Modelling the Human Mental Lexicon Using the Formalism of Quantum Theory

A THESIS SUBMITTED TO

THE SCIENCE AND ENGINEERING FACULTY

OF QUEENSLAND UNIVERSITY OF TECHNOLOGY

IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Mojtaba Aliakbarzadeh

Science and Engineering Faculty

Queensland University of Technology

December 2018

Copyright in Relation to This Thesis

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature: QUT Verified Signature

Date: ~~19~~ 15 Dec 2018

To my Mom and Dad

Abstract

Quantum cognition (QC) is a new branch of science which applies the mathematical structure of quantum mechanics (QM) to better model and understand human behavior for a variety of cognitive phenomena. This thesis focuses on improving the current QC models of language and memory.

QC has delivered a number of models for semantic memory, but to date these have tended to assume projective measurement. Projective measurement for cognitive systems is highly restrictive because it assumes an orthogonal relationship between operators. I intend to relax this assumption, through the use of a positive-operator valued measure (POVM), which is a non-orthogonal measurement. I will make use of the density matrix representation to model ensembles of human subjects in word association experiments. This density matrix will be applicable in the representation of both POVM and projective measurement.

The POVM formulation allows us to reconsider some key terms like compositionality and contextuality of language within a rigorous modern approach. This formulation will facilitate the extension of QC to new conceptual advances. This will create possibilities for new experimental designs based on the generalized Bell inequality, which relates the violations of the inequality to the

complementarity of the observables. This approach allows for a local interpretation of violations of the Bell inequalities, a significant contribution, as to date all explanations have been nonlocal. Based on the POVM formalism, I will use Neumarks dilation theorem to relate the full cognitive state of a subject to a restricted substate which represents only those cognitive processes through which they participate in the experiment.

Moreover, I will use quantum tomography to characterize the unknown state of a cognitive system. Using the insights that I gain from this characterization, I will design new experimental protocols based on repeating projective measurements (or POVM) on similar ensembles of a subject to specify the unknown state of that subject.

In addition to the suggested extensions for the measurement process, this thesis will provide a better technical understanding of contextuality in QC. The progress in this area is being hindered by a considerable ignorance about the fundamental aspects of the original theories and also a lack of new theories. This thesis will highlight some of these fundamental aspects by providing a detailed mathematical description of Bell's inequality and the Kochen-Specker theorem and comparing the operational method and sheaf-theoretical approach of contextuality in physics. This will lead to analyzing the CBD, a generalized probabilistic model, which is mainly known in QC area. I will nominate the operational approach as a plausible candidate to model contextuality in cognition, a new framework that is based on the separate consideration of preparation and measurement processes.

In total, this thesis provides a better understanding of the two main concepts of measurement and contextuality in processes involving memory and language. This helps to consolidate the position of QC as a well-founded branch of mathematical psychology.

Keywords

---

contextuality; operational models; POVM; measurement; preparation; quantum cognition

Acknowledgments

I am truly grateful to my supervisors Dr. Kirsty Kitto and Prof. Peter Bruza for all their professional help and support, their intellectual guidance and advice, and their flexibility for allowing me to not always apply their guidance and advice in my research.

I am also thankful to my teachers and mentors from high school and university who created and accelerated my passion for physics and mathematics in general and for the foundation of quantum mechanics in particular. Notably, Dr. Kurosh Javidan who was the lecturer of modern physics and quantum mechanics in my B.Sc.

I would like to acknowledge the financial support I received during my PhD: the science and engineering faculty scholarship and QUT HDR tuition fee sponsorship.

I would like to gratefully thank Dr. Christian Long (at QUTs academic language and learning service) for his help in writing this thesis.

My deepest gratitude goes to my mother and late father for their endless amounts of love and their constant encouragement and moral support. I am also thankful to my sister and brother and all of my friends for their love and support.

Table of Contents

Nomenclature

Abbreviations

QM	Quantum Mechanics
QC	Quantum Cognition
PVM	Projection-Valued Measure
POVM	Positive-Operator Valued Measure
PI	Parameter Independence
OI	Outcome Independence
KS	Kochen Specker
CBD	Contextuality by Defualt

Symbols		**Chapter**
$\mathbb{A}'$, $\mathbb{A}''$ or ($\mathbb{B}'$, $\mathbb{B}''$)	Observables for two different cases of priming	Ch3
a_k or (b_k)	Eigenvalues	Ch3
$\mathbb{P}_k$	Projectors or (PVM)	Ch3
$\tilde{\mathbb{P}}_k$	POVM	Ch3
M	Measurement	Ch3
P	Preparation	Ch3

$p(k)$	Probability of finding result k	Ch3
$\tilde{R}$	Bivariate POVM	Ch3
(x, y, z)	Cartesian coordinates	Ch3
(r, θ, φ)	The Spherical polar coordinates	Ch3
$(\sigma_x, \sigma_y, \sigma_z)$	Pauli matrices	Ch3
γ	Probability for priming with the sense $\mathbb{A}'$	Ch3
ρ	Density matrix	Ch3
$\wedge$	Sample space	Ch4
$\mathcal{F}$ or $(\Sigma(\wedge))$	Event space	Ch4
ρ or $(\mu_P(\lambda)$, $h_\Lambda(\lambda))$	Probability measure	Ch4
$\xi_\lambda(k\|M)$	Indicator function	Ch4
λ	Ontic state	Ch4
$R(\lambda)$	Random variables	Ch4
$\mathbb{A}_1$, $\mathbb{A}_2$ or ($\mathbb{B}_1$, $\mathbb{B}_2$)	Observables	Ch4
a_1, a_2 or (b_1, b_2)	Eigenvalues (Random variables)	Ch4
P	Preparation procedure	Ch4
M	Measurement procedure	Ch4
$\mathcal{P}$	Set of preparation procedures	Ch4
$\mathcal{M}$	Set of measurement procedures	Ch4
$\mathcal{K}_M$	Set of outcome k related to M	Ch4
X	Set of measurements	Ch4
U	A subset of X	Ch4

m	A measurement event	Ch4
O	Set of possible outcomes	Ch4
s	A section	Ch4
$\mathcal{P}(X)$	The powerset of X	Ch4
$\mathcal{E}$	Sheaf of events	Ch4
$\mathcal{D}_R\mathcal{E}$	A presheaf of R-distribution	Ch4
$\mathcal{D}_{\mathbb{B}}\mathcal{E}$	A presheaf of Boolean-valued distribution	Ch4
C	Measurement context	Ch4
e_C	Empirical distribution	Ch4
d	Global distribution	Ch4
φ	Propositional variable	Ch4
a^c	Bunch for the context c	Ch4
a_q^c	Random variable for the observable q and context c	Ch4

List of Figures

List of Tables

Chapter 1

Introduction

Quantum cognition (QC) is an interdisciplinary field which uses the mathematical formalism of quantum mechanics (QM) to resolve a range of modeling problems in cognitive science. Consider the conceptual combination MAN KIND. In the free word association experiments conducted by Nelson et al. [2004], human subjects attribute the associate "the human race" to this combination which is different from all associates produced by MAN, or KIND individually. The associate is emergent from the conceptual combination; it cannot be obtained from its constituent concepts. More interestingly, when we change the order of our concepts, the new combination KIND MAN has a meaning that aligns with its constituent concepts. In other words, MAN KIND has a different meaning from KIND MAN, which can be displayed mathematically as

$$S_m S_k \neq S_k S_m. \tag{1.1}$$

Here, MAN is denoted by S_m and KIND by S_k. $S_m S_k$ represents a scenario where MAN is the first concept of the pair cued to subject S, and $S_k S_m$ indicates the inverse situation. We cannot model this equation using natural numbers, as they always obey the commutative property. For example

$$2 \cdot 3 - 3 \cdot 2 = 0. \tag{1.2}$$

However, we can find a similar non-commutative property in QM, and this was used by Heisenberg as a founding mathematical structure for his Uncertainty Principle [Heisenberg, 1927]. This principle demonstrates that one cannot know the value of two physical variables at the same time if they do not commute. The property of commutativity can be detected mathematically using a commutator, which has the following form for two operators $\mathbb{A}$ and $\mathbb{B}$

$$[\mathbb{A}, \mathbb{B}] = \mathbb{A}\mathbb{B} - \mathbb{B}\mathbb{A}. \tag{1.3}$$

When Equation (1.3) equals 0, the operators $\mathbb{A}$ and $\mathbb{B}$ are defined to commute, which means that they can be applied in either order with no change in outcome. This does not always happen, for example, momentum and position in QM do not commute

$$[\hat{p}, \hat{x}] = -i\hbar, \tag{1.4}$$

where $\hbar$ is a physical constant called Planck's constant and $i = \sqrt{-1}$ is an imaginary number.

Returning to Equation (1.1), we can add or subtract the same quantity to both sides of the equation and still maintain the inequality. Considering this algebraic property, we can rearrange Equation (1.1) in the form of the commutator

$$[S_k, S_m] \neq 0. \tag{1.5}$$

Cognitive science struggles to model both emergent associations and the non-commutative behaviour exhibited by humans when they combine concepts.

A mental lexicon refers to the structure of language and the relations between its constituent words. "Free association norms" allow us to explore this relation [Nelson et al., 2004]. For example, a single word MAN has a number of associates (e.g. male), each with a different probability of recall. One can find their free association probability using classical free association experiments. Such an experiment cues subjects with the word MAN and asks them to give the first word that comes to mind. By repeating this experiment over several subjects, a probability distribution can be obtained. Figures 1.1(a) and 1.1(b) show this probability distribution for two example words: MAN and KIND. This data comes from the small world of words free association database [de Deyne et al., 2013].

As shown in Figure (1.1(a)), in the case of MAN, we have different associates that can be categorized in different senses. For example, about 58% of subjects selected the associate of WOMAN, which belongs to the SEX sense of MAN. Other associates like (MALE, BOY, DAD,...) also belong to that sense. We can categorize associates like (HUMAN, PERSON,..) using the alternative sense of BEING. The n different senses of a concept like MAN, inspired by a fundamental principle of superposition in QM, can be modeled as a vector basis in n-dimensional Hilbert space [Bruza et al., 2009, Widdows, 2004a]. In QM, the superposition of basis states represents the linear combination of a particle's

(a) MAN (b) KIND (c) MANKIND

Figure 1.1: Bar graph for the 20 most frequent responses to MAN and KIND and MANKIND [de Deyne et al., 2013].

states. Then by performing measurement, we collapse the superposition state onto one of the basis states. The same thing happens when we ask subjects to interpret the word MAN. They can interpret it based on one of the different basis states corresponding to its senses of SEX, BEING or perhaps something else all together.

Pothos and Busemeyer [2013] state that superposition can characterize the uncertainty or fuzziness in a cognitive system. This uncertainty is more comprehensive than the classical uncertainty which indicates a lack of information about the state of the system. Actually, classical uncertainty occurs due to a lack of information about the answers from an experiment, but in QM uncertainty is between two possible answers [Isham, 2001]. For example, for two states $|a\rangle$ and $|b\rangle$ related to two outcomes, a and b, of an experiment, the state $|\psi\rangle$ represents the quantum superposition of those states:

$$|\psi\rangle = c_a|a\rangle + c_b|b\rangle, \tag{1.6}$$

where $|c_a|^2$ and $|c_b|^2$ are the probabilities of each result a and b respectively.

Returning to the case of MAN, instead of using n-dimensional Hilbert space, we can represent its different senses as only two dimensional space. Selecting the dominant sense as one of two dimensions, I denote the rest of the senses using the second dimension. The dominant sense SEX is the sense that is more likely to be recalled. I represent this with the state $|s\rangle$ and all other senses with the state $|o\rangle$:

$$|MAN\rangle = c_s|s\rangle + c_o|o\rangle. \tag{1.7}$$

When subjects select the first associate that comes to their mind, this superposition state collapses onto one of the basis states, SEX or OTHER. In other words, measurement can give a definite value, in spite of the mentioned uncertainty in the superposition state. Figure (1.2) illustrates the superposition scenario of the concept MAN in two dimensional Hilbert space.

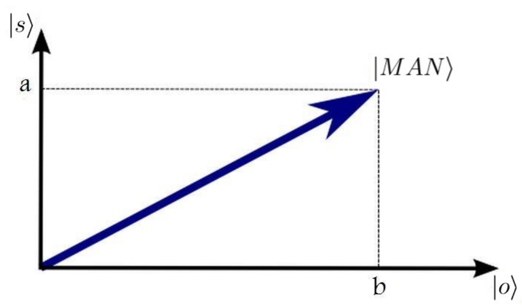

Figure 1.2: Vector representation of MAN with respect to two different bases

Now to measure the probability (membership weight) of a situation (s or o) we can add another question and ask subjects to indicate if they interpret the object in state $|s\ or\ o\rangle$ or not. I predict that this probability is not equal to the sum of probabilities of each state $|s\rangle$ and $|o\rangle$:

$$p(\text{s or o}) = p(s) + p(o) + \text{Interference term} \tag{1.8}$$

Unlike a classical expectation, the probability of (s or o) is not equal to the sum of the probabilities of s and o. This difference is called quantum interference, one of the main properties of QM.

The phenomenon of emergence in a word association experiment will now be illustrated using another quantum property. To describe this phenomenon better, I change the example of conceptual combination to PET HUMAN [1]. This conceptual combination can produces the associate SLAVE which is not produced in PET, or HUMAN separately [Bruza et al., 2012]. So we say the associate SLAVE is emergent from this combination. This emergence property is another aspect of cognitive systems that cannot be described using classical mechanics, in which we cannot decompose composite systems into their subsystems. Based on the vector presentation of these two concepts in 2-dimensional Hilbert space, Bruza et al. [2012] formalize the combination of those two concepts and reveal that it is similar to quantum-entangled (non-separable) particles. In physics, we use the term "entanglement" to describe correlation between particles such that the properties of one state depend on the others; in other words, the particles

[1] We can consider the conceptual combination MAN KIND as a lexicalised word. Hampton [1997] describes the lexicalised words as noun-noun compounds which have converted over time to a single lexical word such as RAILWAY and LIPSTICK. As he states, the study of the relation between the meaning of a lexicalised word and its constituent words can be of interest in historical linguistic rather than in psychology.

lose their individuality. Entanglement is usually epitomized as quantum nonlocality but they are subtly different. Nonlocality arises due to the existence of entanglement phenomena, but there exist some entangled states (like subsets of Werner states) that are local. Nonlocality is completely demonstrated by Bell in his famous theorem [Bell, 1964].

Contextuality is a major reason that we need a new probabilistic model for cognitive science [Bruza et al., 2015b], as there is no analogue for this quantum phenomenon in classical probability theories [Sainz and Wolfe, 2017]. Generally, in both psychology and physics, contextuality refers to the dependence of the measurement result of an observable on the specific experimental situation being used to measure that observable. This experimental situation is known as "context". In psychology, context refers to the effects of external or internal events on a cognitive process [Krank and Wall, 2005]. For example, in the free association experiment of the conceptual combination BOXER BAT, selection of a sense by the subject in interpreting the word BOXER, affects the simultaneous interpretation of the word BAT by that subject. In BOXER BAT, similarly to KIND MAN, each word has different associates with different probabilities of recall. We can categorize these associates for both BOXER and BAT using the "animal" and "sport" senses (for more details see Section 2.1). If BOXER is interpreted for example in the sport sense, there is a higher chance that BAT is interpreted in the same sense as well. So we can relate the interpretation of one word in BOXER BAT to choosing a context for the interpretation of the other word. As Krank and Wall [2005] said "[Choosing a] context implicitly reduces ambiguity, defines the options, and forces choices". Gershenson [2002] similarly believes that all cognitive events happen inside a context and so are determined by that context. On the other hand, context in QM mainly refers to the contextual character of quantum measurements, as demonstrated by the

Kochen-Specker theorem [1967] or Bell's nonlocality [1964]. Nonlocality, which I described earlier, can mostly be considered as a particular case of contextuality, because of its mathematical structure. For example, Acín et al. [2015] constructed a general contextuality model based on the combinatorics of hypergraphs; the special case of this formalism gives rise to the nonlocality of the Bell scenario.

Both Bell's theorem and the Kochen-Specker theorem are considered no-go theorems in QM because they reveal constraints for hidden variable theories. Hidden variable theories were endorsed by physicists such as Einstein et al. [1935], who believed that QM is *incomplete*, meaning that it could not give a complete description of a physical system. They tried to construct a nonrandom or deterministic theory by adding these hidden variables to interpretations of QM; however, Bell and Kochen-Specker (among many others) have put increasingly strict limits on these theories, and shown the impossibility of describing noncontextual hidden variables in QM by retaining the classical probability space [Feintzeig and Fletcher, 2017]. Interestingly, similar frameworks have been used in psychology for different experimental protocols; I will introduce some of them in Chapter 2. However, a new framework has appeared in psychology that interprets contextuality as the impossibility of assigning a single random variable to outcomes of the same observable in different experimental situations [de Barros, J. A., & Oas, 2015, Dzhafarov and Kujala, 2014c]. The definitions of random variables and probability space in this framework differ from the standard definitions of contextuality. I will critically review this approach in Chapter 4.

1.1 Aims

Because of the unique properties of QM, it can be applied to a wide set of cognitive models. In this research project I will focus mainly on cognitive phenomena such as human memory and concepts in language. Topics such as conceptual combination play a fundamental role in the structure of everyday language. Based on currently available quantum models for this domain, I intend to use a more general form of quantum measurement, which will provide a better model of cognition. Moreover I aim to construct a more comprehensive contextuality test for these processes.

So in my research, I will focus upon two broad research aims:

I. To redefine and modify the concept of measurement in quantum cognition.

II. To provide clarity and novel insight regarding contextuality in quantum cognition.

1.2 Research Questions

This thesis will provide a foundation for QC. It will address a number of questions that will elucidate the overall aim by developing the required mathematical structure for modeling cognitive systems. The essential part of this modeling is measurement, which is a tool that assesses a cognitive process. So I ask:

> Research Question 1: *How can the current quantum cognitive models be extended to obtain ensemble data for a collection of subjects?*

In particular,

Research Question 1(a): *How can the current quantum cognitive models include non-ideal and non-orthogonal measurements?*

Much of the work currently occurring in the area of QC assumes a system in a pure state and relies upon projective measurement. This is perhaps not optimal. Cognitive states are not nearly as well behaved as existing cognitive models. Sometimes we need to perform measurements to obtain ensemble data for a collection of subjects. Cognitive states may also exhibit violations of repeatability, and the operators that we use to describe measurements do not appear to be naturally orthogonal in cognitive systems.

Using a representation to model ensembles of human subjects (the preparation process) and clarifying the role of orthogonal and non-orthogonal measurement in QC will help us to deepen the understanding of contextuality.

Research Question 2: *How can the current differences and contradictions between different models of contextuality be resolved?*

To date, none of the research about contextuality in QC has investigated the contextuality in the process of preparation. This limitation can be highlighted as a difference between the current approaches in QC and a recent generalized notation of contextuality in QM [Spekkens, 2005] which is applied to the preparation process. This difference itself emerges from the lack of any clear distinction between preparation and measurement processes in QC. More seriously, there are contradictions between some of these approaches in defining some basic concepts like the probability space and the random variables. My approach will center on the proposition that a better understanding of the measurement process and the probability space of the system will help us to reach a logically consistent realization of contextuality in QC.

1.3 The impact of this research

Quantum cognition is a new research area that helps to explain and understand puzzling aspects of human thinking and memory. The mathematical structure of QM provides a better approach for modeling cognition than traditional probabilistic models.

In this research I focus on conceptual combinations as a fundamental step in the process of understanding language. "Measurement" and "contextuality" are particular aspects of conceptual combination. My research provides a better understanding of these concepts based on the mathematical structures of QM. The general impact of my thesis is to construct a mathematical foundation for describing memory and language. This helps to consolidate the position of QC as a well-founded branch of mathematical psychology.

Chapter 2

Literature Review

As I discussed and demonstrated in Chapter 1, cognitive phenomena manifest processes and properties similar to quantum mechanics (QM). Therefore, several quantum-inspired models have been developed in a wide range of areas to model those cognitive phenomena such as information retrieval or natural language processing (NLP) [Van Rijsbergen, 2004, Widdows, 2004b, Zuccon et al., 2009], decision making [Aerts and Aerts, 1997, Busemeyer and Bruza, 2012, Khrennikov et al., 2014, Pothos and Busemeyer, 2013, Yukalov and Sornette, 2011] and models of language [Aerts and Gabora, 2005, Aerts and Sozzo, 2011, Bruza et al., 2009, 2015a, Clark et al., 2008, Gabora and Aerts, 2002].

As I stated in Section 1.1, this thesis mainly focuses on the cognitive phenomena of human memory and concepts in language. In the last few years, many theories have been proposed to explain this area using QM. As an example, Bruza et al. [2009] suggest a model of the human mental lexicon which is based on word association experiments. I review this approach in the current chapter and Chapter 3. The state-context-property theory ($SCOP$) is another approach to represent concepts in quantum cognition (QC) [Aerts and Gabora, 2005, Aerts and Sozzo, 2011, Gabora and Aerts, 2002]. This theory is inspired by operational QM formalism, in which each concept is associated with sets of "states", "properties" of that concept and "contexts" related to the measurements by which the concept may be observed. $SCOP$ is suitable to model the emergence of meaning when concepts are combined, it is also a proper framework to model quantum effects like "contextuality". In Chapter 1, I briefly introduced Bell's experiment as an important experimental test of contextuality in QM. Bruza et al. [2015a] and Aerts et al. [2013] suggest probabilistic frameworks based on this experiment to model concept combination, these works will be reviewed in Section 2.2. Contextuality-by-default (CBD) [Dzhafarov et al., 2015a, Dzhafarov and Kujala, 2014b, 2015] is another approach proposed in recent years to deal with contextuality in the different branches of cognition, including language. I will provide a critical evaluation of this approach in Chapter 4.

I will divide my literature review into two parts, each of which is based on one of my two research questions[1]: *Part*(I) The needs for more general measurement, *Part*(II) The challenges of current contextual models. This chapter includes all necessary background to establish the significance of this study and identify a place where my contribution will be made. But more details about the existing literature of measurement and contextuality in QC will be provided in Chapters 3 and 4 respectively.

As I mentioned in Chapter 1, there are basic concepts in QM which differentiate it from classical physics and make it suitable to use in cognition. To provide a better idea for readers without a physics background, I will review some of these specifications of QM in boxes to complement my literature review. The topics that I cover in these boxes are quantum interference, POVM, Bell's

[1]Some parts of this chapter have been previously published by the author in a co-authored paper: [Aliakbarzadeh and Kitto, 2016].

theorem and CHSH inequality. Later, in chapters 3 and 4, I will provide detailed descriptions of these concepts in addition to some literature about measurement and contextuality in QM.

2.1 The need for more general measurement

Understanding how words combine to form meaningful phrases and how meanings emerge from such combinations is crucial to understanding the structure of human thinking and memory. As I mentioned in Chapter 1, to investigate the relations between constituent words of language we use ”free association norms”. In this section, I will explain examples of such free association norms with more details, for example the word BAT has different associates (e.g. Ball), each with a different probability of recall which can be found using classical free association experiments. Similar to the example of MAN KIND that I explained in Chapter 1, the experiment cues subjects with the word BAT and asks them to state the first word that comes to mind. By repeating this experiment over several subjects, a probability distribution can be obtained. Tables 2.1(a) and 2.1(b) show this probability distribution for two example words: BAT and BOXER. This data comes from the University of South Florida free association database [Nelson et al., 2004].

As shown in Table 2.1, for BAT, 25 % of subjects selected the associate of BALL, which belongs to the sport sense of BAT. A number of other associates relate to the animal sense (eg. cave, vampire). Bruza et al. [2009, 2015a] associate a word having different senses in memory with a state in Hilbert space where the interaction between different states in memory experiments can be depicted using the features of QM.

Table 2.1: Free association probabilities for the words BAT 2.1(a), and BOXER 2.1(b) [Bruza et al., 2015a].

(a)

Associate	Probability
ball	0.25
cave	0.25
vampire	0.07
fly	0.06
night	0.06
baseball	0.05
blind	0.04
...	...

(b)

Associate	Probability
fighter	0.14
gloves	0.14
fight	0.9
dog	0.8
shorts	0.7
punch	0.5
Tyson	0.5
...	...

Bruza et al. [2009, 2015a] describe the human recall of a concept A as the bivariate variables $\{a_1, a_2\}$ which take values $\{1, -1\}$. Dominant and subordinate *senses* are chosen for A, indicated by the subscripts 1 and 2. When the dominant sense of concept A is first primed, and A is interpreted in that sense by the human subject, then Bruza et al. [2015a] designate $\{a_1 = +1\}$. If A is not interpreted in that sense after priming the dominant sense, we have $\{a_1 = -1\}$. Similarly, $\{a_2 = 1, -1\}$ relates to situations in which the subordinate sense of concept A is primed.

An example can make these notations more clear. As Table 2.1(b) represents the the free association probabilities for the concept BOXER, there is a greater possibility to interpret this concept in the sport sense than the animal sense. So the sport and animal senses are considered as the dominant and subordinate senses, respectively, for that concept. If a subject is first shown the word "glove"and then asked to interpret the concept BOXER, there is a high possibility that the subject will interpret it in a sport sense; this situation is represented as $a_1 = +1$. But if the subject interprets the concept in another sense, the situation is designated as $a_1 = -1$. Conversely, if the subject is first shown the word "vampire", it will awake the animal sense in the mind of the human subject. If the result of subsequent measurement is also in the animal sense, Bruza et al. [2015a] show the situation by $a_2 = +1$. But if the concept is not interpreted in this subordinate sense, they designate it as $a_2 = -1$. This process adopts von Neumann's approach to the quantum measurement of an idealized system. The measurement approach which has been extensively used in QC, e.g., to model human thought [Aerts et al., 2013], to develop a Bell-type experiment for human perception [Conte et al., 2008], or to analyze mental states dynamics [Conte et al., 2007, Khrennikov, 2010].

Von Neumann's approach represents measurement using self-adjoint linear operators, and if we assume that an orthonormal basis exists, then we can write out a Hermitian matrix $\mathbb{A}$ as a series of projection operators $\mathbb{P}_k$ [Bruza et al., 2009]

$$\mathbb{A} = \sum_{k=1}^{n} a_k \mathbb{P}_k. \tag{2.1}$$

where each a_k corresponds to the results of the measurement $\mathbb{A}$ (in the above recall experiment: $k \in \{1, 2\}$).

Repeating this measurement multiple times allows us to calculate an expected value of operator $\mathbb{A}$. However, the state is changed once we have performed the von Neumann measurement. So we need an ensemble of that subject to find a more accurate expected value by repeating the measurement. We try to provide the same experimental conditions in the preparation phase for different subjects, but it is too simplistic to assume that they will all be in exactly the same state $|\psi\rangle$. So we need a mathematical tool to describe measurements on ensembles of subjects with different possible states, which itself needs a better technical description of the von Neumann measurement.

A key problem for projective measurements is that they are too strict in their requirement for repeatability. As an example, consider performing two cueing measurements in a row, using the same word; there is no guarantee that the subject will always respond in the same way. It might be possible that the following sequence occurs:

$$\begin{aligned} &\vdots \\ cue &= SPRING \\ response &= COIL \\ cue &= SPRING \\ response &= LEAF \\ &\vdots \end{aligned}$$

The standard quantum measurement would not be suitable for modeling such a situation. The state has been prepared, and we know that with no intervening measurements to realign the basis states, the probability of returning COIL to the cue SPRING is equal to 1. However, the subject will not always conform to this repeatability requirement; in this case they return LEAF instead of the response predicted by a projective measurement (COIL).

Moreover, we cannot always expect sharp results like the von Neumann measurement for the process of recall. This can happen because of unwanted effects in the process of measurement; e.g. when a human subject does not give a response corresponding to what they actually recalled. To model this situation, that humans are noisy in their responses, we need to use a more generalized form of measurement. A generalized form of measurement has been used in another area of QC (i.e. opinion polling), that can inspire ideas of using generalized measurement in human memory processes. In the following section I will introduce some research in the area of opinion polling that can be a source of this inspiration.

2.1.1 Previous models of POVM

The nature of uncertainty in quantum probabilities, and the special design of QM dealing with random variables, make QM an appropriate choice to apply in opinion polling [Khrennikov et al., 2014]. Although the main interest of this thesis is conceptual combination, in this section, I will review some recent work in the area of opinion polling to present more possible applications of QM in cognition and also to observe how it is possible to use the generalized form of quantum measurement in cognition.

I start with a very simple example, where Pothos and Busemeyer [2013] use probability distributions of outcomes in quantum measurement to model opinion polling experiments. In their experiments a measurement (question) is considered as an observable $\mathbb{A}$ on the states ψ (of the systems on which we perform measurements). This observable $\mathbb{A}$ leads to a set of results $v(\mathbb{A})$ [2]. In this experiment, they consider psychological experiments with dichotomous results (e.g. Yes and No).

Pothos and Busemeyer [2013] provide an example of a hypothetical person who is asked about her state of happiness. For simplicity, they considered the outcomes of the happiness question to be one-dimensional subspaces, meaning, she is definitely happy or is definitely unhappy. Our initial knowledge about the hypothetical person (our knowledge before she makes a decision about the

[2] This is same as the random variable a that we used earlier

question) is indicated by the state vector $|\psi\rangle$; this state vector is a superposition of the two possible decisions $|\psi\rangle = a|happy\rangle + b|unhappy\rangle$

After the person selects "happy" as the answer, the state vector is $|\psi\rangle = |happy\rangle$, and after she selects "unhappy", $|\psi\rangle = |unhappy\rangle$. These two outcomes happen with two probabilities, $|a|^2$ and $|b|^2$ respectively.

Pothos and Busemeyer [2013] extend the situation to when two questions were asked successively; whether the person is happy or not and whether the person is employed or not. Each of these questions has two responses. In classical mechanics, we can associate a joint probability to those four outcomes, but if these two questions are incompatible, we cannot assess them concurrently. Incompatible questions means the results of one question influence the results of the other one. This can be considered as the effect of context or order and can be related to the interference effect in QM.

If we indicate the state of employment of that person as [Pothos and Busemeyer, 2013]

$$|\psi_{employment}\rangle = a'|employed\rangle + b'|notemployed\rangle, \tag{2.2}$$

then we can apply an operator M on this state to detect if the person is happy or not based on her employment state:

$$\begin{aligned} p(\text{happy, unknown employment}) &= \|M.a'|employed\rangle + M.b'|notemployed\rangle\|^2 \\ &= p(happy, employed) + p(happy, nonemployed) \\ &\quad + \text{Interference terms}. \end{aligned} \tag{2.3}$$

The interference term is produced because we square the sum of amplitudes; this property violates the law of total probability (for more details see Box 1).

Box 1. Quantum interference

In physics, interference usually refers to the interaction of two waves that create a superposed wave with greater or weaker amplitude. This property is usually manifested by the double-slit experiment, which was first performed by Thomas Young. In the double-slit experiment, a wave is broken into two parts and each part travels different length paths, then they combine to form a single wave again. But the difference in lengths leads to a phase shift that causes an interference pattern. In the specific version of this experiment, a laser beam is radiated to a plate with two parallel slits, then we can see bright and dark fringes on a screen behind the plate. This interference pattern is because of the wave nature of light as it passes through the two slits. The amplitude of a beam passing each slit is represented by a and b for (ψ_{up} and ψ_{down}) respectively, which is described using the superposition principle: $\psi_{total} = a\psi_{up} + b\psi_{down}$. But the probability of the total beam amplitude is obtained by squaring the modulus of ψ_{total}.

$$p = |\psi_{total}|^2 = |a\psi_{up} + b\psi_{down}|^2 = |a\psi_{up}|^2 + |b\psi_{down}|^2 + \text{Interference term} \tag{2.4}$$

As can be seen, the probability of the total beam is not equal to the sum of probabilities of the beams passing through each slit. This difference is interpreted as interference effect in QM.

Khrennikov and Basieva [2014] investigate the interference effect for two consecutive observables. They defined a quantum-like model based on order probabilities for these two noncommutative observables. The probabilities are expressed in terms of generalized measurement (POVM[3]) (see Box 2). This generalized measurement is used when there are no sharp (yes or no) answers to dichotomous decision observables.

Khrennikov et al. [2014] also analyzed the application of the QM measurement structure in opinion polling. They compared the application of the conventional measurement (PVM[4]) with the POVM for different arrangements of questions. These arrangements were designed in a way that can test "question order effect" and "response (non)replicability". Question order effect entails that the order of asking questions influences the response probabilities frequently. Response replicability indicates a situation in which we repeat a question and receive the same answer for that question every time, even if there are other questions in between.

In this section, I have reviewed two recent works in the area of opinion polling that apply the generalized form of quantum measurement. As I mentioned earlier, these works can be considered as guidelines for the possible application of POVM in human memory. Besides the possible direct application of the existing structures in a new set of problems, it is necessary to modify and extend them. The POVMs used in those reviewed works are restricted to binary outcomes, while it is more natural to consider more than two outcomes for some conditions in cognition (e.g. interpreting a word in more than two senses in the recall experiment). Moreover, neither of these existing works has engaged the generalized measurement of more than one observable. However, there is a joint non-ideal measurement of two observables [de Muynck, 2002] in QM that can be employed to enrich the measurement structures of our cognitive models. In Chapter 3, I will represent an application of extended and modified non-ideal

[3]positive-operator valued measure
[4]projection-valued measure

measurements for human memory.

Box 2. POVM

Measurements in QM have two main roles; they

1. Specify the probabilities related to the different possible measurement outcomes, and
2. Find the post-measurement state of the system.

Quantum measurements are usually represented by a collection of measurement operators $\{\mathbb{A}_k\}$, where the index k indicates possible measurement outcomes for the experiment. If we apply the measurement on a quantum system in the state $|\psi\rangle$, the probability of obtaining result k is

$$p(k) = \langle\psi|\mathbb{A}_k^\dagger \mathbb{A}_k|\psi\rangle, \tag{2.5}$$

and the state after measurement becomes

$$\frac{\mathbb{A}_k|\psi\rangle}{\sqrt{\langle\psi|\mathbb{A}_k^\dagger \mathbb{A}_k|\psi\rangle}}. \tag{2.6}$$

The measurement operators satisfy the completeness equation

$$\sum_k \mathbb{A}_k^\dagger \mathbb{A}_k = I. \tag{2.7}$$

which simply means that their probabilities sum to one. This is the standard law of total probability expressed in terms of linear algebra.

A special important class of the general measurement is known as projective measurement, as I represented in Equation (2.1). In Section 3.2, I will explain projective measurement further, describing its application in a cognitive experiment.

In contrast to projective measurements, the POVM formalism provides a means to obtain the probabilities of a set of measurement outcomes, but it is usually non-projective and non-orthogonal. In Equation (2.5), suppose we define

$$\tilde{\mathbb{P}}_k = \mathbb{A}_k^\dagger \mathbb{A}_k, \qquad \sum_k \tilde{\mathbb{P}}_k = I. \tag{2.8}$$

where $\tilde{\mathbb{P}}$ is a positive operator, then the probability of obtaining the outcome associated with $\tilde{\mathbb{P}}$ would be

$$p(k) = \langle \psi | \tilde{\mathbb{P}} | \psi \rangle, \tag{2.9}$$

The complete set of $\tilde{\mathbb{P}}$ that defines the probabilities of measurement outcomes is known as a POVM. If we use density matrix $\rho = \sum_i |\psi_i\rangle\langle\psi_i|$ (for a pure or a mixed state) the probability of obtaining the outcome is $tr(\tilde{\mathbb{P}}\rho)$.

Those POVMs whose elements are idempotent (meaning that $\tilde{\mathbb{P}}^2 = \tilde{\mathbb{P}}$ for all k), are the subset of the special measurements discussed above (PVM). We can look at the relation between PVM and POVM from the other side; if for our projective measurement we have $\mathbb{P}_i\mathbb{P}_j = \delta_{ij}\mathbb{P}_i$ and $\sum_i \mathbb{P}_i = I$ then $\tilde{\mathbb{P}} = \mathbb{P}_i^\dagger \mathbb{P}_i = \mathbb{P}_i$ [Nielsen and Chuang, 2010].

The post-measurement state of POVM measurement indicates the notion of non-repeatability can be easily incorporated into this framework.

However, POVM has other interesting properties. For example, Naimark's dilation theorem [Gelfand and Neumark, 1943] implies that any POVM can be *lifted* by an operator map to a projection valued measure, which allows us to re-generate the standard representation of measurement. I will talk about an application of this generalized measurement and Naimark's theorem in cognition in Chapter 3.

2.2 The challenge of current contextual models

As I described earlier, to understand the structure of human thinking and semantic networks of language, Bruza et al. [2009, 2015a] associate the state of a subject facing a word to a state in Hilbert space. They explore the relation between these states using the non-local feature of QM (see Boxes 3 and 4). They provide an experimental structure to test the existence of non-local effects between concepts. The effects which indicate the non-compositional behaviors of those concepts. In this model, context is considered as a particular choice of a measuring apparatus as in the quantum realm.

Box 3. Bell theorem

In 1935, Albert Einstein, Boris Podolsky and Nathan Rosen designed an experiment (EPR) claimed to contradict the most common interpretation of QM [Einstein et al., 1935]. The EPR paradox is framed in terms of apparently acceptable assumptions of locality and realism [a]. Later, Bell [1964] provided a mathematical formulation of the assumptions about locality and realism that were used by EPR, the inequality which contradicts the predictions of QM. Experimental tests of the Bell theorem demonstrated that the predictions of QM are correct [Aspect et al., 1982a, Muller et al., 1993, Tittel et al., 1998], which implies that at least one of the locality or realism assumptions is incomplete. Briefly, a violation of Bell's theorem implies that a local realist formalism of QM is impossible.

[a] In physics, realism entails the characterization of a quantum state by some additional hidden parameters (Hidden Variables).

Box 4. CHSH inequality

The CHSH inequality is a form of Bell's theorem that is experimentally realizable. Clauser et al. [1969] derived the CHSH inequality, and its experimental violation [Aspect et al., 1982b] demonstrates that nature cannot be described by local hidden variables theories.

For two distantly separated photons A and B, the CHSH inequality is as follows:

$$-2 \leq E(\mathbb{A}_1, \mathbb{B}_1) + E(\mathbb{A}_1, \mathbb{B}_2) + E(\mathbb{A}_2, \mathbb{B}_1) - E(\mathbb{A}_2, \mathbb{B}_2) \leq 2 \tag{2.10}$$

where $E(\mathbb{A}_i, \mathbb{B}_i)$ is the expectation value of observables $\mathbb{A}_1$ and $\mathbb{A}_2$ of qubit A, and observables $\mathbb{B}_1$ and $\mathbb{B}_2$ of qubit B. The results of measurements with those four observables are labeled by bivalent variables a and b taking values in $\{1, -1\}$. The probability of results for each of these bivalent variables is a function of hidden variables.

This probabilistic statement involves a constraint on the strength of the bipartite statistical correlations, so under an assumption of locality, outcomes of this special function of the probabilities cannot exceed 2. Systems of two entangled quantum particles can violate this inequality; the mathematical formalism of QM predicts this quantity will exceed 2 and can reach a maximum value of $2\sqrt{2}$. This upper limit that bounds quantum correlations was found by mathematician Boris Tsirelson [Cirel'son, 1980]. Some correlations might exist above the Tsirelson bound which are called post-quantum correlations [Popescu and Rohrlich, 1994]. These post-quantum correlations are bounded by the maximum value of 4 for CHSH inequality and are usually described by quantum-like models.

2.2.1 Contextuality and marginal selectivity

Bruza et al. [2009, 2015a] applied the CHSH inequality [(2.10)] to analyze compositionality between two concepts A and B; the behavior of these concepts can be described by bivariate variables $\{a_1, a_2\}$ and $\{b_1, b_2\}$ that take values {1,-1}. Similar to what I described earlier for concept A, we can define the dominant and subordinate *senses* for concepts B, indicated by the numbers 1 and 2. When the dominant sense of concept B is first primed, and B is interpreted in that sense by the human subject, Bruza et al. [2009, 2015a] designate $b_1 = +1$. If B is not interpreted in that sense after priming the dominant sense, we have $b_1 = -1$. Similarly, $b_2 = 1, -1$ relates to situations in which the subordinate sense of concept B is primed.

In the probabilistic view of the conceptual combinational AB, compositionality is concluded if we can have a joint distribution $p(a_1, a_2, b_1, b_2)$, where $p(a_i, b_j)$, $i, j \in \{1, 2\}$ are marginal distributions [Bruza et al., 2015a]. This conclusion is based on Fine's theorem, which defines the necessary and sufficient conditions for the existence of joint probability distribution $p(\mathbb{A}_1, \mathbb{A}_2, \mathbb{B}_1, \mathbb{B}_2)$ for bivalent observables $\mathbb{A}_1, \mathbb{A}_2, \mathbb{B}_1$ and $\mathbb{B}_2$. To fulfill this theorem, the following inequalities (which are considered as Bell/CH inequalities) should be satisfied [Fine, 1982b]:

$$\begin{aligned}
-1 &\leq p(\mathbb{A}_1,\mathbb{B}_1)+p(\mathbb{A}_1,\mathbb{B}_2)+p(\mathbb{A}_2,\mathbb{B}_2)-p(\mathbb{A}_2,\mathbb{B}_1)-p(\mathbb{A}_1)-p(\mathbb{B}_2)\leq 0\\
-1 &\leq p(\mathbb{A}_2,\mathbb{B}_1)+p(\mathbb{A}_2,\mathbb{B}_2)+p(\mathbb{A}_1,\mathbb{B}_2)-p(\mathbb{A}_1,\mathbb{B}_1)-p(\mathbb{A}_2)-p(\mathbb{B}_2)\leq 0\\
-1 &\leq p(\mathbb{A}_1,\mathbb{B}_2)+p(\mathbb{A}_1,\mathbb{B}_1)+p(\mathbb{A}_2,\mathbb{B}_1)-p(\mathbb{A}_2,\mathbb{B}_2)-p(\mathbb{A}_1)-p(\mathbb{B}_1)\leq 0\\
-1 &\leq p(\mathbb{A}_2,\mathbb{B}_2)+p(\mathbb{A}_2,\mathbb{B}_1)+p(\mathbb{A}_1,\mathbb{B}_1)-p(\mathbb{A}_1,\mathbb{B}_2)-p(\mathbb{A}_2)-p(\mathbb{B}_1)\leq 0.
\end{aligned} \tag{2.11}$$

To employ this theorem for the conceptual combinational scenario, Bruza et al. [2015a] substitute the observables with the corresponding random variables of those observables. The results of measurements with observables $\mathbb{A}$ and $\mathbb{B}$ are labeled by bivalent variables a and b respectively, which can take either 1 or -1 as their value. These random variables get values according to the ontic states of the experiment (for more mathematical details see Chapters 3 and 4).

Bruza et al. [2015a] mainly focus on CHSH inequality to test the compositionality of a conceptual combination AB. As with the general form of Fine's theorem, the violation of the CHSH inequality $|CHSH| > 2$ implies the impossibility of constructing the joint probability distribution $p(\mathbb{A}_1,\mathbb{A}_2,\mathbb{B}_1,\mathbb{B}_2)$ from the four joint distributions $p(\mathbb{A}_i,\mathbb{B}_j)$, $i,j \in \{1,2\}$. This violation can be considered as a reliable test for non-compositionality.

Fine's inequalities are based on the assumption of non-signaling in physics. Non-signaling is a result of relativity that states information cannot distribute faster than light. In Bell's experiment (2.10), if qubit A and B are separated by space-like intervals, non-signaling entails that the probability distribution of qubit A's outcomes cannot depend on qubit B's measurement setting, and vice versa. Fine [1982b] makes the same separability assumption; he considers two space-like separated regions S_1 and S_2. This model involves two non-commuting bivalent observables $\mathbb{A}_1$ and $\mathbb{A}_2$ in region S_1 and two non-commuting bivalent observables $\mathbb{B}_1$ and $\mathbb{B}_2$ in region S_2; each $\mathbb{A}_i$ commutes with each $\mathbb{B}_j$.

Dzhafarov and Kujala [2014c] translate non-signaling as marginal selectivity in cognitive systems. They state when marginal selectivity is violated, Fine's inequalities cannot be derived. In other words, Bell-type inequalities violations are irrelevant when marginal selectivity is not satisfied. QM obeys the non-signaling condition. Holding this condition shows that there is no contradiction between QM and special relativity. But the violation of Bell inequality indicated the failure of the stronger form of locality. As Ballentine and Jarrett [1987] states: "The proof of Bell's theorem requires a stronger form of locality than the simple locality principle (non-signaling condition) entailed by special relativity ".

To introduce marginal selectivity, I use an example from [Dzhafarov and Kujala, 2015]; they consider a system with two kind of observables (binary inputs) $\mathbb{A}$, $\mathbb{B}$ [5] related to Alice and Bob respectively, and two corresponding random variables (binary outputs) a, b. Similarly to the Bell experiment, in this example Alice can select between two observables $\mathbb{A}_1$ and $\mathbb{A}_2$ and their corresponding random variables take values $\{1, -1\}$. Bob can also select between observables $\mathbb{B}_1$ and $\mathbb{B}_2$ and the corresponding random variables again take values $\{1, -1\}$. Then Dzhafarov and Kujala [2015] represent the table of joint distribution as

$\mathbb{A}_i, \mathbb{B}_i$	$b_{ij} = +1$	$b_{ij} = -1$	
$a_{ij} = +1$	p_{ij}	...	$p[a_{ij=1}]$
$a_{ij} = -1$	...	...	...
	$p[b_{ij=1}]$	...	

[5] Dzhafarov and Kujala [2015] denote the observables $\mathbb{A}, \mathbb{B}$ as *inputs* and represent them as α, β, and call random variables a and b *outputs* represented by A and B. In their later paper [Dzhafarov and Kujala, 2016], the observables are called *content*s and labeled q.

If Bob's choice of observable $\mathbb{B}$ does not influence Alice's distribution of a and Alice's choice of observable $\mathbb{A}$ does not influence Bob's distribution of b, the marginal selectivity is satisfied. This situation is exhibited in the four tables which comprise Table 2.2.

Table 2.2: The joint probability distributions for marginal selectivity scenario.

$\mathbb{A}_1, \mathbb{B}_1$	$b_{11} = +1$	$b_{11} = -1$	
$a_{11} = +1$	p_{11}		$p_{1\bullet}$
$a_{11} = -1$			
	$p_{\bullet 1}$		

$\mathbb{A}_1, \mathbb{B}_2$	$b_{12} = +1$	$b_{12} = -1$	
$a_{12} = +1$	p_{12}		$p_{1\bullet}$
$a_{12} = -1$			
	$p_{\bullet 2}$		

$\mathbb{A}_2, \mathbb{B}_1$	$b_{21} = +1$	$b_{21} = -1$	
$a_{21} = +1$	p_{21}		$p_{2\bullet}$
$a_{21} = -1$			
	$p_{\bullet 1}$		

$\mathbb{A}_2, \mathbb{B}_2$	$b_{22} = +1$	$b_{22} = -1$	
$a_{22} = +1$	p_{22}		$p_{2\bullet}$
$a_{22} = -1$			
	$p_{\bullet 2}$		

According to these tables, changing $\mathbb{B}_1$ to $\mathbb{B}_2$, for example, does not influence the marginal distribution of $a_{11} = +1$. This property is called marginal selectivity. Dzhafarov and Kujala [2014b] depict the situation that satisfies non-signaling as

$$\begin{array}{cc} \mathbb{A} & \mathbb{B} \\ \downarrow & \downarrow \\ a & b \end{array}$$

Returning to the CHSH test of compositionality suggested by Bruza et al. [2009, 2015a], the expectation values in Equation (2.10) can be calculated for different pairs of words only after analyzing the marginal selectivity rule. As an example, the conceptual combination TOAST GAG satisfies marginal selectivity and does not violate CHSH inequality while its value is in the range of $[-2, 2]$, so this conceptual combination is considered compositional.

However, most of Bruza et al. [2015a]'s conceptual combination examples violate marginal selectivity, like the case of APPLE CHIP. Bruza et al. [2015a]

describe this conceptual combination as having a strong pattern of correlation between senses for different priming cases. One of the combinations, BATTERY CHARGE, appears to satisfy marginal selectivity but violates a Bell-type inequality. More data is required before these results can be considered definitive. However, the question has arisen whether it is possible to have contextuality in the presence of signaling in cognition. Dzhafarov and Kujala [2014b] give a positive answer to this crucial question based on their suggested contextuality model. This model is named Contextuality-by-default (CBD) [Dzhafarov et al., 2016], which is a test of contextuality based on a double indexing scenario [6] of random variables. But as far as I know, in the area of QM, no previous research has confirmed such a possibility in the presence of uncontrolled signaling. This highlights the necessity to compare the CBD approach and its special assumptions with the standard notations of contextuality in QM to reach a further understanding of the non-signaling condition in cognition.

Moreover, Dzhafarov and Kujala [2014c] examine marginal selectivity in another set of experimental results provided by Aerts et al. [2013]. In this experiment, there are two concepts: "A= Animal" and "B= Acts" to make the conceptual combination "Animal Acts". The first concept has two observables "$\mathbb{A}_1$= Horse or Bear?" and "$\mathbb{A}_2$= Tiger or Cat?". The second concept also has two observables "$\mathbb{B}_1$= Growls or Whinnies?" and "$\mathbb{B}_2$= Snorts or Meows?". Aerts et al. [2013] design this experiment to apply CHSH inequality to analyze the compositionality between concepts, but Dzhafarov and Kujala [2014c] show this experiment violates the marginal selectivity, which makes applying CHSH inequality meaningless. Actually, in this experiment there is an obvious correlation between the choice of animals and sounds. In other words, those sounds

[6] In this scenario an observable in different contexts is associated with different random variables.

are obvious attributes of the selected animals, and they rise in mind simultaneously. This correlation completely contradicts the restriction imposed by relativity on signaling. However, the non-signaling (marginal selectivity) condition is not the only way to expose this clear correlation between animals and sounds. Returning to the original definition of non-signaling in physics, it has been established that relativity prohibits faster than light signaling. This non-signaling condition entails that the measurements at space-like separated regions do not influence each other's statistics. Additionally, relativity can be expressed by the condition that the observables associated with the space-like regions commute with each other [Emch, 2009, Prugovecki, 1995]. As described earlier in this section, Dzhafarov and Kujala [2014b,c] highlight the necessity of considering the non-signaling principle for Bell-like cognitive experiments. But the commutativity of observables on the two sides of these experiments can be considered another fundamental principle that needs to be satisfied. This means that in the design stage of the experiment, the observables must be selected in a way that satisfies the commutativity rules of the Bell scenario. By doing so, the clear correlation between the choice of animal and sounds in the Aerts et al. [2013] experiment could be avoided. All of this shows that careful attention to the meaning of relativity and signaling in QC is required.

The need to explicitly investigate the meaning of the CBD approach becomes more important considering other work by Dzhafarov et al. [2015a], in which reviews some behavioral and social experiments including decision making, visual illusions and conceptual combinations using this approach. None of them shows any evidence of contextuality, so this approach rejects the existence of contextuality in behavioral and social systems [7]. This absence of contextuality may be because of a different method of defining random variables in the CBD approach. Or, in a more general view, it may be because of a different understanding of contextuality. So it is necessary to explicitly investigate the meaning of this model and compare it to the standard and accepted existing models of contextuality in physics.

There are recent advanced general models that explain the fundamental concept of contextuality in quantum theory. Some of these models have been designed independently from the Hilbert space, the property that makes them appropriate candidates to apply in areas beyond physics such as cognition. The operational approach of Spekkens [2005] and the sheaf theory of Abramsky and Brandenburger [2011] are two recent leading examples of these models. Studying the possibility of their application in cognition can open completely new avenues in the area of QC, which can cover more sources of contextuality and further develop our understanding of contextuality in cognition. But this better understanding would not be feasible without discovering possible connections between these general models, CBD and the preceding models of contextuality in physics. In other words, it is important to establish whether there is any

[7] However, recently Cervantes and Dzhafarov [2017] claimed that they could present an experimental evidence for contextuality in psychology using the CBD approach. I believe the design of their experiment is not entirely reliable, so their claim is not acceptable. In their Bell's experiment design, they associate a choice between two characters in a story (The Snow Queen by Hans Christian Andersen) with the observable $\mathbb{A}$ and a choice between two characteristics (such as *Beautiful* and *Unattractive*) with the observable $\mathbb{B}$. As specified in this chapter, observables in Bell's experiment are dichotomic (in the other words, their associated random variables are bivalent). This property is completely satisfied in the Bell's experiment of Bruza et al. [2009, 2015a], in which two values $+1$ and -1 relate to situations where a sense of a concept is primed, and either recalled ($+1$) or not (-1). The Bell's experiment of Aerts et al. [2013] also satisfies this requirement, e.g. by considering two clear distinct meanings (*Snorts* and *Meows*) for a dichotomic observable. But the two characteristics (*Beautiful* and *Unattractive*), in [Cervantes and Dzhafarov, 2017], cannot depute values for a dichotomic observable, because of their potential overlap in meaning. I believe, the alternative pairs like (*Beautiful* and *Ugly*) or (*Attractive* and *Unattractive*) are more reasonable choices to satisfy the dichotomic observable requirement. The same problem exists for the other pair (*Kind* and *Evil*) in this paper.

logical consistency between the notations of these different approaches or not.

Chapter 3

Measurement

Quantum cognition (QC) has delivered a number of models for semantic memory, but to date these have tended to assume pure states and projective measurement. Here I relax these assumptions. A quantum inspired model of human word association experiments will be extended using a density matrix representation of human memory and a POVM based upon non-ideal measurements. My formulation allows for a consideration of the key term of measurement within a rigorous modern approach. This approach both provides new conceptual advances and suggests new experimental protocols [1].

3.1 Introduction

How should we model memory? As Shiffrin [2003] states:

> *None of the models we use in psychology or cognitive science, at least for any behavioral tasks I find to be of any interest, are correct. We build*

[1]The content of this chapter has been previously published by the author in two co-authored papers [Aliakbarzadeh and Kitto, 2016, 2018].

> *models to increase our understanding of, and to slightly better approximate, the incredibly complex cognitive systems that determine behavior.*

However, this pragmatism raises an interesting point. What do our models of memory assume? And how do they limit the way in which we can formulate a given memory model?

As I described in section (2.1), currently many models of QC apply a single state vector that assumes a system in a pure state [Aerts, 2011, Bruza et al., 2009, 2015a, Nelson et al., 2013, Pothos and Busemeyer, 2013]. However, when we perform memory experiments we obtain ensemble data for a collection of subjects. This cannot be modeled with a pure state, rather a mixed state is required. In this chapter I will make use of the density matrix representation to model ensembles of human subjects in word association experiments. At first, I will provide a detailed technical description of von Neumann projective valued measurement (PVM). Although PVM measurement has been used in QC before, especially in the recall experiment of Bruza et al. [2009], a better technical description is necessary to describe measurement on ensembles of subjects. Here I will make use of a more precise notation for the specific case of two observables in the recall experiment, and then I will generalize this notation for more possible senses. This will enable me to describe scenarios that have more possible outcomes for each observable.

Another limitation of previous models for semantic memory in QC centers on the use of projective measurement for cognitive systems. This is highly restrictive because QC (i) does not necessarily assume an orthogonal relationship between operators, and (ii) sometimes entails violations of repeatability. An analysis of these restrictions associated with the PVM formalism will lead me to introduce the more modern and general positive operator valued measure

(POVM). I will show that this non-orthogonal measurement provides new understandings and extends the standard advantages of quantum inspired models of memory. I will introduce a generalised Bell inequality in which POVM is used to represent joint nonideal measurement for two observables. This POVM can provide a natural model of the process of conceptual combination. The mathematical structures of density matrix and POVM in this chapter are described using the recall experiment of Bruza et al. [2009] and the generalized versions of that experiment. This consistency provides better technical descriptions for the original experiment and also new possible interpretations for more complex scenarios.

I will also introduce an applications of POVM in the modeling of memory. I will use Neumark's dilation theorem, which shows any POVM measurement can be mathematically reduced to a PVM measurement on a larger space. Using this theorem, I will consider the full cognitive state as a combination of noise and a restricted substate. This substate represents only those cognitive processes through which a subject participates in an experiment.

At the end of the chapter, I will discuss a future direction that I believe will contribute to better understanding of cognitive states. In Section 3.5.1, I will suggest using quantum tomography to reconstruct the unknown density state of a cognitive system. I will suggest this method as a new experimental protocol that could specify the unknown state of a subject based on repeating PVM on similarly prepared ensembles of that subject. In an idealized situation, the whole parameters of an unknown cognitive state could be specified using a single POVM.

3.2 Constructing a density matrix

I start with the example that I described earlier in Chapter 2, but I will explain it with more details here; the example in which a subject might recall an ambiguous word A when cued with a particular prime. In quantum memory models this prime is represented as a basis state (i.e. a measurement context). Here I will use the eventualities $\{a', a''\}$ [2] to describe a subject's responses to a concept A, which can be interpreted according to one of two possible dominant and subordinate *senses*. When the dominant sense of concept A is primed, and A is interpreted in that sense by the human subject, then we designate $a' = +1$. If A is not interpreted in that sense after priming the dominant sense, then we write $a' = -1$. Similarly, $a'' = \{1, -1\}$ relates to situations where the subordinate sense of concept A primed, and either recalled $(+1)$ or not (-1).

An example will help to make this formalism clear. Consider an experimental protocol where a subject cued with a concept A (e.g. BOXER) using a word on a screen "boxer". According to the USF free association norms [Nelson et al., 2004], a subject is more likely to interpret BOXER in the sport sense than the animal sense. We term the sporting sense dominant and the animal sense subordinate. If a subject is first primed with the dominant sense of BOXER using the word "glove", and then asked to interpret the concept BOXER, there is high possibility that they will recall a word that has a sport sense. This measurement process is represented by $\mathbb{A}'$, the result given by the subject is represented with a', and $a' = +1$, as the response agrees with the way in which the subject was primed. If the subject interprets BOXER in

[2] In this chapter, the notations of the bivariate variables $\{a_1, a_2\}$, that I defined in Chapter 2, are changed to $\{a', a''\}$; similarly the observables $\{\mathbb{A}_1, \mathbb{A}_2\}$ changed to $\{\mathbb{A}', \mathbb{A}''\}$. I change the notation only for this chapter and I will use the first defined notion again in the next chapter. These changes help me to provide a more tactical notation for operators such as PVM and POVM.

another sense, then we write $a' = -1$. Conversely, if at first the subject is shown the word "vampire", it will awake the animal sense in the mind of human subject. When the subject responds in a way that agrees with the animal sense of the priming we write $a'' = +1$, but if the concept is not interpreted in this subordinate sense, we use $a'' = -1$.

Adopting von Neumann's approach to the quantum measurement of an idealised system using self-adjoint linear operators, we assume that an orthonormal basis exists. We can now construct a Hermitian matrix $\mathbb{A}$ as a series of projection operators [Bruza et al., 2009]

$$\mathbb{A} = \sum_k a_k \mathbb{P}_k \tag{3.1}$$

where $\mathbb{P}_k$ is the projector onto the eigenspace of $\mathbb{A}$ with eigenvalue a_k, and each a_k corresponds to the results of the measurement $\mathbb{A}$. As an example, for two eigenvalues a_1 and a_{-1}, we can rewrite the von Neumann measurement as

$$\mathbb{A} = a_1 \mathbb{P}_1 + a_{-1} \mathbb{P}_{-1}, \tag{3.2}$$

where $\mathbb{P}_1$ and $\mathbb{P}_{-1}$ are the projectors onto the eigenspace of $\mathbb{A}$ for those two eigenvalues.

For the case of the concept A discussed above, we can therefore write out two noncommuting measurement operators $\{\mathbb{A}', \mathbb{A}''\}$ for the two different cases of priming (dominant and subordinate)

$$\mathbb{A}' = a'_1\mathbb{P}'_1 + a'_{-1}\mathbb{P}'_{-1} \tag{3.3}$$

$$= a'_1|a'_1\rangle\langle a'_1| + a'_{-1}|a'_{-1}\rangle\langle a'_{-1}|, \tag{3.4}$$

$$\mathbb{A}'' = a''_1\mathbb{P}''_1 + a''_{-1}\mathbb{P}''_{-1} \tag{3.5}$$

$$= a''_1|a''_1\rangle\langle a''_1| + a''_{-1}|a''_{-1}\rangle\langle a''_{-1}|. \tag{3.6}$$

where $|a'_k\rangle$ and $|a''_k\rangle$ both represent potentialities of a subject's state of mind after priming. It is reasonable to consider these two operators $\{\mathbb{A}', \mathbb{A}''\}$ as noncommuting because we cannot prime with both dominant and subordinate senses simultaneously. This implies that the related projectors are noncommuting across the two primes. If we consider $|\psi\rangle$ as the cognitive state of a subject, then the probability of obtaining result k is

$$p(k) = \langle\psi|\mathbb{P}_k|\psi\rangle = Tr(\mathbb{P}_k|\psi\rangle\langle\psi|), \tag{3.7}$$

and the subject's post-measurement state is

$$\frac{\mathbb{P}_k|\psi\rangle}{\sqrt{p(k)}}. \tag{3.8}$$

This is known as a projective valued measurement (PVM) in quantum mechanics (QM), a special class of general measurement which has the following properties [Wheeler, 2012]:

I. Hermitian: $\mathbb{P}^\dagger = \mathbb{P}$

   A square matrix $\mathbb{P}$ is Hermitian if it is equal to its transposed complex conjugate. This leads to an important property for operator $\mathbb{P}$, that its eigenvalues are real (not complex).

II. Positive: $\langle\alpha|\mathbb{P}_i|\alpha\rangle \geq 0$ (all α)

Positivity allows us to treat the results of measurements as probabilities when coupled with the next property.

III. Complete: $\sum_i \mathbb{P}_i = \mathbb{I}$

The eigenvalues of a complete set of observables fully specify the state of a system.

IV. Orthogonal: $\mathbb{P}_i\mathbb{P}_j = \delta_{ij}\mathbb{P}_i$

The results of measurement are completely independent from each other.

Returning to the scenario of priming the dominant sense, the probabilities of the measurement of $\mathbb{A}'$ are given as follows:

$$|\psi\rangle \xrightarrow[\text{dominant sense}]{\text{Prime with the}} \begin{cases} a' = +1 \quad \text{with probability} \quad |\langle a'_1|\psi\rangle|^2 = \langle\psi|\mathbb{P}_{+1}|\psi\rangle \\ \\ a' = -1 \quad \text{with probability} \quad |\langle a'_{-1}|\psi\rangle|^2 = \langle\psi|\mathbb{P}_{-1}|\psi\rangle \end{cases} \tag{3.9}$$

Repeating this measurement multiple times allows us to calculate an expected value [Wheeler, 2012]:

Definition 1 *The expected value of operator $\mathbb{A}$ acting on state $|\psi\rangle$ is calculated as*

$$\langle\mathbb{A}\rangle_\psi = \sum_k a_k\langle\psi|\mathbb{P}_k|\psi\rangle = \langle\psi|\mathbb{A}|\psi\rangle, \quad where \quad k \in \{+1, -1\}. \tag{3.10}$$

At this point we should ask whether such an experiment *could* be completed multiple times. The state $|\psi\rangle$ denotes a cognitive state for a subject, and once we have performed the experiment we have irrevocably changed the state of our subject's mind to the state represented by (3.8). We will need an ensemble of subjects to repeat our experiment, but the proposition that even two subjects would share the same initial state $|\psi\rangle$ is highly unlikely. Although we might try to provide the same experimental conditions as we prepare our different subjects, we can not assume that they will all be in exactly the same state $|\psi\rangle$, this can happen because of unwanted priming effects or even the different dynamics of those subjects. Summing up all of our subjects, we can represent a scenario where some proportion of them are in the state $|\psi_1\rangle$, another proportion are in the state $|\psi_2\rangle$ and so on. Averaging these proportions with reference to our total subject pool would give us a scenario where

$$\varepsilon(S(\psi_1, \psi_2, ...)) \begin{cases} |\psi_1\rangle & \text{with probability} \quad p_1 \\ |\psi_2\rangle & \text{with probability} \quad p_2 \\ & \vdots \end{cases} \tag{3.11}$$

where we use ε as an abbreviation for ensemble and S for subject [Wheeler, 2012].

For this expanded scenario, we can now rewrite the expected value for all of our measurements over this ensemble of subjects (using a standard approach that can be found in any QM text e.g. Wheeler [2012]), or indeed in standard texts on QC e.g. Busemeyer and Bruza [2012]).

Definition 2 *The expected value for measurements over ensemble of subjects calculated as*

$$
\begin{aligned}
\langle \mathbb{A} \rangle_{\varepsilon} &= \sum_{v} p_v \langle \mathbb{A} \rangle_{\psi_v} \\
&= \sum_{v} p_v \langle \psi_v | \mathbb{A} | \psi_v \rangle \\
&= \sum_{j} \sum_{v} p_v \langle \psi_v | \mathbb{A} | e_j \rangle \langle e_j | \psi_v \rangle \\
&= \sum_{j} \sum_{v} \langle e_j | \psi_v \rangle p_v \langle \psi_v | \mathbb{A} | e_j \rangle \\
&= Tr(\rho_{\epsilon} \mathbb{A}) \text{ } \textit{where} \text{ } \rho_{\epsilon} = \sum_{v} |\psi_v\rangle p_v \langle \psi_v|.
\end{aligned}
\tag{3.12}
$$

While we started with orthogonal measurements, it is interesting that the states $|\psi_v\rangle$ which are used in the construction of the density matrix ρ_ϵ are not required to be orthogonal. As a result the different ensembles of states can lead to the same density matrix. The density matrix is more convenient way to deal with some scenarios in QM, including the representation of ensembles of states [Nielsen and Chuang, 2010]. The density matrix should have the following properties [Wheeler, 2012]:

I. Hermitian: $\rho^\dagger = \rho$

II. Positive: $\langle \alpha | \rho | \alpha \rangle \geq 0$ (all α)

III. Have unit trace: $Tr\rho = \sum_v p_v = 1$

The density matrix ρ_ϵ in (3.12) becomes a *pure* state ρ_ψ, when one of the probabilities p_v becomes equal to unity and the others vanish. This signifies a return to the scenario where all subjects are prepared in the same initial state, i.e. we have

$$\rho_\psi = |\psi\rangle\langle\psi|. \tag{3.13}$$

When we cannot make this simplifying assumption, we must consider ρ_ϵ to be *mixed*. In general, given a specific density matrix, we can discover if it is mixed or pure using the Trace, as $Tr(\rho^2) = 1$ for pure states and $Tr(\rho^2) < 1$ for mixed ones. When a system is in a pure state, both state vector and density matrix representations of a given system provide the same results [Nielsen and Chuang, 2010].

Until now my description has been based on an idealized assumption that measurements are performed on the pure state $|\psi\rangle$ for each subject. In other words I have assumed that the cognitive state of a subject who recalls a concept A can be represented by a pure state $|\psi\rangle$. But in reality we can not guarantee that the subject will adhere to our designed experimental protocol; it is clear that the human mind can process different concepts or events other than our intended concept A during the experiment. The cognitive state of a human mind does not relate only to the process of recall. If we can represent the general state of the mind with a pure state $|\psi\rangle$, then the state that we use to model the recall process should be a subsystem of that pure state. In physics, as I mentioned earlier, we can use the density matrix to represent ensembles of states, however the density matrix can also be used to represent a subsystem of a pure state [Nielsen and Chuang, 2010].

Elaborating, I will denote this extension of my formalism by rewriting the cognitive state of the subject as a composition of states labeled by R and E, where R denotes that part of the cognitive state that is directly influenced by the recall experiment and E relates to the remainder. In physics, to represent a subsystem of a composite system we can apply the reduced density matrix as

$$\rho^R \equiv Tr_E(\rho^{RE}), \tag{3.14}$$

where ρ^{RE} describes the state of the composite system, and Tr_E is an operator known as a partial trace on operator E. For example, to obtain the density matrix of subsystem R, we use the partial trace over subsystem E [Nielsen and Chuang, 2010]:

$$Tr_E(|r_1\rangle\langle r_2| \otimes |e_1\rangle\langle e_2|) \equiv |r_1\rangle\langle r_2| Tr(|e_1\rangle\langle e_2|). \tag{3.15}$$

where $\{|r_1\rangle, |r_2\rangle\}$ and $\{|e_1\rangle, |e_2\rangle\}$ are spanning vectors in the state space of R, and E respectively, and the standard trace operator $Tr(|e_1\rangle\langle e_2|) = \langle e_2|e_1\rangle$ has been applied on the right hand side.

Now I can rewrite Equations (3.7) and (3.8) using the density matrix ρ of the ensemble of subjects or the subsystem R. The probability of obtaining the result k becomes

$$p(k) = Tr(\rho \mathbb{P}_k), \tag{3.16}$$

and the state after measurement is [Wheeler, 2012]:

$$\frac{{\mathbb{P}_k}^\dagger \rho \mathbb{P}_k}{\sqrt{p(k)}} = \frac{\rho \mathbb{P}_k \mathbb{P}_k}{\sqrt{p(k)}} = \frac{\rho \mathbb{P}_k}{\sqrt{p(k)}}, \tag{3.17}$$

where we have used orthogonal property to write $\mathbb{P}_k = \mathbb{P}_k \mathbb{P}_k$.

I note that the Hermitian matrix $\mathbb{A}$ that was defined in (3.1) and used to construct the density matrix is not particularly large or interesting. However, at this point it is possible to extend the basic approach, which considered only two possible senses.

To this end, I note that it is frequently the case that more than two interpretations are possible for one lexical observable, a situation that can be represented by extending the set of projectors from our PVM measurement (3.2) as

$$\text{Set of projectors for a PVM measurement} \begin{cases} \vdots \\ \mathbb{P}_{k_1} \\ \mathbb{P}_{k_2} \\ \mathbb{P}_{k_3} \\ \vdots \end{cases} \tag{3.18}$$

For example, returning to the BOXER case discussed above, it is possible to interpret this ambiguous word in a third, clothing related, sense (e.g. BOXER SHORTS). This gives us three possible senses: "sport","animal" and "clothing" (but it is important to emphasize that this word could be interpreted in even more than these three senses). Similar to the original example, we can use the observable $\mathbb{A}'$ to represent a measurement process when the subject is first primed with the dominant sense and then asked to interpret the word BOXER. In this case, a'_1 represents a case where the subject's response agrees with the primed sense, while a'_2 and a'_3 relate to two other possible responses in "animal" and "clothing" senses, and a'_4 represents all other possible senses. Then as in (3.3) the von Neumann measurement for the observable $\mathbb{A}'$ becomes

$$\mathbb{A}' = a'_1\mathbb{P}_1 + a'_2\mathbb{P}_2 + a'_3\mathbb{P}_3 + a'_4\mathbb{P}_4, \tag{3.19}$$

where $\mathbb{P}_1$, $\mathbb{P}_2$, $\mathbb{P}_3$ and $\mathbb{P}_4$ are the projectors onto the eigenspace of $\mathbb{A}'$ for four eigenvalues a'_1, a'_2, a'_3 and a'_4. I will use this description of multiple senses to construct the generalized form of non-ideal measurement in the next section.

This section has presented a much more technical introduction to the model presented in [Bruza et al., 2009]. I have shown that it is possible to generalize the standard quantum probability model using density matrices. This allows for the representation of scenarios where we cannot guarantee that an ensemble of subjects have all been prepared with the same pure cognitive state. This is an important consideration in psychology, the assumption that all subjects prepared in the same way are in the same pure cognitive state is a very strong one and does not match with reality. I have also introduced a second application of this density matrix apparatus to describe part of a larger cognitive system. The density matrix operator is capable of dealing with far more complexity in the quantum model of memory, that is, it can fully characterise a cognitive state. This means that once we are given the density matrix we can predict the probabilities for outcomes of any measurement on that state. But how we can access this knowledge about our system? In QM, the standard way to characterize the complete quantum state of a particle is by using quantum state tomography [Thew et al., 2002, Wootters, 2004]. I will briefly describe how this method might be introduced to the field of QC in Section 3.5.1, which identifies an unknown quantum state using a set of measurements. In the next section I will keep using the density matrix, where I will introduce a more general formalism for measurement in QC, demonstrating the specifics of how it can be used in quantum memory models.

3.3 Non-ideal measurement (POVM)

In reality, the process of recall does not always create sharp results like the von Neumann measurement described in the previous section. This happens because of unwanted effects in the process of measurement. For example, there is no guarantee that a human subject will actually give a response corresponding to what they recalled. Despite the best experimental instructions, humans will be noisy in their responses. Thus, in the recall experiment described earlier for the word BOXER, a subject may be primed with the sport sense (the word "glove" was used in the previous example), think of "Muhammad", but censor their response giving a response with an animal sense instead (e.g."dog") with a probability ξ. This can be modelled using inefficient detectors. If we prepare the system in the state $|+1\rangle$ (corresponding to the sport sense), then the result of measurement will be $|+1\rangle$ with the probability $1-\xi$ and $|-1\rangle$ with the probability ξ. To model this inefficiency, physicists often apply an unsharp measurement instead of an ideal von Neumann measurement [Barnett, 2009, de Muynck, 2002, Wheeler, 2012].

Returning to the scenario where we prime the dominant sense (observable $\mathbb{A}'$), an ideal PVM measurement is described by the two projectors $\mathbb{P}_{+1}$ and $\mathbb{P}_{-1}$ introduced in (3.2). The above possibility of imprecision arising in all subjects' recall processes is represented with the probability ξ. This means that if the result of measurement in the ideal situation was $+1$, a non-ideal situation would return $+1$ with the probability $1-\xi$ and -1 with the probability ξ. If the state of our system is described by the density matrix ρ, then the probability that the subjects recall words with the same sense as the prime is given by Barnett [2009]

$$p(+1) = (1-\xi)Tr(\rho\mathbb{P}_{+1}) + \xi Tr(\rho\mathbb{P}_{-1}). \tag{3.20}$$

By defining a new operator $\tilde{\mathbb{P}}_{+1}$ as

$$\tilde{\mathbb{P}}_{+1} = (1-\xi)\mathbb{P}_{+1} + \xi\mathbb{P}_{-1}, \tag{3.21}$$

we can use the technique in Barnett [2009] to write the probability for the measurement outcome $+1$ in a manner similar to the PVM case (3.7)

$$p(+1) = Tr(\rho\tilde{\mathbb{P}}_{+1}). \tag{3.22}$$

This $\tilde{\mathbb{P}}$ is a new type of operator used to describe this non-ideal situation. Any two operators $\tilde{\mathbb{P}}_i$ and $\tilde{\mathbb{P}}_j$ $(i \neq j)$ with this definition are not required to be orthogonal. $\tilde{\mathbb{P}}$ is called a "Positive Operator-Valued Measure" or POVM.

It is possible to extend this situation, and to incorporate the multiple senses described earlier to reach the most general theoretical formulation using non-ideal measurement. As occurred for the case with two possible responses, we have to define probabilities for each non-ideal measurement. In this general case with j possible responses, if the result of measurement in the ideal situation was k, a non-ideal situation would have the result j with the probability $w_{j|k}$ where $\sum_j w_{j|k} = 1$ for all k [Wheeler, 2012]. For a system in a state ρ, the probability of finding result k after a non-ideal measurement would be represented by the following formula [Wheeler, 2012]:

$$\begin{aligned} p(k) &= \sum_j w_{j|k} Tr(\rho\mathbb{P}_j) \\ &= tr(\rho\tilde{\mathbb{P}}_k) \quad where \quad \tilde{\mathbb{P}}_k = \sum_j w_{j|k}\mathbb{P}_j \end{aligned} \tag{3.23}$$

The general form of Equation (3.23) can be reduced to the simple form of Equation (3.22) if we have only two results, $+1$ and -1, where $w_{j|k}$ can take two values p and $1-p$. Thus it is possible to recover the simple scenario discussed above.

Similar to the way in which the set $\{\mathbb{P}_1, \mathbb{P}_2, ...\}$ was a complete set of ideal measurements, the set $\{\tilde{\mathbb{P}}_1, \tilde{\mathbb{P}}_2, ...\}$ is a complete set of non-ideal measurements, but this time with the following properties [Wheeler, 2012]:

I. Hermitian: $\tilde{\mathbb{P}}^\dagger = \tilde{\mathbb{P}}$

II. Positive: $\langle\alpha|\tilde{\mathbb{P}}_i|\alpha\rangle \geq 0$ (all α)

III. Complete: $\sum_i \tilde{\mathbb{P}}_i = \mathbb{I}$

IV. Typically non-projective and non-orthogonal: $\tilde{\mathbb{P}}_i\tilde{\mathbb{P}}_j \neq \delta_{ij}\tilde{\mathbb{P}}_i$

Any operators $\tilde{\mathbb{P}}_1, \tilde{\mathbb{P}}_2, ...$ that satisfy these properties are POVM.

This formalism can bring new opportunities to model psychological phenomena that have not previously been modeled in PVM approaches. For example, it is possible that subjects give different responses to the same repeated cue, a scenario called "non-repeatability". Interestingly, Khrennikov et al. [2014] also used POVM in opinion polling to demonstrate non-repeatability in an evolution-free framework.

To mathematically illustrate how non-repeatability arises within the POVM approach, we can rewrite $\tilde{\mathbb{P}}_k = \mathbb{A}_k^\dagger \mathbb{A}_k$, as we had in Equation (2.8) where $\mathbb{A}_k$ is called a measurement decomposition operator [Jaeger, 2009]. In this case the state after measurement can be written

$$\frac{{\mathbb{A}_k}^\dagger \rho \mathbb{A}_k}{\sqrt{p(k)}}, \tag{3.24}$$

This post-measurement state indicates that unlike projective measurement, a repeated application of the POVM does not lead to the same result. Because if we apply the POVM observable once more on the post-measurement state of

(3.24), the result becomes

$$\frac{\mathbb{A}_k{}^\dagger \mathbb{A}_k{}^\dagger \rho \mathbb{A}_k \mathbb{A}_k}{Tr(\mathbb{A}_k{}^\dagger \mathbb{A}_k{}^\dagger \rho \mathbb{A}_k \mathbb{A}_k)}. \tag{3.25}$$

which is not necessarily equal to the previous state. It will be equal only if the POVM elements are idempotent ($\tilde{\mathbb{P}}^2 = \tilde{\mathbb{P}}$). In this situation, POVM is reduced to a projective measurement. Despite this difference, there is a way to relate these two measurements to each other using Neumark's Theorem, which I will discuss in section 3.4.1.

Experimental scenarios can rapidly become very complex in the case of word association experiments. As I explained above, subjects may not report the word that sprang immediately to mind. A further complexity emerges where we consider the overarching social setting in which experimental priming is carried out. Even if we try our best to design an experiment with equally weighted primes for each possible sense, some primes are more dominant. For example, during an election period the "political" sense of the word PARTY may become stronger. A similar shift in weight towards the alternative sense might occur for a subject who went to a party the night before the experiment. We can apply this complexity to describe a non ideal choice of measurement settings in the generalized Bell-type experiments in cognition. In this case the priming with different senses occurs with different probabilities.

The ideal Bell experiment modeled by Bruza et al. [2015a] assumes an equal choice of the different settings for any given subject (two operators $\{\mathbb{A}', \mathbb{A}''\}$ in (3.3–3.6) which relate to the priming of two senses). It is possible to relax this assumption using the generalized Bell experiment. This is analogous to a situation where a biased interferometer leads a photon arriving at one of two detectors with different probabilities, which can be expressed by a bivariate

POVM [de Muynck, 2002]. This model provides a joint non-ideal measurement of two observables, where for simplicity de Muynck [2002] assumes 100% efficient detectors. This simplification removes the need to consider the possibility that the subject's response is not what first sprang to mind (i.e. the complexity of an inefficient detector). This allows us to assume ideal measurements for each observable separately and non-ideal measurement for the joint observable. We need the following definitions from de Muynck and Martens [1989] to construct POVM as a *jointly non-ideal measurement* of observables.

Definition 3 *A POVM* $\{\tilde{M}_m\}$ *is a* nonideal *measurement of the observable POVM* $\{\tilde{N}_n\}$ *if:*

$$\tilde{M}_m = \sum_n \lambda_{mn} \tilde{N}_n, \qquad \lambda_{mn} \geq 0, \qquad \sum_m \lambda_{mn} = 1. \tag{3.26}$$

Definition 4 *Two observables* $\tilde{M}_m$ *and* $\tilde{N}_n$ *are simultaneously, or* jointly measurable *if a bivariate POVM* $\tilde{R}_{mn}$ *exists such that its marginals* $\{\sum_n \tilde{R}_{mn}\}$ *and* $\{\sum_m \tilde{R}_{mn}\}$ *are POVMs and representing non ideal measurements of* $\tilde{M}_m$ *and* $\tilde{N}_n$ *respectively.*

This approach can be applied to the experiment discussed by Bruza et al. [2015a]. Denoting the probability γ for priming with the sense $\mathbb{A}'$, and the probability $1-\gamma$ for the priming with sense $\mathbb{A}''$, to represent the above scenario. As was the case in (3.3–3.6), our observables $\{\mathbb{A}', \mathbb{A}''\}$ can be represented using two projectors $\mathbb{P}_1, \mathbb{P}_{-1}$. We can write the set of PVMs for the first and second observable as $(\mathbb{P}'_n, \mathbb{P}''_m)$ where n and m take the values in $\{+1, -1\}$. The joint non ideal measurement for PVMs $(\mathbb{P}'_1, \mathbb{P}'_{-1})$ and $(\mathbb{P}''_1, \mathbb{P}''_{-1})$, can be constructed as a bivariate POVM [de Muynck, 2007]

$$\tilde{R}^{\gamma}_{mn} = \begin{pmatrix} 0 & \gamma(\mathbb{P}'_1) \\ (1-\gamma)(\mathbb{P}''_1) & \gamma(\mathbb{P}'_{-1}) + (1-\gamma)(\mathbb{P}''_{-1}) \end{pmatrix}, \tag{3.27}$$

The probability for this joint nonideal measurement is $p_{mn} = Tr\rho\tilde{R}^{\gamma}_{mn}$, as was the case for (3.7). The top left hand corner of this matrix is equal to zero because the subject cannot be primed with two senses for a word at the same time.

The marginals of $\tilde{R}^{\gamma}_{mn}$ are the POVMs $\{\tilde{M}_m\} = \{\gamma\mathbb{P}'_1, I - \gamma\mathbb{P}'_1\}$ and $\{\tilde{N}_n\} = \{(1-\gamma)\mathbb{P}''_1, I - (1-\gamma)\mathbb{P}''_1\}$ which can be represented in matrix form as

$$\begin{pmatrix} \sum_n \tilde{R}^{\gamma}_{1n} \\ \\ \sum_n \tilde{R}^{\gamma}_{-1n} \end{pmatrix} = \begin{pmatrix} \gamma & 0 \\ \\ 1-\gamma & 1 \end{pmatrix} \begin{pmatrix} \mathbb{P}'_1 \\ \\ \mathbb{P}'_{-1} \end{pmatrix} \tag{3.28}$$

$$\begin{pmatrix} \sum_m \tilde{R}^{\gamma}_{m1} \\ \\ \sum_m \tilde{R}^{\gamma}_{m-1} \end{pmatrix} = \begin{pmatrix} 1-\gamma & 0 \\ \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} \mathbb{P}''_1 \\ \\ \mathbb{P}''_{-1} \end{pmatrix} \tag{3.29}$$

These marginals satisfy Definitions 3 and 4. It is clear that the operators $\tilde{M}_m$ are $\tilde{N}_n$ are not commuting because $\mathbb{P}'_n$ and $\mathbb{P}''_m$ are not commuting. So we do not necessarily need commutativity of operators to construct non-ideal joint measurements [de Muynck, 2007]. Note that $\tilde{R}^{\gamma}_{mn}$ only describes one concept (e.g. A or B). This is unlike the scenario that arises for the standard Bell-type inequalities that were constructed using ideal joint measurements on commuting observables for both concepts A and B [Bruza et al., 2015a].

The direct product of the bivariate POVMs (3.27) for two concepts A and B in the Bell-type experiment of Bruza et al. [2015a] leads to a quadrivariate POVM, which can be written as

$$\tilde{R}^{\gamma_A\gamma_B}_{m_A n_A m_B n_B} = \tilde{R}^{\gamma_A}_{m_A n_A}\tilde{R}^{\gamma_B}_{m_B n_B}. \tag{3.30}$$

In this scenario there is no disturbing influence arising on the marginals of one concept when we change the measurement settings for another concept [de Muynck, 2007]. The measurement results of each concept are influenced by the measurement settings of that concept (complementarity). This enables the POVM formalism to model contextual behavior without making use of nonlocality. Complementarity then provides us with a local explanation for violations of the generalized Bell inequality which is expressed using the quadrivariate probability distribution [de Muynck, 2007]

$$p^{\gamma_{AB}}_{m_A n_A m_B n_B} = Tr\rho \tilde{R}^{\gamma_A}_{m_A n_A} \tilde{R}^{\gamma_B}_{m_B n_B}. \tag{3.31}$$

In this scenario there is no disturbing influence arising on the marginals of one concept when we change the measurement settings for another concept [de Muynck, 2007]. The measurement results of each concept are influenced by the measurement settings of that concept (complementarity). This enables the POVM formalism to model contextual behavior without making use of nonlocality. Complementarity then provides us with a local explanation for violations of the generalized Bell inequality which is expressed using the quadrivariate probability distribution [de Muynck, 2007]

$$p^{\gamma_{AB}}_{m_A n_A m_B n_B} = Tr\rho \tilde{R}^{\gamma_A}_{m_A n_A} \tilde{R}^{\gamma_B}_{m_B n_B}. \tag{3.32}$$

Recall in this generalized form of the Bell experiment, each subject comes to the experiment with a different historical context. This context affects the way in which subjects are primed, as the semantic network will activate differently in response to the prime [Nelson et al., 2013]. This activation relates to the probability γ that we used to construct the POVM in (3.27). A violation of the generalized Bell inequality would occur because of each subject's unique historical context.

In this section I have described two methods for constructing a POVM, the first for one observable and the second for two. While I have shown that it is possible to construct different POVMs for these specific experimental scenarios, note that the properties of POVMs mentioned earlier imply that each POVM $\tilde{\mathbb{P}}$ is unique to the relevant experimental context. The empirical evaluation (or analysis) of these two POVMs, is reserved for future work. This will require that we can measure the two probabilities ξ and γ associated with the POVMs, which is not currently possible using word association experiments. However, it may be possible in the future to apply neuroimaging technologies such as the functional magnetic resonance imaging (fMRI) [Glover, 2011, Ogawa et al., 1990] to provide tests of the validity of these new models. Such methods can be used to characterize cognitive phenomena at the level of neural processes, for example, it was used to detect deception by Kozel et al. [2005]. A similar approach could possibly be used to measure probability ξ for the imprecision arose in subjects recall process. As another example, fMRI was used to examine the cognitive phenomenon of attention on a visual tracking task [Culham et al., 2001]; such approach could be used to measure the probability γ for different chances of priming of two senses. Much more work remains to catalog other psychologically plausible mechanisms that can arise in quantum memory experiments, and to demonstrate how they might be modelled using an approach based upon POVM.

3.4 Advantages of using a density matrix and POVM approach

At this point I have described the two processes of preparation and measurement for a quantum memory model. In QM they are considered as separate processes. For example, as Isham [2001, p.154–p.155] states:

> *A measurement is an operation on a system that probes that quantum state immediately before the measurement is made... state preparation is an operation whose aim is to force the system to be in some specified state immediately after the operation.*

In this chapter I introduced the density matrix as a practical tool for describing preparation when dealing with an ensemble of subjects (in section 3.2). This operator ρ most generally represents a mixed state and it contains all the information necessary to predict any possible measurement outcome. For the measurement process we introduced a POVM which gives us a more realistic depiction of word association experiments than an approach based on standard projective measurement (as defined in Section 3.3). Franco [2016] recently constructed a quantum inspired model of decision making which follows a similar methodology; treating preparation as a process where information is provided to a subject, and the measurement stage as a process of testing subjects at the end of this preparation phase.

Preparation and measurement have other more specific applications. For example, as I described in the previous section, Muynck's joint non-ideal measurement can be used to describe the situation of complementarity as it arises in the generalized Bell inequality. I will now introduce a specific application, explaining how they can serve to advance the field of QC. In the next chapter, I will introduce the interesting application in contextuality 4.6.

3.4.1 Neumark's theorem

To draw attention to the different roles of PVM and POVM measurements in a cognitive experiment we can use Neumark's Theorem [Peres, 1990]. This theorem provides a tool for dealing with noise in a cognitive experiment (of the type that was described in Section 3.2). In that experiment I considered the state of a subject's mind as the composition of two states "R" and "E", where "R" denotes that part of the cognitive state related to recall experiment, and "E" is considered as the remainder, which I will refer to as noise. Noise arises from events or thoughts outside the defined bounds of the experiment (e.g. what the subject ate for breakfast or an accident they witnessed on the way to the experiment).

Neumark's theorem relates the POVM of state "R" to a projective measurement on the composition of states "R" and "E" [de Muynck, 2002]. In other words, it extends the Hilbert space of "R" to the tensor product space of "R" and "E".

Theorem 3.4.1 (Neumark) *An arbitrary POVM on a Hilbert space H_R can be expressed using a PVM in a larger Hilbert space H containing H_R.*

The inverse situation arises in similar manner to the partial trace that I introduced in (3.14): Given any PVM on a Hilbert space H, we can find a POVM on a subspace H_R. In fact, we create the POVM on a sub-system when we do not need to consider the extra information contained within the higher dimensional Hilbert space. Tracing out this noise using the mathematical structure of POVM requires that we understand cognitive function well enough to construct the appropriate PVM and POVM for the Hilbert space H and subspace H_R respectively. Thus, we would need to understand the variables involved in shaping states "R" and "E". Current experimental developments may not provide us with this ability; as Khrennikov and Basieva [2014] discuss, the brain can be both a system and the observer in a QC system, which can make it difficult to isolate from its mental environment.

Progressing in this area will require significantly more work to understand how boundaries should be defined in cognitive experiments. For example, we are immediately confronted with the question of: what should be considered noise in a given experiment?

It is essential that the field of QC consider the effects of noise in our models for different cognitive experiments. My work in Section 3.3 was just a first step in this direction. We have so far considered two cognitively motivated effects of noise as: (1) the probability ξ, which represents the imprecision across all subjects in a recall experiment; and (2) the probability γ for priming with the sense $\mathbb{A}'$ and $\mathbb{A}$ in the generalized Bell experiment. However, more work remains to be completed before it will be possible to construct comprehensive models for memory using modern quantum inspired methods.

3.5 What is a quantum cognitive state?

In this chapter, I first generalized the process of preparation and measurement for cognitive systems, and then discussed some possible advantages of this representation. However, it is important to realize that much of the mathematics utilized in QC rests upon rather shaky foundations. To further improve quantum inspired models of cognition we need to advance in our mathematical understanding of the most basic cognitive states. As an example, the model provided in Section 3.2, of the two senses that a subject might associate with the ambiguous word BOXER is constructed using the assumption of a basic two-level cognitive state. The subject can give many possible answers to an observable $\mathbb{A}$ (in this case the cue word BOXER), but I assume in my model that they fall into one of two possible senses (i.e. the same as the priming sense and different from the priming sense), which are denoted with +1 and -1. This basic state can be considered similar to a single qubit (a spin-1/2 particle) system in

QM. It is represented by a pure state

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{3.33}$$

where $|\alpha|^2$ and $|\beta|^2$ are probabilities of spins up and down respectively, and the state of the qubit is a vector in a two-dimensional complex vector space with the orthonormal computational basis $|0\rangle$ and $|1\rangle$. One of the main differences between a qubit and a classical bit is this superposition property. Unlike a classical bit, which acts akin to a coin with only two possible states of "heads" and "tails", a qubit can exist in any weighted continuum of states between $|0\rangle$ and $|1\rangle$. However, when it is measured a qubit only gives measurements results 0 and 1 [Nielsen and Chuang, 2010]. This property has been widely exploited in QC [Aerts, 2011, Asano et al., 2015, Busemeyer and Bruza, 2012, Wang et al., 2013]. As the example I described in Chapter 2, Pothos and Busemeyer [2013] represent the happiness of a hypothetical person using the superposition state $|\psi\rangle = a|happy\rangle + b|unhappy\rangle$. After being asked about her happiness, and the subject deciding upon her answer, the state vector becomes $|\psi\rangle = |happy\rangle$ or $|\psi\rangle = |unhappy\rangle$.

However, an approach like this leaves us with few ideas as to what this representation of $|\psi\rangle$ actually symbolises. What is the underlying cognitive state? And how does it evolve in time as a person moves through their day? Here, I will provide some guidelines that could be considered in future research aimed at clarifying the representation of cognitive states. I note that many more questions are provided in this section than answers, but consider it appropriate to draw attention to what is an underexplored but important area for future research.

Rather than representing qubits using the abstraction of a complex vector space, it is possible to more fully visualize their properties using the geometrical Bloch sphere representation. This method also provides a more explicit representation of the types of operations that we can apply on a qubit. I carry out this transformation by rewriting Equation (3.33) as

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}e^{i\varphi}|1\rangle, \tag{3.34}$$

where spherical coordinates are defined by latitude θ and longitude φ. Each pure state represented by Equation (3.33) associates to a point on the surface of a unit sphere in the Euclidean 3-dimensional space (see figure 3.1). In this

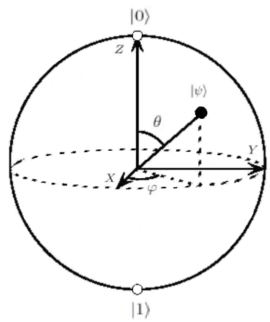

Figure 3.1: Bloch sphere representation of a qubit.

geometrical representation, two orthogonal basis states $|0\rangle = |\uparrow\rangle$ and $|1\rangle = |\downarrow\rangle$ correspond to an orientation of the spin in $+z$ and $-z$ directions respectively. And superposition states can correspond to other orientations of the spin in different spatial directions, e.g. the state $|0_y\rangle$ is oriented in the $+y$ direction. The Spherical polar coordinates (r, θ, φ) can be related to Cartesian coordinates (x, y, z) by

$$\begin{aligned} x &= r\sin\theta\cos\varphi \\ y &= r\sin\theta\sin\varphi \\ z &= r\cos\theta. \end{aligned} \tag{3.35}$$

To measure spin of the qubit in each of these directions we would apply the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.36}$$

In other words, we can visualize the two-level quantum system using this three-dimensional Bloch sphere representation. A cognitive state built from this basic representation brings with it the possibility of modelling its dynamical evolution, but still has a direct mapping to the qubits which have been used in the previous work in QC. However, the visualization of more than two-level quantum systems can have different geometries and usually need higher-dimensional Bloch based vector representation [Bertlmann and Krammer, 2008, Kimura, 2003, Sandeep K Goyal et al., 2016] [3]. Likewise, cognitive systems of more than two-levels (like the example I described in Equation (3.19)) could still be modelled by these higher-dimensional representations.

However, to employ this Bloch sphere representation of a qubit in cognition, we have to provide appropriate meanings for these three directions in phase space. Yearsley and Pothos [2013] provide an example interpretation in a decision making experiment, by projecting a bivariate observable on Z direction. This work employs a Bloch sphere representation to build up an understanding of how a cognitive system might evolve in time to define a test for violations of the temporal-Bell inequalities. Similarly, Broekaert et al. [2017] provide the geometric interpretation for the evolution of states implied by Hamiltonian. Their Hamiltonian is built based on only two Pauli matrices

[3] It is not always necessary to consider a higher dimension, as an example, Kurzyński et al. [2016] suggest a three-dimensional visualization for qutrit.

to describe different dynamical evolution scenarios. However, in both of these approaches, the derivation of the Hamiltonian of the system makes reference only to the mathematical aspects of the Bloch sphere, without providing an exact cognitive meaning for the directions x, y and z that are used in these models.

This is an important point to emphasize; there is little connection in these more advanced models between the physical formalism and the cognitive meaning. This chapter has been careful to associate a strong psychological interpretation to the more advanced models of measurement and preparation as they apply to cognition. However, to move forwards we will require a far stronger connection to the underlying meaning of a cognitive qubit, which I will term a *cobit*. As I emphasized earlier, it is necessary to provide cognitive meanings for the different Cartesian directions in the geometrical representations used with this approach. However, to transform this representation back to the complex vector space, we should still be able to provide cognitive meanings for the orthonormal computational basis $|0\rangle$ and $|1\rangle$ (like the two senses that I assigned to these two basis states in Equation (3.33)). I admit that this geometrical interpretation does not scale to more complex multipartite situations in a straightforward manner, which potentially limits its utility as a general model, however, I consider it necessary that QC place more of an emphasis upon finding the underlying dynamical representation of cognitive processes, and a model based upon a high dimensional Bloch based representation is an immediately plausible option, as has already been recognised by a number of papers previously published in this domain. Providing the cognitive meaning for a cobit as discussed above, helps us to interpret the cognitive meanings of applied operators on that cobit. As an example, we could reach a better understanding of the rotations group $SO(3)$ and $SU(2)$ and their special relation with each other [4]. While an unsatisfactory lack of detail still remains in mapping such models to cognitively plausible representations, my mapping in this paper of the density matrix to interpretable semantic memory tasks gives a new avenue that will help to link this more interpretationally robust class of models to the extensive data sets that have been collected in this the domain of memory and recall already (see e.g. Nelson et al. [2013] for a summary of one such dataset). I leave this contribution for future work.

If we can successfully create a cognitively well justified geometrical representation for our cobit then it will be straightforward to extend the model for multiple cobits. As Nielsen and Chuang [2010] explain, treating qubits as abstract mathematical objects enables us to generalize the concept for more complex situations (e.g. multiple qubits) without depending upon a specific realization. This geometrical representation opens different avenues to make use of the approaches identified in this chapter for using the density matrix and generalized measurement in modelling semantics. As an example of a further approach that this avenue might open up, I will briefly introduce the concept of cognitive tomography, and suggest a way in which it might be used to characterize unknown cognitive states.

3.5.1 Cognitive tomography

In a QC model, the result of measurement gives an indication of the state of a subject's mind with reference to a measurement scenario, or question, just

[4]The special orthogonal group $SO(3)$ represents rotations around the origin of a three-dimensional Euclidean space and the special unitary group $SU(2)$ is the set of 2 by 2 unitary matrices with determinant 1 (like the set of Pauli matrices described in Equation (3.36)). The relation between these two rotations groups is an one-to-two correspondence between any $R \in SO(3)$ and $2U \in SU(2)$ [Miller, 1972].

before the measurement occurred. For instance, in the BOXER example described in Section 3.2, each answer $+1$ or -1 indicates whether BOXER would be interpreted in the same way as the priming occurred (as represented by the operators $\mathbb{A}'$ or $\mathbb{A}''$), or not.

But to fully understand the state of a subject's mind when faced with the word BOXER, we cannot rely on one measurement alone. One possibility would be to repeat the measurement many times on the same subject to get an average of different results. A memory experiment would need to repeat the cuing procedure in a variety of different contexts. However, this demonstrates precisely how difficult it is to construct a reliable measurement in cognition, because the response a subject gives to an experiment can affect the response that they give for the following experiments (see Section 3.2).

So how can we proceed in finding a more precise understanding of the underlying cognitive state? One possibility would be to apply a method inspired by quantum tomography, which specifies an unknown quantum state by performing measurements over an ensemble of equally prepared identical quantum states [Leonhardt, 1997]. To use this method for cognition, we would need to provide the same experimental conditions as we prepare our different subjects.[5] This is similar to the scenario in physics where a device produces a beam of spin-1/2 particles [Wootters, 2004]. It is possible to predict the spin state of particles that the device produces if we perform a set of orthogonal measurements in the x, y and z directions. These three measurements should be "mutually conjugate" [Wootters, 2004] which means that an eigenvector of any one of them must be an equal superposition of the eigenvectors of the two others. For this set of measurements, each different measurement provides

[5] This does not necessarily lead to an ensemble of subjects with the same pure cognitive state, as I mentioned in Section 3.2.

information independent from the information provided by the other measurements [Wootters, 2004].

A detailed mathematical description of tomography for cobits would rely upon having a precise cognitive meaning for the x, y and z directions in their geometrical representation. Then to estimate a cognitive state of cobits we would need to repeatedly apply three linearly independent observables (associated with those three directions) on three sub-ensembles of subjects [Gibbons et al., 2004, Wootters, 1987, 2004]. These observables would need to be "informationally complete", and so completely specify the state of the system. Thus, this method holds promise for being able to help us to estimate the unknown state of a cognitive state. An advantage of this approach is that it allows us to express a cognitive state as a real function on a discrete phase space instead of using the common method of the density matrix. This real function which is known as a Wigner function behaves as a probability distribution, but it can take negative values [Gibbons et al., 2004, Wootters, 2004]. There are recent interests of using negative probabilities in cognition to model a decision making experiment [J Acacio de Barros and G Oas, 2014] or even to model contextuality [de Barros et al., 2016]. As an alternative to these approaches we could consider the Wigner function [Delfosse et al., 2017, Kenfack and Yczkowski, 2004, Raussendorf et al., 2017] which has been widely used in nonclassical calculations in QM. This method could potentially be extended to more complex situations of multiple cobits tomography, with an associated increase in the number of necessary observables. To reduce the number of measurements required to specify the description of multiple cobits, we could use POVM instead of the projective measurements [Lundeen et al., 2008, Wootters, 2004]. There is also a possibility of extending this method to the more ambitious scenarios of "cotrit" (three-level cognitive system) or "codit" (d-level cognitive system) tomography based on qutrit and qudit tomography in physics [Thew et al., 2002]. Thus, the examples like that modelled in (3.19) could be more completely specified using this approach, an avenue that I leave to future work.

3.6 Conclusion

In this chapter I assumed that the measurement process for a cognitive system is separate from its preparation process. I provided a detailed mathematical description of these two processes by introducing density matrices and non-ideal measurement. Having created this more rigorous approach I applied it to existing concepts in QC such as complementary and contextuality, as well as investigating how it might be extended to new concepts like tomography.

I believe these approaches provide a better quantum inspired models of cognition, and so could lead to a better understanding of cognitive systems. As I will describe in next chapter (see Section 4.6), the model of contextuality based on Spekkens' operational method provides a new way to study this important phenomenon in cognition. This general notion of contextuality will be explained based on the precise description of density matrix and POVM provided in the current chapter. This model is more comprehensive than previous studies because of its consideration of the preparation process and non-ideal measurements.

This work provides us with a number of new avenues to follow as we attempt to develop a more detailed understanding of the complex process of cognition, specifically, memory and recall. It suggests a way in which we could start to approach the problem of modelling an underlying cognitive state, and so work towards plausible models of the various ways in which these evolve as a person interacts with the world [Nelson et al., 2013]. The array of episodic events that each of us takes part in every day all influence our underlying cognitive state, and it is essential that we develop modelling methodologies that are capable of capturing the full complexity of this important process. Adopting an operational approach to QC offers precisely this opportunity.

Chapter 4

Contextuality

Contextuality is a very difficult feature to understand in both psychological and physical systems. Over the years, the field of quantum cognition (QC) has seen a number of papers attempting to clarify the nature of this important phenomenon (I reviewed several examples in Chapter 2, for a further selection see e.g. Atmanspacher et al. [2015], Kitto [2008]). However, ongoing research continues into understanding how to define and model contextuality in both physics and psychology. Intriguingly, much work in this area has already been completed by different groups in physics [Abramsky and Brandenburger, 2011, Acín et al., 2015, Cabello et al., 2010a, Liang et al., 2011, Spekkens, 2005] who have attempted to understand how phenomena such as contextuality, nonlocality and complimentarity interrelate, however, many of the more technical results have not to date been incorporated in the field of QC. It is unfortunate that some of these important results are less well known to QC, and call some recent claims made in the QC community into question. This chapter will draw attention to some of these more subtle results and consider what impact they have in QC. This is in alignment with the second aim of my thesis: to provide clarity and novel insights regarding contextuality in QC.

4.1 Introduction

Contextuality is an emerging topic in QC, in which opinions and associated formalizations about it are still forming. As yet, there is no received view on the topic and my thesis will help contribute to this gap by utilizing some recent advanced models of contextuality in QM. In this chapter, I will provide a more complete introduction for the concepts of contextuality and non-locality than in my initial introductions in Chapters 1 and 2; I will focus on three recent approaches to understanding contextuality, two belonging to quantum mechanics (QM), and one mainly known in QC. I will provide consistent descriptions of these approaches which helps to translate their specific notations and meanings to each other.

One of those two approaches in QM is based on the operational method and was originally created by Spekkens in 2005. There are several developments in different philosophical and mathematical aspects of quantum theory based on this approach; for example, new perspectives on arguments for realistic interpretation [Pusey et al., 2012] and epistemic interpretation [Harrigan and Spekkens, 2010] of quantum states or negative probability [Spekkens, 2008]. The aim of this thesis is not to cover all of these perspectives; but it will clearly review the notation of the Spekkens' operational approach especially by comparing with the standard notations of the Bell and KS theorems. This will suggest a possible application of Spekkens' approach to QC (See Section 4.6).

Another major approach to modeling contextuality in physics is based on sheaf theory, which unifies non-locality and contextuality concepts [Abramsky and Brandenburger, 2011]. There is an increasing amounts of research on this approach e.g. its development using Cohomology theory [Abramsky et al., 2015] and its application in database theory Abramsky [2013]. Similar to the case of the Spekkens approach, I will not cover all of these different advances in my thesis; but I will briefly introduce the algebraic structure of sheaf theory and will review connections between this structure and the standard formalism of Bell's inequality and the KS theorem. Based on these connections, I will construct a comparison between this sheaf-theoretical approach and the operational theory of Spekkens.

Contextuality-by-default (CBD) [Dzhafarov et al., 2015a, Dzhafarov and Kujala, 2012, 2014b,c, 2015, Dzhafarov et al., 2016] is the third approach that I incorporate in this chapter. As I described in Chapter 2, Dzhafarov and Kujala [2014c] truly highlight that Bell-type inequalities violations are irrelevant when marginal selectivity (non-signaling condition) is not satisfied. This is similar to what Brask and Chaves [2017] explain: "Clearly, if arbitrary communication is allowed between the parties, any correlations can be explained classically, and there is no nonlocality". However, Dzhafarov and Kujala [2014b] believe the CBD apporach can define contextuality even in the presence of signaling. In Chapter 2, I explained that CBD is a contextuality model arising from quantum cognition that employs a new approach for defining random variables and the probability space. In this chapter I will explain this approach and formally compare it with other approaches to modeling contextuality. This helps us to identify the possible differences and contradictions between these approaches, which is important to reach a logically consistent realization of contextuality in QC.

The outline of this chapter is as follows. In Section 4.2, I will provide a detailed introduction about contextuality in QM, which includes the central topics of the Bell and the Kochen-Specker theorems. In Sections 4.3 and 4.4, I will introduce the operational and the sheaf-theoretical approaches respectively, and I will represent how the notations of these two generalized approaches fit into the Bell and Kochen-Specker theorems. The mathematical descriptions in Sections 4.2, 4.3 and 4.4 will be mostly independent of their possible cognitive meanings. In these sections, I will mainly construct connections between different existing notations to provide a better understanding of contextuality in QM. From this point, I will introduce and evaluate the CBD approach in Section 4.5, where I will highlight how it differs from the former approaches. Section 4.6 constitutes the main part of my contribution to contextuality in QC, where I will apply the described operational approach to model contextuality in a cognitive system. I will conclude in Section 4.7 with some avenues for future research.

4.2 Contextuality in quantum mechanics

Contextuality in QM indicates that the assignment of predetermined outcomes to observables depends on the context or method of observation. Contextuality can be considered a strong signature of nonclassicality that plays a crucial role in quantum information and computation.

The EPR argument (described in BOX 4) claims that the state description of a physical system in QM is incomplete, more complete descriptions have been proposed in terms of hidden variable theories by physicists like Einstein. [1]. No-go theorems like Bell and Kochen-Specker (KS) place a barrier to using this type of hidden variable descriptions in QM. In Bell's theorem, the premise of locality is the source of contradiction between the hidden variable model and the statistical predictions of QM, while in the KS theorem the premise of

[1] These hidden variables in quantum mechanics can be classified in terms of being ψ*-ontic* or ψ*-epistemic* [Harrigan and Spekkens, 2010]; The hidden-variables are ψ*-ontic* when every complete physical state (ontic state) of the system is consistent with only one pure quantum state, and ψ*-epistemic* when ontic states are consistent with more than one pure quantum state. I should note that, an ontic state is a state of reality and an epistemic state is a state of knowledge of reality [Spekkens, 2007].

noncontextuality can play this role. Nonlocality can mostly be considered as a particular case of contextuality: Several research articles reveal the relationship between these two concepts, for example, Fine [1982b] relates both Bell and KS theorems to a marginal property, to determine if it is possible to have a joint distribution for a set of observables when there are joint probability distributions of subsets of those observables. Mermin [1993] shows that it is possible to replace noncontextuality by locality notations for a class of no-hidden variables theorems. In another example, Acín et al. [2015] construct a general contextuality model using the combinatorics of hypergraphs; the special case of this formalism gives rise to the nonlocality of Bell's scenario.

Contextuality is also closely related to the concept of complementarity in QM. As Liang et al. [2011] explain, these two fundamental concepts imply that two objects can not be jointly measured (explained) by noncontextual pre-existing properties. There has been some recent interest in QM, to construct an explicit relationship among the concepts of contextuality, non-locality and complementarity [de Muynck, 2007, Liang et al., 2011, Su et al., 2015]. For example, de Muynck [2007] designs a general form of Bell's inequality that can consider complementarity as a reason for violation instead of the more commonly held consideration of non-locality, the method that I applied to generalize the recall experiment in Chapter 3.

In this section, I review the notions of Bell's inequality and the KS theorem in detail. I clarify mathematical characteristics of Bell's inequality, probability space, observables and random variables. Moreover, I discus some conceptual features of Bell's inequality such as parameter independence, outcome independence and the non-signaling condition. Following these two preparatory steps, I delineate the generalized notations of contextuality in the rest of this chapter.

4.2.1 Bell's inequality

As mentioned in Box 4, the CHSH inequality is a form of Bell's theorem that is experimentally realizable; its violation demonstrates that nature cannot be described by local hidden variables theories. In this section, I provide more details about this inequality, mainly through the consideration of a prominent method suggested by Shimony [1984].

A classical probability space is used to construct deterministic hidden variables theories for observables in Bell's experiment [Fine, 1982b]. I now build up a formal description of what this entails using finite classical probability. Consider an experiment with events that occur randomly; the probability space is a mathematical structure that can model this experiment and consists of three parts ($\wedge$, $\mathcal{F}$, ρ). Sample space $\wedge$ is the set of all possible outcomes (or simple events) that are represented by λ ($\lambda \in \wedge$). Event space $\mathcal{F}$ (or $\Sigma(\wedge)$) is a set all possible events, which themselves can be described as groups of simple events and zero. In other words, $\mathcal{F}$ is a collection of subsets of $\wedge$. And the function ρ assigns probabilities to each event in $\mathcal{F}$, which is called the "probability measure" [2]. In physical or psychological experiments, a random variable quantifies the outcomes of a random process by mapping those outcomes to real numbers. In other words, a random variable $R(\lambda)$ is a function from the sample space $\wedge$ to the real numbers.

I will illustrate how these notations are used, by considering Bell's experiment, which involves measurements on spacelike separated systems S_1 and S_2 (as described in Section 2.2.1). Measurements are performed on the physical

[2] I use the same notation ρ for the "density matrix" in this thesis, but I used it here for the "probability measure" to be consistent with the notations in [Shimony, 1984]. This strategy helps to distinguish the probability measure (ρ) from other probability distributions (p) that are represented later in this chapter.

system S_1 using measuring observables $\mathbb{A}_1$ and $\mathbb{A}_2$, and on the system S_2 using observables $\mathbb{B}_1$ and $\mathbb{B}_2$. The result of measurements on S_1 and S_2 are labeled by bivalent variables a and b respectively which can take either -1 or 1 [3]. The conceptual framework of Bell's inequality relates to an ensemble of pairs of these systems. Each pair of systems can be represented with S $(S_1 + S_2)$ and is characterized by λ, "the complete state of the system" [Fine, 1982a, Shimony, 1984], which is known as an ontic state [4]. The ontic state can alter from pair to pair [Shimony, 2016], and the set of the ontic states that provides the complete specification of the system is represented by Λ. In cognition, an ontic state should refer to the reality of the cognitive system, that is, the presumed features of a cognitive state (of mind) which exist without performing experiments or any other form of observation. My definition of ontic state in cognition is illustrated with reference to the recall experiment in Section 4.6.

Returning to the notation of the Bell experiment, for bivalent observables $\mathbb{A}$ and $\mathbb{B}$, one can define the following probability distribution [Shimony, 2016] [5]

1. $p^1_\lambda(a|\mathbb{A}, \mathbb{B}, b)$ = The probability of obtaining outcome a by performing a measurement on S_1 when λ is the ontic state (b is the result of measurement on S_2),

2. $p^2_\lambda(b|\mathbb{A}, \mathbb{B}, a)$ = The probability of obtaining outcome b by performing a measurement on S_2 when λ is the ontic state,

[3] Although in the original version of Bell's theorem, a and b can take either -1 or 1, in other versions they may take value on any of a discrete set of real numbers in $[-1, 1]$[Shimony, 2016].

[4] A more complete discussion of *ontic sate* and *hidden variables* can be found in [Harrigan and Spekkens, 2010]; for a situation that quantum states do not provide a complete description of reality, they classified ontological models of quantum theory into two categories: (I) *ψ-supplemented*, in which the complete ontic state is given by quantum state ψ supplemented by some additional variables and (II) *ψ-epistemic*, where the quantum state ψ can only define a probability distribution over ontic space.

[5] Shimony [1984] called $p_\lambda(\mathbb{A})$ and $p_\lambda(\mathbb{B})$ "random variables", since one can consider them as functions from the sample space $\wedge$ to the real numbers.

3. $p_\lambda(a, b|\mathbb{A}, \mathbb{B})$ = The probability of obtaining outcomes a and b by the joint measurements $\mathbb{A}$ and $\mathbb{B}$ when λ is the ontic state.

The probability function p is non-negative and summation over all allowed values of a and b, leads to unity (i.e. it is normalized) [Shimony, 2016].

In Bell's experiment, there is a distribution ρ (probability measures) over each λ, which is independent of the mounting of the measurement apparatuses (observables). This situation is called λ-independence by Berkovitz [2016]. But the probability distribution of a quantum mechanical state corresponding to an ensemble of the described pairs (S), depends on the probability distribution of the *complete* states λ. This is called empirical adequacy and shows the probability functions p of outcomes in Bell's experiment can be recovered by averaging over probability measures ρ [Berkovitz, 2016, Shimony, 1984]:

$$p^1(a|\mathbb{A}, \mathbb{B}, b) = \int_\Lambda p^1_\lambda(a|\mathbb{A}, \mathbb{B}, b)d\rho, \tag{4.1a}$$

$$p^2(b|\mathbb{A}, \mathbb{B}, a) = \int_\Lambda p^2_\lambda(b|\mathbb{A}, \mathbb{B}, a)d\rho, \tag{4.1b}$$

$$p(a, b|\mathbb{A}, \mathbb{B}) = \int_\Lambda p_\lambda(a, b|\mathbb{A}, \mathbb{B})d\rho. \tag{4.1c}$$

The Bell argument, in general, indicates that the predictions of QM violate derived inequalities based on a locality assumption [Bell, 1964]. But what precisely is the meaning of locality in Bell's theorem? Locality in Bell's experiment is commonly represented as a factorizability condition [Bell, 1964]; the condition that equates the probability of joint outcomes with the product of the probabilities of the individual outcomes a and b:

$$p_\lambda(a, b|\mathbb{A}, \mathbb{B}) = p^1_\lambda(a|\mathbb{A})p^2_\lambda(b|\mathbb{B}). \tag{4.2}$$

Rephrasing the story of Bell's theorem, we can say the combination of factorization and λ-independence and empirical adequacy connote Bell's inequality [Berkovitz, 2016], which is violated by predictions of QM. Since the predictions of orthodox QM are experimentally accurate and because of the plausibility of λ-independence, Bell [1964] concluded that only factorization can fail in this experiment. Translating the factorization as locality, it is concluded that QM is non-local.

Jarrett [1984] and Shimony [1993] demonstrate the factorization condition as the conjunction of two weaker locality conditions: parameter independence (PI) and outcome independence (OI). This means that the violation of Bell's inequalities requires that either of these two locality conditions is violated. PI can be a barrier for "superluminal signals"(faster-than-light signaling of information) for two sides of the Bell experiment; it entails the result of measurement in one side being statistically independent of the $setting$ (preparation) of the other measuring device [Jarrett, 1984, Maudlin, 2011, Shimony, 1993]:

$$\begin{aligned} &p^1_\lambda(a|\mathbb{A},\mathbb{B}) = p^1_\lambda(a|\mathbb{A}) \quad and \\ &p^2_\lambda(b|\mathbb{A},\mathbb{B}) = p^2_\lambda(b|\mathbb{B}). \end{aligned} \tag{4.3}$$

It is commonly agreed among physicists that superlumnial signaling is impossible in practice; as Maudlin [2011, p.77] said "there is no Bell's telephone"in the experiment.

OI entails the probability of an outcome on one side being independent of the probability of an outcome on the other side [Jarrett, 1984, Maudlin, 2011, Shimony, 1993]:

$$\begin{aligned} p^1_\lambda(a|\mathbb{A},\mathbb{B},b) &= p^1_\lambda(a|\mathbb{A},\mathbb{B}) \quad and \\ p^2_\lambda(b|\mathbb{A},\mathbb{B},a) &= p^2_\lambda(b|\mathbb{A},\mathbb{B}). \end{aligned} \tag{4.4}$$

The failure of this condition which entails some sort of non-separability (or holism [6]) is consistent with relativity and considered as the main reason for violation of Bell's inequality in standard QM [Jarrett, 1984, Shimony, 1993]. To summarize, satisfying the λ-independence and the failure of PI are adequate conditions for the existence of superluminal signaling, but not satisfying the λ-independence and the failure of OI [7]. However, without satisfying the λ-independence, we can have completely different situations, e.g. a deterministic theory such as Bohmian quantum theory [Bohm, 1952] violates PI and satisfies OI [Dickson, 1998, Fano, 2006] [8].

Part of the intention behind this detailed analysis of Bell's inequality in this section is to provide a better understanding of the non-signaling condition in QC, a concern that has arisen recently for this area (see Section 2.2.1). This condition that stems from special relativity, expresses that the measurement of $\mathbb{A}$ should not be able to obtain any information about the measurement setting of $\mathbb{B}$ (and vice versa) by only looking at the statistics of their local measurement outcome. In fact, the design of Bell's experiment involves this non-signaling condition and special relativity. Moreover, to apply Bell's argument in areas outside physics, we cannot exclude an assumption or a requirement in building

[6] For more details, refer to Berkovitz [2016].

[7] In a completely different view, Maudlin [1994] believes that splitting the above factorization condition into PI and OI is not unique and essentially nonsensical. In this thesis I follow Shimony [1993]'s argument, since evaluating the Maudlin's factorization condition in QM requires additional assumptions which make it unnatural [Seevinck, 2008].

[8] We can check these conditions for other interpretations of QM; for example, the spontaneous collapse theory of GRW [Ghirardi et al., 1986] satisfies the PI condition and violates OI condition [Butterfield et al., 1993, Mittelstaedt, 1997] like the standard interpretation.

the theory and expect the remaining parts to still have meaning. The formal definition of non-signaling can be given as

Definition 5 *The joint probability distribution in equations* (4.1) *satisfies non-signaling if*

$$\begin{aligned} p(a|\mathbb{A}) &= \sum_b p(a,b|\mathbb{A},\mathbb{B}), \\ p(b|\mathbb{B}) &= \sum_a p(a,b|\mathbb{A},\mathbb{B}). \end{aligned} \tag{4.5}$$

For example in Table 4.1, the probability of $a_1 = +1$ is independent of the setting for measurement in the other side of the experiment if $p_1 + p_2 = p_3 + p_4$.

Table 4.1: The joint probability distributions for Bell's experiment

$\mathbb{A}, \mathbb{B}$	$b_1 = +1$	$b_1 = -1$	$b_2 = +1$	$b_2 = -1$
$a_1 = +1$	p_1	p_2	p_3	p_4
$a_1 = -1$	p_5	p_6	p_7	p_8
$a_2 = +1$	p_9	p_{10}	p_{11}	p_{12}
$a_2 = -1$	p_{13}	p_{14}	p_{15}	p_{16}

Comparing the PI with the non-signaling conditions gives us a deeper understanding of the signaling concept in Bell's theorem. Both conditions are prohibitions for superluminal signaling in Bell's experiment, but PI can be considered as non-signaling for a given λ. As Bub [2016, p. 76] states "The non-signaling principle is an observational or operational condition characterizing the surface phenomenon and, it does not refer to λ ". The violation of PI in theories like Bohmian quantum mechanics only permits signaling for hidden variables (at the ontological level), whereas the non-signaling condition still prohibits the signaling at the phenomenological [9] level [Bendersky et al., 2017].

[9] To gain an accurate meaning of phenomenology in philosophy, we can compare it with the

The possible representations of the PI and the non-signaling conditions in the CBD approach are discussed in Section 4.5.

4.2.2 The Kochen-Specker theorem

Realistic and deterministic interpretations require that a definite outcome (value) is assigned to a projector $\mathbb{A}$, i.e. $v(\mathbb{A}) \in 0, 1$, even before it is measured [Harrigan and Rudolph, 2007]. The KS theorem considers the possibility that the selection between values 0 or 1 depends on the performed PVM. Such a dependence is termed contextuality which was first introduced in 1967 by Kochen and Specker. This theorem shows that the predictions of quantum theory cannot satisfy a non-contextual framework of the measurements with pre-determined outcomes.

Among different proofs of the KS theorem, I will review a simple proof by Cabello et al. [1996] using the Spekkens' framework in Section 4.3 and later, using the Abramsky approach in Section 4.4

4.3 The operational approach of Spekkens

Spekkens [2005] generalized the standard treatment of contextuality in QM to arbitrary operational theories, which allows us to identify contextuality in QM and provide ways to test it experimentally. Spekkens' generalized notion of contextuality is applicable to unsharp measurement (POVM was described in Section3.3) and is not restricted to deterministic hidden variable models.

---

other disciplines in philosophy such as "ontology" (the study of being) and "epistemology" (the study of knowledge). Here, I provide a definition of phenomenology by Smith [2016, p. 1] "Phenomenology is the study of phenomena: appearances of things, or things as they appear in our experience, or the ways we experience things, thus the meanings things have in our experience".

In this section, I will review the Bell and KS theorems, using Spekkens' approach. This provides a path to compare Spekkens' approach with the sheaf theory and the CBD approaches in Sections 4.4 and 4.5 respectively. In Section 4.6, I will describe the application of this approach in cognition.

4.3.1 Spekkens' definition of measurement and preparation noncontextuality

To introduce Spekkens' notation of contextuality, we need the following definitions of operational theories and ontological models. An operational theory provides general formulations of results for different processes of a system [Abramsky and Heunen, 2016].

Definition 6 *An operational theory can be defined for procedures of preparation* (P) *and measurement* (M)*, which can be denoted by (*$\mathcal{P}$*,*$\mathcal{M}$*,*p*), where* $\mathcal{P}$ *and* $\mathcal{M}$ *are the sets of preparation and measurement procedures respectively, and* p *specifies the probability that the measurement procedure* $M \in \mathcal{M}$ *subsequent to the preparation procedure* $P \in \mathcal{P}$ *leads to the outcome* $k \in \mathcal{K}_M$ *[Harrigan and Spekkens, 2010, Kunjwal, 2016, Spekkens, 2005].*

Definition 7 *An ontological model is defined as (*Λ, μ, ξ*), where* Λ *is the set of ontic states as I defined in Section 4.2.1. The probability distribution of selecting the ontic states* $\lambda \in \Lambda$ *by a preparation procedure* P *is specified by* $\mu_P(\lambda)$ *that takes value in the interval of* $[0, 1]$ *for different* λ*. The probability distribution of the incidence of a measurement outcome* $[k|M]$ *by implementing a measurement procedure* M *for any ontic states* $\lambda \in \Lambda$ *is defined by* $\xi_\lambda(k|M) \in [0, 1]$*.*

The operational theory and the ontological model relate to each other using the equation [Spekkens, 2005]

$$p(k|M,P) = \sum_{\lambda\in\Lambda} \xi_\lambda(k|M)\mu_P(\lambda). \tag{4.6}$$

This empirical adequacy condition states that the ontological model reproduces the predictions of the operational theory. This means, by averaging the probability of a measurement M for a given ontic state λ on the distribution of ontic states sampled by the preparation process P, we can obtain the operational probabilities $p(k|M,P)$.

Having provided the mathematical details of the operational statistics of the preparation and measurement processes, we can now study the definition of the operational equivalence of two experimental procedures:

Definition 8 *Two preparation procedures (P and P') are operationally equivalent if*

$$p(k|M,P) = p(k|M,P') \quad \textit{for all } M. \tag{4.7}$$

This situation, which is denoted by ($P \simeq P'$) [Kunjwal, 2016], shows no subsequent measurement procedures can make a difference between the operational statistics of the two preparation procedures [Spekkens, 2005]. Similarly,

Definition 9 *Two measurement events ($[k|M]$ and $[K'|M']$) are operationally equivalent if*

$$p(k|M,P) = p(k|M',P) \quad \textit{for all } P. \tag{4.8}$$

This is denoted by ($[k|M] \simeq [K'|M']$), and means no preceding preparation procedures can make any difference in the operational statistics of those two measurement events [Kunjwal, 2016, Liang et al., 2011, Spekkens, 2005]. If all measurement events existing in the two measurement procedures M and M' are operationally equivalent ($[k|M] \simeq [K'|M']$), then the two measurement procedures are operationally equivalent as well ($M \simeq M'$). In general, when two experimental procedures are operationally equivalent, we can say they reside in the same operational equivalence class of those experimental procedures. Any feature that can distinguish between two experimental procedures in a certain operational equivalence class is considered as a context for that procedure. So we can define the assumption of noncontextualty for preparation and measurement procedures based on the definition of operational equivalence as

Definition 10 *[Kunjwal, 2016, Liang et al., 2011, Spekkens, 2005] An ontological model is preparation noncontextual if we can represent every preparation procedure independent of context:*

$$P \simeq P' \Rightarrow \mu_P(\lambda) = \mu_{P'}(\lambda) \quad \forall\lambda \in \Lambda, \tag{4.9}$$

And the model is measurement noncontextual if we can represent every measurement event independent of context:

$$[k|M] \simeq [K'|M'] \Rightarrow \xi_\lambda(k|M) = \xi_\lambda(k'|M') \quad \forall\lambda \in \Lambda. \tag{4.10}$$

This noncontextuality is expandable to the level of measurement procedures, if there is one-to-one operational equivalence between all measurement events existing in two measurement procedures.

Actually, the Spekkens' idea of noncontextuality represents the impossibility of drawing distinctions in reality when there is no observable difference in the experience of that reality [Kunjwal, 2016]. The following sections will fit the Bell and KS theorems into this framework, which will contribute to a better understanding of this contextuality model. This will enable me to apply the Spekkens' approach to QC. As mentioned earlier, in Section 4.6, I will use this approach to treat contextuality as it occurs for both preparation and measurement processes in cognition.

4.3.2 Bell's inequality in Spekkens' approach

I relate the probability distribution $\mu_P(\lambda)$ in Equation (4.6) to the probability measure ρ in Equations (4.1), by using the relation $\mu_P(\lambda)d\lambda = d\rho$. Therefore, the difference between these two distributions lies in the choice of the discrete or the continuous probability spaces of ontic states. For example, as a result of a transformation from the continuous to the discrete probability space, equation (4.1c) converts to the operational ontological notation:

$$p(a, b|\mathbb{A}, \mathbb{B}; P) = \sum_{\lambda \in \Lambda} \xi_\lambda(a, b|\mathbb{A}, \mathbb{B})\mu_P(\lambda). \tag{4.11}$$

Using this notation, the PI condition (4.3), OI condition (4.4) and factorization condition (4.2), for each ontic state λ, convert to the following forms that can be found in [Kunjwal, 2016].

- The PI condition:

$$\begin{aligned} &\xi^1_\lambda(a|\mathbb{A}, \mathbb{B}) = \xi^1_\lambda(a|\mathbb{A}) \quad and \\ &\xi^2_\lambda(b|\mathbb{A}, \mathbb{B}) = \xi^2_\lambda(b|\mathbb{B}), \end{aligned} \tag{4.12}$$

- The OI condition:

$$\begin{aligned}\xi^1_\lambda(a|\mathbb{A},\mathbb{B},b) &= \xi^1_\lambda(a|\mathbb{A},\mathbb{B}) \quad and \\ \xi^2_\lambda(b|\mathbb{A},\mathbb{B},a) &= \xi^2_\lambda(b|\mathbb{A},\mathbb{B}),\end{aligned} \tag{4.13}$$

- The factorization condition:

$$\xi_\lambda(a,b|\mathbb{A},\mathbb{B}) = \xi^1_\lambda(a|\mathbb{A})\xi^2_\lambda(b|\mathbb{B}). \tag{4.14}$$

where ξ is the response function specified in Definition 7. Finally, the non-signaling property (4.5) is expressed using this notation as

$$\begin{aligned}p(a|\mathbb{A};P) &= \sum_b p(a,b|\mathbb{A},\mathbb{B};P), \\ p(b|\mathbb{B};P) &= \sum_a p(a,b|\mathbb{A},\mathbb{B};P).\end{aligned} \tag{4.15}$$

Although the experimental statistics of Bell's theorem are suitable to test (non)locality, there are no similar experimental statistics (i.e. (non)contextuality inequalities) to test the traditional notion of noncontextuality [Spekkens et al., 2009]. To achieve this experimental testability we can use the Spekkens' generalized notation of contextuality [Spekkens, 2005] (described in Section 4.3.1). This generalized notation can lead to a noncontextuality inequality that treats any Bell's inequality as a special case [Spekkens et al., 2009]. This operational approach can also provide an experimental test of contextuality beyond the KS theorem [Kunjwal and Spekkens, 2015]. Each of these experimental tests of contextuality can open an avenue of future research in QC. In the next section, I will briefly discuss the relationship between this generalized notation and the KS theorem.

4.3.3 Kochen-Specker theorem in Spekkens' approach

In this section, I review a very simple proof of the KS theorem using 18 vectors in four-dimensional Hilbert space, as suggested by Cabello et al. [1996]. Figure 4.1 depicts the different parts of this proof, in which nodes are associated with measurement events (projectors over 18 vectors [10]). There are 9 loops corresponding to sets of measurement events, each consisting of four nodes. In other words, 18 vectors are divided into 9 orthogonal bases consisting of 4 vectors, in such a way that each vector appears in exactly two orthogonal bases.

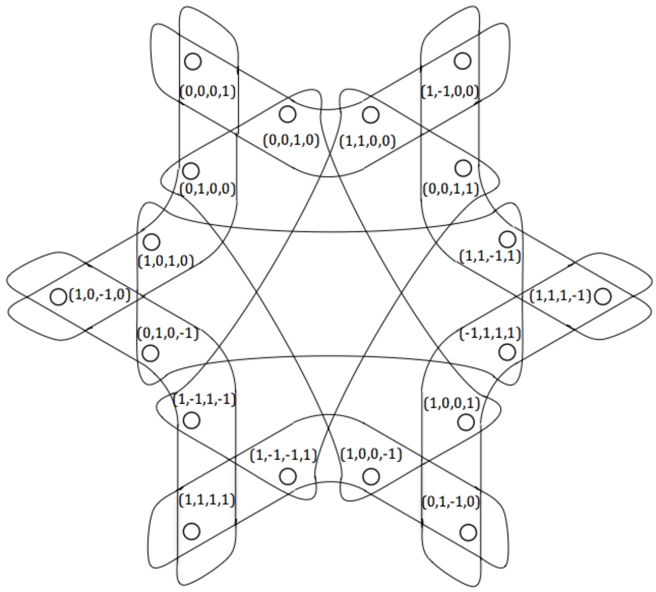

Figure 4.1: An example of the KS theorem with 18 vectors by Cabello et al. [1996].

Before describing the proof, I need to introduce some notations that Cabello et al. [1996] used for this proof. In Figure 4.1, if we denote the projector operator by $\mathbb{P}_k$, then we have projector operators $\{\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3, \mathbb{P}_4\}$ for 4 vectors in each orthogonal bases. These projectors are all mutually commuting (i.e. they

[10] Denoting the row vectors in Figure 4.1 by u, we can represent the associated projectors as $|u\rangle\langle u|$. [Cabello et al., 1996].

correspond to compatible observables). The probability of these observables can be represented by $\{v(\mathbb{P}_1), v(\mathbb{P}_2), v(\mathbb{P}_3), v(\mathbb{P}_4)\}$, which are functions onto real numbers. Since $\mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 + \mathbb{P}_4 = I$, we have

$$v(\mathbb{P}_1) + v(\mathbb{P}_2) + v(\mathbb{P}_3) + v(\mathbb{P}_4) = v(I) = 1. \tag{4.16}$$

Meanwhile, we have the following property for compatible observables:

$$v(\mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 + \mathbb{P}_4) = v(\mathbb{P}_1) + v(\mathbb{P}_2) + v(\mathbb{P}_3) + v(\mathbb{P}_4). \tag{4.17}$$

Considering this property and the outcome determinism assumption that $v(\mathbb{P}_i) \in \{0, 1\}$ ($i \in \{1, 2, 3, 4\}$), we can conclude that only one of the values $\{v(\mathbb{P}_1), v(\mathbb{P}_2), v(\mathbb{P}_3), v(\mathbb{P}_4)\}$ is 1 and the others are 0.

Now to represent the proof, we must look at Figure 4.1 again. The associated projector of each node belongs to two orthogonal bases at the same time. This means each basis is a context for the projector. Obviously there is the operational equivalence between the measurement events in these two contexts (4.7). Because the operational equivalence requires that the probability of occurrence of a measurement event be the same for any quantum state regardless of the context selected.

The condition of measurement noncontextuality is satisfied for this picture if λ (the ontic state of the system) assigns the same probability to operationally equivalent measurements. This is possible by assigning probabilities to nodes in a way that their sum is equal to 1 for each loop (as shown in Equation (4.16)). But the KS contradiction arises when we consider the additional assumption of the outcome determinism. The outcome determinism in the ontological model entails the response functions are deterministic for all ontic states, i.e. $\xi_\lambda(k|M) \in$

$\{0, 1\}$ for all $M \in \mathcal{M}$, $k \in \mathcal{K}_M$, $\lambda \in \Lambda$ [11] [Kunjwal, 2016, Liang et al., 2011, Spekkens, 2005] As a result of this additional assumption, we require a *measurement noncontextual* assignment of values $\{0, 1\}$ to the measurement events by ontic state λ [Kunjwal and Spekkens, 2015, Liang et al., 2011]. Such an attempt fails. As can be seen in Figure 4.1, there are 9 bases, so the total number of projectors given the value of 1 by λ should be odd. However, every vector appears in two bases, which shows that the total number of projectors assigned the value of 1 is even, not odd. Therefore, Spekkens' notation translates the KS theorem as a restriction for either the measurement noncontextuality and the outcome determinism or both.

Kunjwal [2016] shows that the preparation noncontextuality (described in Section 4.3.1) implies both outcome determinism and measurement noncontextuality. So in general, the preparation noncontextuality implies the KS theorem in ontological models of quantum theory. He concludes that the preparation noncontextuality is a stronger (more general) condition than the KS noncontextuality. In other words, it is possible to rule out preparation noncontextuality without dealing with KS theorem. In this thesis, it is not necessary to include the detailed mathematical proofs of this assertion, but I will represent the possible application of this generalized form of contextuality in cognition (see Section 4.6). This framework extends beyond the KS framework, as it can be used with unsharp (or nonprojective) measurements (unlike the KS theorem), and also allows a treatment of contextuality for any arbitrary operational theories and not for quantum theory alone [Spekkens, 2005]. In the next section, I will briefly introduce the sheaf-theoretical structure of contextuality. I will represent the Bell and KS theorems again, but this time using the sheaf theory notation. This helps to compare the sheaf-theoretical approach with the generalized model of

[11]Where $\mathcal{M}$ is a set of measurement procedures and $\mathcal{K}_M$ is a set of outcomes.

Spekkens and reach a better understanding of their notations. Based on this better understanding I can review and evaluate the CBD approach, the other generalized approach which is mainly known in QC (see Section 4.5).

4.4 The sheaf-theoretical approach of Abramsky and Brandenburger

Sheaf theory provides a language that uniquely connects locally defined sets of data; these data should be "compatible" and "gluable". In other words, for a large sheaf of such data, this language reconstructs a general situation by looking at how each set of data behaves locally [Lovering, 2010]. In this section, I will attempt to avoid the technical details of this algebraic structure and convey the meaning only by using the special example of contextuality suggested by Abramsky and Brandenburger [2011].

Abramsky and Brandenburger [2011] provide a unified treatment of contextuality and non-locality based on the mathematical structure of sheaf theory. Elaborating upon their approach requires some more specific notation (which I will keep specific to that paper in order to avoid confusion):

- X: A set of measurements (or observables).
- U: A set of measurements which is a subset of X ($U \subseteq X$). This corresponds to different combinations of observables that can be measured together.
- O: A set of possible outcomes.
- s: A section over U which is a function $s : U \to O$. Each section represents an event which includes measurements in U. In each section, an outcome

s(m) is obtained for each $m \in U$. A set of functions from U to O is written as O^U.

- $\mathcal{P}(X)$: The powerset of X; the set of all subsets of X.

- A poset (or partially ordered set) on a set X is the pair $(X, \leq)$ where $\leq$ is a binary relation between certain pairs of elements in that set. The posets we need in this section is the powerset of X ordered by subset inclusion.

- $\mathcal{E}$: $U \mapsto O^U$ which associates the set of sections on U to each set of measurement U. Considering $\mathcal{E}(U)$ plus a action of restriction we have [12]

$$\begin{aligned} & res_U^{U'} : \mathcal{E}(U') \to \mathcal{E}(U) \qquad \text{for any} \quad U \in U' \\ & \text{be defined by} s \to s|U \\ & \text{where} \quad res_U^U = id_U \\ & \text{and if} \quad U \in U' \in U'', \text{then} \quad res_U^{U'} \circ res_{U'}^{U''} = res_U^{U''} \end{aligned} \tag{4.18}$$

Thus $\mathcal{E}$ is a pre-sheaf; a pre-sheaf of sections over poset $\mathcal{P}(X)$, i.e. $\mathcal{E}$ is a functor $\mathcal{E} : \mathcal{P}(X)^{op} \to \mathbf{Set}$ [13].

$\mathcal{E}$ is a sheaf if local sections can be uniquely glued together. This means for a family of sets $\{U_i\}_{i\in I}$ (denoted as a cover of U) and a family of sections $\{s_i \in \mathcal{E}(U_i)\}_{i\in I}$ which are compatible for all $i, j \in I$ as $s_i|U_i \cap U_j = s_j|U_i \cap U_j$, there is a unique section $s \in \mathcal{E}(U)$ in which $s|U_i = s_i \; \forall i \in I$. Abramsky and Brandenburger [2011] call $\mathcal{E}$ a sheaf of events.

- To consider distribution on events we need commutative semirings R: if

[12]Considering a function $f : X \to Y$, if $X' \subseteq X$, the restriction of f to X' can be demonstrated as $f|X' : X' \to Y$.

[13]The category shaped from the opposite poset is opposite category $\mathcal{P}(X)^{op}$.

we have a function $\phi : X \to R$, the support of ϕ is

$$Supp(\phi) := m \in X : \phi \neq 0, \tag{4.19}$$

for a finite support, an R-distribution on X is defined with a functor $d : X \to R$ such that

$$\sum_{x \in X} d(X) = 1. \tag{4.20}$$

- $\mathcal{D}_R\mathcal{E}$: a presheaf of R-distribution, which assigns the set $\mathcal{D}_R\mathcal{E}(U)$ of distributions on $\mathcal{E}(U)$ to each set of measurements U. Abramsky and Brandenburger [2011] consider R as a ring of reals, positive reals or Booleans to describe different contextuality scenarios. For example, Table 4.3 will represent Bell's argument as an empirical model over the positive reals.

- $\mathcal{M}$: A measurement cover or maximal sets of compatible measurements, which can be considered as an imposed restriction on the poset $\mathcal{P}(X)$, and can be represented as $\mathcal{M} \subset \mathcal{P}(X)$. A measurement cover usually refers to a set of measurement contexts; in QM (or psychology) only some measurements can be performed jointly, each set of such measurements is denoted as a measurement context $C \in \mathcal{M}$ (see Tables 4.3 and 4.4 which display the measurement cover $\mathcal{M}$ in Bell's experiment and a proof of KS theorem respectively). A measurement cover has the following properties

    1. $\cup\mathcal{M} = X$,
    2. $\mathcal{M}$ is an anti-chain, i.e. for measurement contexts $C, C' \in \mathcal{M}$ implies $C = C'$. This condition entails that $\mathcal{M}$ contains only maximal sets of compatible measurements.

The sheaf approach represents contextuality as a situation in which a family of data is *locally* consistent, but *globally* inconsistent [Abramsky et al., 2015]. Suppose we are given a group of contexts each consisting of sets of variables that can be jointly measured or observed. Suppose, moreover, that we are given a group of data or outcomes in a relative position with the group of contexts. *Local* consistency entails the existence of local sections that relate the variables in each context to a set of outcomes. But *global* inconsistency entails the lack of a global section over all the variables to reconcile all the local data. In this section, I will clarify the exact meaning of these terms by using this approach to represent the Bell and the KS theorems.

One of the reasons that I have chosen to introduce this sheaf-theoretical approach here is to compare its notation with the Spekkens' approach that I represented in Section 4.3 (see Table 4.2 and for more details see Sections 4.4.1 and 4.4.2). In this thesis I keep the original notations for both of these two approaches. I believe, comparing these notations provides a better understanding of these two approaches.

Table 4.2: Spekkens versus Abramsky notations. The gray texts denote my suggested translation of the distributions e_c and h_C^{λ} into Spekkens' notation.

	Spekkens		Abramsky
Set of measurements	null	$\sim$	X
Set of measurement procedures	$\mathcal{M}$	$\sim$	null
Measurement cover	null	$\sim$	$\mathcal{M}$
Measurement procedure	$M_{\in \mathcal{M}}$	$\sim$	null
A measurement event	$[K\|M]$	$\sim$	$m_{\in X}$
Probability measure	$\mu_P(\lambda)$	$\sim$	$h_\Lambda(\lambda)$
Empirical distribution for context C	$p(k\|C)$	$\sim$	e_c
Indicator function for context C	$\xi_\lambda(k\|C)$	$\sim$	h_C^{λ}

4.4.1 Bell's inequality in the sheaf-theoretical approach

In this section, I represent some important aspects of Bell's experiment using the sheaf-theoretic approach. The main aim is to compare this representation with the Spekkens' representation of those aspects as provided in Section 4.3.3.

I start with the definition of measurement cover in Bell's experiment. Abramsky and Brandenburger [2011] use I to label the space-like separated parts of the experiment. Considering a disjoint $\{X_i\}_{i\in I}$, where X_i is the set of measurements which can be performed at each part i, Abramsky and Brandenburger [2011] define $\mathcal{M}$ as subsets of X including only one measurement from each part. The performed measurements in different parts of the system are compatible, but it is not possible to have compatible measurements in the same part. For example, in a bipartite situation, we can have contexts $C = \{m_a, m_b\}$ and $C' = \{m_a, m'_b\}$ which are the same as contexts $C = \{\mathbb{A}_1, \mathbb{B}_1\}$ and $C' = \{\mathbb{A}_1, \mathbb{B}_2\}$ defined using Shimony's and Spekkens' notations as given in Sections 4.2.1 and 4.3.3 respectively. Abramsky and Brandenburger [2011] drop m from this notation and show the two possible measurements in the Alice side of the experiment by $\{a, a'\}$ and in the Bob side by $\{b, b'\}$, then the four maximal measurement contexts are represented as

$$\{a, b\}, \{a', b\}, \{a, b'\}, \{a', b'\}.$$

The rows in Table 4.3 are categorized based on these contexts. Each row corresponds to a set of section $\mathcal{E}(C)$, in which different sections assign values {0,1} to measurements a and b according to the context C. Abramsky and Brandenburger [2011] denote the family of probability distributions for different choice of measurements in this table as "an empirical model".

Table 4.3: The joint probability distributions for Bell's experiment [Abramsky and Brandenburger, 2011]. Since there are two space-like separated parts in Bell's experiment, the measurement cover $\mathcal{M}$ is considered as the set of contexts including only one measurement from each part: $\big\{\{a,b\},\{a',b\},\{a,b'\},\{a',b'\}\big\}$.

$\mathbb{A},\mathbb{B}$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
(a,b)	p_1	p_5	p_2	p_6
(a',b)	p_9	p_{13}	p_{10}	p_{14}
(a,b')	p_3	p_7	p_4	p_8
(a',b')	p_{11}	p_{15}	p_{12}	p_{16}

In this notation for each context $C \in \mathcal{M}$, there is a distribution $e_c \in \mathcal{D}_R\mathcal{E}(C)$ (a presheaf $\mathcal{D}_R\mathcal{E}$ is defined over the positive reals). However, there is no specific mathematical representation of such distributions in Spekkens' notation described in Section 4.3. So I suggest to translate this concept into Spekkens' notation as $p(k|C)$ (see Table 4.2). Note that such a difference in notations does not necessarily lead to a difference in meanings described by these two approaches. For example, consider the representation of the non-signaling condition in the sheaf-theoretical approach using this distribution:

$$e_C|C \cap C' = e_{C'}|C \cap C', \tag{4.21}$$

which provides exactly the same meaning as the earlier addressed conditions by Shimony and Spekkens in Equations (4.5) and (4.15) respectively. To show their similarity we can compare Tables 4.3 and 4.1 [14]. The special example in Table 4.1, that the probability of $a_1 = +1$ is independent of the setting for measurement on the other side of the experiment $(p_1+p_2 = p_3+p_4)$, is expressed by Abramsky

[14] Table 4.3 uses random variables $\{a, a', b, b'\}$ instead of $\{a_1, a_2, b_1, b_2\}$ in Table 4.1, and values $\{0, 1\}$ instead of {-1, +1}.

and Brandenburger [2011] in the sheaf-theoretical approach as

$$\sum_{s\in\mathcal{E}(C),s|m_a=s_0} e_C(s) = \sum_{s'\in\mathcal{E}(C'),s'|m_a=s_0} e_{C'}(s'), \tag{4.22}$$

where $s_0 \in \mathcal{E}(\{m_a\})$.

They define distribution h_C^λ, similar to e_C but for each $\lambda \in \Lambda$. There is no specific mathematical representation for this distribution in Spekkens' notation either. So I translate this concept into Spekkens' notation as $\xi_\lambda(k|C)$ (see Table 4.2). Abramsky and Brandenburger [2011] define the PI condition using this distribution as

$$h_C^\lambda|C \cap C' = h_{C'}^\lambda|C \cap C'. \tag{4.23}$$

They recover the probability distributions e, by averaging over the values of the hidden variable h. This is represented using a similar equation to (4.6):

$$e_c(s) = \sum_{\lambda\in\Lambda} h_C^\lambda(s).h_\Lambda(\lambda), \tag{4.24}$$

Here, $h_\Lambda(\lambda)$ ($h_\Lambda \in \mathcal{D}_R(\Lambda)$), like $\mu_P(\lambda)$, indicates the distribution that a preparation process P samples the ontic state $\lambda \in \Lambda$.

As mentioned earlier, the mathematical structure that sheaf theory provides to model contextuality is concerned with the comparison between local and global probability distributions [Abramsky et al., 2015]. To use this approach for Bell's scenario, Abramsky and Brandenburger [2011] associate the non-locality with an obstruction for the existence of a global section, e.g., the restriction to construct one certain joint distribution on the empirical predictions of all allowed joint measurements in Table 4.3. To describe this mathematically, we can say a probabilistic model $(X, \mathcal{M}, \{e_C\}, \mathbb{R}_{\geq 0})$ (for local probability distributions $\{e_C\}$) is noncontextual if there is a global probability distribution $d \in \mathcal{D}_{\mathbb{R}_{\geq 0}}\mathcal{E}$ compatible with the $\{e_C\}$ [Constantin, 2015]. Abramsky and Brandenburger [2011] consider the non-existence of such a global probability distribution as weak contextuality.

Abramsky and Brandenburger [2011] define different levels for contextuality, in which the described Bell non-locality corresponds to the level of probability distributions. There is a stronger form of contextuality that corresponds to the level of the supports of the distributions. The definition of this form of contextuality implies that a local assignment in the support of probability distribution $\{e_C\}$, cannot be built out to a global assignment compatible with the support. To describe this mathematically, we can consider the Boolean-valued distributions $\{e_C^{\mathbb{B}}\}$ as local assignments for the probability distribution $\{e_C\}$, then contextuality entails the non-existence of a global probability distribution $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}$ compatible with the $\{e_C^{\mathbb{B}}\}$ [Constantin, 2015]. An example for this form of contextuality, known as logical (or possibilistic) contextuality, is the Hardy model [Hardy, 1993], which demonstrates an inequality-free proof of Bell's theorem. To use sheaf theory for this model, the empirical model is defined over the Boolean ring. This choice can be compatible with Hardy's conventions, which consider "the possibility" for "positive probability" and "the impossibility" for "probability 0" of measurement outcomes. Abramsky and Brandenburger [2011] demonstrate that Hardy's model satisfies a stronger non-locality feature than Bell's experiment. This means that Hardy's model is both probabilistically and logically contextual. To compare the sheaf-theoretical approach with the Spekkens' operational approach in Section 4.3, we need to precisely translate all levels of contextuality in the sheaf-theoretical approach (including the logical level) into the notation of Spekkens [2005]. This explicit comparison is left for future work.

The other level of contextuality in this sheaf approach is strong contextuality, which represents a situation in which no global assignments are consistent with the support at all. Abramsky and Brandenburger [2011] explain the KS theorem as generic (model-independent) proofs of strong contextuality. In the next section, I will briefly describe this approach.

4.4.2 Kochen-Specker theorem in the sheaf-theoretical approach

The notation of the measurement cover $\mathcal{M}$ fits very well with the proof of KS theorem. In Section 4.3, the 18 vectors proof of the KS theorem was represented using Spekkens' model. Here, I represent the proof using the sheaf-theoretical model. Abramsky and Brandenburger [2011] demonstrate the measurement cover $\mathcal{M}$ using the notation listed in Table 4.4, whose columns list the elements of $\mathcal{M}$. In other words, the 18 nodes (or vectors) in Figure 4.1 correspond to the elements of the set $X = \{m_1, ..., m_{18}\}$ and each loop (or measurement context) in that figure corresponds to a column of this table.

Table 4.4: A table of measurements representing an example of the KS theorem with 18 vectors [Abramsky and Brandenburger, 2011]. Each column of the following table is an element of measurement cover $\mathcal{M}$.

m_1	m_1	m_8	m_8	m_2	m_9	m_{16}	m_{16}	m_{17}
m_2	m_5	m_9	m_{11}	m_5	m_{11}	m_{17}	m_{18}	m_{18}
m_3	m_6	m_3	m_7	m_{13}	m_{14}	m_4	m_6	m_{13}
m_4	m_7	m_{10}	m_{12}	m_{14}	m_{15}	m_{10}	m_{12}	m_{15}

As described in Section 4.3.3, the condition of measurement noncontextuality is satisfied for this situation if a state of system λ assigns the same probability to operationally equivalent measurement events. This is possible by assigning probabilities to measurement events in such a way that they add up to 1 for each column. But to demonstrate the KS contradiction we need to consider the additional requirement of outcome determinism. Here, in the sheaf-theoretical approach, Abramsky and Brandenburger [2011] consider the set of measurements X as a set of Boolean variables, considering each measurement m dichotomic (a measurement with only two outcomes). Using this property they make a connection to logic and associate the two possible outcomes to $True$ and $False$. So for a section $s : C \to O$ of a finite context C, they define a propositional variable as

$$\varphi_s = \bigwedge_{m \in C, s(m)=True} m \wedge \bigwedge_{m \in C, s(m)=False} \neg m. \tag{4.25}$$

Considering the set of dichotomic outcome as $\{0, 1\}$, a section $s : C \to \{0, 1\}$ is a Boolean assignment for these propositional variables. Abramsky and Brandenburger [2011] define φ_C as the set of satisfying assignments

$$\varphi_C := \bigvee_{s \in s(C)} \varphi_s. \tag{4.26}$$

Assigning 1 to φ_C, requires an assignment of 1 to exactly one propositional variable φ_s. This is similar to the situation in Equation (4.16) in which only one of the values $v(\mathbb{P}_i)$ ($i \in \{1, 2, 3, 4\}$) could be equal to 1. But we cannot represent the condition of measurement noncontextuality discussed in Section 4.3.3 for Equation (4.26), since this equation includes the assumption of outcome determinism. Therefore, if we associate each propositional variable φ_s to a node in Figure 4.1, we do not have the freedom of assigning probabilities to nodes in a way that they sum up to 1 for each loop. So we cannot check if λ (the ontic state of the system) assigns the same probability to operationally equivalent measurements or not.

However, the KS contradiction which arises using Equation (4.26) is similar to the KS contradiction described in Section 4.3.3 which holds the assumption of outcome determinism. Here, an empirical model is logically contextual if we cannot satisfy assignments for the formula

$$\varphi_e := \bigwedge_{C \in \mathcal{M}} \varphi_C. \tag{4.27}$$

This is similar to the failure of the measurement non contextual assignment of values $\{0, 1\}$ to the measurement events in Figure 4.1 by ontic state λ.

In this section (Section 4.4), I have only given a general perspective about the sheaf-theoretical approach of contextuality, mainly to highlight its similarities with the operational approach of contextuality described in Section 4.3.3 and the standard notation of Bell's inequality described in Section 4.2.1. Future work could provide a deeper analysis of these similarities at a more rigorous mathematical level e.g. by using representation theories. For example, in a recent work, Wester [2017] generates a categorical isomorphism between some aspects of the operational and sheaf-theoretical approaches. These representation theories might even be used to investigate the association between these general approaches of contextuality and the CBD approach. However, in this chapter, I make all comparisons at a symbolic level and I leave the representation theories for future research. In the next section, I will describe the CBD approach which is constructed using completely different assumptions about random variables and probability space.

4.5 Contextuality-by-default

The last contextuality model that I cover in this chapter is contextuality-by-default (CBD) [Dzhafarov et al., 2015a, Dzhafarov and Kujala, 2014b, 2015, 2016], the model that exploits *contextual random variables* to interpret contextuality. This model is based on the concepts of marginal selectivity and double indexing that I described in Section 2.2.1. Here, I look at these concepts in more detail, by comparing the CBD apporach to the standard forms of contextuality as they are understood in QM and discussed earlier in this chapter.

In this model, a system of random variables comprises *stochastically unrelated* "bunches", each of which is a set of jointly distributed random variables which have the same context [Dzhafarov and Kujala, 2016] [15]. The term "stochastically unrelated" is used in the CBD model to indicate that there is no joint distribution for the random variables, when each random variable belongs to a different bunch. For example, a system with four random variables $\{a_{q_1}^{c_1}, a_{a_2}^{c_1}, a_{q_1}^{c_2}, a_{q_2}^{c_2}\}$ consists of two bunches $a^{c_1} = \{a_{q_1}^{c_1}, a_{q_2}^{c_1}\}$ and $a^{c_2} = \{a_{q_1}^{c_2}, a_{q_2}^{c_2}\}$. In this notation, c_1 and c_2 indicate two different contexts, and q_1 and q_2 represent two different observables[16], the entities that the random variables measure or respond to [Dzhafarov and Kujala, 2016, Dzhafarov et al., 2016]. For example, Dzhafarov and Kujala [2016] consider the two observables q_1 and q_2 as two questions q_1= "Do you like bees?" and q_2= "Do you like to smell flowers?". Dzhafarov and

[15] Each "bunch" in the CBD apporach is like a "measurement context" in the sheaf-theoretical approach (introduced in Section 4.4) is constructed based on a possible set of compatible measurements. In spite of this apparent similarity between "bunches" and a maximal sets of compatible measurements or "a measurement cover", there is a fundamental difference between these two terms according to the different way that CBD defines random variables.

[16] As mentioned in Chapter 2, Dzhafarov and Kujala [2015] denote the observables $\mathbb{A}, \mathbb{B}$ as *inputs* and represent them as α, β, and call random variables a and b *outputs* represented by A and B. In their later paper [Dzhafarov and Kujala, 2016], the observables are called *contents* and labeled q.

Kujala [2016] conduct an experiment by posing these questions to people randomly selected from a population. They consider two controlling conditions for asking these questions. In the first condition (c_1), the randomly selected person is asked these questions following watching a movie about killer bees spreading northwards. And in the second condition (c_2), they are asked after watching a movie about deciphering the waggle dances of the honey bees. The binary answers (Yes or No) to these questions form the four described random variables $\{a_{q_1}^{c_1}, a_{q_2}^{c_1}, a_{q_1}^{c_2}, a_{q_2}^{c_2}\}$.

To clarify this definition of random variables further, we can consider the Bell experiment as a special example. Dzhafarov and Kujala [2014b, 2015] and Dzhafarov et al. [2015a] define random variable $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ in terms of how it exposes the measurement of $\mathbb{A}_1$ in the context of $(\mathbb{A}_1, \mathbb{B}_1)$, which is different from random variable $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}$ in the context of $(\mathbb{A}_1, \mathbb{B}_2)$. This method leads to eight different random variables $\{a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}, a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}, a_{\mathbb{A}_2}^{(\mathbb{A}_2,\mathbb{B}_1)}, a_{\mathbb{A}_2}^{(\mathbb{A}_2,\mathbb{B}_2)}, a_{\mathbb{B}_1}^{(\mathbb{B}_1,\mathbb{A}_1)}, a_{\mathbb{B}_1}^{(\mathbb{B}_1,\mathbb{A}_2)}, a_{\mathbb{B}_2}^{(\mathbb{B}_2,\mathbb{A}_1)}, a_{\mathbb{B}_2}^{(\mathbb{B}_2,\mathbb{A}_2)}\}$ for Bell's experiment. Returning to previous works in the foundations of quantum physics, we can start to see how the assumptions of the CBD approach subtly differ from those in the physics community. As I mentioned in Section 4.2.1, Shimony [1984] defines a probability distribution~~s~~ $p_\lambda(a|\mathbb{A})$ on the set of ontic states $\wedge$ of the system. In that definition, we could not construct a joint probability for non-commuting observables $\mathbb{A}_1$ and $\mathbb{A}_2$. This is similar to what CBD defines as *stochastically unrelated* for two random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_2}^{(\mathbb{A}_2,\mathbb{B}_1)}$. But in the CBD method, two random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}$ are defined as *stochastically unrelated* as well, a situation for which there is no counterpart in Shimony's standard definition.

To make a better comparison to previous standard approaches, we need to look at the probability space of the CBD method. As de Barros et al. [2016] state, to define the double random variables [17], we need a separate probability space for each possible context. Thus, we have a random variable $a_i^j : \Omega_j \to E_i$ [18] where subscripts $i = 1, ..., M$ indicate different observables [19] and $j = 1, ..., N$ indicate different contexts. Here, E is a certain set of possible values, such as $\{-1, 1\}$ in Bell's scenario. And Ω is a probability space, but this time not as the set of ontic states represented by $\wedge$ in Section 4.2.1. As an example, for the observable $\mathbb{A}_1$ in Bell's experiment, Ω can be related to one of the two possible contexts $\{\mathbb{A}_1, \mathbb{B}_1\}$ and $\{\mathbb{A}_1, \mathbb{B}_2\}$. Using this notation, de Barros et al. [2016] declare that random variables a_i^j of different observables i in the same context j are jointly distributed (e.g. two random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{B}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ in Bell's experiment), but random variables a_i^j and $a_{i'}^{j'}$ where $j \neq j'$ and $i, i' \in \{1, ..., n\}$ (equal or not) are stochastically unrelated (e.g. random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}$ or random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_2}^{(\mathbb{A}_2,\mathbb{B}_1)}$ in Bell's experiment).

This consideration of different probability spaces or different random variables for only one observable in different contexts is not allowed within standard quantum models demonstrated earlier in this chapter. For example, in the definition of measurement contextuality suggested by Spekkens (4.10), the measurement procedures which admit contextuality on the ontological level are operationally context-independent. This was explained very well by Simmons et al. [2016]:

> *Discussions of contextuality often focus on scenarios in which an element of an operational theory such as quantum mechanics manifests itself in two different contexts, such as two different decompositions of a density matrix;*

[17]The double random variables stand for random variables with double indexing; one of these two indexes represents the associated observable and the other represents the associated context.

[18]For simplicity, from now on I will use a_i^j instead of $a_{q_i}^{c_j}$.

[19]de Barros et al. [2016] use the term "properties" instead of "observables".

> *or an observable being measured in two different ways, alongside different sets of co-measurable observables. These manifestations are treated identically by the operational theory, always leading to the same probabilities. In fact, this is why the same notation is used for the objects in the first place, as a context-independent symbol is all that is needed to calculate probabilities. However there is no formal argument to be made that these elements which are operationally context-independent should also be ontologically context-independent: this must be taken axiomatically. (p.2)*

The double indexing notation associates e.g. two random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}$ with the observable $\mathbb{A}_1$, where each different random variable is defined based on a different probability space. Substituting these two random variables instead of the two outcomes in the operational equivalence equation (4.8), we obtain: $p(a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}|\{(\mathbb{A}_1,\mathbb{B}_1)\},P) = p(a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}|\{(\mathbb{A}_1,\mathbb{B}_2)\},P)$. This is not completely match with the original operational equivalence. This can explain why Mazurek et al. [2016] call this new equation as: "merely close to operationally equivalent".

In the rest of this section, I will further investigate the consequences of considering these special assumptions of the CBD approach to compare them with the consequences of the standard approaches of contextuality discussed earlier.

4.5.1 Parameter independence and Non-signaling conditions in the CBD approach

In Section 4.2.1, I described PI (4.3) as one of the two weaker locality conditions that works as a barrier for "superluminal signals". Here, I investigate a possible representation of this condition using the CBD notation. First, consider two *stochastically unrelated* random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}$. On a superficial view, we may assume that the PI condition (for each ontic state λ) is satisfied if $\xi_\lambda(a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}|\mathbb{A}_1,\mathbb{B}_1) = \xi_\lambda(a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}|\mathbb{A}_1,\mathbb{B}_2)$. This is similar to $\xi_\lambda(a|\mathbb{A},\mathbb{B}) = \xi_\lambda(a|\mathbb{A})$ in Equation (4.12). But we cannot check the validity of this representation since the CBD approach does not have a clear position about the ontic state. As a result, it is not possible to estimate whether using this notation satisfies the PI condition (similarly to the standard interpretation of QM), or violates it (similarly to the Bohmian interpretation). In general, without the exact meaning of the ontic state, we cannot compare the philosophical position of the CBD approach regarding the nature of reality with the philosophical views behind the forming of the existing quantum interpretations.

The notation of hidden variables was only used by the CBD approach to describe the existence of joint distribution but in an ambiguous way. As Dzhafarov and Kujala [2014a] explained:

> *Being jointly distributed is equivalent to the random outputs being (measurable) functions of one and the same (hidden) random variable* λ *.*

But in Spekkens' notation, a joint probability measurement can be reproduced by averaging response functions $\xi_\lambda(k|M)$ for different (not same) ontic state λ. Each $\xi_\lambda(k|M)$ can be considered as a function of λ (See Section 4.3). For a detailed explanation, we need the definition for existence of a joint distribution [Liang et al., 2011]:

Definition 11 *Consider a set of measurements* $\{M_1, M_2, ..., M_n\}$ *and a set of random variables* $\{k_1, k_2, ..., k_n\}$ *whose values are defined corresponding to those measurements. let* $\{M_s|s \in S\} \subset \{M_1, M_2, ..., M_n\}$*, where* $S \subset \{1, 2, ..., n\}$*. There exist a joint distribution for* $\{M_1, M_2, ..., M_n\}$ *such that the statistics for every joint measurement of a* *subset $\{M_s|s \in S\}$ can be obtained as a marginal of the distribution $p(k_1, k_2, ..., k_n|P)$:*

$$\forall i, \forall P : p(k_s|M_s, P) = \sum_{k_i; i \notin S} p(k_1, k_2, ..., k_n|P). \tag{4.28}$$

This joint probability distribution for the set of measurements $\{M_1, M_2, ..., M_n\}$ could lead to the physical possibility of a joint measurement of them. The joint measurement which can be reproduced by averaging response functions with different probability distributions $\mu_P(\lambda)$ on the set Λ:

$$\begin{aligned} p(k_1, k_2, ..., k_n|M_1, M_2, ..., M_n, P) = \\ \sum_{\lambda \in \Lambda} \xi_\lambda(k_1|M_1)...\xi_\lambda(k_n|M_n)\mu_P(\lambda). \end{aligned} \tag{4.29}$$

This is a more general form of the factorization condition in equation (4.14). Comparison between Equations 4.28 and 4.29 with the CBD definition of joint distribution emphasizes that CBD notation has an unclear position about ontic state λ.

As we reviewed in Section 4.2.1, the non-signaling condition is expressed independent of ontic state λ. This provides a chance to precisely explore the meaning of this condition in the CBD approach. In Table 4.5, we represent the joint probability distributions of Bell's experiment using the CBD notation. We can compare this table with the probability distributions in Table 4.1. This comparison shows that the double indexing scenario does not change the original meaning of non-signaling. Similar to the case of $a_1 = +1$ in Table 4.1, the probability of $a_{\mathbb{A}_1} = +1$ is independent of the setting for measurement in the other side of the experiment if $p_1 + p_2 = p_3 + p_4$. Looking more carefully, we can recognize that the two random variables $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)}$ and $a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)}$ have the same value and the same distribution. This reduces the double indexing random notation to the standard non-contextual representation of random variables (a_i). de Barros et al. [2016] and Dzhafarov et al. [2016] use the term "consistently connected" for the general form of non-signaling condition:

Definition 12 *A system consisting of random variables a_i^j is consistently connected if $a_i^j \sim a_i^{j'}$ for every observables $i \in \{1, ..., m\}$ that belong to different contexts $j, j' \in \{1, ..., n\}$, this notation means a_i has the same distribution in both contexts j and j'. Alternatively this relation is denoted by $Pr[a_i^j = a_i^{j'}] = 1$.*

Dzhafarov et al. [2016] deem that the CBD approach can represent contextuality even if random variables are not consistently connected (signaling exists). This deceptive property can be considered as a proper choice to test contextuality in psychological experiments in which it is not always possible to satisfy the non-signaling condition. However, as mentioned in Section 2.2, Dzhafarov et al. [2015a] apply such an approach in some behavioral and social experiments, including those in which random variables are not consistently connected, and none of them show any evidence of contextuality. In general, any relaxation on the fundamental property of non-signaling in Bell experiment needs a careful justification. For example, Brask and Chaves [2017] suggest novel casual interpretations of CHSH violation allowing communication between two sides of experiment. Their casual structures can simulate quantum and non-signaling correlations. However, it is not possible to compare such approach with CBD, again due to the lack of an exact definition of hidden variables in the CBD approach.

Table 4.5: The joint probability distributions for Bell's experiment using the double indexing scenario

$\mathbb{A}_1, \mathbb{B}_1$	$a_{\mathbb{B}_1}^{(\mathbb{A}_1,\mathbb{B}_1)} = +1$	$a_{\mathbb{B}_1}^{(\mathbb{A}_1,\mathbb{B}_1)} = -1$	$\mathbb{A}_1, \mathbb{B}_2$	$a_{\mathbb{B}_2}^{(\mathbb{A}_1,\mathbb{B}_2)} = +1$	$a_{\mathbb{B}_2}^{(\mathbb{A}_1,\mathbb{B}_2)} = -1$
$a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)} = +1$	p_1	p_2	$a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)} = +1$	p_3	p_4
$a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_1)} = -1$	p_5	p_6	$a_{\mathbb{A}_1}^{(\mathbb{A}_1,\mathbb{B}_2)} = -1$	p_7	p_8
$\mathbb{A}_2, \mathbb{B}_1$	$a_{\mathbb{B}_1}^{(\mathbb{A}_2,\mathbb{B}_1)} = +1$	$a_{\mathbb{B}_1}^{(\mathbb{A}_2,\mathbb{B}_1)} = -1$	$\mathbb{A}_2, \mathbb{B}_2$	$a_{\mathbb{B}_2}^{(\mathbb{A}_2,\mathbb{B}_2)} = +1$	$a_{\mathbb{B}_1}^{(\mathbb{A}_2,\mathbb{B}_2)} = -1$
$a_{\mathbb{A}_2}^{(\mathbb{A}_2,\mathbb{B}_1)} = +1$	p_9	p_{10}	$a_{\mathbb{B}_2}^{(\mathbb{A}_2,\mathbb{B}_2)} = +1$	p_{11}	p_{12}
$a_{\mathbb{A}_2}^{(\mathbb{A}_2,\mathbb{B}_1)} = -1$	p_{13}	p_{14}	$a_{\mathbb{B}_2}^{(\mathbb{A}_2,\mathbb{B}_2)} = -1$	p_{15}	p_{16}

4.5.2 Meaning of contextuality in the CBD apporach

The CBD approach associates the (non)contextuality character of the described system to the possibility or impossibility of imposing a joint distribution on the stochastically unrelated bunches (a joint distribution with a certain property of "maximality"). This imposed joint distribution which is named as "coupling" [Dzhafarov et al., 2016], leads to a new set of random variables. The coupling condition finds a common space for this set of random variables:

Definition 13 *A coupling of a set of random variables* $a_1, ..., a_n$ *is any jointly set of random variables* $b_1, ..., b_n$ *such that* $a_1 \sim b_1, ..., a_n \sim b_n$.

The maximality for coupling of a system of random variables can be defined as [Dzhafarov and Kujala, 2016]:

Definition 14 *let* $a_i^1, ..., a_i^j$ *be a "connection" of a system of random variables (connection is set of random variables with the same observable* i*), an associated coupling* $b_i^1, ..., b_i^j$ *is a maximal coupling if* $Pr(b_i^1 = ... = b_i^j)$ *has the the largest value between all possible couplings. If all the couplings related to the connections of that system are maximal couplings, then the main coupling of the system is maximally connected.*

According to Thereon 3.3 in [Dzhafarov and Kujala, 2016, p. 12], a maximal coupling $b_i^1, ..., b_i^j$ can be constructed for the connection $a_i^1, ..., a_i^j$ if

$$Pr(b_i^1 = ... = b_i^j = v) =$$
$$min(Pr[b_i^1 = v], ..., Pr[b_i^j = v]),$$

where $v \in V$, V is a set of all possible values for random variables in $a_i^1, ..., a_i^j$.

A simple example can clarify these definitions further. Consider a system with three bunches $\{K_1^1, K_2^1\}$, $\{K_2^2, K_3^2\}$ and $\{K_1^3, K_3^3\}$. We can categorize this system of random variables to three connections $\{K_1^1, K_1^3\}$, $\{K_2^2, K_2^1\}$ and $\{K_3^2, K_3^3\}$. We assume that the random variables of this system are bivalent with two possible values of 0 and 1. If the distribution of the first connection is as follow,

$K_1^1 = 0$	$K_1^1 = 1$
0.8	0.2

$K_1^3 = 0$	$K_1^3 = 1$
0.5	0.5

the maximal coupling $\{T_1^1, T_1^3\}$ of this connection can be obtained as

$$Pr[T_1^1, T_1^3 = 0] = min(0.8, 0.5) = 0.5,$$
$$Pr[T_1^1, T_1^3 = 1] = min(0.2, 0.5) = 0.2.$$

The maximal couplings of the other connections can be constructed in the similar way. Considering $S = (S_1^1, S_1^3, S_2^2, S_2^1, S_3^2, S_3^3)$ as the coupling of the system, to investigate contextuality we can compare the subcouplings corresponding to the connections (e.g. $\{S_1^1, S_1^3\}$) with the related maximal couplings (e.g.

$\{T_1^1, T_1^3\}$). I provide more detailed examples in Section 4.5.3.

In the CBD approach, although bunches can be related to empirical meanings in physics (or psychology), the coupling has no empirical meaning. Dzhafarov and Kujala [2016, p. 11] declare that the coupling is merely a mathematical process: "If the bunches are assumed to have links to empirical observations, then the couplings can be said to have no empirical meaning. A coupling forms a base set of its own, consisting of itself". This makes it impossible to compare the exact meaning of contextuality in the CBD approach with the other approaches explained in this chapter. In contrast to the CBD approach, in the sheaf-theoretical approach described in Section 4.4, both local and global (in)consistencies have empirical meanings. The operational approach of Spekkens described in Section 4.3 is also empirically recognizable. In next section, I provide mathematical comparisons between the CBD approach and Spekkens' approach for some cyclic examples.

4.5.3 Cyclic examples

Dzhafarov and Kujala [2016], Kujala et al. [2015] singled out a category of the CBD model for binary random variables denoted as a cyclic class. In this class, each context (or bunch) includes exactly two observables, and each observable is measured in exactly two contexts. The number of observables and the number of contexts are equal to each other and called the rank (n) of the system. In the CBD model, cyclic system of rank 2 forms the simplest (non)contextual inequality which associates the order effect of projective measurements to contextuality.

It is a common belief that we need at least three measurements to derive the simplest scenario of contextuality in quantum mechanics [Kunjwal and Spekkens, 2016, Liang et al., 2011]. This scenario is designed based on the Specker's example of contextuality [Ernst Specker, 2011]. This example requires three bivalent measurements $\{M_1, M_2, M_3\}$ which can be measured jointly in pair but not in triple. The joint measurement of pairs $\{M_1, M_2\}$, $\{M_2, M_3\}$ and $\{M_1, M_3\}$ are denoted by M_{12}, M_{23} and M_{13} respectively. They have statistics that reproduce the measurement statistics of $\{M_1, M_2, M_3\}$ as marginals, for example $\forall P \in \mathcal{P}$:

$$\begin{aligned} p(k_1|M_1, P) &\equiv \sum_{k_2} p(k_1, k_2|M_{12}, P), \\ p(k_1|M_1, P) &\equiv \sum_{k_3} p(k_1, k_3|M_{13}, P). \end{aligned} \tag{4.30}$$

These two coarse-grainings over K_2 (of M_{12}) and K_3 (of M_{13}) imply an equivalence class of measurement for M_1 [Liang et al., 2011]. In other words, marginalizing the statistics of the joint measurements in two contexts $\{M_1, M_2\}$ and $\{M_3, M_1\}$ should lead to the same statistic of M_1. Such kind of operational equivalences for all measurements $\{M_1, M_2, M_3\}$ are considered as "no-disturbance" condition. For spatially separated systems this condition reduces to the no-signaling condition discused in definition (4.15) [Kurzyński et al., 2014, Ramanathan et al., 2012].

Specker's experiment also entails the existence of at least one preparation (e.g. $P*$) which grants negative correlations for all joint measurements M_{ij} ($(ij) \in \{(12), (23), (31)\}$) [Liang et al., 2011]:

$$\begin{aligned} \forall i \neq j : p(K_i = 0, k_j = 1|M_{ij}, P*) &= \frac{1}{2}, \\ p(K_i = 1, k_j = 0|M_{ij}, P*) &= \frac{1}{2}. \end{aligned} \tag{4.31}$$

Specker's contextuality states that there is no joint distribution $p(k_1, k_2, k_3)$, for which all joint measurements M_{ij} are satisfying these anti-correlation conditions [Liang et al., 2011]. A simple proof is represented in Figure 4.2.

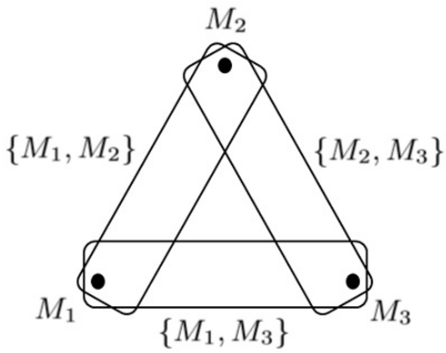

Figure 4.2: Three measurements and three contexts of measurements in Specker's scenario. This picture shows why there is no joint distribution $p(k_1, k_2, k_3)$, for which all joint measurements M_{ij} are satisfying the anti-correlation conditions. Choosing between eight possible valuations of (0, 0, 0), (0, 0, 1), ..., (1, 1, 1) for the triple (k_1, k_2, k_3), at least one of the pairs (k_1, k_2), (k_1, k_3) and (k_2, k_3) should take valuations $(0, 0)$ or $(1, 1)$, which contradicts the anti-correlation conditions [Liang et al., 2011].

The constraint on the triplewise joint measurement cannot be attained using projective measurements in quantum mechanics. Because the pairwise joint measurability of three bivalent projective measurements implies pairwise commutativity of the corresponding observables. And this pairwise commutativity leads to commutativity of all three observables which brings about the joint implementation of those three measurements [Liang et al., 2011]. However this constraint can be achieved using three bivalent non-orthogonal measurements (POVMs), for which joint measurability does not imply commutativity [Kunjwal and Ghosh, 2014, Kunjwal and Spekkens, 2016]. Such approach leads to a generalization of Specker's scenario to theory-independent criteria of contextuality for n-cycle scenarios [Kunjwal and Spekkens, 2016].

The prominent role of random variables in the definition of the CBD approach, does not left any room to identify the role of ideal or non-ideal measurements in that definition. Here, I try to express the classical version of Specker's scenario using this notation. In this notation, as described in Section 4.5, any two random variables from two different contexts are stochastically unrelated. This includes random variables such as $\{k_{M_1}^{(M_1,M_2)}, k_{M_2}^{(M_2,M_3)}\}$ or even $\{k_{M_1}^{(M_1,M_2)}, k_{M_1}^{(M_1,M_3)}\}$. Using this notation, the Specker's scenario should be consistently connected (See Definition 12). This condition entails the associated random variables of a measurement such as M_1 to have the same distributions in two contexts $\{M_1, M_2\}$ and $\{M_1, M_3\}$, which is denoted by $Pr[k_{M_1}^{(M_1,M_2)} = k_{M_1}^{(M_1,M_3)}] = 1$. This can be also represented as $p_1 + p_2 = p_5 + p_6$ (or $p_2 = p_6$) considering three bunches in Table 4.6, which is similar to the marginalization relation for M_1 in equation (4.30):

$$\begin{aligned} p(k_1 = 0, k_2 = 0|M_{12}, P)+ \\ p(k_1 = 0, k_2 = 1|M_{12}, P) = \\ p(k_1 = 0, k_3 = 0|M_{13}, P)+ \\ p(k_1 = 0, k_3 = 1|M_{13}, P). \end{aligned} \tag{4.32}$$

The system is consistently connected if there are similar relations for other random variables as: $Pr[k_{M_2}^{(M_1,M_2)} = k_{M_2}^{(M_2,M_3)}] = 1$ and $Pr[k_{M_3}^{(M_1,M_3)} = k_{M_3}^{(M_2,M_3)}] = 1$. So we can conclude that the "consistently connection" is similar to the "no-disturbance" condition which requires the equivalence relation for all three measurements $\{M_1, M_2, M_3\}$. Moreover, Speckers' scenario requires to satisfy the anti-correlation condition [Dzhafarov et al., 2015b]

$$\begin{aligned} Pr[k_{M_1}^{(M_1,M_2)} &= -k_{M_2}^{(M_1,M_2)}] = 1, \\ Pr[k_{M_2}^{(M_2,M_3)} &= -k_{M_3}^{(M_2,M_3)}] = 1, \\ Pr[k_{M_3}^{(M_1,M_3)} &= -k_{M_1}^{(M_1,M_3)}] = 1, \end{aligned} \tag{4.33}$$

similar to equations (4.31).

Bunch 1	$k_{M_2}^{(M_1,M_2)} = 0$	$k_{M_2}^{(M_1,M_2)} = 1$
$k_{M_1}^{(M_1,M_2)} = 0$	$p_1 = 0$	$p_2 = 0.5$
$k_{M_1}^{(M_1,M_2)} = 1$	$p_3 = 0.5$	$p_4 = 0$

Bunch 2	$k_{M_3}^{(M_2,M_3)} = 0$	$k_{M_3}^{(M_2,M_3)} = 1$
$k_{M_2}^{(M_2,M_3)} = 0$	$p_9 = 0$	$p_{10} = 0.5$
$k_{M_2}^{(M_2,M_3)} = 1$	$p_{11} = 0.5$	$p_{12} = 0$

Bunch 3	$k_{M_3}^{(M_1,M_3)} = 0$	$k_{M_3}^{(M_1,M_3)} = 1$
$k_{M_1}^{(M_1,M_3)} = 0$	$p_5 = 0$	$p_6 = 0.5$
$k_{M_1}^{(M_1,M_3)} = 1$	$p_7 = 0.5$	$p_8 = 0$

Table 4.6: Three bunches in the CBD representation of Specker's scenario. Because of the anti-correlation condition, the probabilities on the diagonal lines are equal to zero and the probabilities on the counter diagonal lines are equal to 0.5 (See Equation (4.31)).

Table 4.7 suggests a possible representation of Specker's scenario using the CBD notation. To demonstrate the contextuality we can assume $a_i^{j'} = a_i^j$ for any measurement i in two different contexts j and j'. With this assumption, two random variables $k_{M_1}^{(M_1,M_2)}$ and $k_{M_1}^{(M_1,M_3)}$ take same value (e.g. 1). I represent these valuations with horizontal hatchings in their corresponding cells. Because of the anti-correlation condition in each pairwise joint measurement, $k_{M_2}^{(M_1,M_2)}$ should be 0. We can conclude that $k_{M_2}^{(M_2,M_3)} = 0$ as well, since it belongs to the same measurement M_2. These are represented by vertical hatching. Continuing this argument, we will reach a contradiction for the value of $k_{M_3}^{(M_2,M_3)}$ which is represented by the grid hatching.

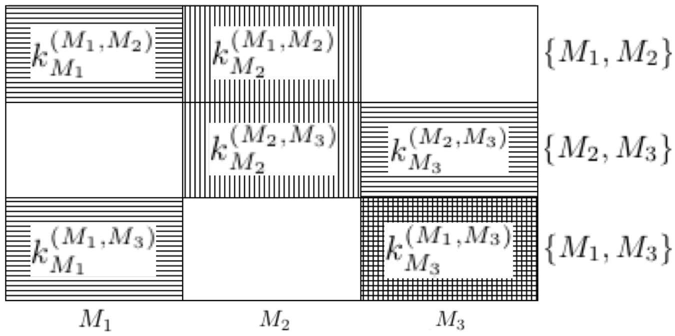

Table 4.7: A representation of Specker's scenario. The horizontal axis represents three measurements and the vertical axis indicates three contexts. There are six random variables which can take values $\{0, 1\}$. To convert this table to the representation in Figure 4.2, one should amuse $a_i^{j'} = a_i^j$. Based on this assumption, different values of random variables are represented by horizontal and vertical hatchings. The raised contradiction is a proof of contextuality which is represented by the grid hatching.

However, this argument seems not to be completely matched with the CBD approach since the equality $a_i^{j'} = a_i^j$ violates the double indexing assumption. In this notation, two random variables like $k_{M_1}^{(M_1,M_2)}$ and $k_{M_1}^{(M_1,M_3)}$ *can* take different values (and two different distributions). Instead of this argument, Dzhafarov et al. [2015b] use the concept of coupling (See Definition 14) to investigate the existence of contextuality in Specker scenario. They claimed the system is contextual since there is no maximally connected coupling for that system. I will show that this approach is completely equal to the above argument since it also requires the equality $a_i^{j'} = a_i^j$, the equality which is concluded from the consistently connected condition.

The maximal couplings of three possible connections are constructed in Table 4.8. There is a restriction to construct a maximally connected coupling based on the three maximal couplings. As illustrated in Figure 4.3, we cannot have a coupling in which all probabilities are achieved together and be compatible with the probabilities in our bunches and connections. In this picture, if we associate 0 to the random variable S_1^1, then S_2^1 should be 1 because of the anti-correlation condition. Moving clockwise we reach the connection 2, in which two random variables S_2^1 and S_2^2 should take a same value and a same distribution because of the consistently connected condition. By moving further clockwise, we will reach a contradiction for the value of S_1^1 in the connection 1.

This proof consider an equality between two random variables of each connection: $a_i^{j'} = a_i^j$. This is similar to what I described earlier for the case of nonsignaling, where the double indexing notation was reduced to the standard non-contextual representation of random variables. Here, if we ignore the double indexing scenario, we can remove the three imaginary connections in Figure 4.3 and convert it to the standard representation of Specker scenario in Figure 4.2.

Coupling 1	$T_{M_1}^{(M_1,M_3)} = 0$	$T_{M_1}^{(M_1,M_3)} = 1$
$T_{M_1}^{(M_1,M_2)} = 0$	0.5	0
$T_{M_1}^{(M_1,M_2)} = 1$	0	0.5

Coupling 2	$T_{M_2}^{(M_2,M_3)} = 0$	$T_{M_2}^{(M_2,M_3)} = 1$
$T_{M_2}^{(M_1,M_2)} = 0$	0.5	0
$T_{M_2}^{(M_1,M_2)} = 1$	0	0.5

Coupling 3	$T_{M_3}^{(M_1,M_3)} = 0$	$T_{M_3}^{(M_1,M_3)} = 1$
$T_{M_3}^{(M_2,M_3)} = 0$	0.5	0
$T_{M_3}^{(M_2,M_3)} = 1$	0	0.5

Table 4.8: Three maximal couplings of the three connections in the CBD representation of Specker's scenario.

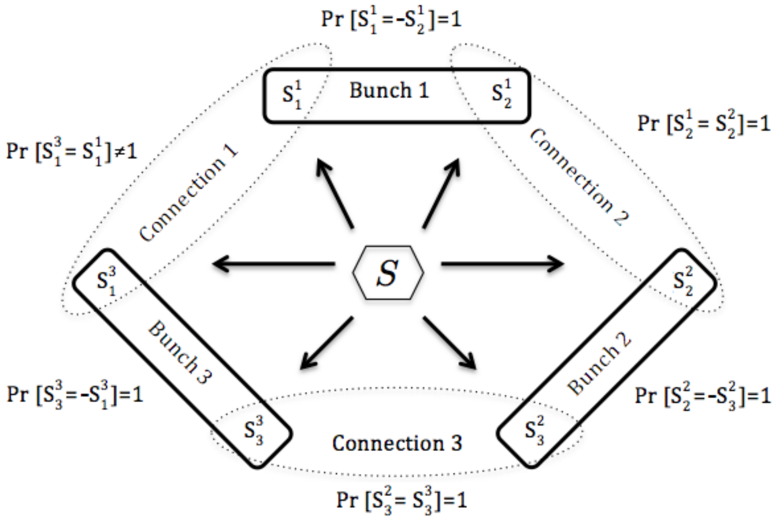

Figure 4.3: S is a possible maximally connected coupling for the Specker scenario based on three maximal couplings in Table 4.8. This shows that all subcouplings $\{\{S_1^1, S_1^3\}, \{S_2^2, S_2^1\}, \{S_3^2, S_3^3\}\}$ related to the connections of this system can be maximal couplings, so the system is noncontextual.

The other cyclic system of rank 3 in the CBD approach is associated with Leggett-Garg (LG) inequality [Leggett and Garg, 1985]. This has a similar structure to Specker's scenario but without the anti correlation condition. LG inequality is generally considered as a temporal version of Bell inequality and is used to test quantum coherence in macroscopic level [Emary et al., 2014]. The affirmative response to this test suggests the existence of a unique joint probability distribution to derive the inequality. To review a simple representation of this inequality, we can consider a macroscopic bivalent variable $k = \pm 1$ in a given system S. Suppose that there are measurements M_1, M_2 and M_3 on S revealing the values of k at three different times t_1,t_2 and t_3 respectively (the values are denoted by k_1, k_2 and k_3). By performing the measurements on many copies of the system S, one can build the correlation function C_{ij} from the joint probability $p(k_i, k_j|M_{ij})$ as

$$C_{ij} = \sum_{k_i,k_j=\pm 1} k_i k_j p(k_i, k_j|M_{ij}). \tag{4.34}$$

The LG inequality can be derived considering all possible correlations for three times t_1,t_2 and t_3:

$$LG \equiv C_{21} + C_{32} - C_{31} \leq 1. \tag{4.35}$$

The violation of this inequality provides a border for quantumness in the sense of the lack of a realistic account of the macroscopic system or the impossibility of a non-disturbant measurement on the system [Emary et al., 2014]. However, CBD interprets such violation in terms of contextuality [Dzhafarov and Kujala, 2014a, 2016, Dzhafarov et al., 2015b, Kujala et al., 2015]. In this thesis, I do not analyze the effect of this notation on the realism or noninvasive measurability assumptions. I will do such analysis and will investigate the possible relation between this notation and the operational approach in my future work. Here, I only provide an example of the LG system to compare the construction of its coupling with the constructed coupling for Specker's scenario in Figure 4.3. Dzhafarov and Kujala [2016] suggest a contextual example of the LG system with the bunches represented in Table 4.9. In this table, unlike Specker's scenario, the probabilities on the diagonal lines are not necessary to be zero, since we do not have the anti-correlation condition.

Bunch 1	$k_{M_2}^{(M_1,M_2)} = 0$	$k_{M_2}^{(M_1,M_2)} = 1$
$k_{M_1}^{(M_1,M_2)} = 0$	0.7	0
$k_{M_1}^{(M_1,M_2)} = 1$	0	0.3

Bunch 2	$k_{M_3}^{(M_2,M_3)} = 0$	$k_{M_3}^{(M_2,M_3)} = 1$
$k_{M_2}^{(M_2,M_3)} = 0$	0.7	0
$k_{M_2}^{(M_2,M_3)} = 1$	0	0.3

Bunch 3	$k_{M_3}^{(M_1,M_3)} = 0$	$k_{M_3}^{(M_1,M_3)} = 1$
$k_{M_1}^{(M_1,M_3)} = 0$	0.4	0.3
$k_{M_1}^{(M_1,M_3)} = 1$	0.3	0

Table 4.9: Three bunches in the CBD representation of LG inequality for a contextual example. This example adapted from [Dzhafarov and Kujala, 2016].

Dzhafarov and Kujala [2016] also represent the maximal couplings for three connections of such system as Table 4.10.

Coupling 1	$T_{M_1}^{(M_1,M_3)} = 0$	$T_{M_1}^{(M_1,M_3)} = 1$
$T_{M_1}^{(M_1,M_2)} = 0$	0.7	0
$T_{M_1}^{(M_1,M_2)} = 1$	0	0.7

Coupling 2	$T_{M_2}^{(M_2,M_3)} = 0$	$T_{M_2}^{(M_2,M_3)} = 1$
$T_{M_2}^{(M_1,M_2)} = 0$	0.7	0
$T_{M_2}^{(M_1,M_2)} = 1$	0	0.7

Coupling 3	$T_{M_3}^{(M_1,M_3)} = 0$	$T_{M_3}^{(M_1,M_3)} = 1$
$T_{M_3}^{(M_2,M_3)} = 0$	0.7	0
$T_{M_3}^{(M_2,M_3)} = 1$	0	0.7

Table 4.10: Three maximal couplings in the LG example adapted from [Dzhafarov and Kujala, 2016].

A possible coupling of the system is illustrated in Figure 4.4, which demonstrates a restriction on constructing a maximally connected coupling. As Dzhafarov and Kujala [2016] describe, moving clockwise from S_1^1 we can conclude $Pr[S_1^1 = S_3^3] = 1$, but moving counterclockwise we can conclude $Pr[S_1^1 = S_1^3] = 1$. As a result we should have $Pr[S_1^3 = S_3^3] = 1$ which does not match with the distribution in bunch 3.

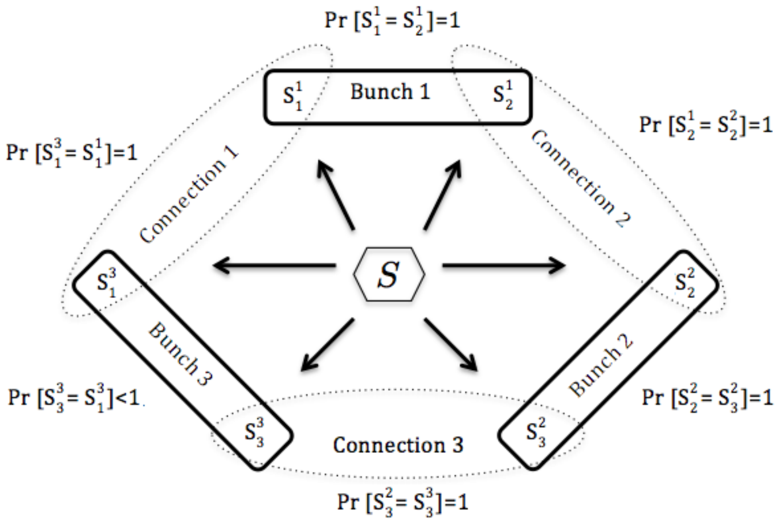

Figure 4.4: A possible maximally connected coupling for an example of a contextual LG system [Dzhafarov and Kujala, 2016]. This shows a restriction on constructing a maximally connected coupling of the system.

In this section, I described some cyclic contextuality scenarios in the CBD approach in comparison with their representations in Spekkens' approach. For example, I explained how the double indexing notation reduces to the standard notation of random variables in order to picture contextuality in Specker scenario. In the earlier sections, I provided more comparisons between the CBD and Spekkens approaches. As a result of these comparisons, I concluded that the CBD approach does not successfully cover all critical concepts in the definition of contextuality. For example, I explained that there is not a clear connection between hidden variables or ontic states and the double indexing notation of random variables in a given system. This limitation leads to the unclear position of this notion about parameter independence condition in Bell scenario. This limitation also entails that it is impossible to evaluate a relaxation of non-signaling condition in this notion using the casual structure. Based on these comparisons, we can consider Spekkens' approach as the more comprehensive contextuality notion. This makes Spekkens' approach as an appropriate candidate for modeling contextuality in cognition. In the next section, I will discuss possible applications of this notion in cognition.

4.6 Application of Spekkens' model in cognition

Much of my contribution in this section was inspired by the work of Spekkens [2005], which was introduced earlier in Section 4.3. I have used his approach to treat contextuality as it occurs for both preparation and measurement processes in cognition. This approach extends existing work on contextuality in QC [Aerts et al., 2013, Bruza et al., 2015a, Bruza, 2016, Dzhafarov and Kujala, 2016, Dzhafarov et al., 2016]. Thus, I have provided: (a) a method for modelling contextuality in preparation, and; (b) refinements in our understanding of contextuality as it arises during measurement.

This approach is formulated in terms of basic operations such as preparation (P) and measurement (M), and the probabilities for various possible measurement outcomes. As I described in Section 4.3.1, in physics, operational theory is defined based on these experimental procedures. This section applies the same approach to cognitive experiments. More specifically, I show how this approach is compatible with the process of recall demonstrated in Section 3.2, and allude to the manner in which the same approach can be used to model contextuality.

The mathematical structures needed for the operational approach were introduced earlier in Sections 3.2, 3.3 and 4.3.1. To define this approach, we apply the density matrix ρ to model the preparation process P, and a POVM$\{\tilde{\mathbb{P}}_k\}$ to the measurement process M using Definitions 6 and 7.

These definitions make it possible to specify the notion of a contextual model, where instead of explicitly considering quantum states and POVMs, we specify probabilistic interpretations of the preparation and measurement procedures that are used to create them ($\mu_P(\lambda)$ and $\xi_{M,k}(\lambda)$). These are the probabilities that determine what can be known and inferred by observers [Harrigan and Spekkens, 2010, Spekkens, 2005].

To fully describe the Spekkens approach for cognition, we require a precise understanding of an ontological model in this approach. As mentioned earlier, the intrinsic properties of a physical system are called its *ontic state* and this is denoted by λ which belongs to a set of all possible ontic states Λ. Now I make use of a similar conception in cognition: as I described in Section 4.2.1, an ontic state in cognition should refer to the reality of the cognitive system, that is, the presumed features of a cognitive state (of mind) which exist without performing experiments or any other form of observation.

The fundamental idea of noncontextuality in this approach is that processes which are operationally equivalent should not be distinguishable in an ontological model [Leifer, 2014]. This means two processes which generate the same observable probabilities should be demonstrated by the same probability distributions over their underlying ontic state. Thus, we can say the two preparation processes of a state are noncontextual when they yield the same probability distributions without changing the intrinsic properties of the system.

We can illustrate this approach in cognition with reference to the recall experiment introduced in Section 3.2. In that example, the preparation scenario leads to some proportion of subjects having different cognitive states (3.11), where the density matrix of ensembles of subjects is represented using the convex composition of pure states ψ_v representing those subjects, with the probability p_v.[20] This density state of ensembles encodes a probability distribution over the ontic space [21]. However, we know that our cognitive reality is not completely controlled by the preparation process, which means that preparing our cognitive system in a specific state ρ does not give us any information about the exact ontic state. Our knowledge of the ontic state can only be described by the probability distribution $\mu(\lambda)$ that we introduced earlier. The assumption of noncontextual preparation entails that the probability distribution of preparation P on an ontic state λ depends only on the ρ related to P [Spekkens, 2005]

$$\mu_P(\lambda) = \mu_\rho(\lambda). \tag{4.36}$$

This implies that for an ensemble of subjects to be noncontextual, the distribution $\mu_P(\lambda)$ should not depend on a specific convex decomposition of ρ (each convex decomposition provides a different context for the preparation P). In other words, consider a hypothetical scenario where two preparation procedures result in the same density matrix. This density matrix of the ensemble of subjects can be implemented based on the two[22] different convex decompositions of ρ, represented as $\rho = \sum_v p_v \rho_v$ (see Equation (3.12)). Each of these

[20]It worth pointing out that we can also consider a more general scenario in which the density matrix of this recall experiment is a convex composition of mixed states.

[21]It is possible to have a one-to-one relationship between the state of a system ψ and the reality, although in most physical models, ψ only indicates a state of incomplete knowledge about reality. A thorough discussion of this can be found in [Harrigan and Spekkens, 2010].

[22]It is clear in a more general scenario, there could be more possible convex decompositions for the density matrix ρ.

possible decompositions can be associated to a member of the equivalence class of preparation procedures for density matrix ρ. These two procedures are non-contextual if the probability distribution $\mu_P(\lambda)$ is independent of each member of that class.

As with the preparation process, we can consider the measurement process as measuring the ontic state of our cognitive system. Performing this measurement does not guarantee access to the ontic state, it only provides different probabilities that the system exists in one of a collection of ontic states. These probabilities are represented by $\xi_\lambda(K|M)$ [Spekkens, 2005]. The assumption of noncontextual measurement entails that the probability distribution of measurement M on an ontic state λ depends only on the POVM $\{\tilde{\mathbb{P}}_k\}$ related to M:

$$\xi_\lambda(K|M) = \xi_\lambda(k|\{\tilde{P}_k\}). \tag{4.37}$$

One way to represent context for measurement in cognition is to consider Neumark's theorem (see Section 3.4.1). This approach suggests that each way of recognizing a POVM on a subsystem H_R, based on its coupling with the environment (noise) and performing a projective measurement on a composite system H, implicitly generates a context. In other words, different sources of noise can lead to different contexts. Examples of these sources of noise might include episodic memories constructed for different subjects throughout their lifetimes (and represented in a semantic network), e.g. subjects who have had their breakfast or not, give different responses to a priming word "food". The measurements of psychology are inherently noisy, but this approach offers ways in which we might start to model this phenomenon.

In this section, I have shown that reconsidering an ensemble of subjects and a generalized measurement for cognitive systems leads us to identify new sources of contextuality. This approach describes, for the first time, the way in which contextuality affects the preparation process in cognition. Moreover, the suggested contextuality model for the measurement process extends the standard measurement approaches traditionally used in quantum cognition with a non-projective measurement. Furthermore, this approach opens up a possibility for extending this model using contextuality models in physics, where every convex decomposition of a POVM $\{\tilde{\mathbb{P}}_k\}$ reveals a context for the measurement process, as described by Spekkens [2005].

4.7 Conclusion

In this chapter, I constructed connections between the notations for the sheaf-theoretical approach of Abramsky and Brandenburger [2011] and the operational theory of Spekkens [2005], mainly through a consideration of Bell's inequality and the KS theorem. This is an important contribution, as while each approach attempts to unify our understanding of contextuality, their use of different notation makes a comparison difficult.

These mathematical comparisons are notable because of their contribution to providing a better understanding of contextuality in QM, although more work is needed to successfully translate all different aspects of these two approaches (e.g. the logical level of contextuality) into each other's terms. As suggested in Section 4.4.2, future work can use the more rigorous mathematical level of representation theories to compare different generalized approaches of contextuality.

These mathematical comparisons are also significant for one of the aims of my thesis which is to provide better insights regarding contextuality in QC. The detailed mathematical descriptions of (non)locality and contextuality in this chapter provide a better understanding of critical concepts such as the ontic state, random variable and probability space. I identified these concepts as the Achilles heel of the CBD approach, a contextuality model arising from QC. For instance, in an specific cyclic example, I explained how the double indexing notation reduces to the stranded representation of random variables in order to illustrate contextuality in Specker scenario. Future research can investigate a compatibility the CBD notion with other (non)contextuality scenarios like Kochen-Specker theorem.

I also highlighted the other obstacle in the CBD approach which is the lack of an empirical meaning, since this approach suggests a set of random variables without any empirical meaning for the coupling process. I believe that any unified theories of contextuality should be both mathematically robust and empirically verifiable. Because of this, I brought up the Spekkens' approach as an appropriate candidate for modeling contextuality in cognition. Its mathematical strength makes it capable of generalizing the notion of contextuality for unsharp measurement and not deterministic hidden variable models. Furthermore, its verifiable empirical content provides us with meaningful formulations and results for the whole space of the theory. The technical aspects of this approach covered in this chapter can offer a number of new avenues to follow as we attempt to model contextuality for more experiments in cognition such as noncontextuality inequalities proposed based on the operational approach in QM [Schmid et al., 2017, Spekkens et al., 2009]. Moreover, the sheaf-theoretical approach can be considered as another appropriate candidate for modeling contextuality in cognition. Its mathematical sophistication makes it difficult to apply at present, but I hope to pursue this approach in my future research.

Chapter 5

Conclusions and future work

This work suggests a number of new avenues that can be pursued in developing a more detailed understanding of the complex processes of cognition, specifically, memory and recall. It provides us with a new treatment of measurement and contextuality as central topics in cognition, which can lead to new conceptual understandings and new empirical designs. In this chapter, I outline the key results of this thesis and bring into focus some potentially interesting directions for future research.

The topic of measurement was investigated in Chapter 3. Previous research in Quantum Cognition (QC) has mainly focused on the ideal situation of performing standard projective measurements on only one subject. But in Chapter 3, I generalized these ideal assumptions to a more realistic situation of performing non-orthogonal and non-projective measurement on an ensemble of subjects. I applied the most general description of quantum states, "a density matrix", and the most general type of measurement, "POVM", to the domain of memory models in cognition. My suggested formalism for cognitive systems separates the preparation process from the measurement process of a cognitive state.

I introduced an application of my generalized measurement formalism by employing Neumarks dilation theorem. To obtain information about the state of a cognitive system, this theorem extends the original Hilbert space into a larger space, and relates a PVM on the enlarged space to a POVM on the restricted space. This mathematical structure can be used to trace out noises from our cognitive experiments.

I concluded Chapter 3 with a discussion of how a better understanding of cognitive states combined with the suggested preparation and measurement processes opens further avenues for using more concepts from Quantum Mechanics (QM) in mathematical models of cognition. As an example of this approach, I suggested using quantum tomography in cognition to characterize the unknown state of a cognitive system. I highlighted that a detailed mathematical description of tomography for cognitive states (cobits) requires a precise cognitive realization of their geometrical representation. Reaching such a realization could be the focus of future studies where the geometrical representation is convertible to the complex vector space representation for cobits.

Another avenue for future research is to use my suggested density matrix to construct the quantum Shannon entropy [Von Neumann, 1996] for cognitive states. Quantum Shannon entropy is a quantitative measure of information and correlation which could be used to study noisy cognitive systems. One of the important challenges in cognition is to understand how correlations in combination of concepts can be modeled, especially if the combination follows a non-classical pattern (see Chapter 1). In this thesis, I described nonlocality (which arises due to the existence of entanglement phenomena) as one of the promising approaches for modeling non-classical correlations in cognition. However, this is not the only possible quantum approach for modeling such correlations. For example, we could use the quantum discord measure [Vedral, 2017] for cognitive correlations. Quantum discord is a form of quantum Shannon entropy inequalities, which does not necessarily entail quantum entanglement.

Distinguishing the preparation process from the measurement process of a cognitive state in Chapter 3 allowed reconsidering the interpretation of context within these two processes separately. As a result, in Chapter 4, I precisely specified the contextuality effects that result from the preparation process and the generalized measurement in a cognitive system. My approach is based on the operational method of contextuality in physics suggested by Spekkens [2005], which I believe is the critical contribution of this work, relating the results of Chapters 3 and 4 to each other. This generalized approach of contextuality unlike Bell's experiment does not necessary require space-like separation. Therefore, its application in modeling cognitive scenarios is not limited by relativity principle or the prohibition of superluminal signaling.

In this thesis I emphasized on the importance of the non-signaling (or non-disturbance) conditions for contextuality models in cognition, when those models are limited to Bell's inequality or other (non)contextuality scenarios in physics which require holding those conditions. But this is unpractical, since none of the existing cognitive experiments are successful to satisfy those conditions. As I discussed in Chapter 4, Spekkens' generalized approach is an alternative way to turn around this limitation. I also briefly introduced the casual interpretations of CHSH violation which can be considered as the other alternative. This model investigates nonlocality in existing of some form of communications. These communications cannot be arbitrary, and I believe designing cognitive experiments with the *controlled* form of communications requires careful attentions in future research.

To provide clarity and novel insights regarding contextuality in QC, in Chapter 4, I compared the Spekkens' operational method with two other main approaches. I provided consistent descriptions of these three generalized approaches of contextuality which helps to translate their specific notations and meanings to each other. Firstly, I compared the Spekkens' approach with the sheaf theory of Abramsky and Brandenburger [2011] which is also based in physics. My comparison revealed an equivalence between this approach and the operational method for Bell's inequality and the Kochen-Specker theorem. Further work is certainly required to investigate if there is a more complete equivalence between these two approaches e.g., to check if an equivalence can be demonstrated by a precise translation of the three levels of contextuality in the sheaf-theoretical approach to the operational method.

The other contextuality approach explored in Chapter 4 was CBD which is mainly known in QC. I pointed out some of the differences between this approach and the standard approaches of contextuality in QM introduced earlier. I described these differences as obstacles in the process of using the CBD in QC, which directed us to using alternative models of contextuality. I explained the application of the Spekkens' operational method in QC as an example of these alternative models, but I kept a possible application of the sheaf-theoretical approach for future research. Moreover, future investigations are necessary to compare the Spekkens' approach with the state context property (SCOP) formalism, the first well developed model of contextuality in QC based on an operational approach in QM (which I briefly introduced in Chapter 2). Future studies could explore the concept of contextuality in cognition further by considering other recent developments in QM such as the graph-theoretical approach of Cabello et al. [2010b] and the hyper graph theory of Acín et al. [2015].

The broad implication of the present research is that it provides a better technical description for existing formalisms used in QC for the topics of measurement and contextuality, in addition to suggesting new frameworks for these two topics. To fit the mathematical formalisms of quantum theory into areas outside physics (such as cognition), we need a deep understanding of these formalisms in their original context. To reach this understanding, I highlighted some key aspects of these formalisms and elucidated associated mathematical notations. In particular, I focused on the important detailed mathematical and physical aspects that have been ignored in some previous researches in QC. This allowed me to extend and modify the existing mathematical structures of QC by applying new formalisms from QM. This process can play an important role in the future development of QC.

Literature Cited

Abramsky, S. (2013). Relational Databases and Bell's Theorem. In Tannen, V., Wong, L., Libkin, L., Fan, W., Tan, W.-C., and Fourman, M., editors, *In Search of Elegance in the Theory and Practice of Computation: Essays Dedicated to Peter Buneman*, pages 13—-35. Springer Berlin Heidelberg, Berlin, Heidelberg.

Abramsky, S., Barbosa, R. S., Kishida, K., Lal, R., and Mansfield, S. (2015). Contextuality, Cohomology and Paradox. *https://arxiv.org/abs/1502.03097*.

Abramsky, S. and Brandenburger, A. (2011). The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036.

Abramsky, S. and Heunen, C. (2016). Operational Theories and Categorical Quantum Mechanics. In Chubb, J., Eskandarian, A., and Harizanov, V., editors, *Logic and Algebraic Structures in Quantum Computing*, pages 88–122. Cambridge University Press.

Acín, A., Fritz, T., Leverrier, A., and Sainz, A. B. (2015). A combinatorial approach to nonlocality and contextuality. *Communications in Mathematical Physics*, 334(2):533–628.

Aerts, D. (2011). Quantum interference and superposition in cognition: A theory for the disjunction of concepts. In Aerts, D., editor, *Worldviews, Science* *and Us: Bridging Knowledge and Its Implications for Our Perspectives of the World*, page 169. World Scientific.

Aerts, D. and Aerts, S. (1997). Applications of Quantum Statistics in Psychological Studies of Decision Processes. In *Topics in the Foundation of Statistics*, pages 85–97. Springer Netherlands, Dordrecht.

Aerts, D. and Gabora, L. (2005). A theory of concepts and their combinations I. *Kybernetes*, 34(1/2):167–191.

Aerts, D., Gabora, L., and Sozzo, S. (2013). Concepts and their dynamics: a quantum-theoretic modeling of human thought. *Topics in cognitive science*, 5(4):737–772.

Aerts, D. and Sozzo, S. (2011). Quantum Structure in Cognition: Why and How Concepts Are Entangled. In *Quantum Interaction: 5th International Symposium, QI 2011, Aberdeen, UK, June 26-29, 2011, Revised Selected Papers*, pages 116–127. Springer, Berlin, Heidelberg.

Aliakbarzadeh, M. and Kitto, K. (2016). Applying POVM to model non-orthogonality in quantum cognition. In Atmanspacher, H., Filk, T., and Pothos, E., editors, *Quantum Interaction*, volume 9535 of *Lecture Notes in Computer Science*, Cham. Springer International Publishing.

Aliakbarzadeh, M. and Kitto, K. (2018). Preparation and measurement in quantum memory models. *Journal of Mathematical Psychology*, 83:24–34.

Asano, M., Khrennikov, A., Ohya, M., Tanaka, Y., and Yamato, I. (2015). *Quantum adaptivity in biology: from genetics to cognition*. Springer.

Aspect, A., Dalibard, J., and Roger, G. (1982a). Experimental test of

Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49(25):1804–1807.

Aspect, A., Grangier, P., and Roger, G. (1982b). Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities. *Physical Review Letters*, 49(2):91–94.

Atmanspacher, H., Bergomi, C., Filk, T., and Kitto, K., editors (2015). *Quantum Interaction*, volume 8951 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham.

Ballentine, L. E. and Jarrett, J. P. (1987). Bell's theorem: Does quantum mechanics contradict relativity? *American Journal of Physics*, 55:5.

Barnett, S. (2009). *Quantum Information*. Oxford University Press.

Bell, J. S. (1964). On the Einstein Podolsky Rosen paradox. *Physics 1*, 1(3):195–200.

Bendersky, A., Senno, G., De La Torre, G., Figueira, S., and Acin, A. (2017). Non-signaling deterministic models for non-local correlations have to be uncomputable. *Physical Review Letters*, 118(13):130401.

Berkovitz, J. (2016). Action at a Distance in Quantum Mechanics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 201 edition.

Bertlmann, R. A. and Krammer, P. (2008). Bloch vectors for qudits. *Journal of Physics A: Mathematical and Theoretical*, 41(23):235303.

Bohm, D. (1952). A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I. *Physical Review*, 85(2):166–179.

Brask, J. B. and Chaves, R. (2017). Bell scenarios with communication. *Journal of Physics A: Mathematical and Theoretical*, 50(9):094001.

Broekaert, J., Basieva, I., Blasiak, P., and Pothos, E. M. (2017). Quantum-like dynamics applied to cognition: a consideration of available options. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 375(2106):20160387.

Bruza, P., Kitto, K., Nelson, D., and McEvoy, C. (2009). Is there something quantum-like about the human mental lexicon? *Journal of mathematical psychology*, 53(5):363–377.

Bruza, P., Kitto, K., Ramm, B., Sitbon, L., Song, D., and Blomberg, S. (2012). Quantum-like non-separability of concept combinations, emergent associates and abduction. *Logic Journal of the IGPL*, 20(2):445–457.

Bruza, P., Kitto, K., Ramm, B. J., and Sitbon, L. (2015a). A probabilistic framework for analysing the compositionality of conceptual combinations. *Journal of Mathematical Psychology*, 67:26–38.

Bruza, P. D. (2016). Syntax and operational semantics of a probabilistic programming language with scopes. *Journal of Mathematical Psychology*, 74:46–57.

Bruza, P. D., Wang, Z., and Busemeyer, J. R. (2015b). Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences*, 19(7):383–393.

Bub, J. (2016). *Bananaworld : quantum mechanics for primates*. Oxford University Press.

Busemeyer, J. R. and Bruza, P. D. (2012). *Quantum Models of Cognition and Decision*. Cambridge University Press.

Butterfield, J., Fleming, G. N., Ghirardi, G. C., and Grassi, R. (1993). Parameter dependence in dynamical models for statevector reduction. *International Journal of Theoretical Physics*, 32(12):2287–2304.

Cabello, A., Estebaranz, J., and García-Alcaine, G. (1996). Bell-Kochen-Specker theorem: A proof with 18 vectors. *Physics Letters A*, 212(4):183–187.

Cabello, A., Severini, S., and Winter, A. (2010a). (Non-)Contextuality of Physical Theories as an Axiom.

Cabello, A., Severini, S., and Winter, A. (2010b). (Non-)Contextuality of Physical Theories as an Axiom. *https://arxiv.org/pdf/1010.2163.pdf*.

Cervantes, V. H. and Dzhafarov, E. N. (2017). Snow Queen is Evil and Beautiful: Experimental Evidence for Probabilistic Contextuality in Human Choices.

Cirel'son, B. S. (1980). Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100.

Clark, S., Coecke, B., and Sadrzadeh, M. (2008). A Compositional Distributional Model of Meaning. In *The Second Symposium on Quantum Interaction*.

Clauser, J. F., Horne, M. A., Shimony, A., and Holt, R. A. (1969). Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884.

Constantin, C. M. (2015). Sheaf-Theoretic Methods in Quantum Mechanics and Quantum Information Theory.

Conte, E., Khrennikov, A., Todarello, O., Federici, A., and Zbilut, J. P. (2008). A Preliminary Experimental Verification On the Possibility of Bell Inequality Violation in Mental States. *NeuroQuantology*, 6(3).

Conte, E., Todarello, O., Federici, A., Vitiello, F., Lopane, M., Khrennikov, A., and Zbilut, J. P. (2007). Some remarks on an experiment suggesting quantum-like behavior of cognitive entities and formulation of an abstract quantum mechanical formalism to describe cognitive entity and its dynamics. *Chaos, Solitons & Fractals*, 31(5):1076–1088.

Culham, J. C., Cavanagh, P., and Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron*, 32(4):737–45.

de Barros, J. A., Kujala, J., and Oas, G. (2016). Negative Probabilities and Contextuality. *Journal of Mathematical Psychology*, 74:34–45.

de Barros, J. A., & Oas, G. (2015). Some Examples of Contextuality in Physics: Implications to Quantum Cognition. In Dzhafarov, E., Jordan, S., Zhang, R., and Victor, C., editors, *Contextuality from Quantum Physics to Psychology*, page 153.

de Deyne, S., Navarro, D. J., and Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, 45(2):480–98.

de Muynck, W. M. (2002). *Foundations of Quantum Mechanics, an Empiricist Approach*. Springer Netherlands, Dordrecht.

de Muynck, W. M. (2007). POVMs: a small but important step beyond standard quantum mechanics. In Nieuwenhuizen, T., Špicka, V., Mehmani, B., Aghdami, M., and Khrennikov, A., editors, *Beyond the Quantum*, pages 69–79. World Scientific.

de Muynck, W. M. and Martens, H. (1989). Nonideal Quantum Measurements, Simultaneous Measurement of Incompatible Observables, and the Bell

Inequalities. In *Proceedings of the 3rd International Symposium Foundations of Quantum Mechanics in the Light of New Technology: Central Research Laboratory, Hitachi, Ltd., Kokubunji, Tokyo, Japan*, pages 171–176.

Delfosse, N., Okay, C., Bermejo-Vega, J., Browne, D. E., and Raussendorf, R. (2017). Equivalence between contextuality and negativity of the Wigner function for qudits. *New Journal of Physics*, 19(12):123024.

Dickson, M. (1998). *Quantum chance and non-locality : probability and non-locality in the interpretations of quantum mechanics*. Cambridge University Press.

Dzhafarov, E., Zhang, R., and Kujala, J. (2015a). Is there contextuality in behavioral and social systems? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2058):20150099.

Dzhafarov, E. N. and Kujala, J. V. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology*, 56(1):54–63.

Dzhafarov, E. N. and Kujala, J. V. (2014a). Generalizing Bell-type and Leggett-Garg-type Inequalities to Systems with Signaling.

Dzhafarov, E. N. and Kujala, J. V. (2014b). No-forcing and no-matching theorems for classical probability applied to quantum mechanics. *Foundations of Physics*, 44(3):248–265.

Dzhafarov, E. N. and Kujala, J. V. (2014c). On selective influences, marginal selectivity, and bell/CHSH inequalities. *Topics in cognitive science*, 6(1):121–128.

Dzhafarov, E. N. and Kujala, J. V. (2015). Probabilistic contextuality in EPR/Bohm-type systems with signaling allowed. In Dzhafarov, E., Jordan, S., Zhang, R., and Cervantes, V., editors, *Contextuality from Quantum Physics to Psychology*, page 287.

Dzhafarov, E. N. and Kujala, J. V. (2016). Context-content systems of random variables: The Contextuality-by-Default theory. *Journal of Mathematical Psychology*, 74:11–33.

Dzhafarov, E. N., Kujala, J. V., and Larsson, J.-Å. (2015b). Contextuality in Three Types of Quantum-Mechanical Systems. *Foundations of Physics*, 45(7):762–782.

Dzhafarov, E. N., Kujala, J. V., and Larsson, J.-Å. (2016). Contextuality-by-default: a brief overview of ideas, concepts, and terminology. In Atmanspacher, H., Filk, T., and Pothos, E., editors, *Quantum Interaction: 9th International Conference, QI 2015, Filzbach, Switzerland, July 15-17, 2015, Revised Selected Papers*, pages 12–23. Springer International Publishing.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Phys. Rev. 47, 777 (1935): Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47(10):777–780.

Emary, C., Lambert, N., and Nori, F. (2014). LeggettGarg inequalities. *Reports on Progress in Physics*, 77(1):016001.

Emch, G. G. (2009). *Algebraic methods in statistical mechanics and quantum field theory*. Courier Corporation.

Ernst Specker, B. (2011). The logic of non-simultaneously decidable propositions.

Fano, V. (2006). quantum non-locality and the mathematical representation of experience. In Garola, C., Rossi, A., and Sozzo, S., editors, *The Foundations* *of Quantum Mechanics, Historical Analysis and Open Questions - Cesena 2004*, pages 142–155. World Scientific.

Feintzeig, B. H. and Fletcher, S. C. (2017). On Noncontextual, Non-Kolmogorovian Hidden Variable Theories. *Foundations of Physics*, 47(2):294–315.

Fine, A. (1982a). Hidden variables, joint probability, and the bell inequalities. *Physical Review Letters*, 48(5):291–295.

Fine, A. (1982b). Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23(7):1306.

Franco, R. (2016). Towards a new quantum cognition model. *https://arxiv.org/pdf/1611.09212.pdf*.

Gabora, L. and Aerts, D. (2002). Contextualizing concepts using a mathematical generalization of the quantum formalism. *Journal of Experimental & Theoretical Artificial Intelligence*, 14(4):327–358.

Gelfand, I. and Neumark, M. (1943). On the imbedding of normed rings into the ring of operators in Hilbert space. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):197 – 217.

Gershenson, C. (2002). Contextuality: a philosophical paradigm, with applications to philosophy of cognitive science. *http://cogprints.org/2621/1/PhilCogSci2-Contextuality.pdf*.

Ghirardi, Rimini, and Weber (1986). Unified dynamics for microscopic and macroscopic systems. *Physical review. D, Particles and fields*, 34(2):470–491.

Gibbons, K. S., Hoffman, M. J., and Wootters, W. K. (2004). Discrete phase space based on finite fields. *Physical Review A*, 70(6):62101.

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery clinics of North America*, 22(2):133–9, vii.

Hampton, J. (1997). Conceptual combination. In Lamberts, K. and David, S., editors, *Knowledge Concepts and Categories*, chapter 4, pages 133–159. psychology press.

Hardy, L. (1993). Nonlocality for two particles without inequalities for almost all entangled states. *Physical Review Letters*, 71(11):1665.

Harrigan, N. and Rudolph, T. (2007). Ontological models and the interpretation of contextuality.

Harrigan, N. and Spekkens, R. W. (2010). Einstein, Incompleteness, and the Epistemic View of Quantum States. *Foundations of Physics*, 40(2):125–157.

Heisenberg, W. (1927). Uber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift fur Physik*, 43(3-4):172–198.

Isham, C. J. (2001). *Lectures on Quantum Theory*. Imperial College Press.

J Acacio de Barros and G Oas, G. (2014). Negative probabilities and counterfactual reasoning in quantum cognition. *Physica Scripta*, T163(T163):014008.

Jaeger, G. (2009). *Entanglement, Information, and the Interpretation of Quantum Mechanics*. The Frontiers Collection. Springer Berlin Heidelberg, Berlin, Heidelberg.

Jarrett, J. P. (1984). On the physical significance of the locality conditions in the Bell arguments. *Noûs*, 18(4):569.

Kenfack, A. and Yczkowski, K. (2004). Negativity of the Wigner function as an indicator of non-classicality. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(10):396–404.

Khrennikov, A. and Basieva, I. (2014). Quantum Model for Psychological Measurements: From the Projection Postulate to Interference of Mental Observables Represented As Positive Operator Valued Measures. *Neuro-Quantology*, 12(3):324–336.

Khrennikov, A., Basieva, I., Dzhafarov, E. N., and Busemeyer, J. R. (2014). Quantum models for psychological measurements: an unsolved problem. *PloS one*, 9(10):e110909.

Khrennikov, A. Y. (2010). *Ubiquitous Quantum Structure: From Psychology to Finance*. Springer Science & Business Media.

Kimura, G. (2003). The Bloch vector for N-level systems. *Physics Letters A*, 314(5-6):339–349.

Kitto, K. (2008). Why quantum theory? In Bruza, P. D., Lawless, W., C., v. R., Sofge, D., Coecke, B., and & Clark, S., editors, *Second Quantum Interaction Symposium*.

Kochen, S. and Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. *In The Logico-Algebraic Approach to Quantum Mechanics*, pages 293–328.

Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., and George, M. S. (2005). Detecting Deception Using Functional Magnetic Resonance Imaging. *Biological Psychiatry*, 58(8):605–613.

Krank, M. D. and Wall, A. (2005). Context and retrieval effects on implicit cognition for substance use. In Wiers, R. W. and Stacy, A. W., editors, *Handbook of Implicit Cognition and Addiction*, pages 281–292. SAGE Publications.

Kujala, J. V., Dzhafarov, E. N., and Larsson, J.-Å. (2015). Necessary and Sufficient Conditions for an Extended Noncontextuality in a Broad Class of Quantum Mechanical Systems. *Physical Review Letters*, 115(15):150401.

Kunjwal, R. (2016). *Contextuality beyond the Kochen-Specker theorem*. PhD thesis, The Institute of Mathematical Sciences, Chennai.

Kunjwal, R. and Ghosh, S. (2014). Minimal state-dependent proof of measurement contextuality for a qubit. *Physical Review A*, 89(4):042118.

Kunjwal, R. and Spekkens, R. W. (2015). From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism. *Physical Review Letters*, 115(11):110403.

Kunjwal, R. and Spekkens, R. W. (2016). Noncontextuality inequalities for Specker's compatibility scenario.

Kurzyński, P., Cabello, A., and Kaszlikowski, D. (2014). Fundamental Monogamy Relation between Contextuality and Nonlocality. *Physical Review Letters*, 112(10):100401.

Kurzyński, P., Kołodziejski, A., Laskowski, W., and Markiewicz, M. (2016). Three-dimensional visualization of a qutrit. *Physical Review A*, 93(6):062126.

Leggett, A. J. and Garg, A. (1985). Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters*, 54(9):857–860.

Leifer, M. S. (2014). Is the Quantum State Real? An Extended Review of ψ-ontology Theorems. *Quanta*, 3(1):67–155.

Leonhardt, U. (1997). *Measuring the Quantum State of Light*. Cambridge University Press.

Liang, Y.-C., Spekkens, R. W., and Wiseman, H. M. (2011). Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Physics Reports*, 506(1):1–39.

Lovering, T. (2010). Sheaf Theory.

Lundeen, J. S., Feito, A., Coldenstrodt-Ronge, H., Pregnell, K. L., Silberhorn, C., Ralph, T. C., Eisert, J., Plenio, M. B., and Walmsley, I. A. (2008). Tomography of quantum detectors. *Nature Physics*, 5(1):27–30.

Maudlin, T. (1994). *Quantum non-locality and relativity: metaphysical intimations of modern physics*, volume 13 of *Aristotelian society series*. Blackwell publishers limited, Oxford, UK.

Maudlin, T. (2011). *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. John Wiley & Sons.

Mazurek, M. D., Pusey, M. F., Kunjwal, R., Resch, K. J., and Spekkens, R. W. (2016). An experimental test of noncontextuality without unphysical idealizations. *Nature Communications*, 7:ncomms11780.

Mermin, N. D. (1993). Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics*, 65(3):803–815.

Miller, W. (1972). *Symmetry groups and their applications*. Academic Press.

Mittelstaedt, P. (1997). Is Quantum Mechanics a Probabilistic Theory? pages 159–175. Springer, Dordrecht.

Muller, A., Breguet, J., and Gisin, N. (1993). Experimental Demonstration of Quantum Cryptography Using Polarized Photons in Optical Fibre over More than 1 km. *Europhysics Letters (EPL)*, 23(6):383–388.

Nelson, D. L., Kitto, K., Galea, D., McEvoy, C. L., and Bruza, P. D. (2013). How activation, entanglement, and search in semantic memory contribute to event memory. *Memory and Cognition*, 41(6):717–819.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407.

Nielsen, M. A. and Chuang, I. L. (2010). *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–72.

Peres, A. (1990). Neumark's theorem and quantum inseparability. *Foundations of Physics*, 20(12):1441–1453.

Popescu, S. and Rohrlich, D. (1994). Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385.

Pothos, E. M. and Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *The Behavioral and brain sciences*, 36(3):255–74.

Prugovecki, E. (1995). *Principles of Quantum General Relativity*. WORLD SCIENTIFIC.

Pusey, M. F., Barrett, J., and Rudolph, T. (2012). On the reality of the quantum state. *Nature Physics*, 8(6):476–479.

Ramanathan, R., Soeda, A., Kurzyński, P., and Kaszlikowski, D. (2012). Generalized Monogamy of Contextual Inequalities from the No-Disturbance Principle. *Physical Review Letters*, 109(5):050404.

Raussendorf, R., Browne, D. E., Delfosse, N., Okay, C., and Bermejo-Vega, J. (2017). Contextuality and Wigner-function negativity in qubit quantum computation. *Physical Review A*, 95(5):052334.

Sainz, A. B. and Wolfe, E. (2017). Multipartite Composition of Contextuality Scenarios.

Sandeep K Goyal, B Neethi Simon, Rajeev Singh, and Sudhavathani Simon (2016). Geometry of the generalized Bloch sphere for qutrits. *Journal of Physics A: Mathematical and Theoretical*, 49(16):165203.

Schmid, D., Spekkens, R. W., and Wolfe, E. (2017). All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed sets of operational equivalences.

Seevinck, M. P. (2008). Parts and Wholes. An Inquiry into Quantum and Classical Correlations.

Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science*, 27(3):341 – 378. 2002 Rumelhart Prize Special Issue Honoring Richard Shiffrin.

Shimony, A. (1984). Contextual hidden variables theories and Bell's inequalities. *The British Journal for the Philosophy of Science*, 35(1):25–45.

Shimony, A. (1993). *Search for a Naturalistic World View*. Cambridge University Press, Cambridge.

Shimony, A. (2016). Bell's Theorem. *The Stanford Encyclopedia of Philosophy*.

Simmons, A. W., Wallman, J. J., Pashayan, H., Bartlett, S. D., and Rudolph, T. (2016). Contextuality under weak assumptions.

Smith, D. W. (2016). Phenomenology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, winter 201 edition.

Spekkens, R. (2007). Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110.

Spekkens, R. W. (2005). Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A*, 71(5):52108.

Spekkens, R. W. (2008). Negativity and contextuality are equivalent notions of nonclassicality. *Physical Review Letters*, 101(2):020401.

Spekkens, R. W., Buzacott, D. H., Keehn, A. J., Toner, B., and Pryde, G. J. (2009). Preparation Contextuality Powers Parity-Oblivious Multiplexing. *Physical Review Letters*, 102(1):010401.

Su, H.-Y., Chen, J.-L., and Liang, Y.-C. (2015). Demonstrating quantum contextuality of indistinguishable particles by a single family of noncontextuality inequalities. *Scientific Reports*, 5(1):11637.

Thew, R. T., Nemoto, K., White, A. G., and Munro, W. J. (2002). Qudit quantum-state tomography. *Physical Review A*, 66(1):012303.

Tittel, W., Brendel, J., Zbinden, H., and Gisin, N. (1998). Violation of Bell Inequalities by Photons More Than 10 km Apart. *Physical Review Letters*, 81(17):3563–3566.

Van Rijsbergen, C. J. (2004). *The geometry of information retrieval*. Cambridge University Press.

Vedral, V. (2017). Foundations of Quantum Discord. In Felipe Fernandes Fanchini, Diogo de Oliveira Soares Pinto, and Gerardo Adesso, editors, *Lectures on General Quantum Correlations and their Applications*, pages 3–7. Springer, Cham.

Von Neumann, J. (1996). *Mathematical foundations of quantum mechanics*. Princeton University Press.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., and Pothos, E. M. (2013). The Potential of Using Quantum Theory to Build Models of Cognition. *Topics in Cognitive Sciences*, 5(4):672–688.

Wester, L. (2017). Almost Equivalent Paradigms of Contextuality.

Wheeler, N. (2012). Generalized Quantum Measurement Imperfect meters and POVMs.

Widdows, D. (2004a). *Geometry and Meaning*. CSLI Publications, Center for the Study of Language and Information.

Widdows, D. (2004b). *Geometry and meaning*. CSLI Publications, University of Chicago Press.

Wootters, W. K. (1987). A Wigner-function formulation of finite-state quantum mechanics. *Annals of Physics*, 176(1):1–21.

Wootters, W. K. (2004). Picturing qubits in phase space. *IBM Journal of Research and Development*, 48(1):99–110.

Yearsley, J. M. and Pothos, E. M. (2013). Challenging the classical notion of time in cognition: a quantum perspective. In *Proc. R. Soc. B*, volume 281, page 20133056.

Yukalov, V. I. and Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision*, 70(3):283–328.

Zuccon, G., Azzopardi, L. A., and Van Rijsbergen, C. J. (2009). The Quantum Probability Ranking Principle for Information Retrieval. In *Advances in Information Retrieval Theory: Second International Conference on the Theory of Information Retrieval, ICTIR 2009 Cambridge, UK, September 10-12, 2009 Proceedings*, pages 232—-240. Springer Berlin Heidelberg, Berlin, Heidelberg.