $|\psi\rangle$ is entangled, we know from [Gis91], that there exist measurements $\{M_\mathrm{i}\}$, $\{N_j\}$ in the simplest scenario $(2, 2, 2)$ such that the corresponding behavior is non-local. By continuity of the probabilities $p(a, b|x, y)$ as a function of the measurements, and continuity of Q as a function of $\mathbf{p}$, there is a ball around $\{M_\mathrm{i}\}, \{N_j\}$ such that, for any choice of measurements inside this ball, Q is strictly positive. Since we are integrating a strictly positive function over a set that is of measure larger than zero, this implies that $V_Q(|\psi\rangle) > 0$.

On the other hand, if $|\psi\rangle$ is separable, every behavior generated with $|\psi\rangle$ is local, and hence $Q(x) = 0$ for every x and $V_Q(|\psi\rangle) = 0$. □

It is also easy to see that the Q-weighted non-local volume is always smaller than the non-weighted version for any faithful quantifier Q that is normalized such that $0 \leq Q \leq 1$. In fact, if Q is faithful, we can write

$$V_Q(\rho) = \frac{1}{V_T} \int_\Lambda Q(x) d^n x \ ,$$

and

$$V(\rho) = \frac{1}{V_T} \int_\Lambda \chi(x) d^n x \ ,$$

where χ is the characteristic function of the non-local set, that is:

$$\chi(\mathbf{p}) = \begin{cases} 1, & \text{if } \mathbf{p} \notin \mathcal{L}, \\ 0, & \text{if } \mathbf{p} \in \mathcal{L} \end{cases} .$$

If $\mathbf{p}$ is local and Q is faithful, $Q(\mathbf{p}) = \chi(\mathbf{p}) = 0$. If $\mathbf{p}$ is non-local, $Q(\mathbf{p}) \leq 1 = \chi(\mathbf{p})$, which implies that $Q(\mathbf{p}) \leq \chi(\mathbf{p})$ for all $\mathbf{p}$ and hence $V_Q(\rho) \leq V(\rho)$ for all ρ.

Reference [Lip+18] also shows that the non-local volume goes to 1 in the limit where both parties have an infinite number of measurements. The proof in this case can not be modified to show that this is also true for V_Q. Numerical results in the figure 4.4 indicate that if we choose $Q = \mathrm{NL}$, V_{NL} seems to increase monotonically with the number of measurements n, which would be the desired behavior. However, it is not possible to claim its limit as n goes to infinity is 1 based on this evidence only. Since V_Q is always smaller than the non-local volume, we can not discard the possibility that V_Q goes to a as n goes to infinity, with $0 < a < 1$. Whether or not this property holds for V_{NL} is an open problem.

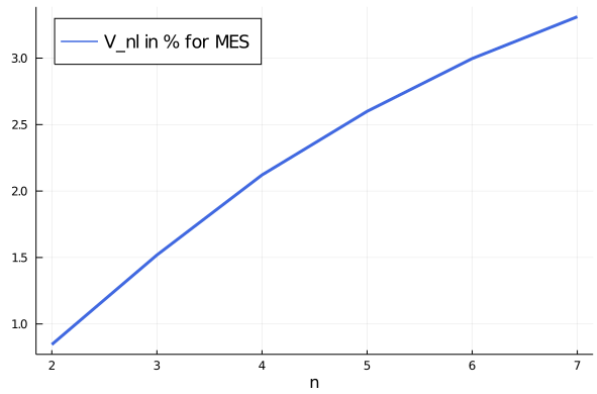


Figure 4.4: Trace-weighted non-local volume in percentage for the maximally entangled two-qubit state as function of the number of measurements detained by each part in the Bell scenario, i.e scenarios of the type $(2, n, 2)$.

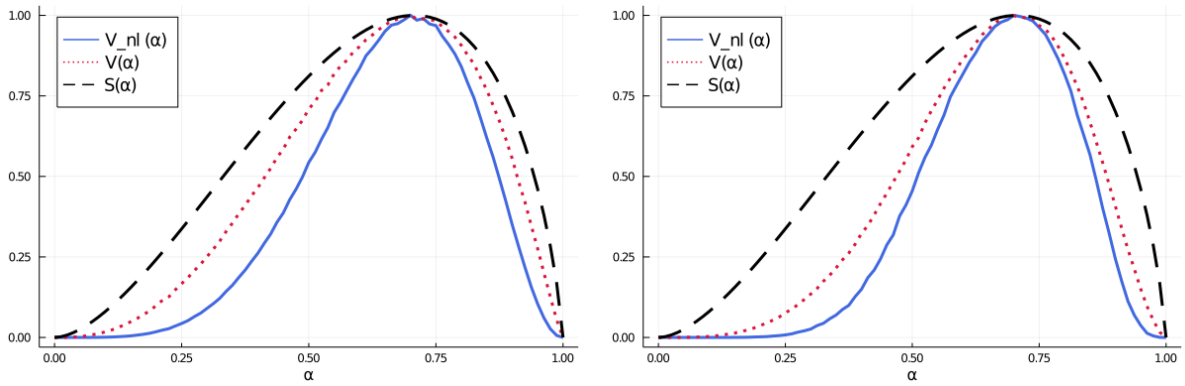
4.4 Anomalies through new lens

In this section we want to revisit known scenarios in order not only to check the functionality of our quantifier V_{NL}, but mainly to show its usefulness to scrutinize the anomalies. Behaviors are decided to be non-local in our simulations for trace distances with magnitude smaller than 10^{-8}.

To start, we check the effect of the weight on the non-local volume for the simplest scenario, the $(2, 2, 2)$ scenario, also known as the CHSH scenario ([Cla+69]). The local set in this case has a simple structure, since there is only one family of Bell inequalities, and we expect that the plots for the entropy of entanglement, the non-local volume, and the Q-weighted non-local volume all have the same comportment if Q is a continuous faithful non-locality quantifier for behaviors. To test that we consider the family of states parameterized by:

$$|\psi(\alpha)\rangle = \alpha|00\rangle + \sqrt{1-\alpha^2}|11\rangle$$

and plot the non-local volume (red) and the trace-weighted non-local volume (blue) in figure 4.5a. Both for the non-local volume and the trace-weighted non-local volume the maximum non-locality is attained at the maximally entangled state. We can also see that the weighted version is smaller or equal the non-weighted version considering a normalization using the maximum value for each curve.



(a) (b)

Figure 4.5: In blue our quantifier lies below its non-weighted version in red, as expected, while maintaining the point of maximum, which coincides with the maximum of the dashed curve representing the entanglement entropy. In (a) for the CHSH scenario and in (b) for the 3322 scenario.

The next interesting scenario, the $(2, 3, 2)$ scenario, is not as trivial as the CHSH scenario. In figure 4.5b the weighted and the non-weighted version are compared for the qubits states with parameters in the same way.

We also compared V_{NL} and the entropy of entanglement for the family of states given in (4.7), see figure 4.6. That is, for the scenario of qutrits. For this family of states, we see that the maximum non-locality is achieved by the maximally entangled state, as in the previous scenarios with qubits. The same behavior is observed for the non-weighted version, as shown in reference [Ros+17]. But in general V_{NL} is not a monotonic function of the entropy of entanglement. This property is known as *weak anomaly of non-locality*.

Still in the same system of qutrits, for the GHZ(α) states parameterized as:

$$\text{GHZ}(\alpha) = \sin(\alpha)|00\rangle + \frac{\cos(\alpha)}{\sqrt{2}}\left(|11\rangle + |22\rangle\right) \ , \tag{4.10}$$

non-locality does not increase monotonically with α. As in the previous cases, it reaches the peak at the maximally entangled state, but it has a local minimum around 6 degrees, similar to what is observed in the non-weighted version. Figure 4.7a shows the weighted

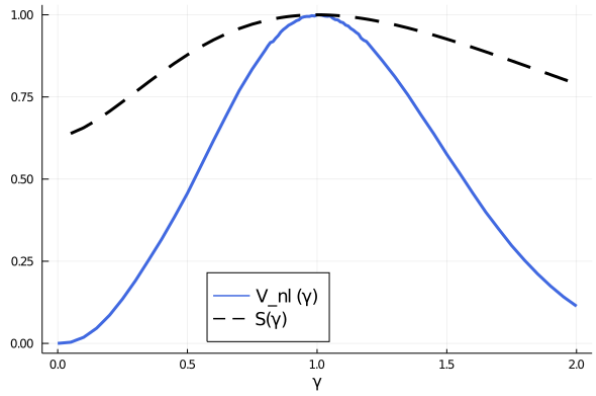
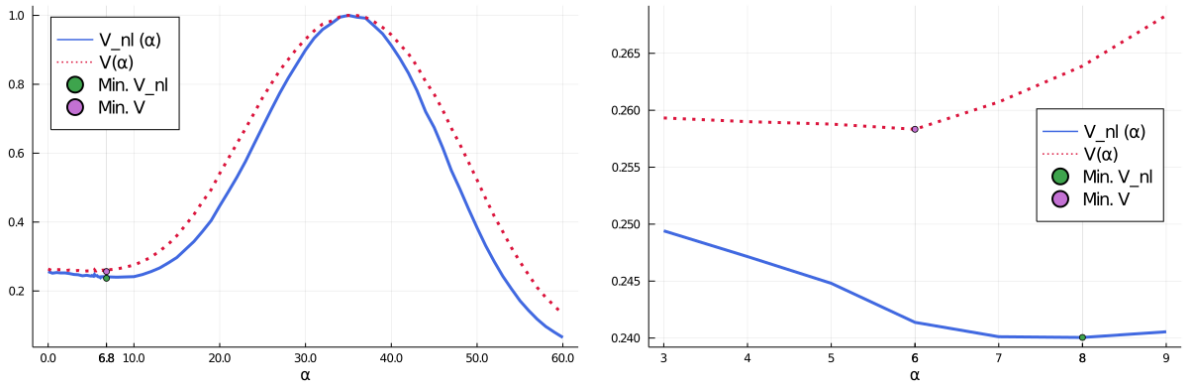


Figure 4.6: The maximum of the trace-weighted non-local volume (in blue) is reached at the state exhibiting the maximum entanglement entropy (in black).

and non-weighted versions of non-local volume as a function of α; figure 4.7b gives a zoom on what is happening very close to the local minimum, emphasizing the different behavior of the quantifiers.



(a) (b)

Figure 4.7: Weak anomaly from the perspective of the trace-weighted non-local volume. In (a) we see how the weight compresses the normalized curve slightly, while in (b) a zoomed view of the region containing the local minimum shows that each quantifier attains the minimum at a different angle (bold ticks).

It is interesting to observe that the local minima are reached for different values of α for the weighted and non-weighted version. This shows that the weak anomaly is not an intrinsic characteristic of the scenario, but is dependent on the choice of quantifier.

To investigate further how entanglement and non-locality can have different forms, we study how these quantities change when we go from a lower rank state to a full rank state. Recall that multi-partite quantum states can be decomposed into the so called Schmidt decomposition[3]; the number of coefficients in the decomposition defines the rank of Schmidt of the state, which is the one we are referring to here. We first show that

[3]For a bipartite system with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, a state $|\psi\rangle$ in $\mathcal{H}_1 \otimes \mathcal{H}_2$ can be written as $|\psi\rangle = \sum_i \alpha_i |\phi_i\rangle \otimes |\phi'_i\rangle\rangle$, where $\{\alpha_i\}$ are the Schmidt coefficients and $\{|\phi_i\}$ and $\{|\phi'_i\rangle\}$ are orthonormal basis in $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The procedure can be recursively repeated for defining the decomposition of multi-partite states.

entanglement never decreases in this situation.

Theorem 4.4.1. *Entanglement always increases when making a continuous transition from a lower rank state $|\psi\rangle$ to a full rank state $|\psi'\rangle$.*

Proof. Let $|\psi\rangle \in \mathcal{H} = \mathcal{H}^{(A)} \otimes \mathcal{H}^{(B)}$, with $\dim \mathcal{H} = D$. That is, we have a Hilbert space $\mathcal{H}$ composed of two Hilbert spaces $\mathcal{H}^{(A)}$ and $\mathcal{H}^{(B)}$. Suppose that the rank of $|\psi\rangle$ is smaller than $\sqrt{D}$, so that its Schmidt decomposition reads:

$$|\psi\rangle = \sum_{\mathrm{i}=0}^{d-1} \alpha_{\mathrm{i}} |\mathrm{ii}\rangle \ ,$$

where $\sum_{\mathrm{i}=0}^{d-1} |\alpha_{\mathrm{i}}|^2 = 1$ and $d < \sqrt{D}$, leading to the entropy of entanglement:

$$E = -\sum_{\mathrm{i}=0}^{d-1} |\alpha_{\mathrm{i}}|^2 \log_2 |\alpha_{\mathrm{i}}|^2 \ .$$

Consider now another state $|\psi'\rangle$ which is full rank, that is, its rank equals $\sqrt{D}$:

$$|\psi'\rangle = \frac{1}{\sqrt{1+|\delta|^2}} (\delta|\phi\rangle + |\psi\rangle) \ ,$$

whose entropy of entanglement is given by:

$$E' = -\frac{|\delta|^2}{1+|\delta|^2} \sum_{\mathrm{i}=d}^{\sqrt{D}} |\beta_{\mathrm{i}}|^2 \log \left(\frac{|\delta|^2 |\beta_{\mathrm{i}}|^2}{1+|\delta|^2} \right) - \frac{1}{1+|\delta|^2} \sum_{\mathrm{i}=0}^{d-1} |\alpha_{\mathrm{i}}|^2 \log \left(\frac{|\alpha_{\mathrm{i}}|^2}{1+|\delta|^2} \right) \ ,$$

which, after massaging the logarithms, simplifies to:

$$E' = \frac{1}{1+|\delta|^2} \big[-|\delta|^2 \log |\delta|^2 + |\delta|^2 E_\phi + |\delta|^2 \log(1+|\delta|^2) + E + \log(1+|\delta|^2) \big] \ ,$$

where E_ϕ is defined for $|\phi\rangle$ in the same way as E was for $|\psi\rangle$.

In the limit of $\delta \to 0$ the quantity above has no Taylor expansion due to the presence of logarithms. However, for positive values of δ in that limit we can write:

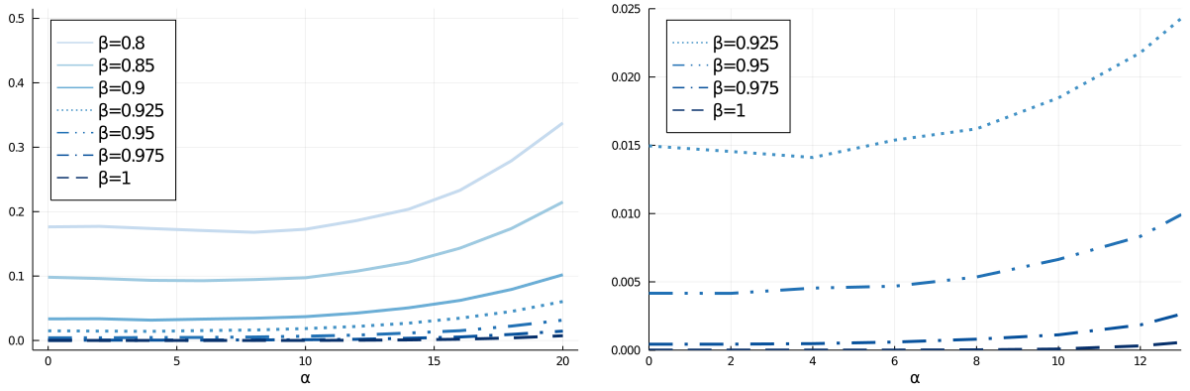
$$E' = E + \delta^2 \left(1 - E + E_\phi - 2 \log \delta\right) + \mathcal{O}(\delta^4) \ .$$

The second term goes to zero (as all other higher order terms), but it is always positive. Not only in view of $\log \delta$, but also because the entanglement entropy is finite and always smaller than 1.

So, entanglement always increases when we make a continuous change from a lower rank state $|\psi\rangle$ to a full rank state $|\psi'\rangle$, although dependent of the "direction" by which the new subspace in penetrated (dependent of $|\phi\rangle$ and of E). □

On the other hand, non-locality can behave very differently when changing from a low rank to a full rank state. Figure 4.8 shows an investigation within regions of states with partial entanglement between states $|11\rangle$ and $|22\rangle$ in the GHZ state with an extra parameter:

$$\mathrm{GHZ}(\alpha, \beta) = \sin(\alpha)|00\rangle + \frac{\cos(\alpha)}{\sqrt{2}} \left(\beta|11\rangle + \sqrt{1-\beta^2}|22\rangle \right) \ . \tag{4.11}$$



(a) (b)

Figure 4.8: Trace-weighted non-local volume for different families of states in function of α; a family in this case is labeled by the value of its β parameter in GHZ(α, β). Lighter curves stand for greater entanglement among states $|11\rangle$ and $|22\rangle$.

Notice in figure 4.8a that the minimum of the anomaly moves towards zero degree as we increase the value of β, i.e. as we consider a starting point closer to a rank-1 state ($|11\rangle$). This suggests that the minimum not only depends on the quantifier for behaviors that we extend to quantify non-locality of states, but also on the region of parameters at which we are looking at. Around the neighborhood of states with different ranks the anomaly shows its unsteadiness.

Figure 4.8b reinforces this observation, while also showing that there is a family of states (with $\beta = 0.975$) for which the quantifier is monotonic as a function of α. Increasing that value would lead to other monotonic curves too. This shows that the relation between entanglement and non-locality is not simple, depending on many aspects of the scenario and the tools we use to quantify these properties. Although this may appear counter intuitive at first sight, since non-locality is a consequence of entanglement, it is already known that entanglement and non-locality are indeed different resources, and the fact that non-locality is not a monotonic function of entanglement is another feature that supports this claim.

4.5 Discussion

We have seen an alternative way for quantifying non-locality of states based on Bell non-locality of a behavior. The key difference from preceding candidates was the introduction of a quantifier of non-locality to weight each contribution from behaviors in the non-local volume.

A new degree of freedom has then been brought to the topic. Here we explored it by considering the simplest possibility among the ones at disposal in the literature when the question is how to compare two quantum probability distributions: the trace distance. We proved that this quantifier has several good properties, including its formulation in terms of a linear programming, but our simulations have shown that it is not a monotonic function of entanglement. More specifically, the weak anomaly of non-locality persists. Nevertheless, the local minimum for non-locality with the trace-weighted non-local volume occurs in a different state than the minimum for the non-weighted version, showing that the weak anomaly is not an intrinsic characteristic of the scenario, but is dependent

on the choice of the quantifier. We conjecture that this non-monotonicity, despite the coincidence of the maxima, is a unavoidable manifestation of the intrinsic inequivalence between entanglement and non-locality. It is a topic to further investigation how the behavior of the weighted version would be for different quantifiers Q and whether we can prove that every non-locality quantifier for states exhibits some kind of anomaly. Moreover, its robustness against noisy systems also remains to be tested.

It should be noted, however, that in the integral defining the non-local volume we considered projective measurements only. The next natural generalization would be to consider the set of POVM's, but two difficulties appear. First, there is no natural definition of uniform measure for POVM's, which makes the quantifier strongly dependent on the choice of measure we choose to sample the measurements. Second, numerical results show that the probability of finding a non-local behavior when sampling over the set of POVM's is very small. In fact, in a first attempt we have made using Neumark's dilation theorem and the Haar measure in the set of unitaries in the larger Hilbert space, typically a non-local behavior is observed in one out of ten million settings for the $(2, 2, 2)$ scenario (even for the maximally entangled state), a simulation which, in addition, takes much more time to be finished than the previous one.

This general idea has come with the seeming necessity for refining known quantifiers in order to extinguish such anomalies against the monotonic relation expected between non-locality and entanglement of states. An interesting venue for further research would be to find out a quantifier of non-locality for behavior that implies monotonicity of the weighted non-local volume and/or presents the desired property at least for projective measurements, and preferably is computationally friendly. However, we conjecture that this is not possible. If no anomaly remains, that is, if non-locality increases monotonically with entanglement in any circumstances, we would conclude that non-locality and entanglement are equivalent resources, which we know not to be true. Moreover, we know that different non-locality quantifiers give different orderings in the set of behaviors, which makes non-locality heavily dependent on the function we use to quantify it. For instance, maximal violation of an inequality given a state and resistance against noise before becoming local are two non-locality quantifiers for state that can be inversely related by the fact that the resistance against detection inefficiency increases as the entanglement of the states decrease [Ebe93].

Chapter 5

Volumes of correlation sets in causal networks

This final chapter is devoted to the presentation of the last contribution made by this author during its PhD. It can be seen as a working example of the content in the first two chapters of this manuscript, but it also represents exploration beyond the main topic towards different aspects of non-classical phenomena. And that is the case especially because it is the product of a collaboration started in a technical visit to Rafael's group at the International Institute of Physics - UFRN (Brazil). To access the online version of the paper, see [Cam+23].

As we are going to see, the work underlying the next sections demands additional frameworks and ideas to those the reader might be aware of from this dissertation at this point. And the reason why we have decided not to cover them in details in previous sections is just to highlight the fact that contextuality has been guiding this author along his studies mostly, and that by studying it one can have the tools and the ability to tackle problems *a priori* unrelated.

The key to understand the relation of this chapter to the initial ones is to notice that the idea passing through the notion of a model describable by means of a common hidden variable λ (see (4.2)) refers to the possibility of having a past classical common cause that determines the outcomes. *Causality* by itself then becomes central in some lines of investigations of non-classical phenomena. And looking at contextuality through its eyes is the precise reason that motivated this author into this work. To understand it, however, we first start presenting *causal models*.

5.1 Causal models

Defining the causal structure for observed phenomena is among the basic goals of scientific research. By using the mathematical formalism of causality, we can formulate causal hypothesis by means of experimental data that can then be tested by new observations [Pea09]. And its application go beyond quantum physics, in which the quantum internet [Kim08; WEH18] and quantum repeaters [Bri+98] are the seminal examples, besides Bell's theorem, that can be seen as a particular case of a causal inference problem [Bel64]. Examples can be found also in economics [Che08; AK01] and in biology and medicine [Fri04; KH11].

The causal modeling framework [Pea09; SGS00] offers a powerful language to describe causal constraints in terms of *directed acyclic graphs* (DAG). In this formalism, each node

$A \in N(G)$ of the DAG G is associated with a random variable and causal relationships are defined by the directed edges $E(G) \subseteq N(G) \times N(G)$ between these nodes. In any real situation, we do not have access to every relevant cause that can influence our system, we then ought to distinguish between nodes that are associated with *observable* variables $\mathcal{O}_G \subseteq N(G)$ and unobserved, or *latent*, ones $\mathcal{L}_G \subseteq N(G)$. Graphically, we will use circles and Latin letters for the former and triangles and Greek letters for the latter (see for example Fig. 5.1). Moreover, the random variables and the nodes will be represented with uppercase letters $A, B, \ldots$ while we will use the corresponding lowercase ones $a, b, \ldots$ to denote their outcomes.

Given a DAG G, we can define the concept of causal parents $\mathrm{Pa}(A)$ (or children $\mathrm{Ch}(A)$) of a given variable A in G, as the set of nodes sharing incoming (or outgoing) edges with A. This notion immediately give us a way to define what it means, for a classical joint distribution $p(\{a_i\}_i)$ on the random variables associated with observable nodes $A_i \in \mathcal{O}_G$, to be compatible with a causal DAG G.

Definition 5.1.1 (Classical compatibility)**.** A distribution $p(\{a_i\}_i)$ on the random variable associated with the nodes $\mathcal{O}(G)$, is compatible with G if it satisfies the following decomposition:

$$p(\{a_i\}_i) = \sum_{\lambda \in \mathcal{L}_G} \prod_{X \in N_G} p(x|\,\mathrm{pa}(X)) \tag{5.1}$$

where $\mathcal{L}_G \subset N_G$ is the set of latent variables in G and $\mathrm{pa}(X)$ the set of outcomes of all the parents of X.

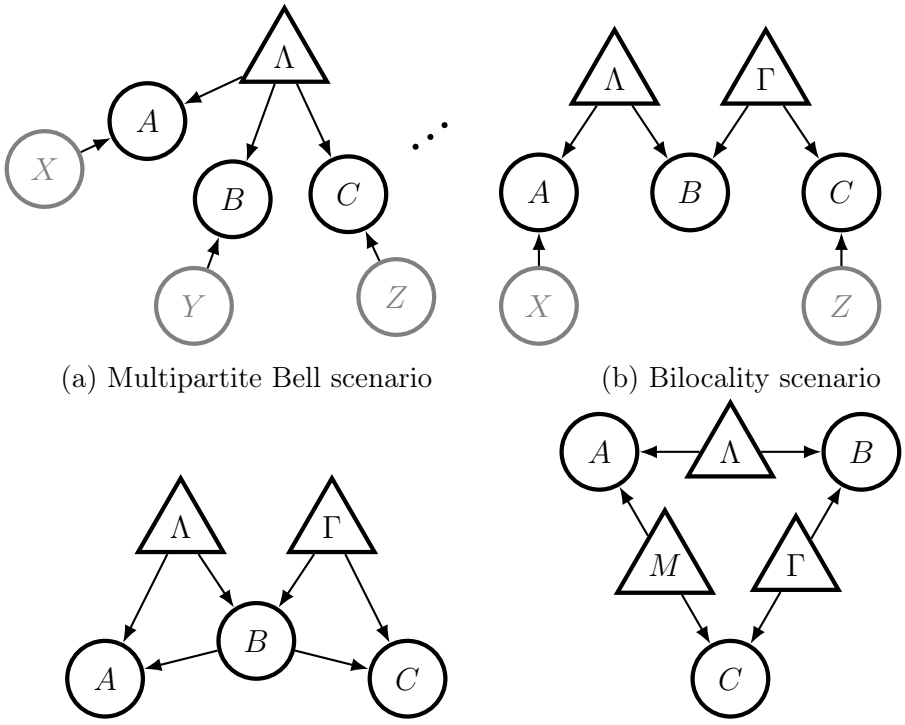
The above decomposition is also called *global Markov condition*. We will use $\mathcal{C}(G)$ to represent the set of distributions compatible in this sense with a DAG G.

The compatibility notion defined above is valid only if we consider that our distribution p arises from models where latent variables can be considered classical systems. If instead, we allow them to be quantum systems, in general, we obtain a strictly larger set of compatible distributions, which we can denote by $\mathcal{Q}(G)$ and for which generalizations of the global Markov condition and the concept of a causal structure have been proposed [CMG15; PB15; CS16; BLO19]

In some cases, it might be interesting to consider post-quantum resources distributed in the network [HLP14]. The distributions that may arise from these models must respect some basic conditions that reflect the natural assumptions of no-signaling and independence of the sources (NSI). The core of the notion of the no-signaling (or no-disturbing, see (1.18)) principle to network scenarios is that the outcomes of one party should be insensitive to whatever the remaining parties do, including any local modifications in the particular arrangement of the topology of their part of the network. The set of correlations that arise from NSI has been explicitly studied primarily in the triangle scenario [Gis+20] and in the Evans scenario [Lau+23]. We denote the set of correlations that is compatible with the principles of no-signaling and independence as $\mathcal{N}(G)$. We remark that the resulting constraints derived from NSI will be valid for all general probabilistic theories [CDP10] (GPTs).

5.1.1 Single-source Bell scenario

Bell's paradigmatic causal structure (and multipartite generalization thereof) is composed by a number of parties, sharing classical correlations described by a random variable Λ, locally measuring different observables parametrized by $X, Y, Z, \ldots$ with corresponding



(a) Multipartite Bell scenario (b) Bilocality scenario

(c) Evans scenario (d) Triangle scenario

Figure 5.1: **Causal DAGs a)** Multipartite Bell scenario where the correlations between the distant parties are mediated by a single source of correlations. **b)** Bilocality scenario, akin to an entanglement swapping experiment [Pan+98], where two independent sources establish the correlations between three spatially separated parties. **c)** The Evans scenario with two independent sources of correlations but with a crucial difference to the bilocality case: the inputs for two of the parties are the measurement outputs of the central node, that is, the correlations are time-like and not space-like separated. **d)** The triangle scenario, where every party shares a bipartite and independent source of correlations with every other party.

measurement outcomes $A, B, C, \ldots$. This classical description is encoded in the class of DAGs L_n represented in Fig. 5.1a, implying that the observed distributions should follow the Markov condition given by

$$p(a, b, c, \cdots | x, y, z, \cdots) = \sum_{\lambda} p(\lambda) p(a|x, \lambda) p(b|y, \lambda) p(c|z, \lambda) \ldots \quad . \tag{5.2}$$

In a quantum description, this probability distribution is given by the Born rule

$$p(a, b, c, \cdots | x, y, z, \cdots) = \mathrm{Tr}\left[\left(A_x^a \otimes B_y^b \otimes C_z^c \otimes \cdots \right) \ \rho_\Lambda \right] \ , \tag{5.3}$$

where the measurement inputs and outputs are associated with POVMs (positive operator valued measurements) and the classical node Λ is replaced by a multipartite (potentially entangled) state ρ_Λ.

Due to the absence of a causal link between nodes associated with different parties, we also expect any compatible distribution to respect some general linear constraints, expressing their independence in terms of the observable distribution, which are called

no-signaling constraints [PR94] defining a set $\mathcal{N}_n$:

$$\begin{aligned} p(a|x,y,z,\ldots) &= p(a|x,y',z',\ldots) \quad \forall a,y,y',z,z',\ldots \\ p(b|x,y,z,\ldots) &= p(b|x',y,z',\ldots) \quad \forall b,x,x',z,z',\ldots \\ &\vdots \end{aligned} \tag{5.4}$$

for each party $A, B, C, \ldots$, where $p(a|x,y,z,\ldots)$ is the marginalization of the distribution over all the other variables different from A.

In the case of a single source of correlations, $\mathcal{C}(L_n), \mathcal{Q}(L_n), \mathcal{N}(L_n)$ are all convex sets, and, in particular, $\mathcal{C}(L_n)$ and $\mathcal{N}(L_n)$ can be described by a finite number of constraints, making them convex polytopes. Moreover, it is known that $\mathcal{C}(L_n) \subset \mathcal{Q}(L_n) \subset \mathcal{N}(L_n)$. More recently, the connection between causal modeling and Bell inequalities prompted the study of the relationships between these correlation sets in more complex causal scenarios that can also include independence between latent variables. In this case, the geometry of the correlation sets becomes considerably more complicated, making them even non-convex in general. In our work, we focused on three such models, the bilocality scenario [BGP10], the triangle scenario [Fri12; Ren+19a] and the Evans Scenario [Eva16; Lau+23], which we will now describe in more detail.

5.1.2 Bilocality scenario

The bilocality scenario [BGP10], represented by the DAG B in figure 5.1b, presents two independent sources Λ and Γ which distribute correlations to three nodes, A, B, and C that can perform measurements chosen by the settings X and Y for A and C respectively, while the central node B has no external settings. As in the Bell case, we can distinguish different sets of compatible distributions associated with such a model. The Markov condition (5.1) for this structure is given by

$$p(a,b,c|x,z) = \sum_{\lambda,\gamma} p(\lambda)\ p(\gamma)\ p(a|x,\lambda)\ p(c|z,\gamma)\ p(b|\lambda,\gamma). \tag{5.5}$$

As anticipated such a condition complicates considerably the characterization of the set of allowed correlation $\mathcal{C}(B)$, which is known to be non-convex, as proved by the existence of polynomial Bell inequalities [Cha16].

Quantum distributions in this scenario $p \in \mathcal{Q}(B)$ are instead defined by

$$p(a,b,c|x,z) = \mathrm{Tr}\left[\left(A_x^a \otimes B^b \otimes C_z^c\right)\left(\rho_\Lambda \otimes \rho_\Gamma\right)\right], \tag{5.6}$$

for any couple of bipartite quantum states $\rho_\Lambda, \rho_\Gamma$ and any set of POVMs with operators A_x^a, B^b, C_z^c.

Similarly to the multipartite Bell scenarios, here we have that some of the correlations in the quantum set $\mathcal{Q}(B)$ are incompatible with a classical description (5.5), even when some of them are still compatible with the one in the tripartite Bell scenario $\mathcal{C}(L_3)$, showing how the assumption of independence between Λ and Γ increases the possibility to detect non-classicality.

Also in this scenario, we expect correlations to respect some basic constraints given in terms of their observable distribution. But differently from the standard multipartite Bell case, besides the linear constraints (5.4), here we have some additional nonlinear

ones due to the conditional independence between the two sources. More specifically, it follows that

$$p(a,c|x,z) = \sum_b p(a,b,c|x,z) = p(a|x)\ p(c|z), \tag{5.7}$$

making this set strictly included in the set of non-signaling correlations in the tripartite Bell scenario $\mathcal{N}_3$. We will denote this set of non-signaling bilocal correlations by $\mathcal{N}(B) \subset \mathcal{N}_3$.

5.1.3 Triangle scenario

The triangle scenario [Fri12] is represented by the DAG in Fig. 5.1d, where three bipartite sources distribute systems to three separate parties forming a triangle-shaped structure. The reason why this network is of particular interest to the community is that it allows for novel non-classical phenomena [Ren+19a; Sup+22; CWR21]. In particular, it allows for non-classicality even in the absence of measurement choices for the parties, that is, when they perform a fixed measurement [Fri12; Cha+21].

Classically, triangle correlations admit models given by

$$p(a,b,c) = \sum_{\lambda,\gamma,\mu} p(\lambda)\ p(\gamma)\ p(\mu)\ p(a|\lambda,\mu)\ p(b|\lambda,\gamma)\ p(c|\gamma,\mu), \tag{5.8}$$

while the quantum description is given by [1]

$$p(a,b,c) = \mathrm{Tr}\left[\left(A^a \otimes B^b \otimes C^c\right)\left(\rho_\Lambda \otimes \rho_\Gamma \otimes \rho_M\right)\right]. \tag{5.9}$$

Differently from the standard Bell or the bilocality cases, the triangle does not have a simple NSI description, even though a few approximations have already been proposed [Gis+20; HLP14; CMG15; BR21; WSF19].

Furthermore, the case where all parts have binary outputs is particularly interesting. Recently, [Poz+23a] has shown that this minimal case supports post-quantum advantage, i.e. the local set and the non-signaling set do not coincide, and also that the set of triangle non-signaling correlations lies outside the quantum set. Strikingly, the conjecture that the local and quantum sets are identical [Tav+22] remains open for this minimal case.

5.1.4 Evans scenario

The last causal structure we will consider is the so-called Evans scenario [Eva16], which was only recently considered from a quantum perspective [Lau+23]. The DAG E for this structure is represented in Fig. 5.1c. Just like in the bilocality scenario, we have three observable nodes A, B, C and two latent ones Λ, Γ, with the difference that now there are no external inputs present, and B can communicate directly its value to both A and C.

Equation (5.1) in this case becomes

$$p(a,b,c) = \sum_{\lambda,\gamma} p(\lambda)\ p(\gamma)\ p(a|b,\lambda)\ p(c|b,\gamma)\ p(b|\lambda,\gamma), \tag{5.10}$$

and the quantum distribution is given by

$$p(a,b,c) = \mathrm{Tr}\left[\left(A^a_b \otimes B^b \otimes C^c_b\right)\left(\rho_\Lambda \otimes \rho_\Gamma\right)\right], \tag{5.11}$$

[1] Note that when evaluating p, one should be attentive to which Hilbert space supports each state and measurements.

for quantum states $\rho_\Lambda, \rho_\Gamma$ and POVMs A_b^a, B^b, C_b^c.

Despite apparent similarities with the bilocal scenario, the characterization of the classical and quantum sets $\mathcal{C}(E), \mathcal{Q}(E)$ turns out to be much more complex [Lau+23], and whether the inclusion of the former in the latter is strict or not is still an open problem.

The set of non-signaling correlations of the Evans scenario is also poorly understood. Although a general route for deriving such theory-independent constraints for this causal structure has been proposed in [Lau+23], it relies on Fourier-Motzkin elimination which can be very costly and out of computational reach even for seemingly simple scenarios. It is known that the non-signaling set is strictly larger than the classical set of correlations $\mathcal{C}(E) \subset \mathcal{N}(E)$, at least for $|A| = 3$ and $|B| = |C| = 2$.

Furthermore, because there is communication between the parts, Evans's scenario allows us to go beyond passive observations of the experiment and ask what would happen if the system is intervened upon [GMC20]. We use do-conditionals $p(a|do(b))$ (and $p(c|do(b))$) to denote the probability of Alice's (Charlie's) outcome a (c) when variable B is set by force to be b. For classical correlations, this can be formalized with the constraint

$$p(a|do(b)) = \sum_\lambda p(\lambda)\ p(a|\lambda, b) \tag{5.12}$$

and similarly for $p(c|do(b))$. Analogously, we can define what the do-conditionals would look like in terms of a quantum strategy

$$p(a|do(b)) = \mathrm{Tr}\left[(A_b^a \otimes \mathbb{I})\, \rho_\Lambda\right] \tag{5.13}$$

and the same can be done for $p(c|do(b))$.

One might be interested in exploring interventions for post-quantum theories in the Evans scenario. In order to do so, notice that the do-conditionals can be expressed in terms of a marginal probability distribution of a particular *interruption* of the original graph, which consists in introducing new independent variables that each inherits one outcoming edge of the original variable, see section IX of [Wol+21] for details. This procedure allows us to map the restrictions due to no-signaling involving do-conditionals of the Evans scenario to the no-signaling conditions on the bilocality scenario and identify the do-conditionals $p(a|do(b))$ and $p(c|do(b))$ with the marginals $p(a|x = b)$ and $p(c|z = b)$.

5.2 The problem

The violation of Bell inequalities [Bru+14] represents the strongest signature of non-classical behavior, as it can be observed in a device-independent context [PSV16]. Specifically, it demonstrates the incompatibility between quantum correlations and classical concepts of cause and effect, without relying on any assumptions about the internal mechanisms involved in the preparation and measurement of the physical system being analyzed.

In the simplest Bell scenario, two distant parties measure two distinct dichotomic observables. However, this scenario has been extended and generalized in various ways. These extensions include incorporating additional measurements [CG04] or expanding the number of possible outcomes [Col+02]. Furthermore, the framework has been expanded to involve multiple parties [WW01a] and relaxations of locality assumptions [Pir03; BC17], as well as measurement independence [Hal10; Cha+15]. Of particular significance are recent

advancements that draw inspiration from causality theory [Pea09]. These generalizations explore networks with diverse topologies, growing in size and complexity, revealing a number of novel non-classical phenomena [BGP10; Fri12; Ren+19a; WC20; Ren+21; Cha+21; Sup+22; Pol+23].

In the context of a specific causal structure, a central inquiry arises: does it exhibit a classical-quantum gap? In other words, if the sources within the network are described by entangled quantum states, can measurements on them produce correlations that lack a classical interpretation? When considering networks with a single source, a classical depiction involves the characterization of a polytope [Pit91]: a convex set defined by a finite number of extremal points or, equivalently, a finite set of linear Bell inequalities. However, even in such cases, the problem is recognized as intrinsically challenging, residing in the NP-hard complexity class [Pit91]. This difficulty is further amplified when independent sources of correlations exist, as is often the case in paradigmatic quantum networks. In such scenarios, the correlations compatible with a given causal structure result in non-convex sets, necessitating computationally intensive algorithms rooted in algebraic geometry [GSS05] or various forms of approximation [Cha+14; Cha16; Kel+19; Åbe+20; Poz+19; WSF19] proposed throughout the years. Not surprisingly, given these difficulties, there is still a very fragmented picture of the set of correlations causal networks can give rise to.

Here we propose an alternative route to gather further insights into the classical description of causal networks and their potential incompatibility with quantum correlations. Using quadratic optimization techniques (leveraging the Gurobi optimizer [Gur22]), already employed to address non-convex constraints originating from causal networks in [Lau+23], as well as other tools such as the inflation technique [WSF19], the covariance approach [Åbe+20] and Finner inequality [Ren+19b], we calculate the volumes [Cab05; WY12] of classical and non-classical probability distributions that various causal structures can yield.

5.2.1 Tools for detecting non-classicality

The compatibility problem consists of answering the question: Are the statistics over the observed variables compatible with the causal structure under scrutiny? For causal structures with a single latent common cause, e.g. the Bell scenario, this amounts to a linear feasibility problem. Indeed, it is known that if we consider a decomposition of the form (5.2) we can, without loss of generality, incorporate any local randomness present in the response functions – i.e. $p(a|x,\lambda)$, $p(b|y,\lambda)$, $p(c|z,\lambda)$ and so on – to the source λ and define a deterministic model. In a deterministic model, each value of λ defines an assignment of one of the possible outputs to each input. The model is a probabilistic mixture of these deterministic assignments of outputs to inputs, with λ specifying which particular assignment is chosen in each run of the experiment. For each assignment, there is a corresponding *local deterministic behavior* $d_\lambda(a,b,c..|x,y,z,..)$ and $p(a,b,c...|x,y,z...)$ is compatible with the standard Bell scenario *if, and only if* it can be expressed as a convex combination of deterministic local points. That is,

$$\begin{aligned}\mathbf{p} \text{ is local} \iff \exists q(\lambda) \text{ s.t. } q(\lambda)\geq 0, \quad \sum_\lambda q(\lambda)=1 \text{ and} \\ p(a,...|x,...)=\sum_\lambda q(\lambda)\; d_\lambda(a,...|x,...).\end{aligned} \tag{5.14}$$

Indeed, determining whether there exist weights $q(\lambda)$ satisfying the linear constraints in Eq. (5.14) is a typical instance of a *linear programming problem* (LP) [BV06].

For a generic causal structure, i.e. more than one source, we can use the inflation technique. Intuitively, the method works by considering the hypothetical situation where one has access to multiple copies of the sources and measurement devices that compose the network and can rearrange them in different configurations. Its core idea is to explore simple (linear) conditions of this inflated network that ultimately translate to polynomial inequalities on the observable probabilities. It has been proven in [NPA08] the existence of a hierarchy of inflations that asymptotically converges to the classical set of correlations of any network and a test of compatibility of a given level of this hierarchy can be done via Linear Programming (LP) or Semi-Definite Programming (SDP) [BV06]. However, for each level n of this hierarchy the memory resources required are superexponential on n. Notably, this hierarchy relies on information broadcasting, a primitive that is not allowed in quantum information.

The inflation technique can constrain not only the set of classical correlations but also the set of quantum correlations a network may give rise to. *Quantum inflation* [Wol+21] can be seen as a quantum analog of the classical inflation technique which avoids the latter's reliance on information broadcasting. This is done by adapting the Navascues-Pironio-Acin (NPA) hierarchy [NPA07], originally developed to characterize quantum correlations in Bell scenarios, in the *inflated* scenario which can be tackled via noncommutative polynomial optimization (NPO) theory [NPA08]. The general goal of NPO theory is to optimize the expectation value of a polynomial over operators subject to a number of polynomial operators and statistical constraints. This optimization is achieved by means of a hierarchy of SDP tests [2]. The types of inflations we have used in our work are shown in Fig. 5.2. For each inflation level, we then study the set varying the NPA levels, and we will denote by $\mathcal{Q}_{n,m}(G)$ the corresponding relaxation associated with n-th order inflation and level m of the NPA hierarchy.

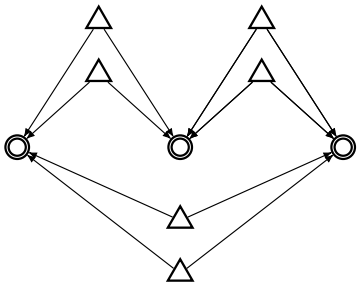
At the core of our numerical approach is the use of quadratic programming (QP) techniques, like the *branch and bound method*, that allows us to extend our optimization problems to include non-linear constraints with reasonable efficiency [Gur22]. These techniques work by iteratively breaking the variables domain down into smaller problems that can each be approximated by a corresponding convex program. This branching subroutine enables primal and dual tasks to define upper and lower bounds that converge, up to computational precision (10^{-9}), to the global optimal solution.

Using this we can assess classicality for the bilocal and the Evans networks by directly imposing the independence of the sources $p(\lambda,\gamma) = p(\lambda)p(\gamma)$. Notice that in both scenarios we can, without loss of generality, make λ determine the outcome a for every x and, similarly, γ determine c for every z, while B has a stochastic response function $p(b|\lambda,\gamma)$. Therefore, we can take $\lambda = \{a_0, ..., a_{|x|-1}\}$, $\gamma = \{c_0, ..., c_{|z|-1}\}$ and $p(a,b,c|x,z)$ is bilocal *if and only if*,

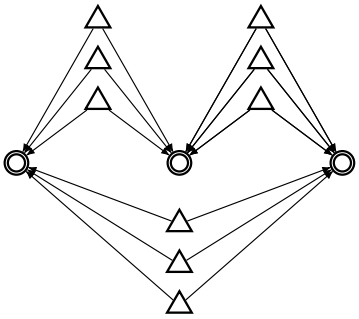
$$\begin{aligned} \exists q(\lambda,b,\gamma) \text{ s.t. } q \geq 0, \quad & \sum_{\lambda,b\gamma} q(\lambda,b,\gamma) = 1, \\ & q(a_x = a, b, c_z = c) = p(a,b,c|x,z), \\ & \text{and} \quad q(\lambda,\gamma) = q(\lambda)q(\gamma). \end{aligned} \tag{5.15}$$

which can be cast as a QP. For compatibility with the Evans scenario, it is sufficient to look at the same conditions but only for $x = z = b$. We also use the arguments given

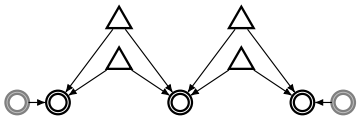
[2] These tests were implemented using [BWP23].



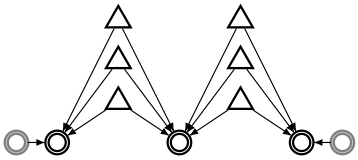
(a) 2nd order quantum inflation for the Triangle scenario.



(b) 3rd order quantum inflation for the Triangle scenario.



(c) 2nd order quantum inflation for the Bilocality scenario.



(d) 3rd order quantum inflation for the Bilocality scenario.

Figure 5.2: The DAGs for the orders used in our analysis for quantum inflation. Here we, similarly to the classical case, considered the first two levels of standard quantum inflation hierarchy [Wol+21], generated by using n copies of each latent variable. As in the other figures, latent nodes are represented by triangles while observable nodes with circles.

in [Lau+23] to extract tailored infeasibility certificates from these quadratic programs.

5.2.2 Sampling non-signaling distributions

The starting point for the analysis of the volumes is generating the data sets to be analyzed. These points must satisfy NSI restrictions, which will be polynomial in the case of the bilocal scenario. For this, we sample separately each coordinate of the behavior vector (or its equivalent representation by correlators) followed by a rejection step to handle the constraints [3]. Specifically, we sampled points uniformly within a hypercube and then selected the subset of interest. We employed the results in [DFM92] to determine how many data points were necessary to sample the whole space of distributions in each situation, that is, at least $(\sim 2.1)^d$ points uniformly generated in a hypercube of dimension d are required to estimate a volume close to the hypercube itself. For example, for a hypercube of dimension eight (the simplest triangle scenario in this work), one should consider data sets containing more than four hundred points to start approximating well the corresponding set of correlations. In view of such uniform distribution, the fraction of points represented by a target subset can be associated with the volume it occupies within the set containing it. More than one data set was used in what follows, therefore we are going to mention them according to their specific uses.

[3] Other ways of sampling were tested, e.g. the Monte Carlo method, which is able to include more complicated constraints to sample from. However, the rejection method was the best in terms of numerical performance; the reason for that being the cardinality of the variables and the number of constraints.

In all these cases, however, the idea behind them is the same: test membership of the generated points in a set of interest via LP, SDP, or QP tests. Such tests can be phrased in terms of feasibility or optimization problems. In what follows, we also investigate how these points are distributed inside the set. To this end, we consider a corresponding optimization in which the objective function to be minimized is the trace distance between the sampled distribution and some distribution inside the set (see [BAC18]).

The trace distance between two probability distributions $\mathbf{p}(x)$ and $\mathbf{q}(x)$, recalling from (4.4), is given by

$$D(\mathbf{p}, \mathbf{q}) := \frac{1}{2} \sum_x |p(x) - q(x)|, \tag{5.16}$$

and the *trace distance-based quantifier* $Q_{\mathcal{X}}(p)$ of the distance of a probability distribution $p(a, b, c|x, y, z)$ to a set $\mathcal{X}$ in a tripartite scenario is computed through

$$Q_{\mathcal{X}}(\mathbf{p}) = \frac{1}{|x||y||z|} \min_{q(a,b,c|x,y,z) \in \mathcal{X}} D(p, q), \tag{5.17}$$

and $|x|$, $|y|$, and $|z|$ stand for the cardinality of the set of measurements for Alice, Bob, and Charlie, respectively. If $\mathcal{X}$ is the local set, then we have a non-local distance, while if it is the bilocal set, we have a non-bilocal distance, and similarly to any other set. Finally, we compare the performance of the aforementioned techniques for different cases.

5.3 Results

As we show, in spite of its wide applicability and convergence in the asymptotic limit [NW20], in computational practice the inflation technique might offer relatively poor performance, since it is unable to reveal a significant portion of non-classical behaviors. Interestingly, similarly to what happens in the standard Bell scenario [Dua+18], we observe a concentration effect where most non-classical points concentrate an average distance from the classical set, the probability of finding points far away decaying exponentially. Finally, we show that the use of interventions [Cha+18; GMC20] can significantly enhance our ability to detect non-classical behaviors.

5.3.1 Bilocality scenario

In the case of the bilocality scenario, we estimated the relative volume among the different sets of correlation $\mathcal{C}(B)$, $\mathcal{Q}(B)$ and $\mathcal{C}(L_3)$ with respect to randomly sampled points inside $\mathcal{N}(B)$ for the case where all variables are binary. Fig. 5.3 schematically illustrates these sets.

For doing so, we parametrize the conditional probability distribution $p(a, b, c|x, z)$ in terms of the single-, two- and three-party correlators $\langle A_x \rangle$, $\langle B \rangle$, $\langle C_z \rangle$, $\langle A_x B \rangle$, $\langle B C_z \rangle$, $\langle A_x C_z \rangle = \langle A_x \rangle \langle C_z \rangle$ and $\langle A_x B C_z \rangle$ for all $x, z \in \{0, 1\}$,

$$\langle A_x B C_z \rangle = \sum_{a,b,c} (-1)^{a+b+c} p(a, b, c|x, z), \tag{5.18}$$

and similiarly for the other correlators. This yields a total of 13 parameters in the interval $[-1, 1]$. We generated a set of uniformly distributed points in the 13-dimensional hypercube and, naturally, considered only the ones inside the region of non-negative probability

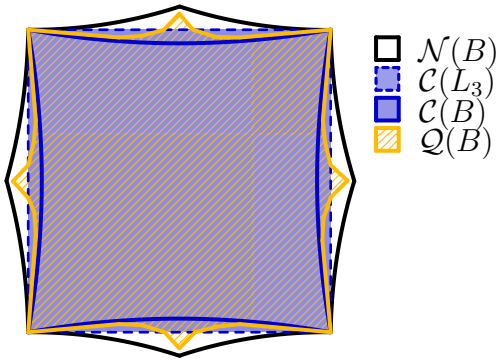


Figure 5.3: Illustration of the different sets of correlations in the bilocal scenario; proportions are not supposed to be faithful. Solid black lines delimit the part of the no-signalling set compatible with the bilocal conditions, while the blue region with a dashed border determines the local set and solid blue lines delimit the bilocal subset within it. The hashed yellow region encompasses quantum behaviors.

distributions, i.e. we excluded the points that do not respect

$$p(a,b,c|x,z) = \frac{1}{8}\Big(1 + (-1)^a\langle A_x\rangle + (-1)^b\langle B\rangle + (-1)^c\langle C_z\rangle + (-1)^{a+b}\langle A_xB\rangle + \\ + (-1)^{b+c}\langle BC_z\rangle + (-1)^{a+c}\langle A_x\rangle\langle C_z\rangle + (-1)^{a+b+c}\langle A_xBC_z\rangle\Big) \geq 0, \tag{5.19}$$

which gives us a set of uniformly distributed points in $\mathcal{N}(B)$. We remark that this can be done without loss of generality for all non-signaling behaviors, i.e. one can always choose some minimal representation or (for the case of binary outcomes) correlators, i.e. expectation values, representation to isolate the free parameters and eliminate all equality constraints and, if necessary, use inequality constraints to filter valid probability distributions from the data set that are uniformly distributed in the region under scrutiny.

First, we analyze which points in our data set violate the standard tripartite Bell locality by computing their trace distance to the local set of distributions $\mathcal{C}(L_3)$. Then, for the remaining points, which have an explicit local decomposition, we analyze different levels of the classical inflation hierarchy $\mathcal{C}_n(B)$ to compute their distances from the bilocal set $\mathcal{C}(B)$, obtained by solving directly the QP problem, and, from them, estimate the relative volumes of these sets. The results for the volumes are presented in Table 5.1. Interestingly, while the set of non-local distributions $\overline{\mathcal{C}}(L_3) = \mathcal{N}(B) \setminus \mathcal{C}(L_3)$ amounts to approximately only 2.39%, the non-bilocal set $\overline{\mathcal{C}}(B) = \mathcal{N}(B) \setminus \mathcal{C}(B)$ occupies 32.6% of the correlation space, a clear signature of the advantage of considering the independence of the sources when testing the non-classicality of the data.

Regarding the use of inflation, the best witness of non-classicality is represented by $\mathcal{C}_3$, i.e. 3rd order inflation, showing that 13.3% of its volume is non-bilocal, still a significantly smaller value than that obtained by QP. Besides being less accurate in detecting non-classicality, the inflation technique is computationally more demanding (around fifty times more as compared with the QP approach), which is the reason why we analyzed two independently generated samples of different sizes: $N = 10^4$ for the inflation and $N = 10^6$ in the QP case.

Additionally, considering $\mathcal{Q}_{2,3}$, i.e. 2nd order of inflation together with the 3rd level of the NPA hierarchy, we estimated the ratio of the quantum volume with respect to

Set	Volume
$\overline{\mathcal{C}}(L_3)$ (Non-local)	0.0239
$\overline{\mathcal{C}}(B)$ (Non-bilocal)	0.326
$\overline{\mathcal{C}}_1(B)$ (Inflation 1st order)	0.0233
$\overline{\mathcal{C}}_2(B)$ (Inflation 2nd order)	0.0999
$\overline{\mathcal{C}}_3(B)$ (Inflation 3rd order)	0.133

Table 5.1: Volumes of different correlation sets in the bilocality scenario. In the first two rows we have the results provided by QP on a data set with 10^6 samples, while in the last three, we have the volume associated with the set of non-bilocal behaviors $\overline{\mathcal{C}}_n(B)$ computed for different orders of inflation n, on a data set with only 10^4 instances. Order $n = 3$ was the highest configuration analyzed. Which is still considerably far from the value obtained by the QP approach.

the other sets. As can be seen in Table 5.2, while the quantum volume (with the NPA level considered) within the non-local correlations is 43%, it increases to about 86% when considering the set of non-bilocal correlations.

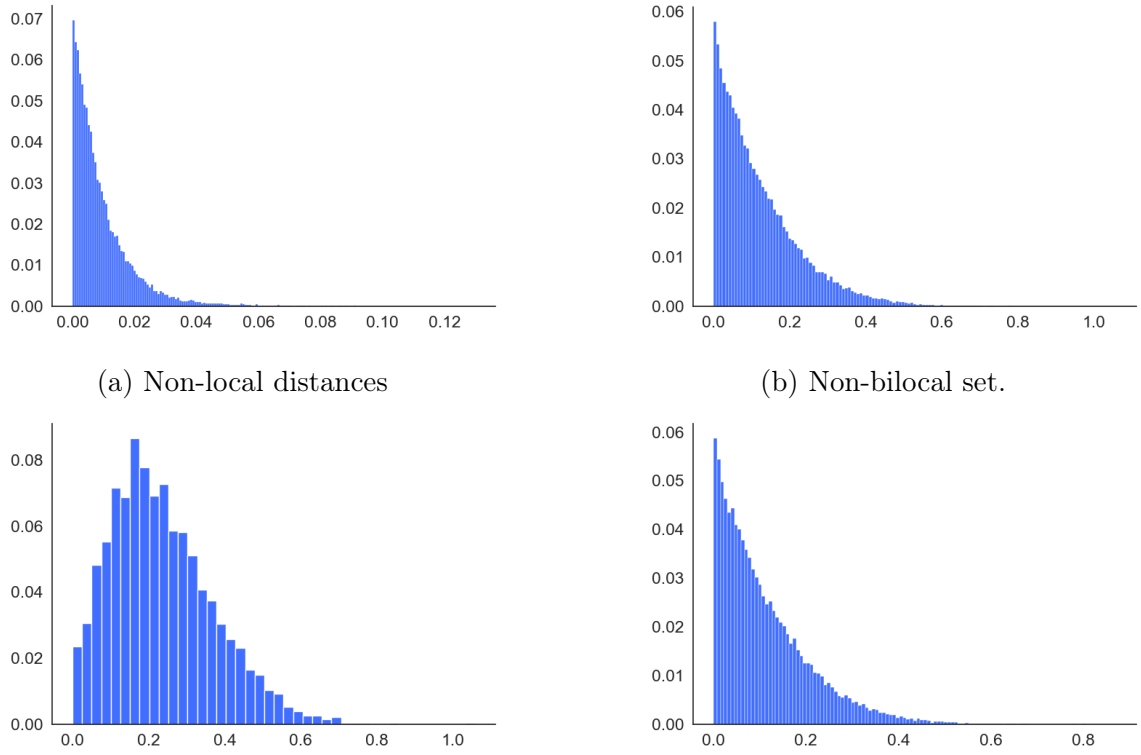
Set	Volume of $\mathcal{Q}_{2,3}(B)$
$\mathcal{N}(B)$ (Non-signaling)	0.952
$\overline{\mathcal{C}}(L_3)$ (Non-local)	0.43
$\mathcal{C}(L_3)$ Local	0.942
$\overline{\mathcal{C}}(B)$ (Non-bilocal)	0.856
$\mathcal{C}(B)$ (Bilocal)	1.0

Table 5.2: Volumes of the behaviors that are compatible with a quantum description, evaluated considering the relaxation $\mathcal{Q}_{2,3}(B)$, i.e. the 2nd order of quantum inflation and NPA level 3. The results were obtained with a data set containing 10^4 behaviors.

To gather more information about the structure of different correlation sets, beyond their relative volumes, we can also estimate how these points are distributed relative to their trace distances inside $\mathcal{N}(B)$. From Fig. 5.4 we can see how the points in $\mathcal{N}(B)$ have their distances from the local $\mathcal{C}(L_3)$ and bilocal $\mathcal{C}(B)$ set distributed according to a Poissonian-like distribution. An interesting observation here is the presence of a concentration of the distances for behaviors that are non-classical, i.e. the distribution is peaked at a small distance from the local and bilocal sets, which can be seen as an instance of the concentration phenomena reported in [Dua+18].

5.3.2 Triangle scenario

Differently from the bilocality and Evans scenarios, for which there is a natural set of NSI correlations to sample from, in the triangle scenario there is no simple way to enforce the no-signaling and independence conditions [Gis+20]. For this reason, we analyze different



(a) Non-local distances

(b) Non-bilocal set.

(c) Non-bilocal within non-local set.

(d) Non-bilocal within local set.

Figure 5.4: Distributions of distances in the bilocal scenario. In **(a)** we have non-local distances for around 2×10^4 behaviors with a gap precision between primal and dual LP tasks equal to 0%. In **(b)** one can see how non-bilocal distances are distributed considering points within the whole non-bilocal set, while in (c) and (d) we split such set between non-local and local, respectively, to offer a different perspective and highlight the concentration observed in (c); results for around 4×10^4 behaviors using precision of 10% for the gap between primal and dual solutions in this case.

volumes of sets of correlations relative to points sampled inside the 8-dimensional simplex, that is, sampling over all well-defined probability distributions $p(a, b, c)$ with $a, b, c = 0, 1$ being dichotomic variables. Also, given that the triangle scenario imposes non-quadratic constraints there is no direct manner to use quadratic optimizers, for this reason we rely on the inflation technique [WSF19; Wol+21] as well as the covariance approach [Kel+19; Åbe+20] and the Finner [Ren+19b] and Shannon-type [FC12; CMG15] inequalities.

First, we compare the set of classical and quantum correlations using the inflation technique on our data points. Table 5.3b shows the results for different orders of classical inflations $\mathcal{C}_m(T)$ with $m = 1, 2, 3$, and for different quantum inflations $\mathcal{Q}_{m,n}(T)$ with $m = 1, 2$ and up to level $n = 5$ of the NPA hierarchy. Notice how we are unable to detect non-quantum behaviors without considering one extra copy of each source at least. Moreover, we have found points that are incompatible with a classical description but are quantum-compatible with $\mathcal{Q}_{2,4}$, i.e. up to the $2nd$ order inflation and NPA level 4. These points offer good candidates to resolve a still open question: whether quantum non-classical correlations can emerge in the triangle scenario with all variables dichotomic [Poz+23b].

Then, we consider the covariance decomposition test — which can only test the network topology and as such are valid for all GPTs [BR21] — and the volume delimited by

the Finner and the Shannon-type entropic inequalities found in [FC12; CMG15]; the first of them are proven to follow for all quantum distributions and conjectured to be satisfied for all non-signaling distributions [Ren+19b] while the entropic ones, as the covariance test, tell us about the topology only. Table 5.3a summarizes these results. Among them, we can see that the covariance test managed to exclude the largest number of behaviors.

Method	Incompatible volume
Covariance	0.0369
Finner	0.0191
Entropic	0.00095

(a) Comparison of other methods to detect incompatibility with the triangle network.

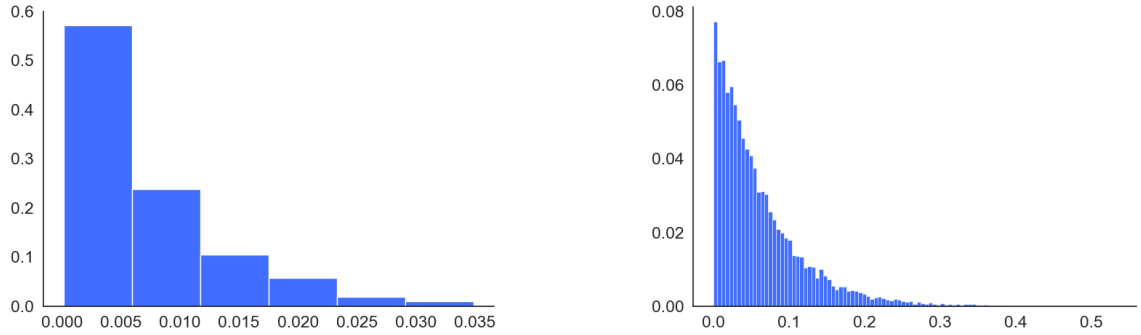
Set	Volume
$\overline{\mathcal{C}}_1(T)$	0.0
$\overline{\mathcal{C}}_2(T)$	0.09592
$\overline{\mathcal{C}}_3(T)$	0.113*
$\overline{\mathcal{Q}}_{1,5}(T)$	0.0
$\overline{\mathcal{Q}}_{2,3}(T)$	0.1098

(b) Volumes of behaviors incompatible with a classical or quantum description. *Here we use about 10% of the data set due to the computational cost.

Table 5.3: **(a)** Volume of incompatible distributions calculated with other known methods, specifically the Finner and the entropic inequalities [FC12] and the covariance method [BR21]. **(b)** Volumes of points incompatible with classical $\mathcal{C}_n(T)$ and quantum inflations $\mathcal{Q}_{m,n}(T)$ for the triangle scenario using different configurations of n, m, on a data set of around 10^5 instances.

5.3.3 Evans scenario

Now we move our attention to the Evans scenario. We start by pointing out that the relationship between the Evans and the bilocality scenario, which was first noticed for classical and quantum correlations [Lau+23], indeed holds to all GPTs. In fact, if one has access to GPT states and measurements one can uniquely define a mapping between Evans and the bilocal scenario by making an identification of the measurement effects of the variables A and C in each scenario, analogously to what was argued in [Lau+23]. Therefore, the probability distribution $p(a, b, c)$ in the Evans scenario is compatible with *any* GPT *if and only if* there exists a bilocal distribution $p_B(a, b, c|x, z)$ compatible with the same GPT which satisfies $p_B(a, b, c|x = z = b) = p(a, b, c)$. Using this mapping, we can explore the set of non-signaling correlations in the Evans scenario by virtue of the non-signaling conditions of the bilocality scenario.



(a) Non-Evans with passive observations.

(b) Non-Evans under intervention.

Figure 5.5: The distribution of the trace-distances of non-classical behaviors to the set of those compatible with a classical description when (a) only passive observations are made and (b) interventional data is also taken into account. While in (a) all the non-classical instances have been used in the plot, in (b) only 10% has been considered.

We can formalize this with the following statement:

$$\begin{aligned} p(a,b,c) \text{ is NSI-compatible } \iff & \exists\ p_B(a,b,c|x,z) \text{ non-signaling distribution,} \\ & \text{s.t. } p_B(a,b,c|x=z=b) = p(a,b,c) \\ & \text{and } p_B(a,c|x,z) = p_B(a|x)p_B(c|z). \end{aligned} \tag{5.20}$$

This tells us how to explore the set of NSI-compatible correlations in the Evans scenario using a single quadratic program. Furthermore, this fact allows our test to be conclusive, i.e. if one can prove that the conditions from (5.20) do not hold for some candidate distribution $p(a,b,c)$ then we can conclude the distribution is NSI-incompatible and, conversely, if one can find a solution then there exists a no-signaling distribution $p_B(a,b,c|x,z)$ that respects the NSI requirements, such that recovers $p(a,b,c)$ by setting $x = z = b$, leaving no ambiguity up to a small ($\sim 10^{-9}$) computational precision. We remark that this is not the case for the inflation tests, as they constitute only necessary conditions for compatibility.

We can, thus, estimate the volume of the Evans non-signaling set $\mathcal{N}(E)$ inside the simplex of probability distributions, see Table 5.4, which includes also results using different numbers of outputs for each party. Naively, one might believe that non-signaling is not relevant in the presence of signals between parties of the network and in the absence of inputs, but this is indeed not true as we can see that, approximately, the non-signaling set represents 84.94% of the volume relative to the simplex of all probability distributions. Notice also how the volume increases as we increase the number of outputs on Bob's side, whilst a similar change on the other parties causes the opposite effect.

Naturally, we also investigated the volume of the classically compatible distribution in the Evans scenario $\mathcal{C}(E)$ and, remarkably, found a very small gap ($\approx 0.12\%$) of NSI-compatible non-classical distributions in the minimal case where all variables are bits. Fig. 5.5a shows the distribution of the distances to the classical set. All non-classical distributions that we have found cannot be ruled out by quantum inflation up to $2nd$ order inflation and NPA level 3, therefore we cannot tell that these points are truly post-quantum with our current techniques. Even the classical version of the inflation technique is unable to exclude any of the non-classical points we found. This opens the interesting possibility (with actual candidate probability distributions) that a classical-quantum gap

$\lvert A\rvert$	$\lvert B\rvert$	$\lvert C\rvert$	Volume
2	2	2	0.8494
2	3	2	0.9823
3	2	3	0.7475

Table 5.4: Volume of the non-signaling set $\mathcal{N}(E)$ within the simplex of valid probability distributions in the Evans scenario for different numbers of outputs for each party. The number of samples considered in each case was of the order of 10^5.

exists in the Evans scenario, a question that remains open [Lau+23].

In particular, starting from these points, we can propose the candidate distribution

$$\begin{aligned} &P_{NS}(1,0,0) = 2/81, \quad P_{NS}(0,0,1) = 1/55 \\ &P_{NS}(0,1,0) = 1/11, \quad P_{NS}(1,0,1) = 1/5 \\ &P_{NS}(1,1,0) = P_{NS}(0,1,1) = 1/81, \\ &P_{NS}(1,1,1) = 1/2\sqrt{2}, \\ &P_{NS}(0,0,0) = 1 - \sum_{a,b,c\neq 0,0,0} P_{NS}(a,b,c) \end{aligned} \tag{5.21}$$

which satisfies the NSI test (5.20) and can be certified to be non-classical with QP with a corresponding witness given by

$$W := \sum_{a,b,c} (p(a,b,c) - P_{NS}(a,b,c))^2 \geq \frac{1}{36.853}. \tag{5.22}$$

Moreover, we analyzed what happens if one considers additional interventional data $p(a,c|do(b))$. For the particular case of the Evans causal structure, it is sufficient to provide only $p(a|do(b))$ and $p(c|do(b))$, since $p(a,c|do(b)) = p(a|do(b))p(c|do(b))$. To do so, we use the interruption technique to map valid non-signaling probability distributions in the bilocal scenario to valid hybrid data tables in the Evans scenario. This is done by considering only $p_B(a,b,c|x=z=b)$ and the marginals $p_B(a|x)$,$p_B(c|z)$ and identifying them with $p(a,b,c)$, $p(a|do(b=x))$ and $p(c|do(b=z))$ respectively.

Similar to the observable case, we can ask what portion of these valid non-signaling hybrid data tables are classically achievable. Remarkably, we found that interventions increase the power to detect non-classicality by two orders of magnitude. Indeed, the non-classical volume, in the case where all variables are binary, increases to 14.6% relative to the number of non-signaling distributions sampled, as opposed to 0.12% using only passive observations. We also look at how these points are distributed relative to their trace distances inside the NSI set, see Fig. 5.5b.

5.4 Discussion

From a modern viewpoint, Bell's theorem can be seen as an instance of a causal inference problem, more precisely causal compatibility where we impose a given causal structure on a quantum experiment and ask whether a classical causal model can explain the observed correlations. This simple yet powerful realization led to a number of generalizations

of Bell's theorem for causal networks of growing size and complexity, showing, for instance, the emergence of non-classical behavior even without the need for measurement choices [Fri12; Ren+19a; Pol+23; Cha+21] or by allowing time-like rather than space-like correlation scenarios [Cha+18]. Of particular relevance, are causal networks composed of independent sources of correlations, scenarios that unveiled new features such as the possibility of self-testing all entangled quantum states [Šup+23] and quantum theory itself [WC20], activation of non-classical behaviour [Poz+19; Pod+20], refined notions of multipartite non-classicality [PGT22; Sup+22] and novel tests of the role of complex numbers in quantum mechanics [Ren+21].

A basic problem within this context is that of characterizing the sets of correlations allowed by each causal structure according to classical, quantum, and non-signaling theories. A problem that, differently from the standard Bell scenario, relies on polynomial causal constraints that impose a non-convex structure to the set of allowed probability distributions compatible with a given causal network. Recently a number of different tools have been proposed to approach this problem but even simple causal networks still have a very fragmented and partial characterization. Moreover, it is unclear how effective those different methods are in witnessing non-classicality. To obtain a more coherent and global picture of both the sets of correlations as well as the tools designed to address them, we analyzed the volumes of such different sets of correlations.

Considering the simplest bilocality scenario and using QP we obtained that only 2.4% of the NSI correlations are non-local while 32.6% are non-bilocal, thus showing that the ability to witness non-classicality is significantly enhanced if we take into account the independence of the sources. Furthermore, the distribution of distances of the non-classical points to the local and bilocal sets show an exponential decay, meaning that most of them are concentrated close to the classical sets. In comparison, the best results we obtained with the inflation technique – corresponding to a 3rd order inflation level – provide a lower bound of 13.3% for the volume of the non-bilocal set, meaning that more than half of the non-bilocal points are not detected by this method.

In the Evans scenario, we have shown that NSI correlations occupy a significant volume of the simplex set (the set of all probability distributions), in some cases surpassing 98%. Surprisingly, only 0.12% of those NSI correlations are actually non-classical. In comparison, the inflation technique again shows relatively poor performance, being unable to detect any of these non-classical points (up to the level of the hierarchy we could handle numerically). Remarkably, however, the volume of non-classical correlations is increased to 14.6% when we consider also the effect of interventions in the Evans scenario, a clear signature that interventional data can enhance significantly our ability to witness non-classicality.

For the triangle network, since it involves third-order polynomial constraints, we cannot directly use QP and, for this reason, we have employed and compared four specific tools: the inflation technique, the covariance test, the Finner inequality, and Shannon-type entropic inequalities. Once more, the 3rd inflation level was the best approximation we achieved with the inflation technique, lower bounding the volume of non-classical correlations to 11.3^*% of the total set of tripartite probability distributions. Interestingly, the lower bound on the volume of post-quantum correlations is 10.98%, pointing out that only a small fraction of the non-classical correlations in the triangle scenario might have a quantum description. In comparison, the Finner inequality detects a volume of 1.91% of non-classical correlations while the covariance approach and the entropic inequalities lead to 3.69% and 0.095%, respectively.

In summary, our approach unveils a few interesting features of the non-convex sets of correlations of causal networks of relevance in the literature. On the positive side, it shows that taking into account the independence of the sources as well as interventional data can greatly improve the volume of non-classical correlations, enhancing our ability to witness them. In turn, the inflation technique, the most general tool at the disposal, cannot detect a significant portion of non-classical correlations, at least the approximation level that was computationally accessible.

This shows that new tools might be needed to advance our understanding of such networks and the non-classical features they entail. One interesting possibility is to adapt the inflation method to generate quadratic constraints only, which could then be efficiently handled by a quadratic optimizer. Another relevant direction is to use these results as the starting point to solve open questions. For instance, we have detected non-classical points in the simplest Evans and triangle scenarios that nonetheless pass the test of quantum inflation. Those are good candidates for possible quantum violations of the causal constraints imposed by such networks and we hope our work motivates further research in those directions.

We provide the codes to reproduce our results at the link [CLP23].

Final Remarks

This dissertation summarizes the results and the learning process of the author along his graduate's program. As it might be clear at this point, the main idea was to present the basic knowledge underlying the research conducted throughout these years and the main contributions derived from it. But not only, while doing so we have made the effort to provide most of the possible connections to the topic that can be found in the literature. In this way, we hope to have provided what could be used by the reader to construct a mental map of contextuality.

Having that in mind, throughout these pages we have tried to cover from the basics aspects of contextuality, including its origins and typical approaches and related debates, as well as the more technical aspects of it, namely mathematical subtleties, experimental challenges, but mainly the characterization side of the phenomenon and the use of computers to approach it.

Moreover, we have discussed three contributions made by this author along the program. One of them, a mostly mathematical result concerning the relation between quantifiers of contextuality for a specific class of scenarios, but that could help with the understanding of the structure of the quantum resource in the future. In the second of these works, we explored the idea of using quantifiers for non-locality of quantum states, following the typical construction from quantifiers for behaviours, to address important questions on the relation between entanglement and non-locality. And the last work, on the volume of correlations sets in causal networks, we addressed the problem of understanding the relative volume among typical sets of correlations in scenarios of great importance in the study of quantum causality nowadays, showing that by considering independent sources and interventions we can improve our ability to detect non-classicality.

To finish, we remark one more time that this dissertation has not been planned to work as a first reading for someone unfamiliar with the subject. On the contrary, we expect that previously motivated readers (probably students) find themselves even more motivated by now, hopefully with a roadmap that could guide them to whatever be the approach or technique of interest they wish to learn more about.

Bibliography

[AB11] Samson Abramsky and Adam Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality". In: *New Journal of Physics* 13.11 (2011), p. 113036.

[Åbe+20] Johan Åberg et al. "Semidefinite tests for quantum network topologies". In: *Physical Review Letters* 125.11 (2020), p. 110505.

[ABM17] Samson Abramsky, Rui Soares Barbosa, and Shane Mansfield. "Contextual fraction as a measure of contextuality". In: *Physical review letters* 119.5 (2017), p. 050504.

[Abr13] Samson Abramsky. "Relational databases and Bell's theorem". In: *In search of elegance in the theory and practice of computation*. Springer, 2013, pp. 13–35.

[AC18] Barbara Amaral and Marcelo Terra Cunha. *On graph approaches to contextuality and their role in quantum theory*. Springer, 2018.

[ACC14] Barbara Amaral, Marcelo Terra Cunha, and Adán Cabello. "Exclusivity principle forbids sets of correlations larger than the quantum set". In: *Physical Review A* 89.3 (2014), p. 030101.

[Aci+02] Antonio Acin et al. "Quantum nonlocality in two three-level systems". In: *Physical Review A* 65.5 (2002), p. 052325.

[AD19a] Barbara Amaral and Cristhiano Duarte. "Characterizing and quantifying extended contextuality". In: *Physical Review A* 100.6 (2019), p. 062103.

[AD19b] Barbara Amaral and Cristhiano Duarte. "Graph Approach to Extended Contextuality". In: *arXiv preprint arXiv:1902.02413* (2019).

[ADO18] Barbara Amaral, Cristhiano Duarte, and Roberto I Oliveira. "Necessary conditions for extended noncontextuality in general sets of random variables". In: *Journal of Mathematical Physics* 59.7 (2018), p. 072202.

[ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger. "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers". In: *Phys. Rev. Lett.* 49 (25 Dec. 1982), pp. 1804–1807.

[AGR81] Alain Aspect, Philippe Grangier, and Gérard Roger. "Experimental Tests of Realistic Local Theories via Bell's Theorem". In: *Phys. Rev. Lett.* 47 (7 Aug. 1981), pp. 460–463.

[AGR82] Alain Aspect, Philippe Grangier, and Gérard Roger. "Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities". In: *Phys. Rev. Lett.* 49 (2 July 1982), pp. 91–94.

[Ahr+13] Johan Ahrens et al. "Two fundamental experimental tests of nonclassicality with qutrits". In: *Scientific Reports* 3.1 (2013), p. 2170.

[AK01] Joshua D Angrist and Alan B Krueger. "Instrumental variables and the search for identification: From supply and demand to natural experiments". In: *Journal of Economic perspectives* 15.4 (2001), pp. 69–85.

[Ama19] Barbara Amaral. "Resource theory of contextuality". In: *Philosophical Transactions of the Royal Society A* 377.2157 (2019), p. 20190010.

[Apo69] Tom M. Apostol. *Calculus Volume II - Multi-Variable Calculus and Linear Algebra, with Applications to Differential Equations and Probability*. John Wiley & Sons, 1969.

[BAC18] Samuraí Gomes de Aguiar Brito, B Amaral, and R Chaves. "Quantifying Bell nonlocality with the trace distance". In: *Physical Review A* 97.2 (2018), p. 022111.

[BC17] Jonatan Bohr Brask and Rafael Chaves. "Bell scenarios with communication". In: *Journal of Physics A: Mathematical and Theoretical* 50.9 (2017), p. 094001.

[BC91] Gilles Brassard and Claude Crépeau. "Quantum bit commitment and coin tossing protocols". In: *Advances in Cryptology-CRYPTO'90: Proceedings 10*. Springer. 1991, pp. 49–61.

[Bel64] John S Bell. "On the einstein podolsky rosen paradox". In: *Physics Physique Fizika* 1.3 (1964), p. 195.

[BEL66] JOHN S. BELL. "On the Problem of Hidden Variables in Quantum Mechanics". In: *Rev. Mod. Phys.* 38 (3 July 1966), pp. 447–452.

[Ber+17] Juan Bermejo-Vega et al. "Contextuality as a resource for models of quantum computation with qubits". In: *Physical review letters* 119.12 (2017), p. 120505.

[Bez+17] Jeff Bezanson et al. "Julia: A fresh approach to numerical computing". In: *SIAM Review* 59.1 (2017), pp. 65–98.

[BGP10] Cyril Branciard, Nicolas Gisin, and Stefano Pironio. "Characterizing the nonlocal correlations created via entanglement swapping". In: *Physical review letters* 104.17 (2010), p. 170401.

[BHK05] Jonathan Barrett, Lucien Hardy, and Adrian Kent. "No Signaling and Quantum Key Distribution". In: *Phys. Rev. Lett.* 95 (1 2005), p. 010503.

[BLM96] Paul Busch, Pekka J Lahti, and Peter Mittelstaedt. *The quantum theory of measurement*. Springer, 1996.

[BLO19] Jonathan Barrett, Robin Lorenz, and Ognyan Oreshkov. "Quantum causal models". In: *arXiv preprint arXiv:1906.10726* (2019).

[BLW14] Paul Busch, Pekka Lahti, and Reinhard F Werner. "Colloquium: Quantum root-mean-square error and measurement uncertainty relations". In: *Reviews of Modern Physics* 86.4 (2014), p. 1261.

[BM10] Costantino Budroni and Giovanni Morchio. "The extension problem for partial Boolean structures in quantum mechanics". In: *Journal of mathematical physics* 51.12 (2010), p. 122205.

[Boh52] David Bohm. "A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I". In: *Phys. Rev.* 85 (2 Jan. 1952), pp. 166–179.

[BR21] Salman Beigi and Marc-Olivier Renou. "Covariance decomposition as a universal limit on correlations in networks". In: *IEEE Transactions on Information Theory* 68.1 (2021), pp. 384–394.

[Bri+98] H-J Briegel et al. "Quantum repeaters: the role of imperfect local operations in quantum communication". In: *Physical Review Letters* 81.26 (1998), p. 5932.

[Bru+14] N. Brunner et al. "Bell nonlocality". In: *Rev. Mod. Phys.* 86 (2 2014), pp. 419–478.

[Bru+15] Peter D Bruza et al. "A probabilistic framework for analysing the compositionality of conceptual combinations". In: *Journal of Mathematical Psychology* 67 (2015), pp. 26–38.

[Bud+21] Costantino Budroni et al. "Quantum contextuality". In: *arXiv preprint arXiv:2102.13036* (2021).

[Bus+16] Paul Busch et al. *Quantum measurement*. Vol. 23. Springer, 2016.

[BV06] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2006.

[BWP23] Emanuel-Cristian Boghiu, Elie Wolfe, and Alejandro Pozas-Kerstjens. "Inflation: a Python library for classical and quantum causal compatibility". In: *Quantum* 7 (May 2023), p. 996. ISSN: 2521-327X.

[Cab05] Adán Cabello. "How much larger quantum correlations are than classical ones". In: *Physical Review A* 72.1 (2005), p. 012113.

[Cab12] Adán Cabello. "Specker's fundamental principle of quantum mechanics". In: *arXiv preprint arXiv:1212.1756* (2012).

[Cab13] Adán Cabello. "Simple explanation of the quantum violation of a fundamental inequality". In: *Physical review letters* 110.6 (2013), p. 060402.

[Cam+23] Giulio Camillo et al. "Estimating the volumes of correlations sets in causal networks". In: *arXiv preprint arXiv:2311.08574* (2023).

[CC23] Giulio Camillo and Víctor H Cervantes. "Measures of contextuality in cyclic systems and the negative probabilities measure CNT3". In: *arXiv preprint arXiv:2305.16574* (2023).

[CD18] Víctor H Cervantes and Ehtibar N Dzhafarov. "Snow queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices." In: *Decision* 5.3 (2018), p. 193.

[CD20] Víctor H. Cervantes and Ehtibar N. Dzhafarov. "Contextuality analysis of impossible figures". In: *Entropy* 22.9 (2020), p. 981. ISSN: 1099-4300.

[CD23] Victor H Cervantes and Ehtibar N Dzhafarov. "Hypercyclic systems of measurements". In: *arXiv preprint arXiv:2304.01155* (2023).

[CDP10] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. "Probabilistic theories with purification". In: *Phys. Rev. A* 81 (6 June 2010), p. 062348.

[CEG96] Adán Cabello, JoséM Estebaranz, and Guillermo García-Alcaine. "Bell-Kochen-Specker theorem: A proof with 18 vectors". In: *Physics Letters A* 212.4 (1996), pp. 183–187.

[Cer23] Víctor H. Cervantes. "A note on the relation between the Contextual Fraction and CNT_2". In: *J. Math. Psychol.* 112 (2023), p. 102726. ISSN: 00222496.

[CG04] Daniel Collins and Nicolas Gisin. "A relevant two qubit Bell inequality inequivalent to the CHSH inequality". In: *Journal of Physics A: Mathematical and General* 37.5 (2004), p. 1775.

[Cha+14] R Chaves et al. In: *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2014.

[Cha+15] Rafael Chaves et al. "Unifying framework for relaxations of the causal assumptions in Bell's theorem". In: *Physical review letters* 114.14 (2015), p. 140403.

[Cha+18] Rafael Chaves et al. "Quantum violation of an instrumental test". In: *Nature Physics* 14.3 (2018), pp. 291–296.

[Cha+21] Rafael Chaves et al. "Causal networks and freedom of choice in bell's theorem". In: *PRX Quantum* 2.4 (2021), p. 040323.

[Cha16] Rafael Chaves. "Polynomial bell inequalities". In: *Physical review letters* 116.1 (2016), p. 010402.

[Che08] Chih-Kai Chen. "Causal modeling of knowledge-based economy". In: *Management Decision* 46.3 (2008), pp. 501–514.

[Cla+69] John F Clauser et al. "Proposed experiment to test local hidden-variable theories". In: *Physical review letters* 23.15 (1969), p. 880.

[CLP23] Giulio Camillo, Pedro Lauand, and Davide Poderini. *Relative Volumes Tripartite*. https://github.com/Giuhcs/relative_volumes_tripartite. 2023.

[CMG15] Rafael Chaves, Christian Majenz, and David Gross. "Information–theoretic implications of quantum causal structures". In: *Nature communications* 6.1 (2015), p. 5766.

[Col+02] Daniel Collins et al. "Bell inequalities for arbitrarily high-dimensional systems". In: *Physical review letters* 88.4 (2002), p. 040404.

[CS16] Fabio Costa and Sally Shrapnel. "Quantum causal modelling". In: *New Journal of Physics* 18.6 (2016), p. 063032.

[CSW10] Adán Cabello, Simone Severini, and Andreas Winter. "(Non-) contextuality of physical theories as an axiom". In: *arXiv preprint arXiv:1010.2163* (2010).

[CSW14] Adán Cabello, Simone Severini, and Andreas Winter. "Graph-theoretic approach to quantum correlations". In: *Physical review letters* 112.4 (2014), p. 040401.

[CWR21] Xavier Coiteux-Roy, Elie Wolfe, and Marc-Olivier Renou. "No bipartite-nonlocal causal theory can explain nature's correlations". In: *Physical review letters* 127.20 (2021), p. 200401.

[DCP17] Giacomo Mauro D'Ariano, Giulio Chiribella, and Paolo Perinotti. *Quantum theory from first principles: an informational approach*. Cambridge University Press, 2017.

[Del+15] Nicolas Delfosse et al. "Wigner function negativity and contextuality in quantum computation on rebits". In: *Physical Review X* 5.2 (2015), p. 021003.

[DEP20a] Giacomo Mauro D'Ariano, Marco Erba, and Paolo Perinotti. "Classical theories with entanglement". In: *Physical Review A* 101.4 (2020), p. 042118.

[DEP20b] Giacomo Mauro D'Ariano, Marco Erba, and Paolo Perinotti. "Classicality without local discriminability: Decoupling entanglement and complementarity". In: *Physical Review A* 102.5 (2020), p. 052216.

[DFM92] Martin E. Dyer, Zoltan Füredi, and Colin McDiarmid. "Volumes spanned by random points in the hypercube". In: *Random Structures & Algorithms* 3.1 (1992), pp. 91–106.

[DHL17] Iain Dunning, Joey Huchette, and Miles Lubin. "JuMP: A Modeling Language for Mathematical Optimization". In: *SIAM Review* 59.2 (2017), pp. 295–320.

[DK16a] Ehtibar N Dzhafarov and Janne V Kujala. "Context–content systems of random variables: The Contextuality-by-Default theory". In: *Journal of Mathematical Psychology* 74 (2016), pp. 11–33.

[DK16b] Ehtibar N Dzhafarov and Janne V Kujala. "Contextuality-by-Default 2.0: Systems with binary random variables". In: *International Symposium on Quantum Interaction*. Springer. 2016, pp. 16–32.

[DK23a] Ehtibar Dzhafarov and Janne V Kujala. "Traditional vs extended contextuality: Impossibility of accepting one while rejecting the other". In: *arXiv preprint arXiv:2302.11995* (2023).

[DK23b] Ehtibar N Dzhafarov and Janne V Kujala. "Contextuality with disturbance and without: Neither can violate substantive requirements the other satisfies". In: *Entropy* 25.4 (2023), p. 581.

[DKC20] Ehtibar N Dzhafarov, Janne V Kujala, and Víctor H Cervantes. "Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems". In: *Physical Review A* 101.4 (2020), p. 042119.

[Dua+18] Cristhiano Duarte et al. "Concentration phenomena in the geometry of Bell correlations". In: *Physical Review A* 98.6 (2018), p. 062114.

[Dzh16] Ehtibar N Dzhafarov. "Stochastic unrelatedness, couplings, and contextuality". In: *Journal of Mathematical Psychology* 75 (2016), pp. 34–41.

[Dzh22] Ehtibar N Dzhafarov. "Contents, contexts, and basics of contextuality". In: *From Electrons to Elephants and Elections*. Springer, 2022, pp. 259–286.

[Dzh23] Ehtibar N. Dzhafarov. "The Contextuality-by-Default view of the Sheaf-Theoretic approach to contextuality". In: *Samson Abramsky on Logic and Structure in Computer Science and Beyond*. Ed. by Alessandra Palmigiano and Mehrnoosh Sadrzadeh. Outstanding Contributions to Logic. Dordrecht: Springer Nature, 2023.

[Ebe93] Philippe H Eberhard. "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment". In: *Physical Review A* 47.2 (1993), R747.

[Eke91] Artur K. Ekert. "Quantum cryptography based on Bell's theorem". In: *Phys. Rev. Lett.* 67 (6 1991), pp. 661–663.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. "Can quantum-mechanical description of physical reality be considered complete?" In: *Physical review* 47.10 (1935), p. 777.

[Eva16] Robin J Evans. "Graphs for margins of Bayesian networks". In: *Scandinavian Journal of Statistics* 43.3 (2016), pp. 625–648.

[Eve57] Hugh Everett. ""Relative State" Formulation of Quantum Mechanics". In: *Rev. Mod. Phys.* 29 (3 July 1957), pp. 454–462.

[FC12] Tobias Fritz and Rafael Chaves. "Entropic inequalities and marginal problems". In: *IEEE transactions on information theory* 59.2 (2012), pp. 803–817.

[FC72] Stuart J. Freedman and John F. Clauser. "Experimental Test of Local Hidden-Variable Theories". In: *Phys. Rev. Lett.* 28 (14 Apr. 1972), pp. 938–941.

[For+22] John Forrest et al. *coin-or/Clp: Release releases/1.17.7*. Version releases/1.17.7. Jan. 2022.

[FP15] EA Fonseca and Fernando Parisio. "Measure of nonlocality which is maximal for maximally entangled qutrits". In: *Physical Review A* 92.3 (2015), p. 030101.

[Fri+13] Tobias Fritz et al. "Local orthogonality as a multipartite principle for quantum correlations". In: *Nature communications* 4.1 (2013), pp. 1–7.

[Fri04] Nir Friedman. "Inferring cellular networks using probabilistic graphical models". In: *Science* 303.5659 (2004), pp. 799–805.

[Fri12] Tobias Fritz. "Beyond Bell's theorem: correlation scenarios". In: *New Journal of Physics* 14.10 (2012), p. 103001.

[Gha10] Sevag Gharibian. "Strong NP-Hardness of the Quantum Separability Problem". In: *Quantum Info. Comput.* 10.3 (Mar. 2010), pp. 343–360. ISSN: 1533-7146.

[Gis+20] Nicolas Gisin et al. "Constraints on nonlocality in networks from no-signaling and independence". In: *Nature communications* 11.1 (2020), p. 2378.

[Gis91] Nicolas Gisin. "Bell's inequality holds for all non-product states". In: *Physics Letters A* 154.5-6 (1991), pp. 201–202.

[GMC20] Mariami Gachechiladze, Nikolai Miklin, and Rafael Chaves. "Quantifying causal influences in the presence of a quantum common cause". In: *Physical Review Letters* 125.23 (2020), p. 230401.

[GO11] Anupam Gupta and Ryan O'Donnell. "Lecture notes for CMU's course on Linear Programming & Semidefinite Programming". In: 2011.

[Gru+14] Andrzej Grudka et al. "Quantifying contextuality". In: *Phys. Rev. Lett.* 112.12 (2014), p. 120401.

[GSS05] Luis David Garcia, Michael Stillman, and Bernd Sturmfels. "Algebraic geometry of Bayesian networks". In: *Journal of Symbolic Computation* 39.3-4 (2005), pp. 331–355.

[Gup+22] Shashank Gupta et al. "Quantum contextuality provides communication complexity advantage". In: *arXiv preprint arXiv:2205.03308* (2022).

[Gur03] Leonid Gurvits. "Classical Deterministic Complexity of Edmonds' Problem and Quantum Entanglement". In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 10–19. ISBN: 1581136749.

[Gur22] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2022. URL: https://www.gurobi.com.

[Hal10] Michael JW Hall. "Local deterministic model of singlet state correlations based on relaxing measurement independence". In: *Physical review letters* 105.25 (2010), p. 250404.

[HHH96] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. "Separability of mixed states: necessary and sufficient conditions". In: *Physics Letters A* 223.1 (1996), pp. 1–8. ISSN: 0375-9601.

[HLP14] Joe Henson, Raymond Lal, and Matthew F Pusey. "Theory-independent limits on correlations from generalized Bayesian networks". In: *New Journal of Physics* 16.11 (2014), p. 113043.

[How+14] Mark Howard et al. "Contextuality supplies the 'magic'for quantum computation". In: *Nature* 510.7505 (2014), pp. 351–355.

[HW10] Teiko Heinosaari and Michael M Wolf. "Nondisturbing quantum measurements". In: *Journal of mathematical physics* 51.9 (2010), p. 092201.

[Jer+16] Markus Jerger et al. "Contextuality without nonlocality in a superconducting quantum system". In: *Nature communications* 7.1 (2016), pp. 1–6.

[Ji+21] Zhengfeng Ji et al. "Mip*= re". In: *Communications of the ACM* 64.11 (2021), pp. 131–138.

[Jon19] Matt Jones. "Relating causal and probabilistic approaches to contextuality". In: *Philosophical Transactions of the Royal Society A* 377.2157 (2019), p. 20190133.

[Kar84] Narendra Karmarkar. "A new polynomial-time algorithm for linear programming". In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. 1984, pp. 302–311.

[KD16] Janne V Kujala and Ehtibar N Dzhafarov. "Proof of a conjecture on contextuality in cyclic systems with binary variables". In: *Foundations of Physics* 46.3 (2016), pp. 282–299.

[KD19] Janne V Kujala and Ehtibar N Dzhafarov. "Measures of contextuality and non-contextuality". In: *Philosophical Transactions of the Royal Society A* 377.2157 (2019), p. 20190149.

[KD22] Janne V Kujala and Ehtibar N Dzhafarov. "Contextuality and dichotomizations of random variables". In: *Foundations of Physics* 52.1 (2022), pp. 1–25.

[KDL15] Janne V Kujala, Ehtibar N Dzhafarov, and Jan-Åke Larsson. "Necessary and sufficient conditions for an extended noncontextuality in a broad class of quantum mechanical systems". In: *Physical Review Letters* 115.15 (2015), p. 150401.

[Kel+19] Aditya Kela et al. "Semidefinite tests for latent causal structures". In: *IEEE Transactions on Information Theory* 66.1 (2019), pp. 339–349.

[KH11] Samantha Kleinberg and George Hripcsak. "A review of causal inference for biomedical informatics". In: *Journal of biomedical informatics* 44.6 (2011), pp. 1102–1112.

[Kim08] H Jeff Kimble. "The quantum internet". In: *Nature* 453.7198 (2008), pp. 1023–1030.

[Kir+09] Gerhard Kirchmair et al. "State-independent experimental test of quantum contextuality". In: *Nature* 460.7254 (2009), pp. 494–497.

[KK92] A Kamath and N Karmarkar. "A continuous approach to compute upper bounds in quadratic maximization problems with integer constraints". In: *Recent Advances in Global Optimization*. 1992, pp. 125–140.

[KK93] Anil P Kamath and Narendra K Karmarkar. "An O (nL) iteration algorithm for computing bounds in quadratic optimization problems". In: *Complexity in Numerical Optimization*. World Scientific, 1993, pp. 254–268.

[Kle+11] Matthias Kleinmann et al. "Memory cost of quantum contextuality". In: *New Journal of Physics* 13.11 (2011), p. 113011.

[Kly+08] Alexander A Klyachko et al. "Simple test for hidden variables in spin-1 systems". In: *Physical review letters* 101.2 (2008), p. 020403.

[KRK12] P Kurzyński, Ravishankar Ramanathan, and Dagomir Kaszlikowski. "Entropic test of quantum contextuality". In: *Physical review letters* 109.2 (2012), p. 020404.

[KS15] Ravi Kunjwal and Robert W Spekkens. "From the Kochen-Specker theorem to noncontextuality inequalities without assuming determinism". In: *Physical review letters* 115.11 (2015), p. 110403.

[KS75] Simon Kochen and Ernst P Specker. "The problem of hidden variables in quantum mechanics". In: *The logico-algebraic approach to quantum mechanics*. Springer, 1975, pp. 293–328.

[Lap+11] Radek Lapkiewicz et al. "Experimental non-classicality of an indivisible quantum system". In: *Nature* 474.7352 (2011), pp. 490–493.

[Lau+23] Pedro Lauand et al. "Witnessing Nonclassicality in a Causal Structure with Three Observable Variables". In: *PRX Quantum* 4.2 (2023), p. 020311.

[Lip+18] Victoria Lipinska et al. "Towards an equivalence between maximal entanglement and maximal quantum nonlocality". In: *New Journal of Physics* 20.6 (2018), p. 063043.

[Lis+14] Petr Lisoněk et al. "Kochen-Specker set with seven contexts". In: *Physical Review A* 89.4 (2014), p. 042101.

[LP86] "7 Matching and Linear Programming". In: *Matching Theory*. Ed. by L. Lovász and M.D. Plummer. Vol. 121. North-Holland Mathematics Studies. North-Holland, 1986, pp. 255–305.

[Maz+16] Michael D Mazurek et al. "An experimental test of noncontextuality without unphysical idealizations". In: *Nature communications* 7.1 (2016), pp. 1–7.

[Mer90] N David Mermin. "Simple unified form for the major no-hidden-variables theorems". In: *Physical review letters* 65.27 (1990), p. 3373.

[Mer93] N David Mermin. "Hidden variables and the two theorems of John Bell". In: *Reviews of Modern Physics* 65.3 (1993), p. 803.

[Moo02] David W Moore. "Measuring new types of question-order effects: Additive and subtractive". In: *The Public Opinion Quarterly* 66.1 (2002), pp. 80–91.

[MS06] André Allan Méthot and Valerio Scarani. "An anomaly of non-locality". In: *arXiv preprint quant-ph/0601210* (2006).

[Naw+13] Mohamed Nawareg et al. "Bounding quantum theory with the exclusivity principle in a two-city experiment". In: *arXiv preprint arXiv:1311.3495* (2013).

[Nes00] Yurii Nesterov. "Squared Functional Systems and Optimization Problems". In: *High Performance Optimization*. Ed. by Hans Frenk et al. Boston, MA: Springer US, 2000, pp. 405–440. ISBN: 978-1-4757-3216-0.

[NN94] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín. "Bounding the Set of Quantum Correlations". In: *Phys. Rev. Lett.* 98 (1 Jan. 2007), p. 010401.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations". In: *New Journal of Physics* 10.7 (2008), p. 073013.

[NW20] Miguel Navascués and Elie Wolfe. "The inflation technique completely solves the causal compatibility problem". In: *Journal of Causal Inference* 8.1 (2020), pp. 70–91.

[Pan+98] Jian-Wei Pan et al. "Experimental entanglement swapping: entangling photons that never interacted". In: *Physical review letters* 80.18 (1998), p. 3891.

[Pat+22] Ari Patrick et al. "From nonlocality quantifiers for behaviors to nonlocality quantifiers for states". In: *arXiv preprint arXiv:2204.06537* (2022).

[Paw+09] Marcin Pawłowski et al. "Information causality as a physical principle". In: *Nature* 461.7267 (2009), pp. 1101–1104.

[PB15] Jacques Pienaar and Časlav Brukner. "A graph-separation theorem for quantum causal models". In: *New Journal of Physics* 17.7 (2015), p. 073020.

[PBR12] Matthew F Pusey, Jonathan Barrett, and Terry Rudolph. "On the reality of the quantum state". In: *Nature Physics* 8.6 (2012), pp. 475–478.

[Pea09] Judea Pearl. *Causality*. Cambridge university press, 2009.

[Per02] Asher Peres. *Quantum theory: concepts and methods*. Springer, 2002.

[Per90] Asher Peres. "Incompatible results of quantum measurements". In: *Physics Letters A* 151.3-4 (1990), pp. 107–108.

[Per91] Asher Peres. "Two simple proofs of the Kochen-Specker theorem". In: *Journal of Physics A: Mathematical and General* 24.4 (1991), p. L175.

[Per92] Asher Peres. "Recursive definition for elements of reality". In: *Foundations of physics* 22.3 (1992), pp. 357–361.

[Per96] Asher Peres. "Separability Criterion for Density Matrices". In: *Phys. Rev. Lett.* 77 (8 Aug. 1996), pp. 1413–1415.

[PGT22] Alejandro Pozas-Kerstjens, Nicolas Gisin, and Armin Tavakoli. "Full network nonlocality". In: *Physical review letters* 128.1 (2022), p. 010403.

[Phy22] The Nobel Committee for Physics. *Scientific Background on the Nobel Prize in Physics 2022*. https://www.nobelprize.org/uploads/2023/10/advanced-physicsprize2022-4.pdf. Accessed on Wed. 23 of Nov. of 2023. 2022.

[Pir03] Stefano Pironio. "Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations". In: *Physical Review A* 68.6 (2003), p. 062102.

[Pit08] Itamar Pitowsky. "Geometry of quantum correlations". In: *Physical Review A* 77.6 (2008), p. 062109.

[Pit91] Itamar Pitowsky. "Correlation polytopes: their geometry and complexity". In: *Mathematical Programming* 50 (1991), pp. 395–414.

[Pod+20] Davide Poderini et al. "Criteria for nonclassicality in the prepare-and-measure scenario". In: *Physical Review Research* 2.4 (2020), p. 043106.

[Pol+23] Emanuele Polino et al. "Experimental nonclassicality in a causal network without assuming freedom of choice". In: *Nature Communications* 14.1 (2023), p. 909.

[Poz+19] Alejandro Pozas-Kerstjens et al. "Bounding the sets of classical and quantum correlations in networks". In: *Physical review letters* 123.14 (2019), p. 140503.

[Poz+23a] Alejandro Pozas-Kerstjens et al. *Post-quantum nonlocality in the minimal triangle scenario*. 2023. arXiv: 2305.03745 [quant-ph].

[Poz+23b] Alejandro Pozas-Kerstjens et al. "Post-quantum nonlocality in the minimal triangle scenario". In: *arXiv preprint arXiv:2305.03745* (2023).

[PR94] Sandu Popescu and Daniel Rohrlich. "Quantum nonlocality as an axiom". In: *Foundations of Physics* 24 (1994), pp. 379–385.

[PSV16] Stefano Pironio, Valerio Scarani, and Thomas Vidick. "Focus on device independent quantum information". In: *New J. Phys* 18.10 (2016), p. 100202.

[PV10] Károly F. Pál and Tamás Vértesi. "Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems". In: *Phys. Rev. A* 82 (2 Aug. 2010), p. 022116.

[Ram+12] Ravishankar Ramanathan et al. "Generalized monogamy of contextual inequalities from the no-disturbance principle". In: *Physical review letters* 109.5 (2012), p. 050404.

[Rau+17] Robert Raussendorf et al. "Contextuality and Wigner-function negativity in qubit quantum computation". In: *Physical Review A* 95.5 (2017), p. 052334.

[Rau13] Robert Raussendorf. "Contextuality in measurement-based quantum computation". In: *Physical Review A* 88.2 (2013), p. 022322.

[Ren+19a] Marc-Olivier Renou et al. "Genuine quantum nonlocality in the triangle network". In: *Physical review letters* 123.14 (2019), p. 140401.

[Ren+19b] Marc-Olivier Renou et al. "Limits on correlations in networks for quantum and no-signaling resources". In: *Physical review letters* 123.7 (2019), p. 070403.

[Ren+21] Marc-Olivier Renou et al. "Quantum theory based on real numbers can be experimentally falsified". In: *Nature* 600.7890 (2021), pp. 625–629.

[Ros+17] Anna de Rosier et al. "Multipartite nonlocality and random measurements". In: *Phys. Rev. A* 96 (1 2017), p. 012101.

[SC23] Paul Skrzypczyk and Daniel Cavalcanti. "Semidefinite Programming in Quantum Information Science". In: *arXiv preprint arXiv:2306.11637* (2023).

[SGS00] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

[SKG13] Jochen Szangolies, Matthias Kleinmann, and Otfried Gühne. "Tests against noncontextual models with measurement disturbances". In: *Physical Review A* 87.5 (2013), p. 050101.

[Sla13] Morton Slater. "Lagrange multipliers revisited". In: *Traces and emergence of nonlinear programming*. Springer, 2013, pp. 293–306.

[Spe+09] Robert W Spekkens et al. "Preparation contextuality powers parity-oblivious multiplexing". In: *Physical review letters* 102.1 (2009), p. 010401.

[Spe05] Robert W Spekkens. "Contextuality for preparations, transformations, and unsharp measurements". In: *Physical Review A* 71.5 (2005), p. 052108.

[Spe07] Robert W Spekkens. "Evidence for the epistemic view of quantum states: A toy theory". In: *Physical Review A* 75.3 (2007), p. 032110.

[Spe08] Robert W Spekkens. "Negativity and contextuality are equivalent notions of nonclassicality". In: *Physical review letters* 101.2 (2008), p. 020401.

[Spe14] Robert W Spekkens. "The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory". In: *Foundations of Physics* 44.11 (2014), pp. 1125–1155.

[Sup+22] Alessia Suprano et al. "Experimental genuine tripartite nonlocality in a quantum triangle network". In: *PRX Quantum* 3.3 (2022), p. 030342.

[Šup+23] Ivan Šupić et al. "Quantum networks self-test all entangled states". In: *Nature Physics* (2023), pp. 1–6.

[Sza15] Jochen Szangolies. *Testing quantum contextuality: the problem of compatibility*. Springer, 2015.

[Tav+22] Armin Tavakoli et al. "Bell nonlocality in networks". In: *Reports on Progress in Physics* 85.5 (Mar. 2022), p. 056001.

[Tav+23] Armin Tavakoli et al. "Semidefinite programming relaxations for quantum correlations". In: *arXiv preprint arXiv:2307.02551* (2023).

[Tez+22] Alisson Tezzin et al. "Impossibility theorem for extending contextuality to disturbing systems". In: *arXiv preprint arXiv:2212.06976* (2022).

[Um+13] Mark Um et al. "Experimental certification of random numbers via quantum contextuality". In: *Scientific reports* 3.1 (2013), p. 1627.

[UVB19] Roope Uola, Giuseppe Vitagliano, and Costantino Budroni. "Leggett-Garg macrorealism and the quantum nondisturbance conditions". In: *Physical Review A* 100.4 (2019), p. 042117.

[Wan+21] Daphne Wang et al. "On the quantum-like contextuality of ambiguous phrases". In: *arXiv preprint arXiv:2107.14589* (2021).

[Wan21] D. Wang. *In search of true contextuality in natural language*. http://www.psych.purdue.edu/~ehtibar/QCQMB2021/files/0AbstractsQCQMB2021.pdf. Presented in the QCQMB Virtual Workshop. 2021.

[WC20] Mirjam Weilenmann and Roger Colbeck. "Self-testing of physical theories, or, is quantum theory optimal with respect to some information-processing task?" In: *Physical Review Letters* 125.6 (2020), p. 060406.

[WEH18] Stephanie Wehner, David Elkouss, and Ronald Hanson. "Quantum internet: A vision for the road ahead". In: *Science* 362.6412 (2018), eaam9288.

[Wei+98] Gregor Weihs et al. "Violation of Bell's Inequality under Strict Einstein Locality Conditions". In: *Phys. Rev. Lett.* 81 (23 Dec. 1998), pp. 5039–5043.

[Wer89] R. F. Werner. "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model". In: *Phys. Rev. A* 40 (8 1989), pp. 4277–4281.

[Wey52] H. Weyl. "The Elementary Theory of Convex Polyhedra". In: *Contributions to the Theory of Games*. Ed. by H. W. Kuhn and A. W. Tucker. Trans. by H. W. Kuhn. Annals of Mathematics Studies 24. (Translation of Weyl H. 1934. Elementare Theorie der konvexen Polyeder, *Comment. Math. Helvetici* **7**, 290–306). Princeton University Press, 1952, pp. 3–18. ISBN: 978-1-4008-8172-7.

[WJD07] H. M. Wiseman, S. J. Jones, and A. C. Doherty. "Steering, Entanglement, Nonlocality, and the Einstein-Podolsky-Rosen Paradox". In: *Phys. Rev. Lett.* 98 (14 Apr. 2007), p. 140402.

[Wol+21] Elie Wolfe et al. "Quantum inflation: A general approach to quantum causal compatibility". In: *Physical Review X* 11.2 (2021), p. 021043.

[WSF19] Elie Wolfe, Robert W Spekkens, and Tobias Fritz. "The inflation technique for causal inference with latent variables". In: *Journal of Causal Inference* 7.2 (2019).

[WW01a] Reinhard F Werner and Michael M Wolf. "All-multipartite Bell-correlation inequalities for two dichotomic observables per site". In: *Physical Review A* 64.3 (2001), p. 032112.

[WW01b] Reinhard F Werner and Michael M Wolf. "Bell inequalities and entanglement". In: *arXiv preprint quant-ph/0107093* (2001).

[WY12] Elie Wolfe and Susanne F Yelin. "Quantum bounds for inequalities involving marginal expectation values". In: *Physical Review A* 86.1 (2012), p. 012123.

[XCG20] Zhen-Peng Xu, Jing-Ling Chen, and Otfried Gühne. "Proof of the Peres conjecture for contextuality". In: *Physical Review Letters* 124.23 (2020), p. 230401.