REDESIGNING QUANTUM THEORY

A Dissertation Presented
by
MATTHEW B. WEISS

Submitted to the Office of Graduate Studies,
University of Massachusetts Boston,
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2026

Computational Sciences Program

© 2026 by Matthew B. Weiss
All rights reserved

REDESIGNING QUANTUM THEORY

A Dissertation Presented
by
MATTHEW B. WEISS

Approved as to style and content by:

Christopher A. Fuchs, Distinguished Professor
Chairperson of Committee

Olga Goulko, Assistant Professor
Member

Akira Sone, Assistant Professor
Member

Kourosh Zarringhalam, Professor
Member

Marcus Appleby, PhD
University of Sydney
Member

Daniel Pomerleano, Program Director
Computational Sciences Program

Rahul Kulkarni, Chair
Department of Physics

ABSTRACT

REDESIGNING QUANTUM THEORY

August 2026

Matthew B. Weiss, B.A., Brown University
M.S., University of Iowa
Ph.D., University of Massachusetts Boston

Directed by Professor Christopher A. Fuchs

QBism understands quantum mechanics to be probability theory supplemented by additional nonclassical coherence conditions. In this dissertation, we develop these nonclassical coherence conditions from first principles, emphasizing the role of a well chosen reference measurement. After treating standard probability on subjective Bayesian lines, we demonstrate an equivalence between the QBist approach and the existing framework of generalized probabilistic theories. We show that the fundamental nonclassical coherence relation may almost always be taken to be a gentle modification of the law of total probability, and give a coherentist account of when an experimental scenario has a classical explanation. Finally, we show that when the reference measurement is chosen to correspond to a complex projective 3-design, the shape of quantum state space is implicit in the probabilities which characterize the reference measurement itself. Thus coherence with this single reference measurement, properly understood, implies coherence with all of finite dimensional quantum mechanics. We then attempt a modest reconstruction of quantum theory along these lines, one practical consequence of which is a method for self-testing complex projective t-designs for $t \geq 3$ in a theory-agnostic way.

ACKNOWLEDGMENTS

None of this work would have been possible without the extraordinary generosity of Chris Fuchs. Studying with him has been the opportunity of a lifetime, and without his inexhaustible support, I would still be out in the wilderness. Five years ago, he welcomed me into the QBist family, and ever since I've had the benefit of not only his wisdom, but also the strength of the community he has built and so carefully and lovingly maintained over decades. I learned so much of the technical spirit of QBism from conversations with the indispensable John DeBrota and Blake Stacey, on whose shoulders I gratefully stand. From the first day of classes in September 2021, Sachin Gupta has been my most reliable friend and collaborator: so many of our ideas could only take form during those long hours at the whiteboard in our little office. Battling through proofs with Gianluca Cuffaro was an honor, and I'll never forget our time soaking in a hot tub on the roof of a hotel in Anaheim speculating wildly about quantum gravity, or wandering around the labyrinth of MIT with Simone Cepollaro. David Llamas will always have my respect for his commitment to the bit, and for the sensitivity of his bullshit detector. I miss the days when Ghi Coulter-de Wit was a part of the group: we couldn't have made it through that first APS March Meeting without them. I am grateful to Jacques Pienaar for bringing the magic of phenomenology to QBism and for his ability to keep the grants flowing: I picture us walking around Lake Växjösjön with the moon overhead. Any time spent with Marcus Appleby is always a genuine pleasure: his energy and curiosity is infectious—not only did he tirelessly induct us all at UMass Boston into the mysteries of algebraic number theory, but the force of his philosophical conviction is an ongoing inspiration. It was my priviledge as well to spend time with Rüdiger Schack, who makes rigorous Bayesianism look easy.

Thanks to Alex Moll, for having quantum office hours that first semester; to Maxim Olshanii for introducing me to the joys of higher dimensional polytopes; Akira Sone for introducing me to the power of stochastic differential equations; Steve Arnason for teaching me how real physics is done; Kourosh Zarringhalam for making sure my linear algebra was up to snuff; to Rahul Kulkarni for his genuine support and interest. I learned so much from the irreplacable Olgo Goulko, and I miss our trips on the Red Line together.

My time at UMass Boston would not have been the same without the kindness of Mary Fries and her one-of-kind QBist art. Thanks to Arjun Dhoot for so many wonderful and exciting conversations; to Jack Davis for making Hanover feel like home; to Gino Elia, for keeping us all philosophically honest; to Austin Monaghan, for bravely continuing the work we started; to Chris Sutton, for his absolute genuineness; to Sai Morapakula, for the flowers; to Money Chanalia, I hope we keep running into each other; to Vajra Badha, for keeping our little club together; to Natalia Freitas, for never giving up.

I'd be remiss not to thank Toby, for his teaching me the old ways. I have so much gratitude for Tim and Sally and Julie and Scotty, for remaining in my life all these years. For Kiesha, for understanding what I was trying to do. For Bubbe, for our weekly chats, full

of counsel and solicitude. To MMJ, for our emails. In a profound way, everything I have done here has its origin in conversations with Jeremy Silver, when we were just two child philosophers, making the circuit around the Hebrew School playground. For Julian and Anna, for Daniel, for Bilan, Liban, Iman, Bashir, Abo and Hooyo.

Thank you to my Mom and Dad, who have always believed in the "timespace interchange," even when it was most difficult. In a real sense, my Dad taught me how to be a scientist in the most loving way. My mom has always been my most devoted reader, and I don't feel like I understand anything until I can explain it to her. One day, we'll figure out what it all means.

And finally, this dissertation would have been impossible without the support of Ladan, my wife and partner in life. She has truly changed my world for the sweeter. The deepest insights in this work I had beside her, and I am so grateful to have a soulmate on this quantum adventure.

This research was supported in part by the National Science Foundation through grants NSF-2210495 and OSI-2328774 as well as through Grant 62424 from the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Disclaimer: I must in good faith disclose that several ideas in this dissertation arose in conversation with ChatGPT versions 5.1 to 5.5. ChatGPT was used to check certain proofs, as well as to automate certain algebraic manipulations, but played no role in the writing of the text. I take full responsibility for the content herein.

TABLE OF CONTENTS

ACKNOWLEDGMENTS v

CHAPTER Page

1. INTRODUCTION 1
1.1 Foundations 1
1.2 Context 4
1.2.1 Generalized probabilistic theories 4
1.2.2 QBism 7
1.3 Chapter summaries 11

2. PROBABILITIES 18
2.1 Introduction 18
2.2 de Finetti style derivation of probability theory 18
2.3 Reflection 21
2.4 Sample spaces and all that 23
2.4.1 Measures 24
2.5 Random variables 25
2.5.1 The algebra of random variables 27
2.6 From composites to the de Finetti representation 28
2.7 Calling the whole thing off 30
2.7.1 Coulda, woulda, shoulda 32
2.8 Pulling out the rug: nonclassicality 34
2.8.1 Hidden variables 34
2.8.2 Classical reference measurements 35
2.8.3 System, reference, state, effect 36
2.8.4 State space 38
2.8.5 Effect space 39
2.8.6 Nonclassical consistency 41
2.8.7 Φ-duality 46
2.9 “Dynamics” 47
2.10 Decomposition 49
2.11 Conclusion 52

3. THE BORN IDENTITY 54
3.1 Introduction 54
3.2 The Born matrix 54
3.3 Subspaces from coherence 57
3.3.1 A word on observables 58

CHAPTER Page

3.4 Rank factorization 59
3.5 Generalized probabilistic theories 63
3.5.1 Examples 66
3.6 Frame theory 67
3.6.1 Dual frames 70
3.6.2 Frame coordinates and dual coordinates 71
3.6.3 Double sided frames 72
3.7 Informational completeness 74
3.7.1 Separation 74
3.8 de Finetti for GPT's 77
3.9 Conclusion 78

4. DIE URGLEICHUNGEN 79
4.1 Introduction 79
4.2 Bloch form 80
4.3 Ubiquitous Urgleichung 82
4.4 Morphophoricity 87
4.4.1 An example from quantum theory 90
4.4.2 Tight IC-POVMs 91
4.5 The LTP Deformation 94
4.5.1 A choice of measures 100
4.6 Conclusion 102
4.7 Appendix: Real Vector Space QM 103

5. (NO) RETURN TO CLASSICALITY 106
5.1 Introduction 106
5.2 Noncontextuality 106
5.2.1 An example 111
5.2.2 Simplex embedding 113
5.2.3 QBist noncontextuality 114
5.3 Computational interlude 117
5.3.1 Alternating projections 117
5.3.2 A linear program for simplex embedding 118
5.4 Stochastic Φ 120
5.4.1 An example 123
5.5 Negativity in quantum theory 124
5.6 QBist Bell Inequalities 125
5.6.1 Local realism? 126
5.6.2 CHSH 127

CHAPTER Page

5.6.3 Bounding $\|I - \Phi\|$ 129
5.6.4 Noncontextuality inequalities 132
5.7 Conclusion 136

6. CHARACTERIZING QUANTUM STATE SPACE WITH A SINGLE QUANTUM MEASUREMENT 137
6.1 Introduction 137
6.2 The reference measurement formalism 139
6.3 Making designs 141
6.4 The shape of quantum state space 143
6.4.1 Bounding agreement 143
6.4.2 The contour of idempotents 146
6.4.3 A variance-based uncertainty principle 148
6.5 The Jordan product 150
6.6 A gentle modification 152
6.7 Conclusion 153
6.8 Appendix: The first three moments 155
6.9 Appendix: The associator and the von Neumann equation 158

7. RECONSTRUCTION 164
7.1 Introduction 164
7.2 The Urgleichung 164
7.3 A nonclassical product 166
7.4 Moment matching 172
7.4.1 Polarization 173
7.4.2 Euclidean Jordan algebras 175
7.5 Lifting the product 177
7.6 Self-duality 182
7.6.1 The uncertainty principle 183
7.6.2 The self-duality constant 185
7.7 Idempotents 186
7.7.1 From valuations to measurements 189
7.7.2 Characterizing idempotents 190
7.7.3 Orthogonality 193
7.7.4 Testing the rank of an EJA 195
7.8 A simple EJA 195
7.9 Full circle 197
7.10 Self-testing 199

CHAPTER Page

7.11 Assume a SIC? 201
7.12 Conclusion 203

LIST OF REFERENCES 211

CHAPTER 1
INTRODUCTION

1.1 Foundations

This dissertation is primarily an exercise in quantum foundations, the ongoing attempt to give a conceptually satisfying account of the lesson that quantum mechanics is trying to teach us about the world and ourselves. It is unapologetically QBist [1, 2]. The QBist school of thought holds that quantum mechanics is best understood as a guidebook for an agent trying to gamble on the consequences of their free actions on a world which is just as free to act back. In this way the agent and the world they act upon are locked in a permeable embrace of mutual cocreation. Already subjective Bayesianism understands probability theory to be of a normative character, a way of organizing one's own beliefs to ensure they are as a whole coherent with each other. QBism views quantum theory in exactly the same way, as a tool any agent may use to tend to their mesh of beliefs about the consequences of their own actions precisely in light of nature's vitality. In particular, the key formal move of the QBist approach is to focus attention on *reference measurements*: such measurements have the property that assigning probabilities to their outcomes is mathematically equivalent to assigning a density matrix, the more familiar guise under which a quantum state appears. Quantum states, effects, unitary evolution: they may all be reexpressed in terms of relations between probability assignments. In this way, QBists do not understand quantum theory by analogy with probability theory: they view it as nothing other than probability theory, properly supplemented with new norms which respect nature's lack of "hidden variables."

In particular, a key part of respecting this absence is the insistence that the agent ought to use quantum mechanics to organize their beliefs about the consequences *for them* of *their own* actions. To gamble on the consequences for another is as much to deny the openness of the world to mutuality: to treat the consequence for another as settled is to treat it as a hidden variable, an already existing property ready to be trotted out. This personalization leads to certain familiar notions, for consistency's sake, to be recast in at first unusual terms: for example, dynamics is understood in terms of an agent's indifference to whether or not they themselves might perform a particular measurement [3]. Moreover, just as probability theory is not "about" coin flips and dice rolls, but rather helps one organize one's beliefs about coins and dice, so too quantum theory in itself is not "about" electrons, hydrogen atoms, or spin glasses; instead it helps the agent organize their beliefs about such phenomena. The beliefs themselves the agent forms in their intercourse with the world, guided by the consequences their world confronts them with. And whereas these consequences are personal to the agent,

any agent is well advised to use quantum mechanics: its universality attests to the common existential situation we all find ourselves in. But what exactly this situation is, of course, remains still up for debate.

Like others in the quantum foundations community who pursue "reconstructions" of quantum mechanics [4], QBists do not wish to take the quantum formalism for granted. We continue to seek a satisfying principled derivation of the nonclassical coherence rules which quantum theory first brought to our attention. In the first part of this dissertation, after developing the usual formalism of probability theory in subjective Bayesian style, I adapt existing arguments in the QBist literature to derive the form of the fundamental nonclassical coherence rule, which in quantum theory corresponds to the Born rule, but which is in reality much more general. The following chapter explores in more detail the mathematical properties of this coherence rule, in particular the nature of the *Born matrix*, which plays a starring role. Ultimately it demonstrates a correspondence between the QBist formalism and the formalism of *generalized probabilistic theories* heavily investigated in the quantum foundations literature [4]. The next chapter studies under what circumstances the fundamental nonclassical rule may take an extraordinarily simple form. Indeed, the general QBist strategy in pursuing foundational questions has been to find those circumstances in which the extra normative rules suggested by quantum mechanics may be made to look as close as possible to the familiar rules of standard probability theory. The gap between the two then witnesses the essential difference between classical and nonclassical ways of approaching the world. For this reason, symmetric informationally complete (SIC) reference measurements in quantum mechanics have been an abiding interest for QBists: such measurements have a host of truly remarkable probabilities, not least of which is that they allow one to rewrite the Born rule $P(E|\rho) = \mathrm{tr}(E\rho)$ as

$$P(E|\rho) = \sum_i P(E|R_i) \left\{ (d+1) P(R_i|\rho) - \frac{d}{n} \right\}, \tag{1.1}$$

which is as gentle a modification of the law of total probability $\sum_i P(E|R_i)P(R_i|\rho)$ as one could imagine, and which QBists have termed the *Urgleichung*, or fundamental equation. I show that in fact this type of rule is ubiquitous in the landscape of generalized probabilistic theories. Thus the form of this rule is not peculiar to quantum theory, and may even be appealed to in situations where classical explanations are available. But then: what do we even mean by classical? We then turn to the question of when an agent's mesh of beliefs is compatible with a classical explanation in terms of underlying properties which condition the results of measurement, giving a QBist spin on Spekkens's notion of a noncontextual ontological model, grounding it in an agent's indifference to whether or not a reference measurement is performed. Moreover, we recast Bell inequalities and noncontextuality inequalities more generally as providing a lower bound on how close the fundamental nonclassical coherence rule is to the law of total probability. For QBists, violations of these inequalities quantify the cost of the failure to adopt the nonclassical coherence rule itself.

The final two chapters, however, form the heart and soul of the dissertation. For QBists, the question of *why quantum theory* amounts in some sense to the question: why do states, in the form of probability distributions on reference outcomes, form just *that* particular subset of the probability simplex? For indeed, even though every quantum state corresponds to a probability distribution on the outcomes of a reference measurement, not every conceivable distribution corresponds to a quantum state. Instead, the set of allowed distributions has a very special geometry. Previous attempts at QBist reconstructions of quantum theory have focused their attention on generalizing the image of quantum state space within the probability simplex induced by a SIC-POVM measurement [5]. Such generalizations are called qplexes, and the goal has been to identify a principle by which the qplex corresponding to quantum theory may be picked out among all others. Similar ideas have been applied more generally to measurements corresponding to complex-projective 2-designs, of which SICs are but the simplest example [6].

Our central innovation is to adopt as a target for generalization not 2-designs, but 3-designs, of contemporary interest for their use in classical shadow estimation procedures [7, 8, 9, 10]. We show that (unbiased) 3-design reference measurements allow the shape of quantum state space to be characterized in a strikingly simple way, first in terms of constraints which govern pure state probability distributions, and then more generally in terms of an uncertainty principle, a lower bound on the variance of any of a natural class of observables. In particular, the entire geometry of the state space is encoded in $P(R|R)$, the conditional probability matrix which characterizes the reference measurement itself. In other words, the probabilities which characterize a 3-design reference measurement characterize the entire theory. The reason is algebraic: the properties of quantum 3-designs imply that the structure coefficients of the Jordan algebra of observables may be extracted from $P(R|R)$ alone. In the end, we find that just as 2-designs make the Born rule look as close as possible to the classical law of total probability, 3-designs make the quantum Jordan product on observables look as close as possible to the classical Jordan product.

In the final chapter, we attempt our own modest reconstruction of finite-dimensional quantum theory. We begin from scratch, with the matrix $P(R|R)$, placing increasingly severe restrictions on its structure, guided by the idea of introducing the gentlest possible modification of the classical rule for multiplying valuations on reference outcomes. In doing so, we retrace the steps in the previous chapter from first principles, motivating a series of assumptions which ultimately imply that the entries of the matrix $P(R|R)$ may be realized as inner products between $d \times d$ Hermitian matrices over $\mathbb{C}$ which constitute a complex projective 3-design. The key step is a "moment-matching" assumption. Any valuation on an arbitrary measurement is equivalent to some valuation on the reference measurement: we assume that for some class of measurements, the mth moment with respect to the original measurement and the mth moment with respect to the reference, calculated according to the nonclassical product, coincide. This leads directly to the terrain of Euclidean Jordan algebras. While our motivations are foundational, the same considerations yield a practical algorithm for self-testing quantum t-designs for $t \geq 3$. We show that relations between the probabilities

$P(R_i|R_j)$ certify that a Hilbert space representation is possible in a way that other measures of "designness" do not: for example, measuring the so-called frame potential and showing that it achieves its minimimum is not sufficient without a guarantee that a Hilbert space exists to begin with. Finally, while we achieve our goal of returning full circle to quantum theory, many deep questions remain regarding both the motivations for our assumptions as well as the geometry they imply, leaving the door open for a great deal of productive work in the future.

1.2 Context

1.2.1 Generalized probabilistic theories

The last twenty-five years have been a golden age for quantum foundations. The turn of the millennium marked a growing consensus that it ought to be possible to rederive or *reconstruct* quantum mechanics from a few sharp, simple principles, ideally of an information-theoretic character [11, 12, 13, 14]. Today a wide variety of reaxiomatizations of quantum theory are on offer: [15, 16, 17, 18, 19, 20, 21, 22] give just a taste. This development coincided with the maturation of quantum information theory and the theory of quantum computation. Before this, quantum mechanics had largely been framed in terms of what the theory did *not* allow one to do, e.g., measure position and momentum with arbitrary accuracy at the same time, or from complete information about a whole perfectly predict the behavior of the parts. The shift to the study of quantum communication protocols, quantum cryptography, and quantum algorithms was a shift in emphasis, to what one *can* do with quantum theory, indeed, what would be impossible or difficult to do classically [23]. There is a long prehistory to these ideas, stretching back into the 20th century, but by the end of the 1990s, due to the indefatigable work of a small band of quantum enthusiasts, it is fair to say that a new age of quantum mechanics dawned.

With it, came new ways of thinking about the quantum mechanical formalism as it was handed down by the early pioneers. Whereas Einstein's postulates for special relativity, or the laws of thermodynamics, appeal to physical principles from which the mathematics flows [24], quantum mechanics is distinctive in that its axioms, as laid down by von Neumann among others, have an almost entirely mathematical flavor. This, combined with the fact that quantum mechanical predictions are often counterintuitive to the classical mind, has led over the decades to a profusion of interpretations of the theory, from pilot waves [25] to many worlds [26]. For most of these interpretations, the strategy is to rewrite the equations in some perspicacious way and then attempt to assign ontological significance to these mathematical expressions. From the beginning, however, another paradigm has been pursued, which recognizes that the comparison between classical and quantum mechanics can only take one so far. This paradigm instead tries to *generalize* quantum mechanics, to place quantum mechanics in a landscape of possible theories, and by comparison throw its distinctiveness into new light. Already in the 1930's, Wigner, Jordan, and von Neumann generalized the algebraic structure of quantum mechanical observables in hopes of understanding how essential matrix

representations are to the theory, in the process classifying the so-called Euclidean Jordan algebras [27, 28], of which we will have much more to say. von Neumann and Birkhoff went on to pursue generalizations of Boolean logic under the heading of "quantum logic" [29]. Later Mackey, Davies, Ludwig, and others began instead from the *convex structure* of quantum theory [30, 31, 32, 33, 34], a theme that would be taken up in the 21st century under the heading of *generalized probabilistic theories*.

The study of *generalized probabilistic theories* [35, 36, 37, 38, 4, 39], or GPTs, starts from an observation about what classical probability theory and quantum theory have in common: the state space is a convex subset of a vector space, and the effect space is a convex set of nonnegative linear functionals on that subset. It is easier to understand what this means with an example. The most familiar GPT is classical probability theory as it is used to formulate a theory of measurement, that is, a theory which attempts to understand observed behavior in terms of the unobserved. Given a set of mutually exclusive and exhaustive events $\{\lambda_i\}$, the familiar law of total probability tells us that we may write the probability of the outcome E of a measurement as

$$P(E) = \sum_i P(E|\lambda_i)P(\lambda_i). \tag{1.2}$$

Gathering up the probabilities $P(\lambda_i)$ into a column vector $|\sigma)$, and gathering up the conditional probabilities $P(E|\lambda_i)$ into a row vector $(E|$, we can write the outcome probability as a linear functional acting on a vector $P(E) = (E|\sigma)$. In the language of GPTs, we would call $|\sigma)$ a *state* and $(E|$ an *effect*. Here the state has an interpretation as a state of uncertainty about which event λ_i occurred, or in language borrowed from quantum foundations, which value a "hidden variable" takes. Another way it is sometimes put is that in classical theory, an observed outcome is always grounded in some underlying properties of a system which preexist the measurement. But not all GPTs are classical! Whereas the states of classical probability theory form a simplex, in general the states of a GPT may form an essentially arbitrary convex subset of a vector space, while the effects are a set of linear functionals which give valid probabilities when they act on states.

Indeed, it turns out quantum mechanics fits exactly this pattern. The Born rule tells us that the probability of an outcome E can be written

$$P(E) = \mathrm{tr}(E\sigma), \tag{1.3}$$

where on the right hand side of the equation, E is a positive semidefinite matrix representing an effect, and σ is a positive semidefinite matrix with trace 1, representing a state. Don't let the fact that these are matrices fool you: one may equivalently think of σ as a vector, and $\mathrm{tr}(E\cdot)$ as a linear functional, just as in classical probability theory. In this sense, for both classical probability theory and quantum mechanics, the structure for calculating probabilities

is the same[1]. The convex sets involved, however, are quite different: for example, the space of states of the simplest quantum system, the qubit, may be identified with the surface and interior of a *sphere*[2] in three dimensions.

Thus far we've encountered a simplex state space, and a spherical state space. But why not any other essentially arbitrary convex shape? For instance, one could develop an analogous theory e.g., where the state space is a square, an icosahedron, or an ice-cream cone, as long as the shape is convex. The corresponding effect space must consist of linear functionals that give valid probabilities on those states, and one may proceed to study the properties of the resulting, generally nonclassical, theory—what it has in common with classical or quantum theory, and how it differs in perhaps surprising ways [40, 41, 42, 43, 44]. In this way, GPTs offer a minimal, but expressive framework for exploring possible theories. Atop this loose structure, one may introduce some well motivated postulates governing the behavior of the theory [45, 46, 47, 48, 49], and study how these postulates constrain the geometry of the state and effect spaces. Alternatively, one may start with a geometry [50] and study the resulting behavior of the theory. In particular, over the years, GPT researchers have constructed many "toy" theories which challenge naive notions about what is the "essence" of quantum mechanics.

For example, it is by now well known that both experiment and quantum mechanics violate Bell inequalities, and this is evidence against the applicability of so-called local realist models in physics [51]. In the classic CHSH scenario as usually described [52], two parties, traditionally called Alice and Bob, are distributed one qubit each (which might be realized as the spins of two electrons, or the polarizations of two photons) which are entangled in the "singlet" state. After being separated sufficiently that no signal could pass between them in the course of the experiment, Alice chooses to perform measurement A_1 or A_2 and Bob chooses to perform measurement B_1 or B_2. The correlations between the outcomes of these measurements according to quantum theory are in general inconsistent with the idea that the two qubits have some prearranged way of responding to measurements. In other words, if one rejects any kind of "spooky action at a distance," one must reject the supposition that the outcomes of measurements reveal preexisting properties of the qubits. The CHSH inequality itself is a constraint on probability assignments compatible with classical assumptions, and quantum mechanics violates this inequality. But one could ask: is this the maximum conceivable violation in *any* theory, compatible with locality? The answer, perhaps surprisingly, turns out to be no. A GPT affectionately known as *boxworld* achieves the maximal violation compatible with the assumption that Alice and Bob's systems cannot signal each other: it consists in a composite GPT where the individual state spaces are not a simplex, nor a sphere, but instead a simple square [36, 4].

In the same vein, in light of entanglement, quantum theory is often described as *holistic*. As Schrödinger put it, one may find that the "best possible knowledge of a whole does not

[1] Although only in the first case can one interpret this structure as the "law of total probability."

[2] The situation is much more complicated in higher dimensions.

include best possible knowledge of its parts—and that is what keeps coming back to haunt us" [53]. One may ask, however: are there theories which are strictly *more holistic* than quantum theory? Again, the answer is yes [54]. It turns out that, entanglement notwithstanding, the quantum state of a composite can be reconstructed solely from separate measurements on its parts, including their correlations (and assuming the same experiment can be run many times). This property is called *local tomography*. One may consider a variant of quantum theory which is defined over the real numbers instead of the complex numbers[3]. The simplest such system, sometimes called a *rebit*, has a state space corresponding to a disk. It turns out such a theory is *bilocally tomographic*: to reconstruct the state of a composite one must perform measurements not only on the parts, but on *pairs* of parts. One may further generalize this idea to construct theories which are n-locally tomographic [55], all of which are in this sense strictly more holistic than standard quantum theory.

These examples challenge qualitative assessments of what quantum mechanics is really about, whether by constructing classical models that reproduce certain aspects of the theory, or by proposing nonclassical models which exemplify yet more dramatically some notion taken to be properly quantum. The study of GPTs allows one to see "around" the givenness of the quantum formalism, and the hope in quantum foundations is that these new vantage points will bring into view ever more incisive ways of clarifying *why quantum theory* among the widest range of potential theories.

1.2.2 QBism

QBism [2, 1] is a research program in the foundations of quantum mechanics which begins by taking a stand on the nature of probabilities. The reader may be familiar with the fact that the history of probability theory is largely the story of the uneasy coexistence of two camps: on the one hand, the frequentists who hold that probabilities represent limiting frequencies (the proportion of times an event would occur given an infinite number of trials), and on the other hand, the Bayesians who hold that probabilities represent degrees of belief [56]. QBism takes a *subjective Bayesian* view of probabilities, following the great Italian probabilist Bruno de Finetti [57, 58]. Remarkably, de Finetti was able to *derive* the laws of probability[4] by defining the probability $P(E)$ as the price at which a gambler would be willing to buy or sell a ticket which pays \$1 if E should occur, and demanding that no sequence of buying and selling tickets leads to a sure loss. In light of this, the laws of probability may be understood entirely in terms of ensuring a gambler's behavior is consistent: the rules protect the gambler from falling into a situation where they are *guaranteed* to lose money, regardless of what events actually occur. In this sense, we may say that the laws of probability are *normative ideals*: they are prescriptions for how one *should* behave, supposing one has a stake in success. For example, the law of total probability $P(A) = \sum_i P(A|B_i)P(B_i)$, where

[3]This provides, in fact, an example of a Euclidean Jordan algebra.

[4]That probabilities should lie between 0 and 1; that $P(A \text{ or } B) = P(A) + P(B)$ for mutually exclusive events; that $P(A \text{ and } B) = P(A|B)P(B)$, and so on.

$\{B_i\}$ is a set of mutually exclusive and exhaustive events, is a *normative criterion* by which one may check consistency between the probability assignments which are implicated in it. In short, for followers of de Finetti, probability theory is *advice* by which an agent may regulate their behavior.

QBism understands quantum theory in precisely the same way. Quantum theory should not be regarded as a *description of physical reality*, but rather as a set of consistency constraints on one's probability assignments in light of nature's vitality [24] as witnessed by the failure of classical models. This view is made the more convincing by rewriting the theory with respect to a *reference measurement*. But what is a reference measurement? Suppose one is confronted with a cloud of particles. Classically, this cloud may be characterized by the positions and momenta of all the particles that make it up. If we consider the outcome E of some measurement on the system, the law of total probability would urge on us the norm

$$P(E) = \sum_{\{x,p\}} P(E|\{x,p\})P(\{x,p\}), \tag{1.4}$$

which relates $P(E)$ to one's conditional probabilities for E given a particular configuration of positions and momenta as well as one's probability distribution over these positions and momenta themselves. What justifies this? First, classically, any other property of the cloud of particles supervenes on, or is grounded in, the positions and momenta. In this sense, a measurement of the positions and momenta serves as a reference measurement. At the same time, classically, it is a matter of indifference whether one actually performs this reference measurement or not. One may appeal to the law of total probability either way, since classically, we suppose that each particle "has" a position and "is" moving in a particular direction at a particular speed: the reference measurement simply reveals what is the case. But quantum mechanics forces us to question this assumption, that we may describe objects as having definite properties independent of measurement.

Indeed, nonclassicality of the form familiar from quantum theory arises by dropping the assumption that a system may be characterized by definite properties independent of measurement, while retaining the assumption that reference measurements exist. To dramatize this, we may then compare two scenarios. In the first, a reference measurement with outcomes $\{R_i\}$ is performed on a system after which an arbitrary measurement with an outcome E is made. By the law of total probability,

$$P_1(E) = \sum_i P(E|R_i)P(R_i). \tag{1.5}$$

In the second scenario, the measurement with outcome E is made directly. Here the reference outcomes $\{R_i\}$ remain hypothetical. We will see, however, that assuming $\{R_i\}$ is a reference measurement implies that we may nevertheless appeal to a new nonclassical norm, which is

in fact a *deformation* of the law of total probability,

$$P_2(E) = \sum_{ij} P(E|R_i)\Phi_{ij}P(R_j). \tag{1.6}$$

There is no inconsistency since the law of total probability gives us no guidance should the intermediate events turn out to *not happen at all*! Nevertheless, the second equation provides us a new consistency constraint to compensate for the loss. In fact this equation is nothing other than the Born rule, if you like, in disguise.

In short, QBism views quantum theory as a kind of "user's manual" [1] for anyone interested in gambling on a world which we may say, in light of the failure of classical models, is undergoing ceaseless creation. From this point of view, QBism does not give up the ideal of a rational agent: rather, nonclassicality brings with it extra coherence relations which a gambler ought to impose on their probability assignments. In a striking inversion, endowing the world with a fundamental openness translates into more restrictive constraints on the agent who wants to gamble well upon it. In fact, the same interpretation can be applied to any nonclassical GPT, although the particular constraints will differ: for the QBist, any GPT is just a convenient way for an agent to take stock of their probability assignments and evaluate them for consistency.

Philosophically, then, QBism takes the freedom of the interested agent as fundamental: upon acting on their world, the agent gets a free response as the world pushes back. This freedom, this vitality, this creative capacity, are the message of nonclassicality. Going still further, for QBists, the free response of nature is taken to be an *experience* of the agent [2], although this should not be understood too narrowly: indeed, the mutual cocreation of agent and world suggests that the traditional boundaries between subject and object must flex and bend. By taking the experience of the agent seriously, by positing that quantum theory is in fact a first person theory[5], QBism is able to resolve to some degree the interpretative difficulties e.g., raised by entanglement. In the CHSH scenario, rather than viewing Alice and Bob from a third person "objective" point of view from which vantage the correlations between the results of their measurements imply some kind of "spooky *context-dependence* at a distance," QBism urges one for example, to take the point of view of some third person, Charlie, who will perform a measurement on Alice, eliciting her result, and a measurement on Bob, eliciting his result, and who wants to gamble on the outcomes. The correlations in a Bell experiment are precisely correlations for Charlie, for whom the results of measurement are created, locally, on the spot.

More broadly, for QBists, nonclassicality urges us to grapple with a notion of experience which is not so much localized in someone's brain, but which is instead found at the threshold as self and other freely negotiate a boundary. This idea has led QBists to associate with phenomenologists and enactivists for whom experience and action are not separate [59, 60]. What is real is on the surface: the world presents itself to us as invitations to act; cognition

[5] As is any user's manual.

itself is embodied, embedded into an environment that cocreates it, enacted not only through neurons firing, but also through everything an organism does, so that we may regard mind itself to be extended out into what was traditionally called the world [61]. This too is part of the message of nonclassicality, the failure of classical models to adequately provide good advice.

As we have observed, from a technical point of view, the QBist reconstruction project begins from the observation that distributions on reference outcomes ought to be restricted in some way. In GPT language, the valid reference distributions ought to live in the special subset of the probability simplex which corresponds to the image of the GPT state space induced by the choice of reference measurement. Wandering out of this subset will generally yield negative probabilities. The question for QBists has therefore been: how can one motivate the particular shape of quantum state space within the probability simplex on the outcomes of a well-chosen reference measurement? As we noted earlier, this led to the introduction of the notion of a *qplex*, a generalization of the image of quantum state space within the simplex induced by a symmetric informationally complete (SIC) measurement. Just as the landscape of GPTs allows one to place quantum theory in a broader context, so too the landscape of qplexes allows the particular features of quantum theory to be brought to the fore. [5] showed how one can pick out the qplex corresponding to quantum theory on grounds of symmetry. In this dissertation, however, we take a different approach, instead generalizing the image of quantum state space induced by a so-called complex projective 3-design reference measurement [62].

Complex projective or quantum state t-designs are ensembles of states whose tth moment reproduces the tth moment of quantum state space as a whole, and are widely used in quantum information theory. In particular, the structure of 3-designs is intimately tied to the Jordan algebraic structure of quantum theory [63], and so we will make much use of the pioneering work of Jordan, von Neumann and Wigner in formalizing and classifying the Euclidean Jordan algebras [22, 28]. Jordan began by noting that quantum observables form a commutative, but not associative algebra under the symmetric product $A \odot B = (AB + BA)/2$. Generalizing the essential properties of this product, he gave birth to the notion of a Jordan algebra, and already by 1934, von Neumann, Wigner, and Jordan demonstrated that any so-called Euclidean Jordan algebra must be a composite of simple Jordan algebras, corresponding essentially to quantum mechanics over different number fields. Later the work of Koecher and Vinberg [64] showed that such Jordan algebras are in correspondence with *symmetric cones*, giving them an entirely geometric interpretation [27]. Work on Jordan algebras continues to this day: they are of interest to pure mathematicians, as well as those working in convex optimization [65], not to mention those working in quantum foundations [22, 66]. In particular, we show that for QBists, the import of Euclidean Jordan algebras lies in the interplay between valuations on the outcomes of arbitrary measurements and their equivalent valuations on reference outcomes. If one minimally deforms the classical rule for multiplying valuations in a way analogous to how one deforms the classical law of total probability into the Born rule, then demanding that there exists a class of measurements for which the mth

moment of a reference valuation with respect to the nonclassical product reproduces the mth moment of the original valuation, one is led almost inevitably to a Euclidean Jordan algebra. By giving a QBist motivation for the Jordan algebraic structure of quantum theory, we hope to open up a new chapter in the QBist reconstruction project.

1.3 Chapter summaries

With this background established, we offer detailed chapter summaries to give a sense of the general flow of argument in the dissertation.

Chapter 2: Probabilities. This chapter lays the conceptual and formal groundwork for everything that follows. We begin in subjective Bayesian style by presuming that there is an agent who is trying to make better decisions. In accordance with the philosophy of QBism, we regard the agent as free to act on a world beyond them while the world has the freedom to return back consequences beyond the agent's control. We first derive the familiar laws of probability theory in the style of de Finetti [67], interpreting probabilities as bets, and demanding that no sequence of bets leads the agent to a sure loss. This demand implies that the gambler's bets must satisfy certain standards of consistency, which turn out to be just the rules of probability theory. We also discuss the use of van Fraassen's reflection principle to make sense of how an agent should think of their probability assignments across time [68]. We extend the analysis to random variables more generally [69, 70], and show that valuations on events form an algebra. We then discuss how to treat gambles on the outcomes of several experiments, and introduce the de Finetti representation theorem which gives subjective Bayesian meaning to the idea of "estimating unknown probabilities," e.g., inferring the probability that a coin will come up heads or tails from the results of flipping the coin many times. If a probability is the fair price the gambler assigns to a ticket, how can a probability be "unknown"? The theorem provides an answer: if the joint distribution on many coin flips is so-called exchangeable, it may be expressed in terms of a underlying distribution on the bias of the coin, and so the gambler may act *as if* there were an unknown probability they are trying to estimate. Next, we emphasize how the law of total probability $P(E) = \sum_i P(E|R_i)P(R_i)$ holds precisely because the events $\{R_i\}$ are mutually exclusive and exhaustive so that the gambler believes that exactly one of them will occur. What if, however, the events $\{R_i\}$ are called off? Is there any coherence condition which yet relates the probabilities $P(E)$, $P(E|R_i)$ and $P(R_i)$? In the special case that these events can be related to another set of mutually exclusive and exhaustive events $\{\lambda_i\}$ of a certain kind, then all is not lost. We show that we may then rewrite the law of total probability in terms of a matrix Φ, giving us our first taste of a formal structure which will be thoroughly investigated in this work. We then discuss the idea of a reference measurement, another central theme, at first grounding it in classical notions. We make the central nonclassical assumption: we drop the idea that a system is characterized by definite properties independent of measurement, but retain the idea of a reference measurement. We develop a series of definitions for system, state, and effect, grounded solely in the relations

between the probabilities that the gambler assigns to the outcomes of measurement, so that a notion of state and effect spaces with convex geometry emerges. We then come to the centerpiece of this chapter. We compare two scenarios. In the first, a measurement A is performed, followed by the reference measurement, followed by a measurement B; in the second, A is performed and then B directly. We prove that under very general assumptions, the probabilities the agent assigns in the two scenarios may be related by a nonclassical coherence condition: this is our core representation theorem. In the first scenario, the agent ought to use the familiar law of total probability $P(B|R, A) = P(B|R)P(R|A)$ whereas in the second they ought to use $P(B|A) = P(B|R)\Phi P(R|A)$, where we have employed here a matrix notation which we will frequently use throughout the text. In other words, the nonclassical coherence rule takes the form of a deformation of the law of total probability by the presence of the Born matrix Φ, where Φ satisfies $P(R|R)\Phi P(R|R) = P(R|R)$: here $P(R|R)$ is the conditional probability matrix which characterizes the reference measurement itself. We then briefly discuss how to incorporate dynamics into this formalism, in QBist spirit appealing to van Fraassen's reflection principle to argue that assigning a particular dynamics is equivalent to the agent's being indifferent to whether an intermediate measurement is performed. Finally we treat briefly the subject of when the gambler might be led to carve subsystems out of a whole.

Chapter 3: The Born Identity. Having established the fundamental nonclassical coherence condition in the previous chapter, we turn to discussing the mathematical properties of the Born matrix Φ, identifying it as a particular kind of generalized inverse of $P(R|R)$, namely, a $\{1\}$-inverse [71]. Indeed, previous accounts of the Born matrix focused their attention on the case that $P(R|R)$ is invertible; an important innovation of the present account is a systematic treatment in the case that $P(R|R)$ is singular. We show that the fundamental coherence condition in fact restricts states $P(R|\rho)$ to live in the column space of $P(R|R)$ and effects $P(E|R)$ to live in the row space of $P(R|R)$. We then introduce a central tool of analysis: the rank factorization of a matrix, which allows us to represent states and effects as r-dimensional vectors, where $r = \text{rank}(P(R|R))$. The $\{1\}$-inverse condition, that $P(R|R)\Phi P(R|R) = P(R|R)$ can be reinterpreted as a resolution of the identity $\mathbf{S}\Phi\mathbf{R} = I_r$, where the rows of $\mathbf{R}$ are compact representatives of the reference effects and the columns of $\mathbf{S}$ are compact representatives of the reference states. This representation makes clear that in fact our framework is essentially equivalent to the framework of generalized probabilistic theories [4]. Our derivation of the nonclassical coherence condition $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ amounts to a proof of a simplified version of Ludwig's embedding theorem [38], which ensures that any GPT has a vector space representation. We discuss the two core examples of GPTs: classical probability theory and quantum mechanics. We then develop the key ideas and constructions of frame theory, showing that if one starts with a GPT, then our reference measurement formalism arises as a "double-sided" frame representation. Indeed, we define more rigorously what a reference measurement is from the GPT point of view, introducing the notion of informational completeness and its ability to separate states or effects. We reproduce a classic proof that reference measurements whose effects form a basis,

that is, minimal informationally complete (MIC) measurements, exist in any unrestricted GPT (which allows all mathematically possible states and effects), and explain how the de Finetti representation theorem can be adapted to a wide class of GPTs[6] to make sense of the tomographic procedure of "estimating unknown states."

Chapter 4: Die Urgleichungen[7]. In this chapter, we discuss the conditions under which the Born matrix may be chosen to take a very simple form, $\Phi = \alpha I + (1-\alpha) w u^\dagger$ so that $P(E|\rho) = \sum_i P(E|R_i)\big\{\alpha P(R_i|\rho) + (1-\alpha)P(R_i|\mu)\big\}$, where $|\mu)$ is a distinguished state which plays the role of the state of complete uncertainty and J is the matrix of all 1's. We call it the *Protourgleichung*. We first show that any GPT can be brought into a standard "Bloch" form, introducing a block matrix notation which facilitates the proofs of this chapter. Our first theorem is that in any unrestricted GPT, given a reference measurement, one may always construct a set of reference states which allows Φ to take the abovementioned form: we conclude therefore the Protourgleichung is ubiquitous. An alternative characterization of such reference measurements may be given in terms of the corresponding channel operator: Φ may take Protourgleichung form iff the channel operator is depolarizing. In this case, the matrix $P(R|R)$ acts simply on vectors in its column space, depolarizing them in just such a way as to invert the action of the Protourgleichung Φ. We then adapt the recently introduced notion of morphophoricity [6, 76] to our setting, in fact extending the definition to include weighted morphophoric measurements. Such measurements imply that the reference measurement maps the GPT state space into the probability simplex in a shape-preserving way. We show that weighted morphophoric measurements always exist in an unrestricted GPT, and that for a MIC, the choice of reference states which give a depolarizing channel is fixed up to scale/sign. A consequence: the effects of any unbiased morphophoric MIC measurement must form a regular simplex, that is, a symmetric informationally complete measurement, and we give an example from quantum theory. Continuing to connect the result to quantum theory, and elaborating on the famous result of Andrew Scott [77, 6], we provide an alternative proof that unweighted morphophoric quantum reference measurements whose effects are proportional to pure states must correspond to complex projective 2-designs. Finally, we turn to a quantitative measure of how much the law of total probability is deformed, $\|I - \Phi\|$, with respect to any unitarily invariant norm. This measure has been discussed primarily for MIC measurements in quantum theory [78]. We discuss the challenges that come from defining this measure for reference measurements which form overcomplete frames. In particular, we show that there are many choices of Φ matrix which act identically on the relevant subspace $\mathrm{col}(P)$ as a Protourgleichung but which have different values for the LTP deformation $\|I - \Phi\|$.

[6] There is an interesting analogy here. The quantum de Finetti theorem was originally proven in the 1970's [72], but a much more straightforward and conceptually satisfying proof was provided in [73] by introducing a reference measurement and appealing to the classical de Finetti representation theorem. The proof for GPTs follows the same pattern. Similarly, Ludwig proved his original theorem in [32, 74] building on his work from the 1970's. The introduction of a reference measurement makes the finite dimensional version of the proof straightforward.

[7] Material in this chapter is based on [75].

Moreover, we show by the method of Lagrange multipliers that for the Frobenius norm, the unique Born matrix which minimizes the LTP deformation over the choice of Φ matrix does not necessarily act as a Protourgleichung in general: we discuss the special case that it does. We thus conclude with a word of caution about the general applicability of this measure.

Chapter 5: (No) return to classicality. When does a set of probabilities across different experiments have a classical explanation? As we saw in chapter 2, merely being able to write a coherence relation in terms of a Born matrix is not enough. In order to answer this question, we adapt Spekkens's notion of a noncontextual ontological model [79, 80, 81] to the QBist way of thought, emphasizing that the key notion of classicality is the indifference of the gambler to whether a reference measurement is performed or not. We discuss the notion of contextuality from a linear algebraic point of view, relating it to the notion of a rank factorization, and following recent work [81], show how noncontextual classical models may be understood in terms of nonnegative matrix factorizations. We then give a QBist spin on the equivalent notion of a simplex embedding, a linear embedding of states and effects into the probability simplex and the dual space of response functions in a probability preserving way. In particular, we show how the noncontextuality of an ontological model of a scenario which includes a reference measurement $\{R_i\}$ is equivalent to the assumption that $\{R_i\}$ forms a reference for the ultimate classical reference measurement $\{\lambda_i\}$. We discuss the computational details of calculating simplex embeddings, in terms of alternating projections, as well as a more sophisticated scheme using linear programming supplemented with the ability to switch from the vertex to the halfspace representation of a convex set or cone [82]. While the result is not original, we show how the construction may be understood in terms of a variant of the Born matrix, and in a special circumstance the Born matrix itself. In such a case the question turns on whether a nonnegative Born matrix can be found. We then adapt an old result [83] characterizing the conditions under which Φ may be taken to be a stochastic matrix. We then turn to the subject of Bell inequalities and more generally noncontextuality inequalities. For the QBist, the key point is the gambler is well advised to use the nonclassical coherence condition $P(E|R) = P(E|R)\Phi P(R|\rho)$. Attempts to build so-called Bell local models simply defy this rule despite the fact that performing a reference measurement or not makes a difference. After recalling the general setup of Bell inequalities, focusing on the CHSH scenario, we show that in fact Bell inequality violations may be reinterpreted as providing lower bounds on $\|I - \Phi\|$. In this way, such violations may be understood as quantifying the cost of abandoning the fundamental nonclassical coherence condition. We show that the same analysis may be applied to noncontextuality inequalities more generally.

Chapter 6: Characterizing quantum state space with a single quantum measurement[8]. So far we have discussed nonclassicality in very general terms, and in some sense that's all we need: given enough experimentation, adherence to the fundamental nonclassical coherence rule is enough to guarantee consistency with e.g., quantum mechanics. But we desire something more: to characterize the state and effect spaces of quantum theory

[8]Material in this chapter is based on [84].

specifically in coherentist terms. In this chapter, we show that remarkably the state space of d-dimensional quantum theory can be derived from studying the behavior of a single reference measurement—if the reference measurement corresponds to a complex projective 3-design. In this privileged case, not only does each quantum state correspond to a probability distribution over the outcomes of a single measurement, but also the probability distributions which correspond to quantum states can be elegantly characterized as those which respect a generalized uncertainty principle. The latter takes the form of a lower bound on the variance of a natural class of observables as measured by the reference. We give simple equations which pure state probability distributions must satisfy, and contextualize these results by showing how 3-designs allow the structure coefficients of the Jordan algebra of observables to be extracted from the probabilities which characterize the reference measurement itself. In fact, just as 2-designs make the Born rule appear as a gentle modification of the law of total probability, 3-designs make the quantum Jordan product appear as a gentle modification of the classical Jordan product on valuations. Taken as a whole, this is in line with the QBist view that quantum theory can be primarily understood as a set of normative constraints on probability assignments on reference outcomes which reflect nature's lack of hidden variables. Moreover, our result further cements the significance of 3-designs, already of interest e.g., for their use in classical shadow estimation protocols, in quantum information science. In an appendix, we also discuss how unitary maps corresponding to symmetries of the state space manifest in Jordan algebraic terms.

Chapter 7: Reconstruction: This chapter forms a pair with the previous. There, we showed how 3-design reference measurements give rise to a remarkably simple characterization of quantum mechanics in probabilistic terms. The central insight is that just as 2-designs make the Born rule look as close as possible to the law of total probability, 3-designs make the quantum rule for multiplying valuations on the reference measurement—the Jordan product on $d \times d$ Hermitian matrices over $\mathbb{C}$—look as close to the classical rule as possible, the latter being the Hadamard or elementwise product on valuations. Indeed, this is the Jordan product on $\mathbb{R} \oplus \cdots \oplus \mathbb{R}$. In this chapter, we start from scratch and attempt to climb our way back to quantum mechanics, and a 3-design reference measurement in particular, by placing ever more severe restrictions on $P(R|R)$, the conditional probability matrix which characterizes the reference measurement itself. On the one hand, this is an exercise in quantum foundations in the spirit of the QBist reconstruction of quantum mechanics inspired by SIC-POVMs; on the other hand, it provides a practical means of self-testing (unbiased) quantum t-designs for $t \geq 3$, since the results of the chapter show precisely what restrictions on $P(R|R)$ guarantee a Hilbert space representation. We begin by assuming that $P \equiv P(R|R)$ takes a simple form: symmetric, with a constant along the diagonal. The assumption that $\Phi = \alpha I + (1-\alpha)J/n$ is a $\{1\}$-inverse of P restricts the eigenstructure of P and implies that P depolarizes vectors in its column space. We then discuss how valuations on arbitrary measurements can be transferred to equivalent valuations on the reference measurement. We prove that any deformation of the classical rule for multiplying valuations which satisfies a handful of desiderata motivated by the idea that we seek the gentlest possible modification of the classical rule must take the form

$x \odot y = \gamma(x \circ y) + (1-\gamma)(\overline{y}x + \overline{x}y - (x \cdot y)u)/n$ for some constant γ and where u is the vector of all 1's. The hope is that on the one hand, the classical product $x \circ y$, the entrywise product of valuations, has an equivalent valuation on some measurement, as does the nonclassical product of valuations $x \odot y$. The difference between them, we suppose, is that $x \odot y$ ought not to depend on the choice of reference measurement. Defining an inner product which reproduces the nonclassical probability rule, and demanding that $\odot$ be consistent with this inner product, fixes the value of γ in terms of α: moreover, it shows that the algebra defined by the nonclassical product is so-called formally real. We then demand that for some class of measurements, the mth power of a reference valuation with respect to our product reproduces the mth moment of the original measurement's valuation. To make this coherent, we must assume that the product is power associative, so that the mth power has an unambiguous meaning. By a classic result of Jordan, von Neumann, and Wigner [28], imposing power associativity (along with our other assumptions) is equivalent to imposing that the nonclassical product satisfies the so-called Jordan identity. This provides a means of implementing power associativity in terms of constraints on the matrix entries of $P(R|R)$ itself, which we work out in terms of the third order polarization of our product. Now that the algebra of valuations must be a Euclidean Jordan algebra, we can appeal to many powerful results. First, we lift our product on valuations to a product on probability distributions and response functions, recasting the Jordan identity in those terms. We then discuss how state and effect spaces arise from a Euclidean Jordan algebra in terms of the self-dual cone of squares with respect to the Jordan product. In fact, this leads to a means of characterizing the state space in terms of an uncertainty principle: a distribution $P(R|\rho)$ is valid iff it satisfies this uncertainty principle, and this constraint can be reformulated in terms of the postive semidefiniteness of a matrix $\mathcal{L}_{P(R|\rho)}$, which in fact is the linear operator which performs the (lifted) Jordan product itself. Further appealing to self-duality, we rescale our product by a self-duality constant κ into a form convenient for discussing the idempotents of the Jordan algebra. We show how many key linear algebraic results port to the Jordan algebraic setting. In fact, we make good on our original ambition: the spectral theorem for Euclidean Jordan algebras guarantees that any reference valuation is equivalent to a valuation on a measurement whose effects form mutually orthogonal idempotents. We give both vector and scalar characterizations of the primitive idempotents, as well as for their orthogonality, and show that our assumptions imply that the reference states themselves must be primitive. Finally, after explaining how to check whether $P(R|R)$ gives rise to a simple Euclidean Jordan algebra, we appeal to the famous classification of the latter to nail down quantum theory: the rank of $P(R|R)$ must be the square of the maximum number of mutually orthogonal primitive idempotents. This fixes the values of κ and α, and we then show that our assumptions imply that the reference measurement forms a complex projective 3-design. We show how the above considerations, while motivated by foundational concerns, amount to an algorithm for self-testing complex projective t-designs for $t \geq 3$: indeed, the constraints that we put on $P(R|R)$ guarantee that the states and effects have a Hilbert space representation, and once this representation is secured, one can rely on the minimization of the frame potential to certify higher order designs. Finally, recapitulating

our results, we give an overview of what questions remain open about our construction, and speculate about future directions for research into the subject.

An open-source python library, `redesigning`*, implementing many of the constructions developed throughout this dissertation has been made available [85].*

CHAPTER 2

PROBABILITIES

2.1 Introduction

We begin from the presumption that there is an agent—You—who wants to make better decisions. The agent is free to act on a world beyond them, but the world returns back consequences beyond the control of the agent: that same freedom the agent has, the world has too. How should the agent organize their behavior in light of this mutual freedom? We begin by developing standard probability theory along subjective Bayesian lines. Then, after defining the notion of reference measurement, state, and effect, we prove a fundamental representation theorem which narrows down the form of the fundamental nonclassical coherence condition necessary to respect nature's openness.

2.2 de Finetti style derivation of probability theory

In the 1930's, the great Italian probabilist Bruno de Finetti taught us that if we interpret probabilities as bets, the laws of probability can be derived from the simple demand that a gambler ought to agree only to buy or sell tickets in such a way that avoids a guaranteed loss [69, 70, 86]. Of course, you might lose any given bet: your horse simply may not come in. But you might find yourself making a series of transactions such that *regardless of what happens*, you end up losing money. This is what probability theory helps you prevent from happening by teaching you how your probability assignments ought to mesh together. In this sense, the laws of probability are not laws in a Newtonian sense, laws which "govern," but are instead essentially good advice, norms which one ought to follow to the extent that one has a stake in success. This is the starting point of subjective Bayesianism [87], and in what follows we hew closely to the account given in [67].

Definition 2.2.1 (Probability)**.** *The probability $P(E)$ of an event E is the price at which a gambler is willing to buy or sell a lottery ticket promising "Pay \$1 if E."*

Although this definition may seem arbitrary, it is not hard to show that we may take such a ticket to be fundamental. The following proof is illustrative of the general style of argument, which proceeds by appealing to the equivalence of holding different ensembles of tickets.

Theorem 2.2.1. *If a gambler is willing to buy or sell a ticket "Pay \$1 if E" at $P(E)$, then they ought to be willing to buy or sell a ticket "Pay \$$x$ if E" at $xP(E)$.*

Proof. Suppose $x = \frac{p}{q}$, a rational number, and consider the following three tickets,

(1) "Pay \$1 if E"

(2) "Pay $\$\frac{1}{q}$ if E"

(3) "Pay $\$\frac{p}{q}$ if E."

Clearly, holding q tickets of type (2) is equivalent to holding a single ticket of type (1). The gambler therefore ought to be willing to buy or sell at the same price for both: $P_1(E) = qP_2(E)$, so that $P_2(E) = \frac{1}{q}P_1(E)$. But holding p tickets of type (2) is equivalent to holding a single ticket of type (3): $P_3(E) = pP_2(E)$. Thus $P_3(E) = \frac{p}{q}P_1(E)$. By the idealization of continuity, the gambler ought to be willing to buy or sell a ticket "Pay \$$x$ if E" at $xP(E)$ for any real number x. □

We now proceed to derive the basic laws of probability.

Theorem 2.2.2. *A probability $P(E)$ ought to satisfy $0 \leq P(E) \leq 1$.*

Proof. On the one hand, suppose the gambler is willing to sell a ticket "Pay \$1 if E" for a price $P(E) < 0$. Equivalently, they are willing to pay someone to take it off their hands, which leads to a sure loss. We therefore require $P(E) \geq 0$. On the other hand, suppose the gambler is willing to buy such a ticket for a price $P(E) > 1$. Equivalently, they are willing to pay more for the ticket than it will ever pay off, guaranteeing a loss: so we require $P(E) \leq 1$. □

Remark. *If a gambler is certain that an event E will occur, they ought to assign $P(E) = 1$, since otherwise they must expect to lose money. Conversely, if a gambler sets $P(E) = 1$, then they must expect to lose money unless they are certain that E will occur.*

Theorem 2.2.3. *$P(A \text{ or } B) = P(A) + P(B)$, for mutually exclusive events A, B.*

Proof. A gambler judges events $\{A_i\}$ to be *mutually exclusive* if they believe only one of them will occur. Consider the following three tickets:

(1) "Pay \$1 if A or B"

(2) "Pay \$1 if A"

(3) "Pay \$1 if B."

Suppose the gambler buys (1) and sells (2) and (3). From buying (1), if A or B occurs, the gambler makes $1 - P(A \text{ or } B)$. Suppose A occurs. From selling (2), the gambler makes $P(A) - 1$. From selling (3), since only one of A or B can occur and B did not occur, the

gambler makes $P(B)$, for a total of $P(A) + P(B) - 1$. Moreover, this is just the same in the case that B occurs. To avoid a sure loss, we require

$$\big(1 - P(A \text{ or } B)\big) + \big(P(A) + P(B) - 1\big) \geq 0 \Longrightarrow P(A \text{ or } B) \leq P(A) + P(B). \tag{2.1}$$

Alternatively, suppose the gambler sells (1) and buys (2) and (3). From selling (1), if A or B occurs, the gambler makes $P(A \text{ or } B) - 1$, and from buying (2) and (3), the gambler makes $1 - P(A) - P(B)$. To avoid a sure loss, then

$$\big(P(A \text{ or } B) - 1\big) + \big(1 - P(A) - P(B)\big) \geq 0 \Longrightarrow P(A \text{ or } B) \geq P(A) + P(B). \tag{2.2}$$

We conclude that the gambler ought to set $P(A \text{ or } B) = P(A) + P(B)$. □

Remark. *Clearly the above proof can be extended to any finite number of mutually exclusive events. Under the idealization that the gambler can buy or sell a countably infinite number of tickets whose payout depends on a countable infinity of mutually exclusive events, one arrives at the countable additivity of probabilities. This however was no small matter of controversy for de Finetti himself [86].*

Theorem 2.2.4. *$P(A \text{ and } B) = P(A|B)P(B)$ for events A, B.*

Proof. Consider the following three tickets:

(1) "Pay \$1 if A and B; pay $\$P(A|B)$ if not B"

(2) "Pay \$1 if A and B"

(3) "Pay $\$P(A|B)$ if not B."

Let $P(A|B)$ be the price at which the gambler is willing to buy or sell (1), so that the gambler will get a refund if B does not occur. Holding the ticket at that price means they will win \$1 if A occurs given that B occurs, and so we can think of $P(A|B)$ as a conditional probability. Meanwhile, let $P(A \text{ and } B)$ be the price the gambler is willing to buy or sell (2), and for (3), notice that by Theorem 2.2, the gambler ought to set price $P(A|B)P(\text{not B})$. Clearly, holding ticket (1) is equivalent to holding (2) and (3), and so they should have the same price:

$$\begin{aligned} P(A|B) &= P(A \text{ and } B) + P(A|B)P(\text{not } B) && (2.3)\\ &= P(A \text{ and } B) + P(A|B)\big(1 - P(B)\big) && (2.4)\\ &= P(A \text{ and } B) + P(A|B) - P(A|B)P(B), && (2.5) \end{aligned}$$

from which we conclude $P(A \text{ and } B) = P(A|B)P(B)$. □

Remark. *If the gambler judges that A does not depend on B, then $P(A|B) = P(A)$, and $P(A \text{ and } B) = P(A)P(B)$.*

Corollary 2.2.4.1 (Bayes's Rule)**.** *Since* $P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$*, if* $P(A) \neq 0$*,*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{2.6}$$

With these theorems established, we have already erected essentially the whole structure of probability theory as usually formalized by Kolmogorov's axioms. We now prove a useful lemma, the law of total probability (LTP).

Lemma 2.2.5. $P(A) = \sum_i P(A|B_i)P(B_i)$ *for an event A and a collection of mutually exclusive and exhaustive events* $\{B_i\}_{i=1}^n$.

Proof. A gambler judges events $\{B_i\}$ to be *exhaustive* if they believe at least one of them must occur. From the relationship between joint and conditional probabilities and from the additivity of probabilities, we have

$$\sum_i P(A|B_i)P(B_i) = \sum_i P(A \text{ and } B_i) \tag{2.7}$$

$$= P\big((A \text{ and } B_1) \text{ or } \ldots \text{ or } (A \text{ and } B_n)\big) \tag{2.8}$$

$$= P\big(A \text{ and } (B_1 \text{ or } \ldots \text{ or } B_n)\big). \tag{2.9}$$

The event $(B_1 \text{ or } \ldots \text{ or } B_n)$ is just the event that some B_i occurs. But since the gambler believes $\{B_i\}$ is a set of mutually exclusive and exhaustive events, they are certain that exactly one of the B_i's will occur. Thus they ought to buy or sell a ticket "Pay \$1 if A and $(B_1 \text{ or } \ldots \text{ or } B_n)$" at the same price as a ticket "Pay \$1 if A." We conclude that $P(A) = \sum_i P(A|B_i)P(B_i)$. □

Corollary 2.2.5.1 (Bayes's Rule: LTP version)**.** *Given an event A such that* $P(A) \neq 0$*, and a collection of mutually exclusive and exhaustive events* $\{B_i\}$*,*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}. \tag{2.10}$$

2.3 Reflection

Bayes's rule tells us that a gambler ought to set $P(A|B) = P(A, B)/P(B)$. Here $P(A|B)$ expresses the fair price the gambler has assigned at the present moment (call it $t = 0$), that is, before the events A and B have been resolved. But what about later, say at some time $t = \tau$? Suppose event B befalls the gambler. Must they set $P_\tau(A) = P_0(A|B)$, where we have tagged probabilities with subscripts denoting the time? This is the standard Bayesian update rule, but in fact, without further assuptions, there is no coherence argument which would compel the gambler to do so [88, 89]. Nevertheless, one may argue from coherence that a gambler's probability assignments at different times ought to be constrained in a particular

way, summarized by van Fraassen's *reflection principle* [68, 90, 91, 92]. We follow closely the exposition in [89]. Crucially, in order to formulate the reflection principle we must consider the gambler's beliefs now *about their own future probability assignments*. For example, $P_0(P_\tau(A) = q)$ signifies how strongly the gambler believes now (at $t = 0$) that at time $t = \tau$ they will assign probability q to event A. The reflection principle constrains the probabilities the gambler assigns now, given their beliefs about how they will assign probabilities in the future.

Theorem 2.3.1 (Reflection principle)**.** *Let $P_0(A)$ be the probability the gambler assigns to event A at $t = 0$, and let $P_\tau(A)$ be the probability the gambler assigns to event A at $t = \tau$. Suppose that $P_0(P_\tau(A) = q) = 1$. Then coherence demands that $P_0(A) = P_\tau(A) = q$.*

Proof. Suppose at $t = 0$, the gambler sets $P_0(P_\tau(A) = q) = 1$: that is, the gambler at $t = 0$ is certain that at time τ they will assign probability q to event E. This means that if $q < P_0(A)$, then at $t = 0$ the gambler is willing to buy a ticket at price $\$P_0(A)$ even as they also believe that later they will be willing to sell it for the lower price $\$q$. Similarly, if $q > P_0(A)$, then at $t = 0$, the gambler is willing to sell a ticket at $\$P_0(A)$ even as they also believe that later they will be willing to buy it for a higher price $\$q$. Either way, unless $P_0(A) = P_\tau(A) = q$, the gambler faces a sure loss. □

Theorem 2.3.2 (Reflection principle: conditional version)**.** *Let $P_0(A|P_\tau(A) = q)$ be the fair price the gambler assigns at $t = 0$ to a conditional ticket on A which gives a refund if at $t = \tau$, the gambler assigns $P_\tau(A) \neq q$. We assume that $P_0(P_\tau(A) = q) > 0$ so that the conditional probability is well defined. Coherence then demands that*

$$P_0(A|P_\tau(A) = q) = q. \tag{2.11}$$

Proof. We want to show that $P_0(A|P_\tau(A) = q) = q$. $P_0(A|P_\tau(A) = q)$ represents the fair price at which the gambler would buy or sell a ticket "Pay \$1 if A and $P_\tau(A) = q$; pay $P_0(A|P_\tau(A) = q)$ if $P_\tau(A) \neq q$." In other words, $P_0(A|P_\tau(A) = q)$ is the fair price the gambler assigns for a conditional ticket on A, whose cost is refunded unless $P_\tau(A) = q$. In particular, the gambler is willing at $t = 0$ to buy such a ticket at price $P_0(A|P_\tau(A) = q)$. Suppose $P_0(A|P_\tau(A) = q) > q$. If in the end $P_\tau(A) = q$, this means that at time τ, the gambler is willing to sell the ticket "Pay \$1 if A" at price q. But if they buy the original ticket for $P_0(A|P_\tau(A) = q)$ and later sell the A-ticket at q, they lose $P_0(A|P_\tau(A) = q) - q$. Of course, if in the end $P_\tau(A) \neq q$, they get a refund: thus conditional on $P_\tau(A) = q$, setting $P_0(A|P_\tau(A) = q) > q$ exposes the gambler to a sure loss. (Notice that this is independent of whether A itself occurs.) By the same token, the gambler is willing at $t = 0$ to sell the conditional ticket at price $P_0(A|P_\tau(A) = q)$. Suppose $P_0(A|P_\tau(A) = q) < q$. If in the end $P_\tau(A) = q$, the gambler is willing at time τ to buy the ticket "Pay \$1 if A" at price q. Since $P_0(A|P_\tau(A) = q) < q$, the gambler certainly loses $q - P_0(A|P_\tau(A) = q)$ conditional on $P_\tau(A) = q$. Thus coherence demands that $P_0(A|P_\tau(A) = q) = q$. □

Corollary 2.3.2.1. *By reflection, $P_0(A|P_\tau(A) = q) = q$, so that the law of total probability demands*

$$P_0(A) = \sum_i P_0(A|P_\tau(A) = q_i)P_0(P_\tau(A) = q_i) = \sum_i q_i P_0(P_\tau(A) = q_i). \tag{2.12}$$

Remark. *In the present, the gambler may contemplate their future beliefs about event A. There are several possibilities: they might assign probabilities $q_1, q_2, \ldots$ to event A, e.g., given their other beliefs about what might happen between now and then. In particular, still in the present, they assign probability $P_0(P_\tau(A) = q_1)$ that in the future they will judge q_1 to be a fair price for an A ticket, $P_0(P_\tau(A) = q_2)$ that they will judge q_2 as a fair price for an A ticket, and so on. In order for their present belief about A to be coherent with their beliefs now about their future beliefs, the gambler ought to adopt $P_0(A) = \sum_i q_i P_0(P_\tau(A) = q_i)$: the probability they assign now to A should be a mixture of the probabilities they might assign in the future, weighted by the probability that they will make that assignment.*

2.4 Sample spaces and all that

In our arguments, we have implicitly made use of the idea that an agent may assign probabilities not just to isolated events, but to also more general propositions involving events, e.g., to the event that *A and B* occur, or the event that *A or B* occurs, or event that *A* does *not* occur at all. We can formalize this logic using a *Boolean algebra*. In particular, following de Finetti [70], we may consider a notion of "constituent" or "atomic" events into which arbitrary events can be refined. Given a set of n events $\{E\}$, we may consider the 2^n possible conjunctions of these n events and their negations. For example, for two events A, B, we may consider the events

(1) A and B
(2) A and not B
(3) not A and B
(4) not A and not B.

If we consider any pair of these atomic events, they must differ in at least one event. For example, (1) requires that B, but (2) requires that not B; (1) requires that A, but (3) requires that not A, and so forth. Thus the 2^n atomic events so constructed are all mutually exclusive. Moreover, if we consider the disjunction of all the atomic events that implicate one of the original events, we recover that original event. For example, if we consider the disjunction (1) or (2), we have

$$(A \text{ and } B) \text{ or } (A \text{ and not } B) = A \text{ and } (B \text{ or not } B) = A. \tag{2.13}$$

Thus given any set of events, we may resolve them into a set of mutually exclusive atomic events Ω. For n events, we require no more than 2^n atomic events, but possibly less since we may remove atomic events which are equivalent or impossible. More general events may then

be identified with subsets of Ω, logical disjunction (or) with set union ($\cup$), logical conjunction (and) with set intersection ($\cap$), and logical negation with set complement ($\neg$). The empty subset may be regarded as an event the agent regards as impossible, and the full set Ω may be regarded as an event the agent regards as necessary.

From these considerations, we can see that de Finetti probability theory is equivalent to Kolmogorov's axiomatic definition [93] in terms of probability spaces.

Definition 2.4.1 (Probability space)**.** *A finitely additive probability space consists of a* sample space Ω*, the set of atomic events; an* event space E*, the set of all subsets (the powerset) of* Ω*; and a* probability measure P*, which assigns to each event in the event space a probability. The probability measure must satisfy* $\forall A \in E : P(A) \geq 0$*,* $P(\Omega) = 1$*, and* $P(\bigcup_{i=1}^{N} A_i) = \sum_{i=1}^{N} P(A_i)$ *for a finite family of disjoint events* A_i*. One can work more generally with event spaces forming a nontrivial* σ*-algebra, and impose that the final axiom holds for any countable sequence of disjoint events: this is a countably additive probability space.*

Using set theoretic notation, we can write a somewhat more compact proof of the law of total probability (LTP). Let $A = \{A_0, A_1, \dots\}$ be a partition of the sample space Ω into disjoint subsets whose union is Ω. Since A is therefore a mutually exclusive and exhaustive set of events, we have for any other event B

$$P(B) = P(B \cap \Omega) = P((B \cap A_0) \cup (B \cap A_1) \cup \dots) \tag{2.14}$$

$$= \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i). \tag{2.15}$$

In particular, if we consider the ultimate partition given by Ω itself, we have

$$P(B) = \sum_{\omega \in \Omega} \delta_{\omega \in B} P(\omega), \tag{2.16}$$

that is, the conditional probabilities in the LTP are just so-called indicator functions.

2.4.1 Measures

For de Finetti, countable additivity was not forced upon the gambler by coherence alone: assuming it is a mathematical convenience which however does not exhaust all the possibilities open to the gambler [86]. But let us see what it buys us. Consider for example, the case where the sample space Ω is uncountably infinite: let Ω be the real line $\mathbb{R}$. In the Kolmogorovian measure theoretic treatment of probabilities, one doesn't assign probabilities to any arbitrary subsets of $\mathbb{R}$, but specifically to *measurable subsets*. For example, one might choose the Borel σ-algebra on $\mathbb{R}$ which is generated by the open intervals of $\mathbb{R}$, and which is the smallest collection of subsets of $\mathbb{R}$ which contains the open intervals, is closed under complements, countable unions, and countable intersections.

A *probability measure* on the σ-algebra assigns probabilities to events. Under certain regularity conditions, a probability measure dP can be expressed in terms of a probability density $p(x)$ so that one may express the probability of falling within an interval as

$$P(x \in [a,b]) = \int_{[a,b]} dP = \int_a^b p(x)dx \tag{2.17}$$

where $p(x) \geq 0$ and integrates over the line to 1. For such continuous distributions, individual points have probability 0: there is no contradiction here since an interval is an uncountable union of points. This same procedure described above can be generalized to more exotic sample spaces. In this dissertation, however, we will almost entirely focus on discrete events, e.g., measurements with a finite number of outcomes.

2.5 Random variables

Intuitively, a random quantity X is a quantity that could take different values $\{x_i\}$ conditional on this or that event occurring: it is a valuation on outcomes. In betting terms, we can regard X as a ticket or contract such that if event E_i occurs, one gains (or owes) some amount x_i. (We will always refer to this as a gain, with the understanding that a negative gain means a loss.) In his *Theory of Probability* [69], de Finetti defines the *prevision*, or fair price, $P(X)$ of a random quantity X in the following way: a gambler picks $P(X)$ with the understanding that it commits them to accepting any bet with gain $c(X - P(X))$ where c is arbitrary and "at the choice of an opponent."

To understand the logic, suppose $c = 1$. Then $X - P(X)$ is just the gain or loss the gambler realizes after having paid $P(X)$ for a ticket, X now taking one of its values x_i. $c = 2$ corresponds to the case of the gambler having purchased two such tickets. If $c = -1$, then $P(X) - X$ is the gain or loss the gambler realizes after selling a ticket at price $P(X)$. The point is that c can be chosen adversarially: in declaring $P(X)$, the gambler has specified the *certain gain* $P(X)$ that they consider equivalent to the *uncertain gain* X. Moreover, we say $P(X)$ is the *fair price* the gambler has assigned to X: the gambler is willing to take either side of a gamble at that rate.

de Finetti's coherence principle amounts to a single normative injunction: "It is assumed that You do not wish to lay down bets which will with *certainty* result in a loss for You" [69]. A set of previsions is coherent, then, if there is no combination of bets whose gains are all uniformly negative. In particular, given a collection of random quantities X_i, coherence demands that we must choose previsions $P(X_i)$ such that there is no linear combination

$$\sum_i c_i(X_i - P(X_i)) < 0. \tag{2.18}$$

Otherwise, if the gambler announces previsions $\{P(X_i)\}$, an adversary could induce them to buy or sell tickets in amounts $\{c_i\}$, forcing upon the gambler a certain loss. That is, the adversary could pick $\{c_i\}$ such that $\sum_i c_i X_i < \sum_i c_i P(X_i)$: no matter what values are realized for each X_i, the gambler has overpaid.

For example, consider the random quantities X, Y, and $X+Y$, and the linear combination

$$\big((X+Y) - P(X+Y)\big) - \big(X - P(X)\big) - \big(Y - P(Y)\big) = P(X) + P(Y) - P(X+Y) \geq 0. \tag{2.19}$$

By the same token we can consider the linear combination

$$-\big((X+Y) - P(X+Y)\big) + \big(X - P(X)\big) + \big(Y - P(Y)\big) = P(X+Y) - P(X) - P(Y) \geq 0. \tag{2.20}$$

Putting these two inequalities together, we find that coherence demands $P(X+Y) = P(X) + P(Y)$. In fact, it would have sufficed to observe that the initial linear combination is a constant *regardless* of the actual payoffs of X, Y, and $X+Y$. If it were a negative constant, the gambler would always lose money; if it were a positive constant, flipping the signs in the linear combination, swapping buying for selling, one would end up again with a negative constant. Thus the only constant that is coherent is 0. Similarly, consider random quantities X and aX for some real constant a, and the linear combination

$$\big(aX - P(aX)\big) - a\big(X - P(X)\big) = aP(X) - P(aX) \geq 0. \tag{2.21}$$

Again this is a constant regardless of outcome, and so the only possibility is that $P(aX) = aP(X)$. Thus in general, previsions behave linearly

$$P\left(\sum_i a_i X_i\right) = \sum_i a_i P(X_i). \tag{2.22}$$

With this in hand, suppose we have a ticket X such that one gains x_i conditional on event E_i occurring. At the same time, we can consider tickets E_i which pay \$1 if a particular event E_i occurs, and 0 otherwise. (We conflate the symbol for the ticket and the event.) As we have seen, $P(E_i)$ is just the probability of the event E_i. Since we can decompose the random quantity as $X = \sum_i x_i E_i$, by the linearity of previsions,

$$P(X) = P\left(\sum_i x_i E_i\right) = \sum_i x_i P(E_i). \tag{2.23}$$

We conclude: the fair price for a random quantity X is just what we normally think of as the expectation value, denoted $\mathbb{E}[X]$ or $\langle X \rangle$. But here the interpretation is that $\langle X \rangle$ is the certain quantity which, if it were swapped for the uncertain quantity X, the gambler would be

indifferent to the exchange. Similarly, the variance,

$$\mathrm{Var}[X] = P\big((X - P(X))^2\big) = P(X^2) - P(X)^2, \tag{2.24}$$

may be interpreted as the fair price of a ticket that pays more when your stated prevision of X is badly off. Finally, by the homogeneity of previsions, e.g., $P(aX) = aP(X)$, we may contemplate previsions of random quantities which cannot in themselves be interpreted as gains or losses of money. We simply pick a conversion factor a such that aX *is* a monetary value: then let $P(X) = (1/a)P(aX)$.

2.5.1 The algebra of random variables

Appealing to the set of atomic events Ω, we can more formally define a random variable X to be a map from Ω to a set of valuations. In other words, a random variable assigns a valuation to atomic events. We focus here on the particular case of real valued random variables, $X : \Omega \to \mathbb{R}$, which we now show form an algebra. Indeed, suppose now we have two random variables $G : \Omega \to \mathbb{R}$ and $H : \Omega \to \mathbb{R}$. We can add them as $(H + G)(\omega) = H(\omega) + G(\omega)$; we can multiply them by scalars as $(\lambda H)(\omega) = \lambda H(\omega)$; and finally we can multiply them as $(HG)(\omega) = H(\omega)G(\omega)$. This latter operation may be viewed as the pointwise or Hadamard product on $\mathbb{R}^n$, where $|\Omega| = n$. (As usual, we assume n is finite, but really it need not be.)

Let $A = \{A_0, A_1, \dots\}$ be a partition of Ω. We say a random variable X is A-measurable if X is constant on each block of A: $\forall \omega \in A_i : X(\omega) = x_i$. The intuition is that X assigns a particular payoff depending only upon which composite event A_i occurs. We then have $P(X = x_i) = \sum_{\omega \in A_i} P(\omega)$, and

$$\langle X \rangle = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_i x_i \sum_{\omega \in A_i} P(\omega) = \sum_i x_i P(X = x_i). \tag{2.25}$$

Moreover, given X which is A-measurable and Y which is B-measurable, we have

$$\langle XY \rangle = \sum_{\omega \in \Omega} X(\omega)Y(\omega)P(\omega) = \sum_{ij} x_i y_j \sum_{\omega \in A_i \cap B_j} P(\omega) \tag{2.26}$$

$$= \sum_{ij} x_i y_j P(X = x_i, Y = y_j). \tag{2.27}$$

Here $P(X = x_i, Y = y_j)$ is shorthand for $P(X = x_i \text{ and } Y = y_j)$. From this we can see that the familiar formula for the expectation value of the product of two random variables $\langle XY \rangle$ can be derived from the pointwise product of valuations on the underlying sample space.

2.6 From composites to the de Finetti representation

In the preceding, we used the fact that X and Y correspond to alternative coarse grainings $\{A_i\}$ and $\{B_i\}$ of the same underlying atomic events. For example, consider rolling a single die. The atomic events Ω correspond to the different outcomes of the die roll, and e.g., a random variable Z might map "1 pip shows" $\to 1$, "2 pips show" $\to 2$ and so on. For convenience, we might therefore write the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$[1]. We might then consider the partition $A_0 = \{2, 4, 6\}, A_1 = \{1, 3, 5\}$ and call the event corresponding to the former "even pips," and the event corresponding to the latter "odd pips." We could distinguish them with an A-measurable random variable, for example, which takes $+1$ for $\omega \in A_0$ and -1 for $\omega \in A_1$. We could also consider the partition $B_0 = \{1, 2, 3\}, B_1 = \{4, 5, 6\}$, and distinguish them with an B-measurable random variable which takes $+1$ on B_0 and -1 on B_1. We could then contemplate the probability that the pip is both even and in the first half, that is, $P(A = +1, B = +1) = P(2)$: this turns out to be an atomic event.

If however we'd like to gamble on two rolls of the die, we will need to expand our sample space. We can take the Cartesian product of sample spaces $\Omega \times \Omega = \{(1, 1), (1, 2), \ldots, (2, 1), (2, 2), \ldots, (6, 1), (6, 2), \ldots\}$. Now the atomic events correspond to getting particular pairs of rolls. Going in reverse, define random variables $D_1(i, j) = i$ and $D_2(i, j) = j$. Then $D_1 = i$ means the event that the first die comes up i regardless of the second and corresponds to the subset $\{(i, j) : j = 1, \ldots, 6\}$—and similarly for D_2. Then

$$P(\text{first die shows } 1) = P(D_1 = 1) = \sum_j P(D_1 = 1, D_2 = j) = \sum_j P((1, j)). \tag{2.28}$$

In the special case that the two rolls are independent, we have $P(D_1 = i, D_2 = j) = P(D_1 = i)P(D_2 = j)$. But how can we capture the intuition that we are rolling "the same die" in each trial? One strong assumption is that the rolls are not just independent but identically distributed, so that

$$P(D_1 = i, D_2 = j, D_3 = k, \ldots) = P(D = i)P(D = j)P(D = k)\ldots, \tag{2.29}$$

where a single distribution $P(D = \cdot)$ is used for every roll. But where does that single distribution come from? In particular, how can we justify the idea that precisely by rolling the dice many times, we may sharpen our judgement about what probabilities to assign to each outcome in *any* roll of the die? de Finetti recognized that the more basic judgement is that when we assign joint probabilities to repeated trials the order of the rolls should not matter. This led to one of his great accomplishments: the de Finetti representation theorem, which shows that an infinitely exchangeable joint distribution over many trials can be represented as a mixture of independent and identically distributed trials. In particular, it justifies acting

[1] From this point of view, any semantic meaning applied to the atomic events may be recovered from the random variables that implicate them.

"as if" there is an unknown single roll distribution which may be sharpened by repeated rolling. The "as if" matters: for a subjective Bayesian, there are no such thing as an "unknown distribution," as probabilities are by definition beliefs on the part of the gambler.

Suppose the gambler assigns a joint probability distribution to the outcomes of n trials, e.g., n rolls of the dice. Further suppose that the gambler is indifferent to the order of the die rolls, and could imagine rolling the die indefinitely. In this case, the gambler ought to assign a joint probability distribution which is (infinitely) *exchangeable*. What does this mean? A distribution is symmetric (finitely exchangeable) if it is invariant under permutations of its arguments, corresponding to the different trials. A distribution is (infinitely) exchangeable if it is symmetric and extendible to a symmetric distribution on arbitrarily many random trials: formally, the distribution over n trials can be obtained from a distribution over $n + m$ trials by marginalizing over the last m trials[2]. de Finetti proves the following theorem [86, 73], which is a special case of a broader result:

Theorem 2.6.1 (discrete classical de Finetti representation theorem)**.** *Suppose $P(x_1, \ldots, x_n)$ is the n-trial marginal of an infinitely exchangeable sequence of k-valued random variables X_i where x_i denotes which of the k values X_i takes. Then there exists a unique probability measure $P(\mathbf{p})$ on Δ_k such that*

$$P(x_1, x_2, \ldots, x_n) = \int_{\Delta_k} P(x_1|\boldsymbol{p}) \ldots P(x_n|\boldsymbol{p}) P(\boldsymbol{p}) d\boldsymbol{p} \tag{2.30}$$

$$= \int_{\Delta_k} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k} P(\boldsymbol{p}) d\boldsymbol{p}, \tag{2.31}$$

where $\boldsymbol{p} = (p_1, \ldots, p_k)$; Δ_k is the probability simplex over k outcomes; $P(\boldsymbol{p})$ is a probability density function on the simplex; and n_j is the number of times the outcome $j \in \{1, \ldots, k\}$ occurred. Note that $P(x_i|\boldsymbol{p})$ just pulls out the probability p_i corresponding to the outcome denoted by x_i.

Upon obtaining outcomes on m trials, denoting them T_m, one may use Bayes's rule to pass from the prior $P(\mathbf{p})$ to the posterior

$$P(\mathbf{p}|T_m) = \frac{P(T_m|\mathbf{p})P(\mathbf{p})}{P(T_m)} = \frac{P(T_m|\mathbf{p})P(\mathbf{p})}{\int_{\Delta_k} P(T_m|\mathbf{p})P(\mathbf{p})d\mathbf{p}}, \tag{2.32}$$

where $P(T_m|\mathbf{p}) = p_1^{n_1} \ldots p_k^{n_k}$, where the $\{n_j\}$ refer to the number of times the jth outcome actually occurred. This justifies the usual idea that to determine the "unknown probability distribution" over the outcomes of e.g., a roll of the dice, one collects the frequency data

[2]We note that exchangeability is a weaker notion that the independence of trials. A famous example of a distribution which is the former but not the latter is provided by Polya's urn model. We also note that one may prove finite de Finetti theorems where only finite exchangeability is demanded: the representation then proves to be approximate with a certain quantifiable error.

over many trials. But for de Finetti, this is an "as if": probabilities do not exist in nature, unknown and waiting to be known: probabilities are judgements of an individual gambler. de Finetti therefore begins with the gambler who has assigned joint probabilities over many trials, who is committed to an exchangeable distribution, and who can then write their joint distribution in terms of a probability distribution over which probabilities to assign to one of the six pips coming up. As data comes in from the trials, the theorem justifies the gambler in using Bayes's rule to update their probabilities over probabilities, $P(\mathbf{p})$, using the outcome frequencies. In principle, as the number of trials increases, $P(\mathbf{p})$ may become highly peaked around a particular distribution, a single assignment of probabilities to each of the k outcomes of a single trial, e.g., each roll of the dice. The point is that even if several gamblers start with different priors, assuming there are no pathologies (assignments of zero probabilities) if they update on the same data for long enough, they will converge on the same distribution $P(\mathbf{p})$. Thus the theorem is really about how through repeated experiments, different gamblers ought to come to agreement in their beliefs.

2.7 Calling the whole thing off

The law of total probability $P(E) = \sum_i P(E|R_i)P(R_i)$ urges on the gambler a particular relationship between the probabilities they assign to events E and $\{R_i\}$. The key assumption is that exactly one of the $\{R_i\}$'s will in fact occur. Suppose however that after assigning probabilities $P(E|R_i)$ and $P(R_i)$, the gambler learns that no R_i will occur after all. Can we derive any coherence condition that $P(E)$, $P(E|R_i)$, and $P(R_i)$ must satisfy—or are all bets off?

In general, yes, all bets are off. But it may be that the gambler can relate events E and $\{R_j\}$ to another set of mutually exclusive and exhaustive events $\{\lambda_k\}$. We assume that the λ_k's mediate the relevance of the R_j's to E in the sense that $P(E|R_j, \lambda_k) = P(E|\lambda_k)$. Indeed, this implies

$$P(E|R_j) = \sum_k P(E, \lambda_k|R_j) = \sum_k P(E|\lambda_k, R_j)P(\lambda_k|R_j) = \sum_k P(E|\lambda_k)P(\lambda_k|R_j). \quad (2.33)$$

Now since the events $\{\lambda_k\}$ are mutually exclusive and exhaustive, if it turns out that no R_j will occur, there remains a norm that the gambler can appeal to, namely

$$P(E) = \sum_k P(E|\lambda_k)P(\lambda_k). \quad (2.34)$$

Under a special circumstance, we can take an alternative perspective on this same norm. First, we note that according to Bayes' rule,

$$P(\lambda_k|R_j) = \frac{P(R_j|\lambda_k)P(\lambda_k)}{\sum_i P(R_j|\lambda_i)P(\lambda_i)}. \quad (2.35)$$

We then ask: is it possible to find a matrix Φ which satisfies

$$\sum_{kl} P(\lambda_j|R_k)\Phi_{kl}P(R_l|\lambda_m) = \delta_{jm}? \tag{2.36}$$

In more compact linear algebraic notation, we require

$$P(\lambda|R)\Phi P(R|\lambda) = I, \tag{2.37}$$

where $P(\lambda|R), P(R|\lambda)$ denote conditional probability matrices. (We shall also use $P(R), P(\lambda)$ as a compact notation for the corresponding probability vectors.)

When can one find such a Φ? We may appeal to some linear algebra. Let $P(\lambda|R) \in \mathbb{R}^{r\times m}$ and $P(R|\lambda) \in \mathbb{R}^{m\times r}$ so that I is the $r \times r$ identity matrix. We will later show in Theorem (3.6.4) that such a Φ can be found iff $P(\lambda|R), P(R|\lambda)$ are both of rank r. Explicitly, for example, we could take $\Phi = P(\lambda|R)^+P(R|\lambda)^+$ where A^+ denotes the Moore-Penrose pseudoinverse of A. Now if it is possible to find such a Φ, we say that gambler believes that the events $\{R_i\}$ constitute a *reference* for events $\{\lambda_i\}$. To appreciate this, we may insert the resolution of the identity provided by Eq. (2.37) into Eq. (2.34),

$$P(E) = P(E|\lambda)P(\lambda) = P(E|\lambda)\Big[P(\lambda|R)\Phi P(R|\lambda)\Big]P(\lambda) = P(E|R)\Phi P(R), \tag{2.38}$$

which looks like the law of total probability save for the presence of this matrix Φ.

To recapitulate, if the gambler believes that some R_i will occur, in order to be coherent, they ought to follow a norm: the law of total probability,

$$P_1(E) = \sum_i P(E|R_i)P(R_i). \tag{2.39}$$

If instead it turns out some event R_i will *not* occur, but nevertheless the gambler can relate the events $\{R_i\}$ to another set of mutually exclusive and exhaustive events $\{\lambda_i\}$ which they believe *will* occur, and moreover they judge that the events $\{R_i\}$ constitute a reference for the events $\{\lambda_i\}$ in the linear algebraic sense we have described, then the gambler may appeal to a *deformation* of the law of total probability,

$$P_2(E) = \sum_{ij} P(E|R_i)\Phi_{ij}P(R_j), \tag{2.40}$$

where Φ satisfies $P(\lambda|R)\Phi P(R|\lambda) = I$. Despite the fact that the events $\{R_i\}$ are counterfactual in the second scenario, nevertheless if the agent believes that the events $\{R_i\}$ form a reference for the events $\{\lambda_i\}$, they can reuse the probabilities they would assign in the first scenario to deal with the second.

It may seem that there is something contradictory here. In general the matrix Φ may contain negative entries in defiance of our earlier demonstration that assigning negative probabilities is incoherent! Of course, the inconsistency is merely apparent: after all, we have just shown that adopting $P_2(E) = P(E|R)\Phi P(R)$ is equivalent to adopting $P_2(E) = P(E|\lambda)P(\lambda)$. Furthermore, recall that we derived the conditions for coherence on the assumption that one of the events R_i would really happen: in the derivation of the law of total probability, it is assumed that $\{R_i\}$ constitute a mutually exclusive and exhaustive set of events. But if that is no longer the case, then the law of total probability no longer applies.

We can also give a de Finetti style justification for Eq. (2.40). Using the resolution of the identity from Eq. (2.37), we have in particular

$$P(\lambda) = \Big[P(\lambda|R)\Phi P(R|\lambda)\Big]P(\lambda) = P(\lambda|R)\Phi P(R). \tag{2.41}$$

In other words, since the gambler judges $\{R_i\}$ to be a reference for $\{\lambda_i\}$, we can write the probabilities $P(\lambda_i)$ as

$$P(\lambda_i) = \sum_j x_j^{(i)} P(R_j), \tag{2.42}$$

where $x_j^{(i)} = \sum_k P(\lambda_i|R_k)\Phi_{kj}$. We can then compare on the one hand, tickets (1_i), "Pay \$1 if λ_i"; and on the other hand, the bundle of tickets (2_i),

- "Pay $\$x_1^{(i)}$ if R_1"
- $\vdots$
- "Pay $\$x_m^{(i)}$ if R_m."

The price at which the gambler ought to buy or sell (1_i) is $P(\lambda_i)$, while the price at which the gambler ought to buy or sell the bundle of tickets (2_i) is $\sum_j x_j^{(i)} P(R_j)$. Because $\{R_i\}$ is a reference for $\{\lambda_i\}$, the fair price that the gambler assigns to (1_i) and (2_i) are the same. Thus the gambler ought to be willing to exchange the one for the other. If the gambler suspects that one of the R_i's will not occur after all, they ought to aggregate all their $\{R_i\}$-related tickets into a bundle of the form (2_i), which they can convincingly market as equivalent to (1_i). Having converted their (2_i)'s into (1_i)'s, should the $\{R_i\}$'s in fact be called off (nothing is certain), their fortunes will be secured.

2.7.1 Coulda, woulda, shoulda

The possibility of reselling R-tickets notwithstanding, one might still worry: if adopting $P_2(E) = P(E|R)\Phi P(R)$ is equivalent to adopting $P_2(E) = P(E|\lambda)P(\lambda)$, what is the usefulness of the former, if the latter is available? Suppose in fact that the gambler

believes some $\{\lambda_i\}$ will occur, but they do not believe they can assign informed probabilities to them. In that case, $P_2(E) = P(E|\lambda)P(\lambda)$ cannot offer them much advice, nor apparently can $P_2(E) = P(E|R)\Phi P(R)$, since we defined Φ to satisfy $P(\lambda|R)\Phi P(R|\lambda) = I$, whereas we have stipulated the gambler feels they cannot in good conscience assign good conditional probabilities to these events.

Suppose however that there is (yet) another set of mutually exclusive and exhaustive events $\{R'_i\}$ to which the gambler *has* assigned probabilities, and which are presumed related to the events $\{\lambda_i\}$. Even though the gambler is not informed enough to assign specific probabilities $P(R'|\lambda)$ or $P(\lambda|R)$, if they nevertheless believed that if they *did* assign such probabilities, that $P(R'|\lambda)$ and $P(\lambda|R)$ would both be rank r, that is, that there *would* exist a Φ such that

$$P(\lambda|R)\Phi P(R'|\lambda) = I, \tag{2.43}$$

then multiplying by $P(R'|\lambda)$ on the left and $P(\lambda|R)$ on the right, we find that this Φ must satisfy

$$P(R'|R)\Phi P(R'|R) = P(R'|R). \tag{2.44}$$

In fact, this is the defining equation for a $\{1\}$-inverse of $P(R'|R)$, and as we will see, it turns out that *every* matrix has at least one $\{1\}$-inverse. In other words, if the gambler believes (a) that $\{R_i\}$ and $\{R'_i\}$ form references for $\{\lambda_i\}$ and (b) that they have assigned probabilities $P(R'|R)$, then they can appeal to

$$P_2(E) = P(E|\lambda)\Big[P(\lambda|R)\Phi P(R'|\lambda)\Big]P(\lambda) = P(E|R)\Phi P(R'), \tag{2.45}$$

even if they have not actually assigned probabilities $P(\lambda|R)$ or $P(R'|\lambda)$ since Φ may now be defined as any $\{1\}$-inverse of $P(R'|R)$. We will later explore the equivalence between these two definitions: the key is that $P(R'|R)$ is itself rank r.

In practice, it may be that we can in some sense identify the events $\{R_i\}$ and $\{R'_i\}$. For example, the events $\{R_i\}$ might refer to an initial measurement on a system, after which "the same measurement" is repeated giving outcome $\{R'_i\}$. To identify the two means that all things being equal (e.g., if the order of the measurements were reversed) the gambler would assign the same probabilities in both instances, $P(R_i) = P(R'_i)$. In such a circumstance, we can write, as before,

$$P_2(E) = P(E|R)\Phi P(R), \tag{2.46}$$

where $P(R|R)\Phi P(R|R) = P(R|R)$. Crucially, neither expression depends on any probabilities assigned to $\{\lambda_i\}$: instead we rely upon the gambler's judgements $P(R_i|R_j)$, and their judgement that $\{R_i\}$ forms a reference for the events $\{\lambda_i\}$.

Nevertheless, in the absence of reasons to assign probabilities to the events $\{\lambda_i\}$, why should the gambler believe that the events $\{R_i\}$ form a reference for them? They may well have legitimate, independent reasons for believing so. But even if they don't, as we have observed, a $\{1\}$-inverse Φ exists for *any* matrix so that regardless of whether $\{R_i\}$ forms a reference, we can always calculate $P_2(E) = P(E|R)\Phi P(R)$ for Φ any $\{1\}$-inverse of $P(R|R)$ in the case that the events $\{R_i\}$ are called off. On the one hand, the gambler may hope that the $\{R_i\}$ form at least a partial reference for $\{\lambda_i\}$, so that $P_2(E) = P(E|R)\Phi P(R)$ would still have some advice to offer, albeit unreliable advice: think of $P(R|\lambda)$ having an approximate left inverse. On the other hand, if the gambler has independent reasons for assigning $P_2(E)$ directly, to the extent that $P_2(E) = P(E|R)\Phi P(R)$ is in fact satisfied by their mesh of beliefs, the gambler is the more justified in saying that $\{R_i\}$ is a good reference after all. Of course, just because the gambler assigns $P_2(E) \neq P(E|R)\Phi P(R)$ does not imply that $\{R_i\}$ is *not* a good reference: it might also be that the gambler simply ought to adjust their assignments $P_2(E)$, $P(E|R_j)$, and $P(R_i)$. In what follows, we will be able to treat this sort of question more systematically.

2.8 Pulling out the rug: nonclassicality

2.8.1 Hidden variables

So far we have been a bit fast and loose with our language, talking about events and outcomes which might occur or might be called off, like a horse race on a stormy day. There is nothing in itself wrong with this: there are many brands of subjective Bayesianism to which the development of probability theory we have given here would be acceptable. For example, in interpreting the construction of the previous section, we might gloss the events $\{\lambda_i\}$ as referring to whether or not something has this or that underlying property, which may not be accessible to observation, even as it conditions the results of observation. In other words, in the language of quantum foundations, we could consider the λ_i's to be *hidden variables*. The assumption that $P(R|R) = P(R|\lambda)P(\lambda|R)$ is in fact a nontrivial constraint, even if the λ_i's are "hidden." This is because we assume $\{R_i\}$ forms a reference for $\{\lambda_i\}$, which linear algebraically means that all three matrices must have rank r: $P(R|R) \in \mathbb{R}_r^{n\times n}$, $P(R|\lambda) \in \mathbb{R}_r^{n\times r}$ and $P(\lambda|R) \in \mathbb{R}_r^{r\times n}$. Put another way, the assumption is that $P(R|R)$, a stochastic matrix, has a *rank factorization* into the product of two stochastic matrices of the same rank as $P(R|R)$. We will return to this subject in Chapter 5. For now we simply observe that such a factorization cannot be found for just any stochastic matrix! The assumption that some $\{\lambda_i\}$ really occurs is thus a strong constraint on the probabilities $P(R|R)$ the gambler might assigns.

In a similar vein, we may interpret the measurement $\{R_i\}$ as disturbing these hidden variables, even as it forms a reference for them. On the one hand,

$$P(E|R)P(R) = P(E|\lambda)P(\lambda|R)P(R|\lambda)P(\lambda), \tag{2.47}$$

shows that $P(\lambda|R)P(R|\lambda)$ encodes the disturbance to the hidden variables as a result of the measurement, while

$$P(E|R)\Phi P(R) = P(E|\lambda)\Big[P(\lambda|R)\Phi P(R|\lambda)\Big]P(\lambda) = P(E|\lambda)P(\lambda) \tag{2.48}$$

allows us to appreciate that the role of Φ is precisely to take this disturbance into account and undo it.

We will return to this theme in a later chapter when we discuss notions of classicality. As discussed in the introduction, however, in light of quantum mechanics, the assumption that there are "hidden variables" of this sort is impossible in general to maintain. We therefore begin from an entirely different premise. The agent or gambler identifies something external to themselves as a system, external in the sense of beyond their control. They freely act upon that system, and their action has a consequence: the system is as free to respond to the agent's grasp as the agent was free to reach out. In particular, there is no presumption that the consequences are conditioned by some underlying properties of the system. From this point of view, measurements are nothing other than actions the agent takes on their world, and outcomes or events are nothing other than the consequences of these acts. So going forward, we will admit no events which are not consequences, or outcomes, of an act.

2.8.2 Classical reference measurements

Before we begin, however, let us review classical probability theory once more, specifically from the point of view of measurement theory. Having designated a part of the world as *the system*, one supposes that the system is "in" one of some number of states defined by some conjunction of properties, states which are mutually exclusive and exhaust all possibilities. We may denote them $\{\lambda_i\}$. For example, in a classical Newtonian system, each λ_i might correspond to the positions and momenta of some number of particles. This set may be finite or infinite: for simplicity of exposition, let us for now assume there are some finite number n of underlying states. Out of ignorance, one may assign probabilities $P(\lambda_i)$ to these underlying possible states of the system. A measurement with outcomes $\{E_i\}$ is formalized as a set of conditional probabilities $P(E_i|\lambda_j)$: given that the system is in the state λ_j, what probabilities ought one to assign to the different outcomes of the measurement? To say that $\{\lambda_i\}$ form a mutually exclusive and exhaustive set means that we believe exactly one of them is the case. Then we can invoke the law of total probability,

$$P(E_i) = \sum_j P(E_i|\lambda_j)P(\lambda_j). \tag{2.49}$$

By a reductionist principle, the $\{\lambda_i\}$ ought to be identified with the set Ω of atomic events upon which more general events $\{E_i\}$ supervene: each E_i is identified with a subset of Ω with the stipulation that subsets $\{E_i\}$ partition Ω into disjoint subsets. From this point of view, the different measurements one may perform upon the system correspond to different partitions

$\{E_i\}$ of Ω, different ways of carving up Ω into mutually disjoint subsets. It is assumed that exactly one of the λ_i's is true: and the outcome of any measurement is simply whichever outcome corresponds to a subset containing λ_i, the true state. But most measurements are coarse instruments for determining this λ_i.

Thus in classical probability theory there is an ultimate *reference measurement*, the one which distinguishes all the atomic events $\{\lambda_i\}$. For example, in classical physics, one may ideally read off the positions and momenta of a cloud of point particles without disturbing them in any way so that $P(\lambda|\lambda) = I$. In the formalism of the previous section, we have

$$P_1(E) = P(E|\lambda)P(\lambda) \qquad\qquad P_2(E) = P(E|\lambda)\Phi P(\lambda), \tag{2.50}$$

where Φ satisfies $P(\lambda|\lambda)\Phi P(\lambda|\lambda) = P(\lambda|\lambda)$. But $P(\lambda|\lambda) = I$ and so $\Phi = I$, and $P_2(E) = P_1(E)$. In other words, classically, one need not necessarily distinguish between the properties of the system and the outcomes of a reference measurement since the latter just reads off the former.

But what if we drop the assumption that a system is characterized by some definite properties independent of measurement? This may seem contradictory: how can a system be characterized at all if not by its properties? The resolution is that one may nevertheless retain the idea of a *reference measurement* and the Φ-formalism we have already developed can provide a series of norms to which a gambler can appeal.

2.8.3 System, reference, state, effect

Definition 2.8.1 (System)**.** *A* system *is identified by the agent as precisely something beyond their control but which they can nevertheless act upon.*

Definition 2.8.2 (Reference measurement)**.** *Let ρ_1 and ρ_2 be two outcomes of an agent's actions on a system. R is a reference measurement with outcomes $\{R_i\}$ iff*

$$\forall i : P(R_i|\rho_1) = P(R_i|\rho_2) \iff \forall E : P(E|\rho_1) = P(E|\rho_2). \tag{2.51}$$

In other words, if the agent assigns the same reference probabilities conditional on outcome ρ_1 or ρ_2, then they ought to assign the same probabilities to any *outcome E conditional on outcome ρ_1 or ρ_2.*

Definition 2.8.3 (State)**.** *A* state *is an equivalence class of measurement outcomes conditional upon all of which a gambler would assign the same probabilities to a subsequent reference measurement–and so to all measurements.*

Remark. *In the definition of a reference measurement, ρ_1 and ρ_2 would belong to the equivalence class ρ. The gambler calls ρ a* state *since for them each outcome in the equivalence class implies the same "state of expectation" with respect to the reference measurement. We will then write simply $P(E|\rho)$: in practice ρ will be implemented by*

some particular member of the equivalence class, and we will often interchangeably refer to ρ and $P(R|\rho)$ as a state. Indeed, if the gambler has a particular "state of expectation" with respect to the reference measurement, it must be the result of some constellation of circumstances, the consequences of many actions. Considering these as a whole, this single collective consequence must lie by definition in the equivalence class.

Remark. *From the definition, a reference measurement* separates *distinct states: conversely, if two states yield all the same probabilities with respect to reference measurement, they are really one state after all. Although the outcomes in the equivalence class might be genuinely different experiences for the agent, the gambler believes these differences make no difference at all to their intercourse with the system.*

Remark. *A* state preparation *e.g., of ρ, is an idealization: the agent takes an action among whose consequences is a desired member of the equivalence class ρ, and to which consequence the agent assigns probability 1.*

Definition 2.8.4 (Preparatory measurement)**.** *A* preparatory measurement *is any measurement with outcomes $\{\rho_i\}$ such that the gambler assigns*

$$\forall i, j : P(R_i|\rho_j) = P(R_i|\rho_j, \textit{any other consequences}). \tag{2.52}$$

In other words, the gambler would assign the same reference probabilities conditional on outcome ρ_i regardless of any other consequences: for the gambler, the outcome ρ_i is all that matters as far as the system is concerned.

Definition 2.8.5 (Reference states)**.** *Let E_1 and E_2 be two different outcomes, not necessarily of the same measurement. $\{R_i\}$ is a set of* reference states *iff*

$$\forall i : P(E_1|R_i) = P(E_2|R_i) \iff \forall \rho : P(E_1|\rho) = P(E_2|\rho). \tag{2.53}$$

In other words, if the agent assigns the same probabilities to outcomes E_1 and E_2 conditional on the reference states, then they ought to assign the same probabilities to outcomes E_1 and E_2 conditional on any *state.*

Definition 2.8.6 (Effect)**.** *An* effect *is an equivalence class of measurement outcomes to which the gambler would assign the same probabilities conditional on a set of reference states—and thus they would assign the same probabilities conditional on any state.*

Remark. *If E_1 and E_2 lead to the same "effect" conditional on reference outcomes, they ought to lead to the same "effect" conditional on any outcome. Thus we call E an* effect*, and talk of measuring E as a short hand for some member of the equivalence class of which E_1 and E_2 are a part. A set of reference states* separate *distinct effects: conversely, if two effects yield all the same probabilities with respect to the reference states, they must be the same effect after all. We will often interchangably refer to E and $P(E|R)$ as an effect.*

Remark. *We assume that the reference measurement is a preparatory measurement, and that the reference outcomes are reference states. Thus conditioning on a reference outcome leaves subsequent measurement outcomes independent of any other consequences, and the same reference measurement separates states and effects.*

2.8.4 State space

Given a reference measurement, let us now consider the set $\mathcal{S}$ of probability distributions $\{P(R|\rho)\}$ which the gambler may consistently assign to its outcomes. Equivalently $\mathcal{S}$ is the set of states $\{\rho\}$, and we shall call the set $\mathcal{S}$ the *state space*. Suppose the agent performs a preparatory measurement with outcomes $\{\pi_i\}$ followed by the reference measurement. By the law of total probability,

$$P(R|\rho) = \sum_i P(R|\pi_i)P(\pi_i|\rho). \tag{2.54}$$

Since the preparatory measurement can be chosen arbitrarily, we conclude that $\mathcal{S}$ is a *convex set*: if $\{P(R|\pi_i)\}$ is a set of distributions the gambler is willing to assign to the reference measurement, then the gambler ought to also be willing to assign $P(R|\rho) = \sum_i P(R|\pi_i)p_i$ for probabilities p_i summing to unity[3].

Definition 2.8.7 (Convex set)**.** *A* convex set *S is a subset of a vector space such that* $\forall x, y \in S, p \in [0,1] : px + (1-p)y \in S$.

For convenience, we recall here some elementary facts [94] about convex sets. An *extreme point* of a convex set is any point which cannot be written as a nontrivial convex combination of different points in S: x is extreme whenever $x = py + (1-p)z$ for $y, z \in S$ and $p \in (0,1)$ implies $y = z = x$. By the Heine-Borel theorem, in $\mathbb{R}^n$ a set S is *compact* if and only if it is *closed* and *bounded*. A set is *closed* if it includes its boundary: if we have a sequence of points $\lim_{n\to\infty} x_n = x$, then the limit point $x \in S$. A set is *bounded* if it can be enclosed in a sphere of finite radius. We note that continuous functions on compact sets are bounded and attain their minimum and maximum values [95, 96]. By the Krein-Milman theorem, a compact convex set is the *convex hull* of its extreme points [97], and given a set X,

[3]Alternatively, as in [24], one may appeal to the reflection principle. We saw that $P_0(A) = \sum_i q_i P_0(P_\tau(A) = q_i)$. In particular,

$$P_0(R) = \sum_j P(R)P_0(P_\tau(R) = P(R)) = \sum_j P(R)p_j, \tag{2.55}$$

for probabilities p_j. Here the convexity of the state space comes from fact that the gambler ought to adopt now mixtures of state assignments they might make in the future.

its convex hull is

$$\mathrm{conv}(X) = \left\{ \sum_i p_i x_i : x_i \in X, p_i \geq 0, \sum_i p_i = 1 \right\}. \tag{2.56}$$

Since $\mathcal{S} \subseteq \Delta_n$, where Δ_n is the probability simplex on n outcomes, the state space is clearly bounded, and since we define $\mathcal{S}$ in terms of the convex hull of the distributions the gambler is willing to assign to the reference measurement, we may take $\mathcal{S}$ to be closed, and therefore compact. Finally, it is often convenient to work with unnormalized probability vectors, including even the 0 vector, giving us what we shall call the *state cone* $\hat{\mathcal{S}}$, a proper convex cone.

Definition 2.8.8 (Convex cone)**.** *A* cone *C is a subset of a vector space V such that if $x \in C$, then $\lambda x \in C$ for $\lambda > 0$. A cone is* convex *if for any $x, y \in C$ and $\lambda, \mu > 0$, $\lambda x + \mu y \in C$. A convex cone is* pointed *if $C \cap (-C) = \{0\}$ [27]: geometrically this means C contains no full line through the origin. A convex cone is* proper *if it is pointed and closed. A convex cone is* spanning *if $C - C = V$.*

We recall some more useful facts. A ray of a cone $\{\lambda x : \lambda \geq 0\}$ generated by a point $x \in X$ is *extremal* whenever $x = y + z$ for $y, z \in C$ implies there exist $\alpha, \beta \geq 0$ such that $y = \alpha x$ and $z = \beta x$. The *conic hull* of a set X is

$$\mathrm{cone}(X) = \left\{ \sum_i \lambda_i x_i : x_i \in X, \lambda_i \geq 0 \right\}, \tag{2.57}$$

and finally, a closed pointed finite dimensional convex cone is the conic hull of its extreme rays [94]. A *base* of a cone is any nonempty convex set B within the cone not containing the origin such that every nonzero point in the cone can be written as $x = \lambda b$ for $\lambda > 0$ and $b \in B$. We conclude that the state space itself is the base of the state cone, picked out by the hyperplane on which $\sum_i P(R_i|\rho) = 1$.

2.8.5 Effect space

Let us now consider the set $\mathcal{E}$ of response functions $\{P(E|R)\}$ which the gambler may consistently assign conditional on the outcomes of the reference measurement. Equivalently, $\mathcal{E}$ is the set of effects $\{E\}$, and we shall call $\mathcal{E}$ the *effect space*. Now on the one hand, the gambler may always coarse grain outcomes of a measurement,

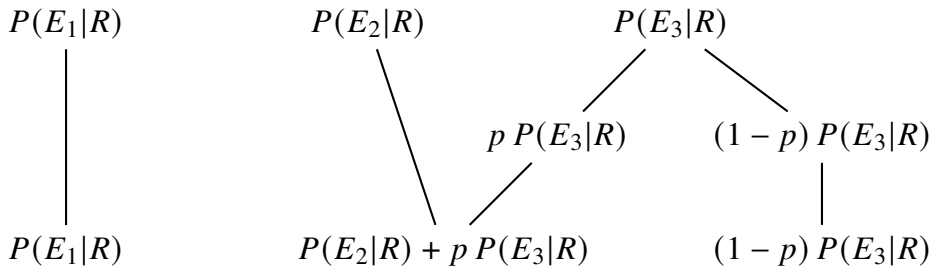
$$P(E_1 \cup E_2|R) = P(E_1|R) + P(E_2|R), \tag{2.58}$$

treating two outcomes as equivalent. On the other hand, the gambler may always fine grain outcomes of a measurement e.g., by performing another measurement conditional on an

outcome of the first, which the gambler believes only depends upon that outcome, and treating two step measurement as a single measurement. Clearly,

$$P(E_1|R) = \sum_i P(E_i^{(1)}|E_1)P(E_1|R), \tag{2.59}$$

since $\sum_i P(E_i^{(1)}|E_1) = 1$. By alternating fine graining and coarse graining, from a single measurement, we may obtain arbitrary combinations $\sum_i P(E_i|R)p_i$ for $p_i \in [0, 1]$ but where $\{p_i\}$ need not sum to 1. For example:



As before when we extended the state space to a state cone, and then restricted ourselves to the base of normalized states, let us consider for mathematical convenience an *effect cone* $\hat{\mathcal{E}}$, where we allow any conic combination of effects. In particular, two effects play a special role. On the one hand, let 0 denote the equivalence class of outcomes the gambler judges to be impossible. Then $P(0|R) = 0$ must be the zero vector. On the other hand, let 1 denote the equivalence class of outcomes the gambler judges to be necessary. Then $P(1|R) = (1, \ldots, 1)$ must be the row vector of all 1's. Since $\hat{\mathcal{E}}$ is a subset of the nonnegative orthant and contains the zero vector $P(0|R)$, the effect cone must be a pointed convex cone.

But we can say more. Given any outcome E, we may always consider the outcome that E does not occur, that is, the complement of E, with probability $P(\neg E) = 1 - P(E)$. Thus $\forall E : P(\neg E|R) = P(1|R) - P(E|R) \in \mathcal{E}$: the "opposite effect" $P(\neg E|R)$ must also be an effect. Now given any proper convex cone C, we can define a partial order[4] $x \leq y$ iff $y - x \in C$. In particular, the effect cone defines a partial order on effects, and

$$P(\neg E|R) = P(1|R) - P(E|R) \in \hat{\mathcal{E}} \iff P(E|R) \leq P(1|R). \tag{2.60}$$

Moreover, $P(E|R) - P(0|R) \in \hat{\mathcal{E}}$ so that in terms of the partial ordering induced by the effect cone

$$\forall E : P(0|R) \leq P(E|R) \leq P(1|R). \tag{2.61}$$

[4]Recall that a partial order satisfies reflexivity, antisymmetry, and transitivity.

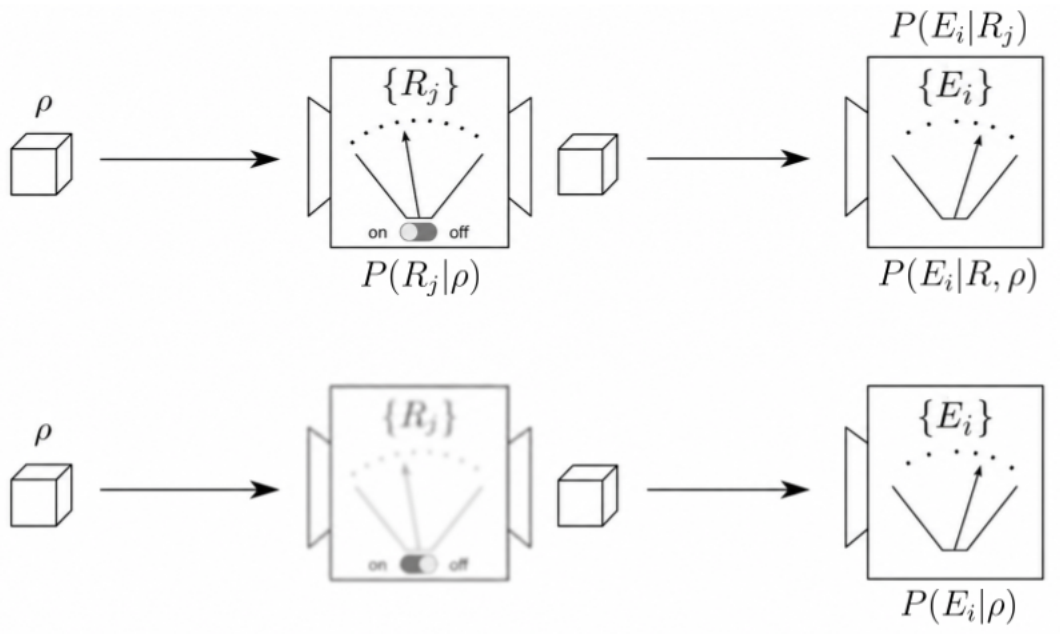


Figure 2.1: Two scenarios [2].

We thus recover the idea that we may take nonnegative combinations of effects only insofar as the sum can appear in a measurement, whose effects must sum to $P(1|R) = (1, \ldots, 1)$. The effect space $\mathcal{E}$ is thus itself a compact convex set, a subset of the hypercube Δ_n^* of response functions, and the intersection of two cones, one opening upwards from $P(0|R)$ and the other opening downwards from $P(1|R)$.

The picture that has thus emerged is that normalized states $P(R|\rho)$ live in the state cone $\hat{\mathcal{S}}$, a proper closed convex cone, and in particular, on the base picked out by $P(1|R) = (1, \ldots, 1)$. All this is taking place within the probability simplex Δ_n on n outcomes. Meanwhile, the effects $P(E|R)$ live in the effect cone $\hat{\mathcal{E}}$, also a proper closed convex cone. Effects in the effect space itself satisfy $P(0|R) \leq P(E|R) \leq P(1|R)$ with respect to the partial ordering defined by the effect cone, and all this takes place within the space of response functions on n outcomes Δ_n^*.

2.8.6 Nonclassical consistency

Let us now consider the two scenarios depicted schematically in Fig. 2.1 in terms of "measuring devices." The box in the middle is a reference device, which can be turned on or off. In the first scenario, the reference device is "on" signifying that after the outcome ρ of a preparatory measurement, the agent performs a reference measurement $\{R_i\}$ followed by an arbitrary measurement $\{E_i\}$. Since the outcomes of both measurements really occur, a

gambler betting on their outcomes is urged to adopt the law of total probability,

$$P(E_i|R,\rho) = \sum_j P(E_i|R_j)P(R_j|\rho), \tag{2.62}$$

to check the consistency of their assignments. In the second scenario, after the preparatory measurement yields outcome ρ, the agent measures $\{E_i\}$ directly: the reference measurement is "turned off." We ask again: is there any norm that we can appeal to that constrains $P(E_i|\rho)$ in terms of $P(R_j|\rho)$ and $P(E_i|R_j)$? There ought to be, precisely because $\{R_i\}$ is a reference measurement: by definition, the probabilities $P(R_i|\rho)$ fully characterize the state ρ, and the conditional probabilities $P(E_i|R_j)$ fully characterize the effects $\{E_i\}$. Thus $P(E_i|\rho)$ ought to be *some* function of these two distributions,

$$P(E_i|\rho) = \mathcal{F}\big(P(E_i|R), P(R|\rho)\big). \tag{2.63}$$

We now show that $\mathcal{F}(\cdot,\cdot)$ must take a particular form. Our argument is indebted to the one given in [57] which itself was inspired by [49, 98, 99].

Theorem 2.8.1. *Let $\{R_i\}_{i=1}^n$ be a reference measurement. The function $\mathcal{F}(\cdot,\cdot)$ in $P(E|\rho) = \mathcal{F}\big(P(E|R), P(R|\rho)\big)$ which expresses the coherence condition the gambler ought to impose on their probability assignments when the reference measurement remains counterfactual, is bilinear. In particular,*

$$P(E|\rho) = P(E|R)\Phi P(R|\rho), \tag{2.64}$$

where Φ is any matrix satisfying $P(R|R)\Phi P(R|R) = P(R|R)$, $P(R|R)$ being the conditional probability matrix which characterizes the reference measurement itself.

Proof. Suppose that before the reference measurement, the gambler performs a preparatory measurement $\{\rho_i\}$. If conditional on ρ_i, the gambler would assign $P(R|\rho_i)$ to the reference measurement, then they ought also to be willing to assign $P(R|\rho) = \sum_i P(R|\rho_i)P(\rho_i)$ to the reference measurement. The same applies should the gambler perform some arbitrary measurement with outcome E instead: $P(E|\rho) = \sum_i P(E|\rho_i)P(\rho_i)$. From

$$P(E|\rho) = \mathcal{F}\big(P(E|R), P(R|\rho)\big) = \mathcal{F}\big(P(E|R), \sum_i P(R|\rho_i)P(\rho_i)\big) \tag{2.65}$$

$$= \sum_i P(E|\rho_i)P(\rho_i) = \sum_i \mathcal{F}\big(P(E|R), P(R|\rho_i)\big)P(\rho_i), \tag{2.66}$$

we conclude that $f(x) = \mathcal{F}(P(E|R), x)$ for some fixed $P(E|R)$ preserves convex combinations on the state space $\mathcal{S}$,

$$f\big(px + (1-p)y\big) = pf(x) + (1-p)f(y) \qquad p \in [0,1] \qquad x, y \in \mathcal{S}. \tag{2.67}$$

We shall now extend f to a series of functions defined on progressively more general domains. First, we define f_1 not just on the state space $\mathcal{S}$ but also the state cone $\hat{\mathcal{S}}$. Now in general, any vector x in a cone can be written uniquely as λy for $\lambda > 0$ and y in a base of the cone: the only exception to uniqueness is when $x = 0$. In our case, since the base of the state cone is picked out by the normalization constraint on probability vectors, any $x \in \hat{\mathcal{S}}$ can be written $x = \overline{x}y$ for $y \in \mathcal{S}$ itself, where $\overline{x} = \sum_i x_i$. We thus define $f_1(x) = \overline{x} f(x/\overline{x})$. Clearly this agrees with f on $\mathcal{S}$. Further, we assume that $f_1(0) = P(E|0) = \mathcal{F}\big(P(E|R), P(R|0)\big) = \mathcal{F}\big(P(E|R), 0\big) = 0$ so that f_1 is centered since any event conditioned on the impossible event ought to be assigned zero probability.

From the definition $f_1(x) = \overline{x} f(x/\overline{x})$, we have for $t > 0$,

$$f_1(tx) = \overline{tx} f\left(\frac{tx}{\overline{tx}}\right) = t\overline{x} f\left(\frac{x}{\overline{x}}\right) = t f_1(x), \tag{2.68}$$

so that in fact f_1 is homogeneous for $t \geq 0$. Moreover, f_1 is also additive on $\hat{\mathcal{S}}$:

$$\begin{aligned} f_1(x+y) &= \overline{x+y} f\left(\frac{x+y}{\overline{x+y}}\right) = (\overline{x}+\overline{y}) f\left(\frac{\overline{x}}{\overline{x}+\overline{y}}\frac{x}{\overline{x}} + \frac{\overline{y}}{\overline{x}+\overline{y}}\frac{y}{\overline{y}}\right) && (2.69)\\ &= (\overline{x}+\overline{y})\left[\frac{\overline{x}}{\overline{x}+\overline{y}} f\left(\frac{x}{\overline{x}}\right) + \frac{\overline{y}}{\overline{x}+\overline{y}} f\left(\frac{y}{\overline{y}}\right)\right] = \overline{x} f\left(\frac{x}{\overline{x}}\right) + \overline{y} f\left(\frac{y}{\overline{y}}\right) && (2.70)\\ &= f_1(x) + f_1(y). && (2.71) \end{aligned}$$

We now extend the domain to include $\hat{\mathcal{S}} - \hat{\mathcal{S}} = \mathrm{span}(\mathcal{S})$. By definition, any vector $x \in \mathrm{span}(\mathcal{S})$ can be written $x = x^+ - x^-$ for $x^+, x^- \in \hat{\mathcal{S}}$. Thus let $\forall x \in \mathrm{span}(\mathcal{S}) : f_2(x) = f_1(x^+) - f_1(x^-)$. Clearly, for vectors in the cone, f_2 agrees with f_1. In fact, we could use *any* decomposition of x into the difference of two vectors in $\hat{\mathcal{S}}$. Let $x = x^+ - x^- = y^+ - y^-$ be two such decompositions. Then $x^+ + y^- = y^+ + x^- \in \hat{\mathcal{S}}$ so that

$$\begin{aligned} f_1(x^+ + y^-) &= f_1(y^+ + x^-) && (2.72)\\ f_1(x^+) + f_1(y^-) &= f_1(y^+) + f_1(x^-) && (2.73)\\ f_1(x^+) - f_1(x^-) &= f_1(y^+) - f_1(y^-) && (2.74)\\ f_2(x^+ - x^-) &= f_2(y^+ - y^-) = f_2(x). && (2.75) \end{aligned}$$

It is straightforward to see that f_2 is also homogeneous for $t \geq 0$ and centered. It is also additive on $\hat{\mathcal{S}} - \hat{\mathcal{S}}$. Let $x = x^+ - x^-$ and $z = z^+ - z^-$. Since $x + z = (x^+ + z^+) - (x^- + z^-)$,

$$f_2(x+z) = f_1(x^+ + z^+) - f_1(x^- + z^-) = f_1(x^+) + f_1(z^+) - f_1(x^-) - f_1(z^-) = f_2(x) + f_2(z). \tag{2.76}$$

Since f_2 is additive on $\hat{\mathcal{S}} - \hat{\mathcal{S}}$, we have $f_2(0) = f_2(x + (-x)) = f_2(x) + f_2(-x) = 0$ showing that $f_2(-x) = -f_2(x)$. Let $t = -s < 0$. Then $f_2(tx) = f_2(-sx) = -f_2(sx) = -s f_2(x) = t f_2(x)$ so

that f_2 is in fact homogeneous over all of $\mathbb{R}$. We have thus established that f_2 is homogeneous over $\mathbb{R}$ and additive on $\hat{\mathcal{S}} - \hat{\mathcal{S}}$: in other words, it is linear on span($\mathcal{S}$). Finally, we extend our function (not uniquely) to a linear function on the entirety of the n-dimensional vector space V. Let $\{b_i\}$ be an orthonormal basis for span($\mathcal{S}$) and $\{c_i\}$ be an orthonormal basis for its complement: together they form a basis for V. Let $f_3(b_i) = f_2(b_i)$ while $f_3(c_i) = a_i$ may be arbitrary, e.g., 0. We may decompose any vector $x = x_b + x_c$, where x_b lies in span($\mathcal{S}$) and x_c lies in its complement. Since f_3 is assumed linear,

$$f_3(x) = f_3(x_b + x_c) = f_3(x_b) + f_3(x_c) = \sum_i (b_i \cdot x_b) f_3(b_i) + \sum_i (c_i \cdot x_c) f_3(c_i) \tag{2.77}$$

$$= \sum_i (b_i \cdot x_b) f_2(b_i) + \sum_i (c_i \cdot x_c) a_i. \tag{2.78}$$

We have thus constructed a function $f_3(x)$ such that when x is fully supported on span($\mathcal{S}$), it agrees with $f_2(x)$; and when x is in the state cone $\hat{\mathcal{S}}$, $f_2(x)$ agrees with $f_1(x)$; and when x is in the state space $\mathcal{S}$, $f_1(x)$ agrees with our original function $f(x) = \mathcal{F}(P(E|R), x)$. Since f_3 is linear on the entire vector space, and its restriction to $\mathcal{S}$ agrees with f, we conclude that f itself is linear on its domain. Thus $\mathcal{F}(\cdot, \cdot)$ is linear in its second argument.

Let us now consider the first argument. Since we may always coarse grain measurements, letting E_{12} be the event that E_1 or E_2 befalls the agent, we must have

$$P(E_{12}|\rho) = \mathcal{F}\big(P(E_{12}|R), P(R|\rho)\big) = \mathcal{F}\big(P(E_1|R) + P(E_2|R), P(R|\rho)\big) \tag{2.79}$$

$$= P(E_1|\rho) + P(E_2|\rho) = \mathcal{F}\big(P(E_1|R), P(R|\rho)\big) + \mathcal{F}\big(P(E_2|R), P(R|\rho)\big), \tag{2.80}$$

so that $\mathcal{F}(\cdot, \cdot)$ is additive in its first argument on the effect space $\mathcal{E}$. Notice that we assume that E_1 and and E_2 are part of the same measurement so that $P(E_1|R) + P(E_2|R) \leq P(1|R)$. But we may also fine grain measurements, performing a subsequent measurement conditional on the outcome of the first which the gambler believes depends only upon the outcome of the first. Letting E_k denote the outcome of the first measurement and F_j the outcome of the second, we have $P(E_{jk}|\rho) = P(F_j, E_k|\rho) = P(F_j|E_k)P(E_k|\rho)$ where by writing $P(E_{jk}|\rho)$ we have treated the two measurements as one. Thus

$$P(E_{jk}|\rho) = \mathcal{F}\big(P(E_{jk}|R), P(R|\rho)\big) = \mathcal{F}\big(P(F_j|E_k)P(E_k|R), P(R|\rho)\big) \tag{2.81}$$

$$= P(F_j|E_k)P(E_k|\rho) = P(F_j|E_k)\mathcal{F}\big(P(E_k|R), P(R|\rho)\big), \tag{2.82}$$

so that $\mathcal{F}(\cdot, \cdot)$ is homogeneous for $p \in [0, 1]$. Let $g(x) = \mathcal{F}(x, P(R|\rho))$ for some fixed $P(R|\rho)$: we have shown that g is additive on the effect space $\mathcal{E}$ and homogeneous for $p \in [0, 1]$, where we recall that the effect space is the subset of the effect cone satisfying $P(0|R) \leq P(E|R) \leq P(1|R)$.

Now any vector in the effect cone can be written not uniquely $x = \lambda_x \tilde{x}$ for $\lambda_x > 0$ and $\tilde{x} \in \mathcal{E}$. We define $g_1(x)$ on the entire effect cone by taking $g_1(x) = \lambda_x g(\tilde{x})$. In fact it

doesn't matter which decomposition of x we choose. Let $x = \lambda_x \tilde{x} = \lambda'_x \tilde{x}'$ and define $g_1(x)$ in terms of the former and $g'_1(x)$ in terms of the latter. We have $\tilde{x}' = (\lambda_x/\lambda'_x)\tilde{x}$, and without loss of generality we may assume $\lambda_x < \lambda'_x$. Then $g(\tilde{x}') = g\left((\lambda_x/\lambda'_x)\tilde{x}\right) = (\lambda_x/\lambda'_x)g(\tilde{x})$ so that $\lambda'_x g(\tilde{x}') = \lambda_x g(\tilde{x})$, and thus $g'_1(x) = g_1(x)$. We now show that the extension is homogeneous for $t \geq 0$. Let $t > 1$. If $x = \lambda_x \tilde{x}$ for $\tilde{x} \in \mathcal{E}$, then $tx = t\lambda_x \tilde{x} = \lambda_{tx}\tilde{x}$, so that we may take $\lambda_{tx} = t\lambda_x$. Since $1/t \in [0, 1]$, we have $g_1(tx) = \lambda_{tx} g(tx/\lambda_{tx}) = t\lambda_x g(x/\lambda_x) = t g_1(x)$ so that g_1 is homogeneous for $t \geq 0$. Finally, we show that the extension is additive on $\hat{\mathcal{E}}$. Let $x + y = \lambda_x \tilde{x} + \lambda_y \tilde{y}$ for $\tilde{x}, \tilde{y} \in \mathcal{E}$ and $\lambda_x, \lambda_y > 0$. Let

$$\tilde{z} = \frac{\lambda_x}{\lambda_x + \lambda_y}\tilde{x} + \frac{\lambda_y}{\lambda_x + \lambda_y}\tilde{y} \in \mathcal{E}, \tag{2.83}$$

so that $z = (\lambda_x + \lambda_y)\tilde{z} = x + y$. Then

$$g(\tilde{z}) = \frac{\lambda_x}{\lambda_x + \lambda_y} g(\tilde{x}) + \frac{\lambda_y}{\lambda_x + \lambda_y} g(\tilde{y}), \tag{2.84}$$

so that $(\lambda_x + \lambda_y)g(\tilde{z}) = \lambda_x g(\tilde{x}) + \lambda_y g(\tilde{y})$ or $g_1(x + y) = g_1(x) + g_1(y)$. We conclude g_1 is also additive on $\hat{\mathcal{E}}$. We now extend the domain to include $\hat{\mathcal{E}} - \hat{\mathcal{E}} = \mathrm{span}(\mathcal{E})$. Again we decompose $x = x^+ - x^-$ for $x^+, x^- \in \hat{\mathcal{E}}$ and define $g_2(x) = g_1(x^+) - g_1(x^-)$. By the same argument we used in examining the state dependence, g_2 must be homogeneous over $\mathbb{R}$ and additive on $\hat{\mathcal{E}} - \hat{\mathcal{E}}$, that is, linear on $\mathrm{span}(\mathcal{E})$. Finally, as before, we may then extend g_2 to a linear function g_3 defined on the whole vector space which agrees with g_2 on $\mathrm{span}(\mathcal{E})$. Since g_3 is linear, and the restriction of g_3 to the effect space $\mathcal{E}$ agrees with g, we conclude that $\mathcal{F}(\cdot, \cdot)$ is linear in its first argument.

Putting these two results together, we conclude that $\mathcal{F}(\cdot, \cdot)$ is a bilinear function, and therefore can be represented as a matrix Φ,

$$P(E|\rho) = \mathcal{F}\big(P(E|R), P(R|\rho)\big) = P(E|R)\Phi P(R|\rho). \tag{2.85}$$

Suppose that the arbitrary measurement $\{E_i\}$ is in fact $\{R_i\}$, the reference measurement itself. For consistency, we require

$$P(R|\rho) = P(R|R)\Phi P(R|\rho), \tag{2.86}$$

and specializing to the reference states, we require

$$P(R|R) = P(R|R)\Phi P(R|R). \tag{2.87}$$

But this is the defining equation of a $\{1\}$-inverse of the conditional probability matrix $P(R|R)$, which characterizes the reference states in terms of the reference measurement, or equivalently, the reference measurement in terms of reference states. □

Remark. *In fact,* $P(R|\rho) = P(R|R)\Phi P(R|\rho)$ *and* $P(E|R) = P(E|R)\Phi P(R|R)$ *imply that any* $P(R|\rho)$ *must lie in the column space of* $P(R|R)$ *and any* $P(E|R)$ *must lie in the row space of* $P(R|R)$*. We will discuss this in more detail in the next chapter. Only if these column and row space conditions are met will* $P(R|\rho) = P(E|R)\Phi P(R|\rho)$ *give identical results for any choice of* Φ*.*

We conclude that in our two scenarios, a gambler ought to check their probability judgements against the following two norms,

$$P(E|R,\rho) = P(E|R)P(R|\rho) \qquad\qquad P(E|\rho) = P(E|R)\Phi P(R|\rho), \tag{2.88}$$

where Φ is any $\{1\}$-inverse of the conditional probability matrix $P(R|R)$ which characterizes the reference measurement itself. In the first scenario, the reference measurement is performed in between the preparatory measurement (with outcome ρ) and the final measurement. $P(E|R,\rho)$ is no different than in the classical case: it is given simply by the law of total probability. In the second scenario, the reference measurement is counterfactual. But since the reference probabilities characterize the states and effects completely, there must be some way of expressing the probability $P(E|\rho)$ in terms of them. We have shown that the right normative rule is given by a deformation of the law of total probability by the interposition of the Born matrix Φ.

In this way, we can supplement standard probability theory to handle situations where systems cannot be defined by properties independent of measurement but which *can* nevertheless be fully characterized by the probabilities an agent assigns to a reference measurement, in the sense that if two preparations yield the same probabilities on the reference, they must yield the same probabilities on all measurements. In this way, we weaken the classical notion of a system, without sacrificing the idea that an agent is served well by imposing on themselves a notion of consistency. In a sense, the outcomes of the reference play the role of the atomic sample space. But now other events, e.g., the outcomes of other measurements, are not simple stochastic coarse grainings or subsets of the atomic events, as we saw in Eq. (2.16), $P(E) = \sum_{\omega\in\Omega} \delta_{\omega\in E} P(\omega)$. Instead, we have *quasistochastic* regraining as in $P(E|\rho) = P(E|R)\Phi P(R|\rho)$. Why "regraining"? Because the measurement $\{E_i\}$ may have not just fewer, but even more outcomes than the reference! This is quite unlike the classical case.

2.8.7 Φ-duality

Now that we have established our nonclassical coherence rule $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ as well as the basic geometry of the state and effect cones, we can define a notion of Φ-duality. Call $\hat{\mathcal{S}}^\star$, the cone Φ-dual to the state cone, in the following sense,

$$\hat{\mathcal{S}}^\star = \{y \in V^* |\ \forall x \in \hat{\mathcal{S}} : y\Phi x \geq 0\}. \tag{2.89}$$

Clearly, $\hat{\mathcal{E}} \subseteq \hat{\mathcal{S}}^\star$. Notice that this is not the usual notion of duality for cones due to the presence of the Born matrix Φ [27]. Similarly, we may consider the cone Φ-dual to the effect cone,

$$\hat{\mathcal{E}}^\star = \{x \in V | \; \forall y \in \hat{\mathcal{E}} : y\Phi x \geq 0\}, \tag{2.90}$$

so that so that $\hat{\mathcal{S}} \subseteq \hat{\mathcal{E}}^\star$.

2.9 "Dynamics"

So far we have focused on the case of preparatory measurements $\{\rho_i\}$ such that $P(R|\rho_i) = P(R|\rho_i, \text{any other events})$: the agent believes only the outcome ρ_i is relevant as far as the reference measurement is concerned. We have focused on this case because the reference measurement itself is taken to be preparatory, and we have only considered the probabilities for some final measurement $\{E_i\}$. But the gambler may update their state assignment conditional on the result of a measurement in a more general way. For example, consider the following sequence of events: the gambler performs a preparatory measurement with outcome ρ, followed by a measurement $\{E_i\}$, which is then followed by a reference measurement $\{R_i\}$. We can consider the latter two to form a composite measurement with outcomes $\{E_i \text{ and } R_j\}$. We can then compare this to the situation where after obtaining outcome ρ, the gambler first performs the reference measurement $\{R_i\}$, followed by $\{E_i\}$, followed by $\{R_i\}$ again. We may relate the two situations, as we have seen, using the Born matrix. We have

$$P(R_i, E_j|\rho) = \sum_{lm} P(R_i, E_j|R_l)\Phi_{lm}P(R_m|\rho), \tag{2.91}$$

so that

$$P(R_i|E_j, \rho) = \frac{1}{P(E_j|\rho)} \sum_{lm} P(R_i, E_j|R_l)\Phi_{lm}P(R_m|\rho). \tag{2.92}$$

Suppose now that we are interested in how the gambler should update their state assignment in time, or better said, how the gambler ought to harmonize their beliefs now about the present and their beliefs now about the future. In order to formalize this, we again invoke the reflection principle. We saw earlier that $P_0(E) = \sum_i q_i P_0(P_\tau(E) = q_i)$, where $P_0(E)$ is the probability the gambler assigns at $t = 0$ to E and $P_0(P_\tau(E) = q_i)$ is the probability the gambler assigns at $t = 0$ to the possibility that they will assign probability q_i to event E at time $t = \tau$. Now suppose the agent plans to undertake a sequence of measurements: first, a preparatory measurement with outcome ρ at t_0; at t_1, a measurement with outcomes $\{E_i\}$; and at t_2, a measurements with outcomes $\{F_i\}$. Let τ be a time after the measurement at t_1

but before the measurement at t_2. By the law of total probability,

$$P_0(F_j) = \sum_i P_0\Big(F_j\Big|P_\tau(F_j) = P(F_j|E_i,\rho)\Big)P_0\Big(P_\tau(F_j) = P(F_j|E_i,\rho)\Big). \tag{2.93}$$

On the one hand, by reflection,

$$P_0\Big(F_j\Big|P_\tau(F_j) = P(F_j|E_i,\rho)\Big) = P(F_j|E_i,\rho). \tag{2.94}$$

On the other hand,

$$P_0\Big(P_\tau(F_j) = P(F_j|E_i,\rho)\Big) = P(E_i|\rho), \tag{2.95}$$

since now at t_0 the agent naturally believes that when $t = \tau$ rolls around they will assign $P(F_j|E_i,\rho)$ to getting outcome F_j conditional to E_i with probability $P(E_i|\rho)$. Putting these two statements together, and expressing all probabilities with respect to the reference measurement,

$$\begin{aligned} P_0(F_j) &= \sum_i P(F_j|E_i,\rho)P(E_i|\rho) && (2.96)\\ &= \sum_i \big\{P(F_j|R)\Phi P(R|E_i,\rho)\big\}P(E_i|\rho) && (2.97)\\ &= \sum_i P(F_j|R)\Phi\left\{\frac{1}{P(E_i|\rho)}P(R,E_i|R)\Phi P(R|\rho)\right\}P(E_i|\rho) && (2.98)\\ &= P(F_j|R)\Phi\left\{\sum_i P(R,E_i|R)\right\}\Phi P(R|\rho), && (2.99) \end{aligned}$$

where in the second to last line we've used Eq. (2.92). Thus reflection implies that at t_0, the agent should assign a probability $P_0(F_j)$ which is obtained from the initial state $P(R|\rho)$ by averaging over the updating maps associated to each outcome E_i of the intermediate measurement.

Let us now suppose that the agent is in fact indifferent to whether they perform the intermediate measurement $\{E_i\}$ or not. Indifference means that they would assign the same probabilities at $t = \tau$, just before the measurement at t_2, regardless of whether they end up performing the measurement or not. But this just means that they should assign

$$P_0(F_j) = P(F_j|R)\Phi\left\{\sum_i P(R,E_i|R)\right\}\Phi P(R|\rho), \tag{2.100}$$

even if the $\{E_i\}$ measurement is not performed. Because of the agent's indifference, even if they don't perform the measurement, they should update their state $P(R|\rho)$ using $\sum_i P(R, E_i|R)$. This is principle of dynamics in our framework which takes an agent's actions and their consequences to be primary: to determine the dynamics, that is, how the gambler ought to update their state assignment in time *in the absence of measurement*, the gambler ought to identify precisely those measurements which according to their mesh of probabilities they are indifferent to. Because of this indifference, they ought to use the same update rule even in the measurement's absence.

We note that this reasoning implies that the set of possible dynamics $\{P(R, E|R)\}$ is a subset of substochastic matrices: this subset must be convex, and in fact a subset of a convex cone satisfying $0 \leq P(R, E|R) \leq J/n$, where 0 denotes the matrix of all 0's, J the matrix of all 1's, and the inequality is understood in the conic sense. Further, let us consider the special case that the gambler performs a measurement with a single certain outcome, and then updates their state. What could this mean? Following [3], we take such an action to correspond to carving the system out of its background, the consequence being "the sense of the objects existence for the agent." Indifference to this kind of measurement means that even in the absence of the system being present to the agent, the agent ought to update their state according to some $P(R, 1|R) \equiv P_\chi(R|R)$, which we note need not be expressible as a mixture: it may correspond to an extreme ray of the cone. Given two such stochastic matrices $P_\chi(R|R)$ and $P_{\chi'}(R|R)$, it may be that

$$P_\chi(R|R)\Phi P_{\chi'}(R|R)\Phi = P_{\chi'}(R|R)\Phi P_\chi(R|R)\Phi = P(R|R)\Phi, \tag{2.101}$$

in which case we say that the two are inverses of each other, and $P_\chi(R|R)$ along with its inverse correspond to symmetries of the state space.

2.10 Decomposition

Finally, let us discuss briefly what it might mean for the gambler to carve out more than one nonclassical system? Certainly, the gambler may introduce a reference measurement for each system they are free to act upon. Let us at first suppose the gambler assigns probabilities independently to the measurements on the different systems. Clearly, for two systems A and B treated independently, since $P(E^{(A)}, E^{(B)}|\rho^{(A)}, \rho^{(B)}) = P(E^{(A)}|\rho^{(A)}) \otimes P(E^{(B)}|\rho^{(B)})$, where $\otimes$ denotes the tensor product, we have

$$\begin{aligned} &P(E^{(A)}, E^{(B)}|\rho^{(A)}, \rho^{(B)}) \\ &= \left\{P(E^{(A)}|R^{(A)}) \otimes P(E^{(B)}|R^{(B)})\right\}\left\{\Phi^{(A)} \otimes \Phi^{(B)}\right\}\left\{P(R^{(A)}|\rho^{(A)}) \otimes P(R^{(B)}|\rho^{(B)})\right\}, \end{aligned} \tag{2.102}$$

where $\{R_i^{(A)}\}$ and $\{R_i^{(B)}\}$ But what if we assign more general distributions to the reference measurements? In particular, what if we allow for actions and consequences that implicate

the whole collective, not just separate measurements on the parts? For the moment, let us assume that if $\{R_i^{(A)}\}$ is a reference for A and $\{R_i^{(B)}\}$ is a reference for B, then $\{R_i^{(A)} \cap R_j^{(B)}\}$ is a reference for A and B. Since $P(R^{(A)}, R^{(B)}|R^{(A)}, R^{(B)}) = P(R^{(A)}|R^{(A)}) \otimes P(R^{(B)}|R^{(B)})$, we may take $\Phi = \Phi_A \otimes \Phi_B$ as this will be a $\{1\}$ inverse of $P(R^{(A)}, R^{(B)}|R^{(A)}, R^{(B)})$. Thus more generally,

$$P(E|\rho) = P(E|R^{(A)}, R^{(B)})\Big\{\Phi^{(A)} \otimes \Phi^{(B)}\Big\}P(R^{(A)}, R^{(B)}|\rho), \tag{2.103}$$

where E and ρ denote outcomes of measurements on the collective, which as a special case may reduce to the conjunction of outcomes of measurements on the two systems, $E = E_i^{(A)} \cap E_j^{(B)}$.

Now given the state $\mathcal{S}^{(A)}$ and effect $\mathcal{E}^{(A)}$ spaces of A, and the same for B, what can we say about the state and effect spaces of the collective? Since a state space must be convex, the smallest collective state space $\mathcal{S}^{(AB)}$ would consist solely of mixtures of separate state assignments for system A and system B,

$$\mathcal{S}_{\min}^{(AB)} = \Big\{P(R^{(A)}, R^{(B)}|\rho) = \sum_i p_i P(R^{(A)}|\rho_i^{(A)}) \otimes P(R^{(B)}|\rho_i^{(B)})\Big\}. \tag{2.104}$$

We call such states *separable*. In contrast, the largest possible collective state space consistent with the gambler's beliefs about A and B would consist of those assignments to $\{R_i^{(A)} \cap R_j^{(B)}\}$ which imply valid probabilities on all effects in $\mathcal{E}_A$ and in $\mathcal{E}_B$.

$$\begin{aligned}\mathcal{S}_{\max}^{(AB)} = \Big\{P(R^{(A)}, R^{(B)}|\rho)\Big|&\forall E^{(A)} \in \mathcal{E}^{(A)}, E^{(B)} \in \mathcal{E}^{(B)} : \\ &0 \leq \big(P(E^{(A)}|R^{(A)}) \otimes P(E^{(B)}|R^{(B)})\big)\big(\Phi^{(A)} \otimes \Phi^{(B)}\big)P(R^{(A)}, R^{(B)}|\rho) \leq 1\Big\}.\end{aligned} \tag{2.105}$$

Any state in $\mathcal{S}_{\max}^{(AB)}$ which is not a part of $\mathcal{S}_{\min}^{(AB)}$, that is, which is not separable, we call *entangled*. Symmetrically, the smallest collective effect space would consist of conic combinations of effects on A and effects on B less than $P(1|R^{(A)}, R^{(B)})$, while the largest would consist of effects $P(E|R^{(A)}, R^{(B)})$ which satisfy $0 \leq P(E|R^{(A)}, R^{(B)})\{\Phi^{(A)} \otimes \Phi^{(B)}\}\{P(R^{(A)}|\rho^{(A)}) \otimes P(R^{(B)}|\rho^{(B)})\} \leq 1$ for all $P(R^{(A)}|\rho^{(A)}) \in \mathcal{S}^{(A)}$ and $P(R^{(B)}|\rho^{(B)}) \in \mathcal{S}^{(B)}$. So coherence with the gambler's assignments to independent measurements on A and B demands that the collective state and effect spaces satisfy

$$\mathcal{S}_{\min}^{(AB)} \subseteq \mathcal{S}^{(AB)} \subseteq \mathcal{S}_{\max}^{(AB)} \qquad\qquad \mathcal{E}_{\min}^{(AB)} \subseteq \mathcal{E}^{(AB)} \subseteq \mathcal{E}_{\max}^{(AB)}. \tag{2.106}$$

Of course, we may phrase all this in terms of the state and effect cones as well—and say a little more, in fact. It will be helpful, however, to establish the following lemma first.

Lemma 2.10.1 (Φ-duality reverses inclusion)**.** *Suppose cone $K^\star$ is Φ-dual to cone K, and $L^\star$ is Φ-dual to cone L, and that $K \subseteq L$. It follows that $L^\star \subseteq K^\star$.*

Proof. By definition,

$$K^\star = \left\{ y \;\middle|\; \forall x \in K : y\Phi x \geq 0 \right\} \qquad L^\star = \left\{ y \;\middle|\; \forall x \in L : y\Phi x \geq 0 \right\}. \tag{2.107}$$

Since $K \subseteq L$, if y satisfies $\forall x \in L : y\Phi x \geq 0$, then certainly $\forall x \in K \subseteq L : y\Phi x \geq 0$. But this means that if $y \in L^\star$, then $y \in K^\star$. We conclude $L^\star \subseteq K^\star$. □

On the one hand, we have already established the following cone sandwiches: $\hat{\mathcal{S}}^{\min}_{AB} \subseteq \hat{\mathcal{S}}_{AB} \subseteq \hat{\mathcal{S}}^{\max}_{AB}$ and $\hat{\mathcal{E}}^{\min}_{AB} \subseteq \hat{\mathcal{E}}_{AB} \subseteq \hat{\mathcal{E}}^{\max}_{AB}$. But also $\hat{\mathcal{S}}_{AB} \subseteq \hat{\mathcal{E}}^*_{AB}$ and $\hat{\mathcal{E}}_{AB} \subseteq \hat{\mathcal{S}}^*_{AB}$: any state and effect cones must be subsets of the set of mathematically possible states (and effects, respectively). Finally, we note that since

$$\hat{\mathcal{E}}^{\min}_{AB} = \text{cone}\{a \otimes b, a \in \hat{\mathcal{E}}_A, b \in \hat{\mathcal{E}}_B\}, \tag{2.108}$$

we have

$$\left(\hat{\mathcal{E}}^{\min}_{AB}\right)^* = \left\{ x \;\middle|\; (a \otimes b)\Phi x \geq 0, a \in \hat{\mathcal{E}}_A, b \in \hat{\mathcal{E}}_B \right\} = \hat{\mathcal{S}}^{\max}_{AB}, \tag{2.109}$$

and similarly $(\hat{\mathcal{S}}^{\min}_{AB})^* = \hat{\mathcal{E}}^{\max}_{AB}$. Putting all this together, we have the following inclusions:

$$\hat{\mathcal{S}}^{\min}_{AB} \subseteq \hat{\mathcal{S}}_{AB} \subseteq \hat{\mathcal{E}}^*_{AB} \subseteq (\hat{\mathcal{E}}^{\min}_{AB})^* = \hat{\mathcal{S}}^{\max}_{AB} \tag{2.110}$$

$$\hat{\mathcal{E}}^{\min}_{AB} \subseteq \hat{\mathcal{E}}_{AB} \subseteq \hat{\mathcal{S}}^*_{AB} \subseteq (\hat{\mathcal{S}}^{\min}_{AB})^* = \hat{\mathcal{E}}^{\max}_{AB}. \tag{2.111}$$

These mathematical niceties aside, let us return to our assumption that $\{R^{(A)}_i \cap R^{(B)}_j\}$ forms a reference for the collective of A and B: in the quantum foundations literature, this assumption is called *local tomography*. In general, there is no reason to think this ought to be satisfied. It may be that there is strictly more vitality in the systems together than apart. Whenever multiple systems are involved, therefore, in general, one must take care to identify reference measurements at each level of collectivity, taking into account reference measurements $\{R^{(AB)}_i\}$ upon the collective as a whole. Indeed, it is ideal to start from the utmost whole, and decompose the whole into parts, rather than build up the whole from the parts. On what principle may one decompose a whole? Precisely in a principle of indifference, that as regards the gambler's judgements about measurements in some set A, they are indifferent to what measurements in some other set B are performed. This amounts to saying the gambler has carved out two systems which may be addressed separately. Formally,

this amounts to

$$a, a' \in A; b \in B : \sum_i P(a_i, b_j|\rho) = \sum_i P(a'_i, b_j|\rho) = P(b_j|\rho) \tag{2.112}$$

$$a \in A; b, b' \in B : \sum_j P(a_i, b_j|\rho) = \sum_j P(a_i, b'_j|\rho) = P(a_i|\rho). \tag{2.113}$$

In other words, the marginal probability $P(b|\rho)$ the gambler assigns to a measurement in B is independent of the choice of which measurement the gambler might also make in A: indeed, this is a prerequisite for considering a marginal state of B separate from A at all—that is, considering the whole as consisting of parts to begin with. In the quantum foundations literature, these are usually called the "no signalling" conditions. From there, the gambler may consider what comes of restricting themselves to measurements in $\mathcal{A}$, identifying a reference measurement, and building up a coherent mesh of probabilities—and do the same for measurements in $\mathcal{B}$. Then given e.g., $\mathcal{S}^{(A)}$ and $\mathcal{S}^{(B)}$ they may identify which composite state space $\mathcal{S}^{(AB)}$ recovers (at least part of) the original state space $\mathcal{S}$ of the whole.

Going in the opposite direction, supposing we can identify reference measurements $\{R_i^{(A)}\}$ on system A and $\{R_i^{(B)}\}$ on system B, then we can characterize a collective with a joint distribution $P(R_i^{(A)}, R_j^{(B)}|\rho)$. Marginalizing over either measurement, we obtain $P(R^{(A)}|\rho)$ and $P(R^{(B)}|\rho)$, and since each is assumed to be a reference measurement, we have

$$P(E^{(A)}|\rho) = P(E^{(A)}|R^{(A)})\Phi^{(A)}P(R^{(A)}|\rho) \tag{2.114}$$

$$P(E^{(B)}|\rho) = P(E^{(B)}|R^{(B)})\Phi^{(B)}P(R^{(B)}|\rho). \tag{2.115}$$

At the same time,

$$P(E^{(A)}, E^{(B)}|\rho) = \left\{P(E^{(A)}|R^{(A)}) \otimes P(E^{(B)}|R^{(B)})\right\}\left\{\Phi^{(A)} \otimes \Phi^{(B)}\right\}P(R^{(A)}, R^{(B)}|\rho), \tag{2.116}$$

whose marginals clearly agree so that "no signalling" is satisfied.

2.11 Conclusion

In this chapter, we have developed the essentials of probability theory from a subjective Bayesian point of view, emphasizing throughout the coherence of an individual agent's mesh of beliefs. Taking up the QBist point of view, that the agent, their own actions, and the consequences of those actions for them ought to be taken as primary, we introduced a notion of system, state, and effect that formalizes a particular notion of nonclassicality. Central to this notion is the idea of a reference measurement. We showed that our assumptions imply a fundamental nonclassical norm which the gambler ought to abide by which supplements the familiar law of total probability. The formalism we have developed here is expansive enough

to aid a gambler grappling with an open-ended, creative nature whose unfolding the agent themselves is implicated in. In some sense, this chapter contains everything such a gambler really needs. In the next chapter, however, we will bring an extra, linear algebraic layer of sophistication to the formalism, indeed, showing its equivalence to the standard framework of *generalized probabilistic theories*, of which quantum theory is an example.

CHAPTER 3

THE BORN IDENTITY

3.1 Introduction

Armed with the fundamental nonclassical coherence condition, expressed in terms of the Born matrix Φ, we now turn to studying its properties linear algebraically. In the course of doing so, we show that the conditional probability matrix $P \equiv P(R|R)$ which characterizes the reference measurement defines two fundamental subspaces: consistency requires that distributions on reference outcomes $P(R|\rho)$ must live in $\mathrm{col}(P)$ and response functions $P(E|R)$, as well as valuations on reference outcomes, ought to live in $\mathrm{row}(P)$. Introducing bases for these spaces, we pass to a compact representation of states and effects which shows the essential equivalence between the QBist framework and the framework of generalized probabilistic theories or GPTs. In the other direction, we show how given a GPT, one can develop a "double sided" frame representation which reproduces the nonclassical coherence condition we began with. Relying on already existing results in the literature, we define what properties a measurement must have in a GPT to be a proper reference measurement, showing that generally speaking such measurements always exist, and moreover, allow one to prove a nonclassical de Finetti representation theorem.

3.2 The Born matrix

As we have seen, having identified a reference measurement, the gambler may consider two scenarios. In the first scenario, the gambler performs a preparatory measurement, followed by a reference measurement, followed by some arbitrary measurement, in which case they may invoke the familiar classical coherence condition, the law of total probability,

$$P(E|R,\rho) = P(E|R)P(R|\rho). \tag{3.1}$$

In the second scenario, they skip the reference measurement. Despite the fact that in the latter case, the reference measurement is hypothetical, the gambler may appeal to a nonclassical coherence condition, whose form we derived in the last chapter to be

$$P(E|\rho) = P(E|R)\Phi P(R|\rho), \tag{3.2}$$

where Φ satisfies $P(R|R)\Phi P(R|R) = P(R|R)$. Here $P(R|R)$ is the conditional probability matrix which characterizes the reference measurement itself. In fact, this is the defining

equation of a $\{1\}$-inverse of $P(R|R)$, and we shall now study this linear algebraic object more closely.

We first establish that *every* matrix over $\mathbb{R}$ has a $\{1\}$-inverse. In fact, our proof holds over $\mathbb{C}$, and the result itself holds over any field $\mathbb{F}$, as we will show later. We begin with this proof, however, which relies on the singular value decomposition, in order to build an intuition as to how one might calculate one numerically.

Theorem 3.2.1 (Existence and form of $\{1\}$-inverse)**.** *Every matrix $M \in \mathbb{C}^{m\times n}$ has a $\{1\}$-inverse, i.e., a matrix $M^{(1)}$ such that*

$$MM^{(1)}M = M. \tag{3.3}$$

Proof. Every matrix has a singular value decomposition $M = U\Sigma V^\dagger$. Let

$$M^{(1)} = V \left(\begin{array}{c|c} \sigma^{-1} & A \\ \hline B & C \end{array}\right) U^\dagger, \tag{3.4}$$

where σ is a diagonal matrix of the $r = \operatorname{rank}(M)$ non-zero singular values, and where the blocks A, B, C are completely arbitrary. Observe that

$$MM^{(1)}M = \left[U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger\right] \left[V \left(\begin{array}{c|c} \sigma^{-1} & A \\ \hline B & C \end{array}\right) U^\dagger\right] \left[U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger\right] \tag{3.5}$$

$$= U \left(\begin{array}{c|c} I & \sigma A \\ \hline 0 & 0 \end{array}\right) \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger = U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger = M, \tag{3.6}$$

showing that $M^{(1)}$ is a $\{1\}$-inverse. Conversely, suppose $MM^{(1)}M = M$, that is,

$$\left[U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger\right] M^{(1)} \left[U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger\right] = U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger. \tag{3.7}$$

Multiplying on the left by $U^\dagger$ and on the right by V yields

$$\left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) V^\dagger M^{(1)} U \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right) = \left(\begin{array}{c|c} \sigma & 0 \\ \hline 0 & 0 \end{array}\right). \tag{3.8}$$

This gives us one constraint: writing the blocks of $V^\dagger M^{(1)} U$ as $\left(\begin{array}{c|c} X & A \\ \hline B & C \end{array}\right)$, we see that $\sigma X \sigma = \sigma$, which implies $X = \sigma^{-1}$. Then multiplying $V^\dagger M^{(1)} U$ by V from the left and U from the right, we arrive at $M^{(1)} = V \left(\begin{array}{c|c} \sigma^{-1} & A \\ \hline B & C \end{array}\right) U^\dagger$, as desired. □

Remark. *Suppose that M is in fact invertible. Then $M^{-1}\left[MM^{(1)}M\right]M^{-1} = M^{-1}MM^{-1}$ implies that $M^{(1)} = M^{-1}$, the usual inverse. In the case that $A = B = C = 0$, $M^{(1)} = M^{+}$, the Moore-Penrose pseudoinverse.*

In fact, there is a whole zoo of generalized inverses X for singular matrices M which are classified according to a scheme due to Penrose. Each species is specified by a subset of the following list of axioms they satisfy [71]:

$$
\begin{aligned}
&(1)\ \ MXM = M\\
&(2)\ \ XMX = X\\
&(3)\ \ (MX)^{\dagger} = MX\\
&(4)\ \ (XM)^{\dagger} = XM\\
&(5)\ \ MX = XM\\
&(5^{k})\ \ M^{k}X = XM^{k}\\
&(6^{k})\ \ MX^{k} = X^{k}M
\end{aligned}
$$

Notice that all of these are satisfied by the true matrix inverse. The only restriction on a Born matrix per se is that it is a $\{1\}$-inverse: but it may satisfy other axioms. For instance, the Moore-Penrose pseudoinverse is the unique $\{1, 2, 3, 4\}$-inverse. Later we will consider reasons one might prefer one choice over another. As a taste, we offer one proof in the interim. One might wish $W(R|\rho) = \Phi P(R|\rho)$ to be a *quasiprobability vector*, that is, not necessarily nonnegative, but nevertheless summing to 1. We can guarantee this if the Born matrix Φ is itself quasistochastic, that is, satisfying $u^{\dagger}\Phi = u^{\dagger}$ where $u = (1, \ldots, 1)^{\dagger}$.

Lemma 3.2.2. *Let Φ be the $\{1\}$-inverse of a stochastic matrix P. Then Φ can be chosen to be quasistochastic, so that it takes probability vectors to quasiprobability vectors.*

Proof. Let $u = (1, \ldots, 1)^{\dagger}$, and pick any vector v such that $u^{\dagger}v = 1$ and any $\{1\}$-inverse Φ_0 not necessarily quasistochastic. Then let

$$\Phi = \Phi_0 + v(u^{\dagger} - u^{\dagger}\Phi_0). \tag{3.9}$$

On the one hand, $u^{\dagger}\Phi = u^{\dagger}$, so that Φ is quasistochastic. Using the stochasticity of P and $\{1\}$-inverse axiom, we have $u^{\dagger}\Phi_0 P = u^{\dagger}P\Phi_0 P = u^{\dagger}P = u^{\dagger}$, and so

$$P\Phi P = P\Phi_0 P + Pv(u^{\dagger} - u^{\dagger}\Phi_0)P = P + Pv(u^{\dagger} - u^{\dagger}) = P. \tag{3.10}$$

□

3.3 Subspaces from coherence

We derived that Φ ought to be a $\{1\}$-inverse of $P(R|R)$ by considering the special case that both the preparatory and final measurements are the reference measurement itself: $P(R|R) = P(R|R)\Phi P(R|R)$. But we can actually say a bit more. By the same logic, for any state ρ, we have by coherence with the reference measurement,

$$P(R|\rho) = P(R|R)\Phi P(R|\rho). \tag{3.11}$$

In fact, this puts a strong restriction on the gambler's assignments $P(R|\rho)$. To see this, we must understand the action of $P(R|R)\Phi$.

Lemma 3.3.1. *Let $M^{(1)}$ be a $\{1\}$-inverse of M. Then $MM^{(1)}$ projects onto the range of M, that is, the column space of M [71].*

Proof. Let $\Pi = MM^{(1)}$. Then by axiom (1), $\Pi^2 = MM^{(1)}MM^{(1)} = M^{(1)}M = \Pi$: Π is idempotent, and so a (not necessarily orthogonal) projector. On the one hand, $\text{range}(\Pi) = \text{range}(MM^{(1)}) \subseteq \text{range}(M)$. On the other hand, for $x \in \text{range}(M)$, we can write $x = My$: x is some linear combination of the columns of M. Then $\Pi x = \Pi My = MM^{(1)}My = My = x$ so that Π fixes all $x \in \text{range}(M)$, from which we conclude $\text{range}(\Pi) = \text{range}(M)$. □

Corollary 3.3.1.1. *Recognizing that $P(R|R)\Phi$ is a projector onto range$(P(R|R))$, from $P(R|\rho) = P(R|R)\Phi P(R|\rho)$ we arrive at another coherence condition:* the set $\mathcal{S}$ of probability distributions on the reference measurement must lie in $\text{range}(P(R|R))$. *Equivalently, the agent should only assign distributions $P(R|\rho)$ to the reference measurement which are linear combinations of the columns of $P(R|R)$.*

Similarly, for any effect E, we have by coherence with the reference measurement,

$$P(E|R) = P(E|R)\Phi P(R|R), \tag{3.12}$$

which leads us to ask about the action of $\Phi P(R|R)$ on row vectors.

Lemma 3.3.2. *Let $M^{(1)}$ be a $\{1\}$-inverse of M. Then $M^{(1)}M$ projects row vectors onto the rowspace of M.*

Proof. $(M^{(1)}M)^2 = M^{(1)}MM^{(1)}M = M^{(1)}M$ so $M^{(1)}M$ is a projector. On the one hand, if $(xM^{(1)})M = x$, then $x \in \text{row}(M)$, since x is a linear combination of the rows of M. Conversely, if $x = yM$ for some row vector y, then $xM^{(1)}M = yMM^{(1)}M = yM = x$. □

Corollary 3.3.2.1. *We thus arrive at another coherence condition: since $P(E|R) = P(E|R)\Phi P(R|R)$,* the effect space $\mathcal{E}$ of conditional probability distributions $P(E|R)$ must lie in $\text{row}(P(R|R))$.

In particular, if $P(R|R)$ is symmetric, then, identifying vectors and covectors, both states and effects must live in the same subspace.

Lemma 3.3.3. *Let $P(E|R) \in row(P(R|R)$ and $P(R|\rho) \in col(P(R|R))$ and suppose Φ and Φ' are both $\{1\}$-inverses of $P(R|R)$. Then*

$$P(E|R)\Phi P(R|\rho) = P(E|R)\Phi' P(R|\rho). \tag{3.13}$$

Proof. Since $P(E|R) \in \mathrm{row}(P(R|R))$, we can write $P(E|R) = a^\dagger P(R|R)$ for some a; since $P(R|\rho) \in \mathrm{col}(P(R|R))$, we can write $P(R|\rho) = P(R|R)b$ for some b. Then

$$P(E|R)\Phi P(R|\rho) = a^\dagger P\Phi Pb = a^\dagger Pb = a^\dagger P\Phi' Pb = P(E|R)\Phi' P(R|\rho). \tag{3.14}$$

□

3.3.1 A word on observables

Classically, we defined a random variable as a valuation e.g., $X : \Omega \to \mathbb{R}$ on the atomic events, or in the language of measurements, on the outcomes of the finest grained reference measurement. In the nonclassical setting, the role of Ω is played by the outcomes of a reference measurement, but these do not correspond to "atomic events" whose coarse grainings give all possible composite events: this difference is witnessed by the presence of the Born matrix Φ. Nevertheless, we may still take a random variable to be $X : \{R_i\} \to \mathbb{R}$, a valuation of reference outcomes. To see this, denote by $\mathfrak{x}$ a row vector of valuations on a measurement with outcomes X_i. We may always rewrite it as a vector of valuations on the reference measurement,

$$x = \mathfrak{x}P(X|R)\Phi, \tag{3.15}$$

so that $\langle \mathfrak{x} \rangle_\rho = \mathfrak{x}P(X|\rho) = \mathfrak{x}P(X|R)\Phi P(R|\rho) = xP(R|\rho) = \langle x \rangle_\rho$. Indeed, since $\langle \mathfrak{x} \rangle_\rho = \langle x \rangle_\rho$, the agent should value them at the same price.

Moreover, since $P(R|\rho) \in \mathrm{col}(P)$, any component of x orthogonal to $\mathrm{col}(P)$ will not contribute to any expectation values. Indeed, if we write $x = x^{(\|)} + x^{(\perp)}$, then

$$xP(R|\rho) = (x^{(\|)} + x^{(\perp)})P(R|\rho) = x^{(\|)}P(|\rho), \tag{3.16}$$

so that the gambler ought to assign the same price regardless. Thus without loss of generality, the gambler need only consider valuations on reference outcomes which lie in $\mathrm{col}(P)$. At the same time, if $x = \mathfrak{x}P(X|R)\Phi$, then since $P(X_i|R) \in \mathrm{row}(P)$, $x \in \mathrm{row}(P)\Phi$. We will often take Φ to be invertible, so that $x \in \mathrm{row}(P)$ itself. If $P = P^T$, then such an x will automatically be in $\mathrm{col}(P)$.

3.4 Rank factorization

For convenience let $P \equiv P(R|R)$. Let n be the number of outcomes of the reference measurement and let $r = \text{rank}(P) = \dim(\text{col}(P)) = \dim(\text{row}(P))$. We have just shown that our effects $P(E|R)$ live in $\text{row}(P)$ while our states $P(R|\rho)$ live in $\text{col}(P)$. Thus rather than work in the n-dimensional space of probability vectors, we may work directly in the r-dimensional spaces in which the states and effects live. To this end, we may introduce bases for the row and column spaces of P, and one nice way of thinking about this is in terms of a *rank factorization* of $P(R|R)$.

Theorem 3.4.1. *Suppose $P \in \mathbb{F}_r^{m\times n}$. There always exists a full rank factorization $P = \boldsymbol{ES}$ where $\boldsymbol{E} \in \mathbb{F}_r^{m\times r}$ and $\boldsymbol{S} \in \mathbb{F}_r^{r\times n}$. Here r is the rank of P.*

Proof. Since P has rank r, it has r linearly independent columns. Let column vectors $e_1, \ldots, e_r$ be a basis for the column space of P, and arrange them into an $m \times r$ matrix

$$\mathbf{E} = \begin{pmatrix} e_1 & | & \ldots & | & e_r \end{pmatrix}. \tag{3.17}$$

By construction, every column $p_1, \ldots, p_n$ of P is a linear combination of the columns of $\mathbf{E}$. Arrange the expansion coefficients for each column of P into column vectors themselves $s_1, \ldots, s_n$. Let

$$\mathbf{S} = \begin{pmatrix} s_1 & | & \ldots & | & s_n \end{pmatrix}, \tag{3.18}$$

which is an $r \times n$ matrix. Then $P = \mathbf{ES}$ as desired. □

Remark. *We could have equally well begun with a basis for the row space of P, constructing $\boldsymbol{S}$ first, and then $\boldsymbol{E}$. A straightforward way of obtaining a full rank factorization is by means of the singular value decomposition,*

$$P = U\Sigma V^\dagger = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\dagger \\ V_2^\dagger \end{bmatrix} = U_1(\Sigma_r V_1^\dagger) = \boldsymbol{ES}, \tag{3.19}$$

where r is the rank of P.

In particular, let $P(R|R) = \mathbf{RS}$ be a rank decomposition of the conditional probability matrix which characterizes the reference measurement itself. If we write

$$\mathbf{R} = \sum_i |i)(R_i| = \begin{pmatrix} (R_1| \\ \vdots \\ (R_n| \end{pmatrix} \qquad \mathbf{S} = \sum_j |S_j)(j| = \begin{pmatrix} |S_1) & \cdots & |S_n) \end{pmatrix}, \tag{3.20}$$

we may consider the rows of $\mathbf{R}$ to be a compact representation of the effects $(R_i|$ and the columns of $\mathbf{S}$ to be a compact representation of the states $|S_i)$ since by construction

$$P(R_i|R_j) = (R_i|S_j). \tag{3.21}$$

In fact, this will lead to an alternative perspective on the Born matrix itself. To attain it, we first reprove a standard lemma: we have seen that any matrix has a $\{1\}$-inverse, but if a matrix has full row rank or full column rank, we can say even more.

Lemma 3.4.2. *Let $M \in \mathbb{F}_r^{r\times n}$: then $M^{(1)}$ is a right inverse of M. Similarly, let $M \in \mathbb{F}_r^{n\times r}$: then $M^{(1)}$ is a left inverse of M [71].*

Proof. If M has full row rank, its n columns span $\mathbb{F}^r$. Thus we can express basis vectors e_i in $\mathbb{F}^r$ as linear combinations of the columns of M: $e_i = Mr_i$ for some coefficient vectors r_i. But then arrange these r_i's into the columns of a matrix R. Clearly, $I = MR$, so M has a right inverse. But then multiply the $\{1\}$-inverse axiom by R: $MM^{(1)}MR = MR \implies MM^{(1)} = I$, as desired. On the other hand, if M has full column rank, its n rows span $\mathbb{F}^r$, so we can write the basis covectors $e_i = l_iM$, which amounts to $I = LM$, so M has a left inverse. Then $LMM^{(1)}M = LM \implies M^{(1)}M = I$, as desired. □

Corollary 3.4.2.1. *The Moore-Penrose pseudoinverse [71] may be calculated, in the case that $\boldsymbol{R}$ has full column rank and in the case that $\boldsymbol{S}$ has full row rank respectively, as*

$$\boldsymbol{R}^L = (\boldsymbol{R}^\dagger\boldsymbol{R})^{-1}\boldsymbol{R}^\dagger \qquad\qquad \boldsymbol{S}^R = \boldsymbol{S}^\dagger(\boldsymbol{S}\boldsymbol{S}^\dagger)^{-1} \tag{3.22}$$

Proof. We first must establish the invertibility $\mathbf{R}^\dagger\mathbf{R}$ and $\mathbf{S}\mathbf{S}^\dagger$. Suppose $\mathbf{R}^\dagger\mathbf{R}$ were not invertible. Then $\mathbf{R}^\dagger\mathbf{R}x = 0$ for some non-zero vector x. But then $x^\dagger\mathbf{R}^\dagger\mathbf{R}x = ||\mathbf{R}x||^2 = 0$ which can only be true if $\mathbf{R}x$ is the zero vector. But then $\mathbf{R}$ can't have a left inverse. We conclude that $\mathbf{R}^\dagger\mathbf{R}$ is invertible after all. The argument is the same for the invertibility $\mathbf{S}\mathbf{S}^\dagger$. Then clearly $\mathbf{R}^L\mathbf{R} = \left[(\mathbf{R}^\dagger\mathbf{R})^{-1}\mathbf{R}^\dagger\right]\mathbf{R} = I = \mathbf{S}\left[\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\right] = \mathbf{S}\mathbf{S}^R$. □

But once we have one left or right inverse, we actually can obtain them all. More generally, given any $\{1\}$-inverse, we have the following parameterization of the space of all $\{1\}$-inverses[1].

Lemma 3.4.3. *Suppose $M^{(1)}$ is a $\{1\}$-inverse of $M \in \mathbb{F}^{n\times m}$ so that $MM^{(1)}M = M$. Then any other $\{1\}$-inverse can be obtained via*

$$\tilde{M}^{(1)} = M^{(1)} + A - M^{(1)}MAMM^{(1)}. \tag{3.23}$$

where $A \in \mathbb{F}^{m\times n}$ is completely arbitrary [71, 100, 101, 102].

[1]We observe in passing that all these constructions are now independent of the choice of scalars.

Proof. On the one hand,

$$M\tilde{M}^{(1)}M = M\big(M^{(1)} + A - M^{(1)}MAMM^{(1)}\big)M \tag{3.24}$$
$$= MM^{(1)}M + MAM - (MM^{(1)}M)A(MM^{(1)}M) \tag{3.25}$$
$$= M. \tag{3.26}$$

Conversely, suppose $\tilde{M}^{(1)}$ is a 1-inverse. We can obtain $\tilde{M}^{(1)}$ from $M^{(1)}$ by taking $A = \tilde{M}^{(1)} - M^{(1)}$. Then the RHS of Eq. (3.23) becomes

$$= M^{(1)} + (\tilde{M}^{(1)} - M^{(1)}) - M^{(1)}M(\tilde{M}^{(1)} - M^{(1)})MM^{(1)} \tag{3.27}$$
$$= \tilde{M}^{(1)} - M^{(1)}(M\tilde{M}^{(1)}M)M^{(1)} + M^{(1)}(MM^{(1)}M)M^{(1)} \tag{3.28}$$
$$= \tilde{M}^{(1)}. \tag{3.29}$$

□

Let $P(R|\rho)$ be a state: by coherence, it must live in $\mathrm{col}(P)$. Let $P = \mathbf{RS}$ be a rank decomposition of P. By construction, the columns of $\mathbf{R}$ form a basis for $\mathrm{col}(P)$, and so we must have $P(R|\rho) = \mathbf{R}|\rho)$ for some $|\rho)$. Similarly, by coherence, any effect $P(E|R)$ must live in $\mathrm{row}(P)$, but since the rows of $\mathbf{S}$ form a basis for $\mathrm{row}(P)$, we must have $P(E|R) = (E|\mathbf{S}$ for some $(E|$. We then come to the following lemma.

Lemma 3.4.4 (The Born identity)**.** *Let $P = \boldsymbol{RS}$ be full rank factorization such that $\boldsymbol{R} \in \mathbb{F}_r^{n\times r}$ and $\boldsymbol{S} \in \mathbb{F}_r^{r\times n}$. Then*

$$P\Phi P = P \iff \boldsymbol{S}\Phi\boldsymbol{R} = I. \tag{3.30}$$

Proof. Clearly, acting on $\mathbf{S}\Phi\mathbf{R} = I$ from the left with $\mathbf{R}$ and from the right with $\mathbf{S}$ yields $P\Phi P = P$. Conversely, by Lemma (3.4.2), $\mathbf{R}$ has a left inverse $\mathbf{R}^L$ and $\mathbf{S}$ has a right inverse $\mathbf{S}^R$. Acting from the left with $\mathbf{R}^L$ and from the right with $\mathbf{S}^R$ on $P\Phi P = P$, we conclude $\mathbf{R}^L(\mathbf{RS})\Phi(\mathbf{RS})\mathbf{S}^R = \mathbf{R}^L(\mathbf{RS})\mathbf{S}^R$, that is, $\mathbf{S}\Phi\mathbf{R} = I$. □

Immediately, we have

$$P(E|\rho) = P(E|R)\Phi P(R|\rho) = (E|\mathbf{S}\Phi\mathbf{R}|\rho) = (E|\rho). \tag{3.31}$$

We conclude that for *any* effect E and state ρ, we may write $P(E|\rho) = (E|\rho)$. Going in the other direction, we have explicitly

$$|\rho) = \mathbf{R}^L P(R|\rho), \tag{3.32}$$

where $\mathbf{R}^L$ is some left inverse of $\mathbf{R}$ and

$$(E| = P(E|R)\mathbf{S}^R, \tag{3.33}$$

where $\mathbf{S}^R$ is some right inverse of $\mathbf{S}$. Putting them together, we have

$$(E|\rho) = P(E|R)\mathbf{S}^R\mathbf{R}^L P(R|\rho). \tag{3.34}$$

Indeed, $P(\mathbf{S}^R\mathbf{R}^L)P = \mathbf{RS}(\mathbf{S}^R\mathbf{R}^L)\mathbf{RS} = P$ shows that we may always take $\Phi = \mathbf{S}^R\mathbf{R}^L$ as a Born matrix, a $\{1\}$-inverse of P. If we do so, however, such a Born matrix enjoys another generalized inverse axiom, namely,

$$\Phi P \Phi = \mathbf{S}^R\mathbf{R}^L(\mathbf{RS})\mathbf{S}^R\mathbf{R}^L = \mathbf{S}^R\mathbf{R}^L = \Phi. \tag{3.35}$$

In other words, such a Φ is a $\{1, 2\}$-inverse. We note but do not prove the following theorem of Bjerhammar [71],

Theorem 3.4.5. *Let X be a $\{1\}$-inverse of a matrix M. Then X is a $\{1, 2\}$ inverse if and only if $rank(X) = rank(M)$.*

In this way, coherence with the reference implies that we can express $P(E|\rho)$ as the action of a covector $(E| = P(E|R)\mathbf{S}^R$ on a vector $|\rho) = \mathbf{R}^L P(R|\rho)$ in an $r = \text{rank}(P)$ dimensional vector space. As such we will often refer to these objects as states and effects as well. Moreover, consider conditional probability matrices $P(R|E, R)$ which capture how the gambler ought to update their state assignments conditional on a subsequent measurement result E: $P(R|E, \rho) = P(E|\rho)^{-1}P(R|E, R)\Phi P(R|\rho)$. By coherence the row space of $P(R|E, R)$ ought to be $\subseteq \text{row}(P)$ and the column space of such a matrix ought to be $\subseteq \text{col}(P)$. Thus we may construct $T = \mathbf{R}^L P(R|E, R)\mathbf{S}^R$ which acts natively on states $|\rho)$ or effects $(E|$.

Finally, let us make a brief comment about normalization. Suppose we have a measurement with outcomes $\{E_i\}$. The gambler assigns to each outcome an effect $P(E_i|R)$ so that together they form the rows of a conditional probability matrix $P(E|R)$. To guarantee that probabilities sum to 1, such a matrix ought to be column stochastic: equivalently, $\sum_i P(E_i|R) = P(1|R) = (1, \ldots, 1)$. We see then that the effects in a measurement must sum to the effect $P(1|R)$ which assigns probability 1 to all states. Conceptually, this is like saying the gambler believes one of the outcomes must certainly occur; mathematically, it guarantees that the probabilities $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ sum to 1. Picking a rank decomposition $P(R|R) = \mathbf{RS}$, we have

$$(1| = P(1|R)\mathbf{S}^R \qquad \forall \rho : (1|\rho) = 1 \qquad \sum_i (E_i| = (1|, \tag{3.36}$$

where the latter holds for any measurement.

Of course, the representation of states, effects, and update maps so obtained is not at all unique, as the following theorem, whose proof we reproduce here, shows.

Theorem 3.4.6. *Any two full rank factorizations $P = \boldsymbol{R}_1\boldsymbol{S}_1 = \boldsymbol{R}_2\boldsymbol{S}_2$ are related by some invertible matrix T so that $\boldsymbol{R}_2 = \boldsymbol{R}_1T^{-1}$ and $\boldsymbol{S}_2 = T\boldsymbol{S}_1$ [103].*

Proof. Since $\mathbf{R}_2$ has full column rank, it has a left inverse; since $\mathbf{S}_2$ has full row rank, it has a right inverse. Consider the two matrices $\mathbf{R}_2^L\mathbf{R}_1$ and $\mathbf{S}_1\mathbf{S}_2^R$. We have $(\mathbf{R}_2^L\mathbf{R}_1)(\mathbf{S}_1\mathbf{S}_2^R) = \mathbf{R}_2^L P \mathbf{S}_2^R = \mathbf{R}_2^L\mathbf{R}_2\mathbf{S}_2\mathbf{S}_2^R = I_r$. Moreover,

$$r = \text{rank}(\mathbf{S}_2) = \text{rank}(\mathbf{R}_2^L\mathbf{R}_2\mathbf{S}_2) = \text{rank}(\mathbf{R}_2^L\mathbf{R}_1\mathbf{S}_1) = \text{rank}(\mathbf{R}_2^L\mathbf{R}_1), \tag{3.37}$$

where the last follows from the fact that $\mathbf{S}_1$ has full row rank. Similarly,

$$r = \text{rank}(\mathbf{R}_2) = \text{rank}(\mathbf{R}_2\mathbf{S}_2\mathbf{S}_2^R) = \text{rank}(\mathbf{R}_1\mathbf{S}_1\mathbf{S}_2^R) = \text{rank}(\mathbf{S}_1\mathbf{S}_2^R), \tag{3.38}$$

where the last follows from the fact that $\mathbf{R}_1$ has full column rank. Thus $\mathbf{R}_2^L\mathbf{R}_1$ and $\mathbf{S}_1\mathbf{S}_2^R$ are both full rank, and in fact they must be inverses of each other. Let $T = \mathbf{R}_2^L\mathbf{R}_1$ and $T^{-1} = \mathbf{S}_1\mathbf{S}_2^R$. We then have

$$\mathbf{R}_1 T^{-1} = \mathbf{R}_1[\mathbf{S}_1\mathbf{S}_2^R] = \mathbf{R}_2\mathbf{S}_2\mathbf{S}_2^R = \mathbf{R}_2 \qquad T\mathbf{S}_1 = [\mathbf{R}_2^L\mathbf{R}_1]\mathbf{S}_1 = \mathbf{R}_2^L\mathbf{R}_2\mathbf{S}_2 = \mathbf{S}_2, \tag{3.39}$$

as desired. □

Remark. *Our freedom in choosing states $|\rho)$ and effects $(E|$ amounts to a choice of invertible T. But if a Φ matrix works for one choice, it works for them all: $\boldsymbol{S}'\Phi\boldsymbol{R}' = T\boldsymbol{S}\Phi\boldsymbol{R}T^{-1} = TT^{-1} = I$.*

3.5 Generalized probabilistic theories

We defined a *state* to be an equivalence class of consequences $\rho = \{\rho_i\}$ such that

$$\forall j, k : P(R|\rho_j) = P(R|\rho_k), \tag{3.40}$$

that is, after any measurement with an outcome in the equivalence class, the gambler would assign the same probabilities to a subsequent reference measurement. A state ρ may be identified on the one hand with a particular member of the equivalence class of outcomes, e.g., which is actually realized, and on the other hand, with a unique state of expectation about the outcomes of the reference measurement $P(R|\rho)$. Similarly, a *effect* is an equivalence class of measurement outcomes $E = \{E_i\}$ such that

$$\forall j, k : P(E_j|R) = P(E_k|R), \tag{3.41}$$

that is, conditional on a reference outcome, the gambler would assign the same probabilities to any outcome in the equivalence class upon a subsequent measurement. We may identify an effect E with a particular member of the equivalence class, and at the same time the conditional distribution $P(E|R)$. Finally, we defined the *reference measurement* itself to be a

preparatory measurement such that

$$P(R|\rho) = P(R|\rho') \iff \forall E \in \mathcal{E} : P(E|\rho) = P(E|\rho') \iff \rho \simeq \rho' \tag{3.42}$$

$$P(E|R) = P(E'|R) \iff \forall \rho \in \mathcal{S} : P(E|\rho) = P(E'|\rho) \iff E \simeq E', \tag{3.43}$$

that is, if two states are equivalent with respect to the reference measurement, they are equivalent with respect to any measurement, and similarly for effects.

This framework established, we derived a norm a gambler may appeal to in the case that the reference measurement remains hypothetical,

$$P(E|\rho) = \mathcal{F}(P(E|R), P(R|\rho)) = P(E|R)\Phi P(R|\rho). \tag{3.44}$$

Moreover, we developed a geometrical picture of the states and effects. The state space $\mathcal{S}$ ought to be a convex set since we can always consider convex combinations of reference distributions: allowing for arbitrary rescalings, we arrive at the state cone $\hat{\mathcal{S}}$, recovering $\mathcal{S}$ as a base cut by the effect $P(1|R) = (1, \ldots, 1)$ which ensures normalization of probability vectors. Similarly, by contemplating coarse and fine graining of effects, we consider another convex set: the effect space $\mathcal{E}$, and its associated cone $\hat{\mathcal{E}}$. The effect space itself must satisfy $P(0|R) \leq P(E|R) \leq P(1|R)$ with respect to the ordering induced by the effect cone: this comes from the fact that if $P(E|R)$ is an effect then so must $P(\neg E|R) = P(1|R) - P(E|R)$. We also introduced a notion of Φ-duality, so that e.g., $\hat{\mathcal{S}}^\star$ is the cone Φ-dual to the state cone, and consists of all mathematically possible effects. None of these cones were assumed to be spanning of the n dimensional vector space $\mathcal{V}$ in which the probability vectors live. In fact, we have just shown that $\mathcal{S} \subseteq \mathrm{col}(P(R|R))$ and $\mathcal{E} \subseteq \mathrm{row}(P(R|R))$.

Introducing bases for the column and row spaces of $P(R|R)$, we were able to rewrite

$$P(E|\rho) = (E|\rho), \tag{3.45}$$

where now we work in an $r = \mathrm{rank}(P(R|R))$ dimensional space V. In this setting, $P(E|\rho)$ is given by a linear functional which we may identify with the effect acting directly on a vector which we may identify with the state. Measurements become sets of effects $\{(E_i|\}$ which sum to the image of $P(1|R)$, that is, $\sum_i (E_i| = (1|$, such that $\forall \rho : (1|\rho) = 1$. State updates conditional on measurement outcomes can be represented by linear operators T acting on this same space. In this representation, the state and effect cones become by construction *generating cones*, that is, they satisfy $C - C = V$. Moreover, the effect cone is now a subset of the dual of the state cone in the usual sense of convex duality [94], i.e., $C^* = \{y \in V^* : \forall x \in C : (y|x) \geq 0\}$, and similarly for the state cone. The "cone sandwiches" we exhibited to better understand possible collectives may also be reexpressed in terms of standard convex duality.

In fact, what we have arrived at is a *convex operational theory* or *generalized probabilistic theory* (GPT) [4, 36], and in essence, we have proven a version of Ludwig's

embedding theorem for finite dimensional GPTs [74, 32]. For comparison, we reproduce a somewhat informal statement of that theorem as summarized in [104].

Theorem 3.5.1 (Finite dimensional Ludwig's embedding theorem, informal)**.** *Let $\mathcal{S}$ be the set of states, and $\mathcal{E}$ be the set of effects. A* theory *is a function $f : \mathcal{E} \times \mathcal{S} \to [0, 1]$ which takes a state and an effect and returns a probability. Let f satisfy the following axioms:*

1. *The function f separates states and effects: if $\rho_1, \rho_2 \in \mathcal{S}$ satisfy $\forall E \in \mathcal{E} : f(E, \rho_1) = f(E, \rho_2)$, then $\rho_1 = \rho_2$. Similarly, if $E_1, E_2 \in \mathcal{E}$ satisfy $\forall \rho \in \mathcal{S} : f(E_1, \rho) = f(E_2, \rho)$, then $E_1 = E_2$.*

2. *There is a trivial effect u such that $\forall \rho \in \mathcal{S} : f(u, \rho) = 1$, and for every $E \in \mathcal{E}$ there is an opposite effect $\neg E$ such that $\forall \rho \in \mathcal{S} : f(\neg E, \rho) = 1 - f(E, \rho)$.*

3. *Probabilistic mixtures of states (respectively, effects) are again states (respectively, effects). For all states $\rho_1, \rho_2 \in \mathcal{S}$ and probabilities $p \in [0, 1]$, there exists a state τ such that $\forall E \in \mathcal{E} : f(E, \tau) = p f(E, \rho_1) + (1 - p) f(E, \rho_2)$, and similarly for effects.*

Then there exists a vector space V such that

- *$\mathcal{S} \subseteq V, \mathcal{E} \subseteq V^*$ are convex sets with $0, u \in \mathcal{E}$ and $\mathcal{E} = u - \mathcal{E}$.*
- *$\mathbb{R}^+ \cdot \mathcal{S} = \hat{\mathcal{S}}$ is a cone, $\mathbb{R}^+ \cdot \mathcal{E} \subseteq \hat{\mathcal{S}}^* = \{y \in V^* : \forall x \in \hat{\mathcal{S}}, y(x) \geq 0\} \subseteq V^*$ is the corresponding dual cone, and the set of states satisfies $\mathcal{S} = \{x \in \hat{\mathcal{S}} : e(x) = 1\}$.*
- *The generalized Born rule holds: $f(E, \rho) = E(\rho)$.*

Here V^* is the dual of the vector space V, the space of all linear functionals acting on V. We call e the normalization functional.

Remark. *The essential difference in our development is that we begin from the assumption of a reference measurement, and admit conditioning only on the consequences of an agent's own actions. The basic properties of the state and effect spaces for us derive from demanding coherence with the reference measurement.*

Remark. *We merely assume that the states are a* subset *of the cone dual to the effects, and the effects are a* subset *of the cone dual to the states. In the language of GPTs, we do not assume the "no-restriction hypothesis."*

Remark. *Another subtlety is that in the standard GPT presentation, one takes states to live in an ordered real vector space V and effects to live in the dual space of linear functionals V^* so that $P(E|\rho) = E(\rho)$, where E is a linear functional acting on a state ρ—without necessarily equipping V with an inner product. But every finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ may be turned into an inner product space simply by choosing a basis and defining an inner product on the coordinates (identifying V with $\mathbb{R}^n$ or $\mathbb{C}^n$). The finite dimensional Riesz*

representation theorem [105] guarantees that one may identify each linear functional $E \in V^*$ with a unique vector $E \in V$ such that $P(E|\rho) = E(\rho) = \langle E, \rho \rangle = (E|\rho)$. For us, the inner product is inherited from the probability representation, and thus in essence we already work in the framework of "geometric GPT's" introduced in [76].

3.5.1 Examples

Classical probability theory is an example of a GPT: the state space is the space of probability distributions on n outcomes, i.e., the probability simplex Δ_n, and the effect space is the hypercube Δ_n^* dual to the simplex, the space of response functions. In other words, $P(E|\rho) = \sum_i P(E|\lambda_i)P(\lambda_i|\rho)$. The normalization functional is $(1| = (1, \ldots, 1)$; measurements correspond to (column) stochastic matrices; and general transformations correspond to (sub)stochastic matrices.

Quantum theory is also a GPT. The state space of quantum theory is the space of $d \times d$ density matrices, i.e., positive semidefinite Hermitian matrices $\rho \geq 0$ with $\mathrm{tr}(\rho) = 1$, which tells us that the normalization functional is $\mathrm{tr}(\cdot)$. The effect space consists of positive semidefinite Hermitian matrices $0 \leq E \leq I$ with no restriction on their trace. Measurements in quantum theory correspond to *positive operator valued measures* (POVMs), which map events to positive semidefinite operators on Hilbert space or, more simply, which are collections of effects such that $\sum_i E_i = I$. Both the state space and the effect space consist of PSD matrices: quantum theory is what is known as a *self-dual* theory, for which effects can always be rescaled into states. Transformations correspond to completely positive trace preserving (CPTP) maps, of which unitary transformations, which preserve pure states, are a special case. Quantum theory provides an example of a Euclidean Jordan algebra, about which we will have much to say in the sequel.

Notice that we can consider quantum states and effects as elements of vector spaces over $\mathbb{R}$: $\mathrm{Herm}(d)$ is a real vector space of dimension d^2. For example, for a qubit, the state space corresponds to the closed unit ball in $\mathbb{R}^3$, which we can see by expanding in the orthogonal basis provided by the Pauli matrices

$$\rho = \frac{1}{2}\Big(I + \mathrm{tr}(\sigma_x\rho)\sigma_x + \mathrm{tr}(\sigma_y\rho)\sigma_y + \mathrm{tr}(\sigma_z\rho)\sigma_z\Big). \tag{3.46}$$

In this representation, the normalization functional is $(1| = (1, 0, 0, 0)$, and we have

$$(E| = \frac{1}{2}\begin{pmatrix}\mathrm{tr}(E) & \mathrm{tr}(\sigma_x E) & \mathrm{tr}(\sigma_y E) & \mathrm{tr}(\sigma_z E)\end{pmatrix} \qquad |\rho) = \begin{pmatrix}1 \\ \mathrm{tr}(\sigma_x\rho) \\ \mathrm{tr}(\sigma_y\rho) \\ \mathrm{tr}(\sigma_z\rho)\end{pmatrix} \tag{3.47}$$

so that

$$P(E) = \mathrm{tr}(E\rho) = (E|\rho). \tag{3.48}$$

This same "Bloch sphere" construction can be generalized to any dimension d by employing, for example, the generalized Gell-Mann matrices, which are similarly Hermitian, traceless, and orthogonal—and we will take recourse to the analogous representation in nonquantum theories.

More generally, one can consider constructing GPTs from first principles, with more exotic state and effect spaces, e.g., theories with square state spaces, state spaces corresponding to higher dimensional spheres, state spaces shaped like ice cream cones, and deduce many illuminating theorems relating their geometries to a diverse set of operationally grounded axioms.

3.6 Frame theory

Let us now work in the opposite direction, seeing how we can begin from the GPT framework and arrive at the Born matrix formalism. The bridge is *frame theory* [62].

Definition 3.6.1 (Frame)**.** *A* frame *is a set of vectors* $\{|\phi_i)\}$ *in an inner product space* V *over a field* $\mathbb{F}$ *such that there exist constants* $0 < a \leq b < \infty$ *such that*

$$\forall x \in V : a(x|x) \leq \sum_i |(\phi_i|x)|^2 \leq b(x|x). \tag{3.49}$$

Remark. *If all the frame vectors have equal norm, this is called a* equal norm*,* unweighted*, or* unbiased *frame.*

One may begin from this definition, but it is easier to motivate it in the following way. Suppose V to be finite dimensional, with dimension r. Then let $\mathcal{A} \in \mathbb{F}_r^{n\times r}$ be an *analysis operator*,

$$\mathcal{A} = \sum_i |i)(\phi_i| : \tag{3.50}$$

$\mathcal{A}$ analyzes a vector $|x) \in V$ in terms of the frame vectors,

$$|y) = \mathcal{A}|x) = \sum_i (\phi_i|x)|i). \tag{3.51}$$

Since $\mathcal{A}$ has full column rank, it has a left inverse. Another way of saying this is that the frame vectors span V. A *synthesis operator* $\mathcal{S}$ is any choice of left inverse for $\mathcal{A}$. Writing it

$$\mathcal{S} = \sum_i |\tilde{\phi}_i)(i|, \tag{3.52}$$

where $\{|\tilde{\phi}_i)\}$ are called *dual vectors*, we have the resolution of the identity

$$I = \mathcal{S}\mathcal{A} = \sum_i |\tilde{\phi}_i)(\phi_i|. \tag{3.53}$$

Notice that we may rewrite

$$\sum_i |(\phi_i|x)|^2 = \sum_i (x|\phi_i)(\phi_i|x) = (x|\left[\sum_i |\phi_i)(\phi_i|\right]|x) = (x|F|x), \tag{3.54}$$

where

$$F = \mathcal{A}^\dagger\mathcal{A} = \sum_i |\phi_i)(\phi_i| \tag{3.55}$$

is called the *frame operator*, which is clearly not only Hermitian but also positive semidefinite. The frame condition then reads

$$\forall|x) \in V : a(x|x) \leq (x|F|x) \leq b(x|x), \tag{3.56}$$

so that a, b are the smallest and largest eigenvalues of F, respectively. As we saw in Corollary (3.4.2.1), F must be invertible. Suppose it were not. Then for some $|x)$, $(x|F|x) = (x|\mathcal{A}^\dagger\mathcal{A}|x) = \|\mathcal{A}|x)\|^2 = 0$ which can only be true if $\mathcal{A}|x)$ is the zero vector. But then $\mathcal{A}$ couldn't have a left inverse, a contradiction. Conversely, if F is invertible, then the corresponding analysis operator has a left inverse, and the frame condition is satisfied. Of course, one can consider frames for *subspaces* of a vector space, in which case F must be invertible on that subspace, and the restriction of $\mathcal{A}$ to that subspace must have a left inverse.

A *tight frame* is one for which $a = b$, implying

$$I = \frac{1}{a}F = \sum_i \frac{1}{a}|\phi_i)(\phi_i|, \tag{3.57}$$

that is, the dual vectors are simply the frame vectors rescaled. When $a = 1$, the tight frame is known as a Parseval frame. We note that

$$G = \mathcal{A}\mathcal{A}^\dagger = \sum_{ij} (\phi_i|\phi_j)|i)(j| \tag{3.58}$$

is the Gram matrix of the frame. Calculating the Gram matrix is a useful way of checking whether the frame vectors really span the space: they will if and only if the rank of G is r, the dimension of V. To see this, recall the following familiar lemma:

Lemma 3.6.1. *Let $A = XX^\dagger$ and $B = X^\dagger X$. Then A and B have the same non-zero eigenvalues.*

Proof. $Av = \lambda v \implies XX^\dagger v = \lambda v \implies X^\dagger X(X^\dagger v) = \lambda(X^\dagger v) \implies Bv' = \lambda v'$. Thus if λ, v is an eigenvalue/eigenvector pair of $A = XX^\dagger$, then $\lambda, X^\dagger v$ is an eigenvalue/eigenvector pair of $B = X^\dagger X$. □

Corollary 3.6.1.1. *For a Parseval frame, the $r \times r$ frame operator $F = \mathcal{A}^\dagger\mathcal{A} = I$ has r eigenvalues equal to 1. Thus the $n \times n$ Gram matrix $G = \mathcal{A}\mathcal{A}^\dagger$ also has r eigenvalues equal to 1, and is consequently a rank-r projector. More generally, since the frame operator is invertible iff $\{|\phi_i)\}$ form a frame, and F and G share the same nonzero spectrum, then the Gram matrix will be rank-r if and only if the n frame vectors span the r-dimensional space V.*

Finally, call

$$\mathcal{F}(\{|\phi_i)\}) = \mathrm{tr}(F^2) = \sum_{ij} |(\phi_i|\phi_j)|^2 \tag{3.59}$$

the *frame potential*. The frame potential allows one to characterize tight frames as the following lemma shows [62].

Lemma 3.6.2. *Let $\{|\phi_i)\}$ be a frame for a vector space of dimension r with frame operator $F = \sum_i |\phi_i)(\phi_i|$. The frame potential satisfies*

$$\mathcal{F}(\{|\phi_i)\}) \geq \frac{1}{r} tr(F)^2 \tag{3.60}$$

with equality iff $\{|\phi_i)\}$ forms a tight frame.

Proof. Since F is positive semidefinite and invertible, it will have r nonnegative eigenvalues. Recalling

$$\mathrm{tr}(F) = \sum_{k=1}^{r} \lambda_k \qquad\qquad \mathrm{tr}(F^2) = \sum_{k=1}^{r} \lambda_k^2, \tag{3.61}$$

and applying the Cauchy-Schwartz inequality,

$$\left(\sum_{k=1}^{r} \lambda_k \cdot 1\right)^2 \leq \left(\sum_{k=1}^{r} \lambda_k^2\right)\left(\sum_{k=1}^{r} 1^2\right), \tag{3.62}$$

we conclude

$$\mathrm{tr}(F^2) \geq \frac{1}{r}\mathrm{tr}(F)^2. \tag{3.63}$$

Equality is given when the vectors $(\lambda_1, \ldots, \lambda_r) = a(1, \ldots, 1)$, that is, they are proportional to each other so that all the eigenvalues are the same. We conclude that $\mathrm{tr}(F) = ra$, and

$$F = \frac{\mathrm{tr}(F)}{r} I. \tag{3.64}$$

Thus the frame must be tight with frame constant $a = \mathrm{tr}(F)/r$. □

Corollary 3.6.2.1. *In the special case that the frame vectors are normalized, $tr(F) = \sum_i (\phi_i|\phi_i) = n$ so that the lower bound becomes $\mathcal{F}(\{|\phi_i)\}) \geq n^2/r$.*

3.6.1 Dual frames

Taking the adjoint of the resolution of the identity

$$I = \mathcal{A}^\dagger \mathcal{S}^\dagger = \sum_i |\phi_i)(\tilde{\phi}_i|, \tag{3.65}$$

we see that

$$\mathcal{S}^\dagger = \sum_i |i)(\tilde{\phi}_i| \tag{3.66}$$

may be viewed as the analysis operator of a *dual frame*, with frame vectors $\{|\tilde{\phi}_i)\}$, dual frame operator $\tilde{F} = \mathcal{S}\mathcal{S}^\dagger = \sum_i |\tilde{\phi}_i)(\tilde{\phi}_i|$, and with a choice of synthesis operator,

$$\mathcal{A}^\dagger = \sum_i |\phi_i)(i|. \tag{3.67}$$

For any frame, there is a *canonical dual frame*. One way of characterizing it, useful to us, is the following.

Lemma 3.6.3. *The range of the dual analysis operator $\mathcal{S}^\dagger$ to be the same as the range of the original analysis operator $\mathcal{A}$ iff $\mathcal{S} = \mathcal{A}^+$, the Moore-Penrose pseudoinverse.*

Proof. If $\mathrm{col}(\mathcal{S}^\dagger) = \mathrm{col}(\mathcal{A})$, then we can write

$$\mathcal{S}^\dagger = \mathcal{A}X, \tag{3.68}$$

for some X. But then $I = \mathcal{S}\mathcal{A} = X^\dagger \mathcal{A}^\dagger \mathcal{A} = X^\dagger F$. Since F is invertible, this means that $X^\dagger = F^{-1} = X$, since F is Hermitian. We conclude that canonical synthesis operator is given by

$$\mathcal{S}_{\mathrm{can}} = F^{-1}\mathcal{A}^\dagger = (\mathcal{A}^\dagger \mathcal{A})^{-1}\mathcal{A}^\dagger = \mathcal{A}^+, \tag{3.69}$$

that is, the canonical synthesis operator is Moore-Penrose pseudoinverse $\mathcal{A}^+$ of the analysis operator. Conversely, if $\mathcal{S} = \mathcal{A}^+ = F^{-1}A^\dagger$, then $\mathcal{S}^\dagger = \mathcal{A}F^{-1}$ so that $\mathrm{col}(\mathcal{S}^\dagger) = \mathrm{col}(\mathcal{A}F^{-1}) = \mathrm{col}(\mathcal{A})$ since F^{-1} is invertible[2]. □

By the above argument, the canonical dual frame operator $\tilde{F} = F^{-1}$, and so the canonical dual frame satisfies

$$\forall |x) \in V : \frac{1}{b}(x|x) \leq (x|F^{-1}|x) \leq \frac{1}{a}(x|x). \tag{3.70}$$

Notice that since $\mathcal{S}$ is a left inverse of $\mathcal{A}$, $\mathcal{A}\mathcal{S}\mathcal{A} = \mathcal{A}$: we can equivalently say that $\mathcal{S}$ is any $\{1\}$-inverse of $\mathcal{A}$. From Lemma 3.3.1, we know therefore that $\mathcal{A}\mathcal{S}$ is a projector onto $\mathrm{range}(\mathcal{A})$. We may therefore understand this now familiar fact in a new way: we take a vector in the n-dimensional space where the frame representation lives, project it into the original r-dimensional space V, and then return back to the n-dimensional space. Since in general a frame is overcomplete, there may be many different n-dimensional vectors which project to the same r-dimensional vector. But there is a special choice: that given by the analysis operator itself, that is, the one that lives in $\mathrm{range}(\mathcal{A})$ and $\Pi = \mathcal{A}\mathcal{S}$ is the projector onto that subspace. We note in the special case that $\mathcal{S} = \mathcal{A}^+$, we have $\Pi = \mathcal{A}\mathcal{A}^+ = \mathcal{A}(\mathcal{A}^\dagger\mathcal{A})^{-1}\mathcal{A}^\dagger = \Pi^\dagger$: the projection is in fact an orthogonal projection.

Finally, as in Lemma (3.4.3), notice that given one choice of synthesis operator, we can write any other choice of synthesis operator

$$\mathcal{S} = \mathcal{S}_0 + X(I - \Pi), \tag{3.71}$$

since $\mathcal{S}\mathcal{A} = \mathcal{S}_0\mathcal{A} + X(I - \Pi)\mathcal{A} = I$, where X is completely arbitrary, as $(I - \Pi)\mathcal{A} = 0$.

3.6.2 Frame coordinates and dual coordinates

Given a frame with analysis and synthesis operators

$$\mathcal{A} = \sum_i |i)(\phi_i| \qquad\qquad \mathcal{S} = \sum_i |\tilde{\phi}_i)(i|, \tag{3.72}$$

and a choice of dual frame with analysis and synthesis operators

$$\mathcal{S}^\dagger = \sum_i |i)(\tilde{\phi}_i| \qquad\qquad \mathcal{A}^\dagger = \sum_i |\phi_i)(i|, \tag{3.73}$$

[2]We can understand this as a basic property of the pseudoinverse itself. To see it, recall that $X = U\Sigma V^\dagger$ is the singular value decomposition of X, then $X^+ = V\Sigma^+U^\dagger$, where Σ^+ is the diagonal matrix of the reciprocals of the nonzero singular values. Since the first $r = \mathrm{rank}(X)$ columns of U form a basis for $\mathrm{col}(X)$ and the first r rows of $V^\dagger$ form a basis for $\mathrm{row}(X)$, we have $\mathrm{range}(X^+) = \mathrm{range}(X^\dagger)$. If X is Hermitian, then $\mathrm{range}(X^+) = \mathrm{range}(X)$.

we have two resolutions of the identity $\mathcal{S}\mathcal{A} = I = \mathcal{A}^\dagger\mathcal{S}^\dagger$, providing two representations of a vector $|x)$,

$$|x) = \mathcal{S}\mathcal{A}|x) = \sum_i (\phi_i|x)|\tilde{\phi}_i) = \mathcal{A}^\dagger\mathcal{S}^\dagger|x) = \sum_i (\tilde{\phi}_i|x)|\phi_i). \tag{3.74}$$

Let $|y)$ denote the vector with components $(\phi_i|x)$: we will call this the *frame representation* in terms of *regular coordinates*. Meanwhile, let $|\tilde{y})$ denote the vector with components $(\tilde{\phi}_i|x)$: the *dual frame representation* in terms of *dual coordinates*.

Suppose we seek a *transfer matrix* $\mathcal{T}$ which takes us between the two representations. Unless the frame forms a basis, there will be multiple choices of dual frame, and moreover, for a given choice of dual frame, there will be multiple choices of transfer matrices $\mathcal{T}$. As you might suspect, $\mathcal{T}$ must be a $\{1\}$-inverse of the Gram matrix G, satisfying $G\mathcal{T}G = G$. Since $G = \mathcal{A}\mathcal{A}^\dagger$ and $\mathcal{A}$ and $\mathcal{A}^\dagger$ have left and right inverses respectively, this implies $\mathcal{A}^\dagger\mathcal{T}\mathcal{A} = I$. Then

$$\mathcal{A}^\dagger\mathcal{T}\mathcal{A}|x) = \mathcal{A}^\dagger\mathcal{T}|y) = \mathcal{A}^\dagger|\tilde{y}') = |x), \tag{3.75}$$

that is, the dual synthesis operator acts on $|\tilde{y}')$ to give us back $|x)$. In this sense, $\mathcal{T}$ delivers perfectly good dual coordinates, but they need not lie in the range of the dual analysis operator $\mathcal{S}^\dagger$. If we select the canonical dual frame, however, then the range of $\mathcal{S}^\dagger$ is the same as the range of $\mathcal{A}$. Moreover, $\text{range}(G) \subseteq \text{range}(\mathcal{A})$, and taking $\mathcal{T} = G^+$, $\text{range}(\mathcal{T}) = \text{range}(G)$ so that $\mathcal{T}|y) = |\tilde{y})$ for $|\tilde{y}) = \mathcal{S}^\dagger|x)$. Finally, $G|\tilde{y}) = \mathcal{A}\mathcal{A}^\dagger\mathcal{T}\mathcal{A}|x) = \mathcal{A}|x) = |y)$ takes us back to dual coordinates to regular frame coordinates.

3.6.3 Double sided frames

Suppose we introduce two interlocking frames, one for vectors and another for covectors: we shall call such a construction a *double sided frame*, a notion implicit in the literature, but original to us. In the context of a GPT, this would be like introducing a frame for states consisting of reference effects $\{(R_i|\}$, and a frame for effects consisting of reference states $\{|S_i)\}$. In other words, the frame vectors (or covectors) are in fact states and effects in the GPT. If the GPT dimension is r, then we demand that the reference states and effects both span the r-dimensional space: and in this context, we call them *informationally* or *tomographically* complete. (We will have more to say on this notion in the following section.) Of course, we would also like the reference effects to form a legitimate measurement, that is, $\sum_i (R_i| = (1|$.

In other words, let

$$\mathbf{R} = \sum_i |i)(R_i| \qquad\qquad \mathbf{S} = \sum_i |S_i)(i| \tag{3.76}$$

be the analysis operators for states and effects, respectively. If the reference effects are spanning, then $\mathbf{R}$ will have a left inverse, $\mathbf{R}^L\mathbf{R} = I$ and similarly if the reference states are spanning, then $\mathbf{S}$ will have a right inverse $\mathbf{S}\mathbf{S}^R = I$. Putting these two resolutions of the identity together, we have

$$\mathbf{S}(\mathbf{S}^R\mathbf{R}^L)\mathbf{R} = \mathbf{S}\Phi\mathbf{R} = I. \tag{3.77}$$

Indeed, this is a necessary and sufficient condition for such a Φ to exist:

Theorem 3.6.4. *Let $A \in \mathbb{F}_r^{r\times m}$ and $B \in \mathbb{F}^{m\times r}$. Then there exists a $\Phi \in \mathbb{F}^{m\times n}$ such that $A\Phi B = I$ iff $\mathrm{rank}(A) = \mathrm{rank}(B) = r$.*

Proof. If $A\Phi B = I$, then from

$$r = \mathrm{rank}(I) = \mathrm{rank}(A\Phi B) \leq \mathrm{rank}(A) \leq r \tag{3.78}$$

$$r = \mathrm{rank}(I) = \mathrm{rank}(A\Phi B) \leq \mathrm{rank}(B) \leq r, \tag{3.79}$$

we conclude that $\mathrm{rank}(A) = \mathrm{rank}(B) = r$. Conversely, if $\mathrm{rank}(A) = r$, then A has a right inverse A^R, and if $\mathrm{rank}(B) = r$, then B has a left inverse B^L. Then we may take $\Phi = A^R B^L$: in fact, by Theorem (3.4.5) such a Φ is a $\{1, 2\}$-inverse. □

Now if we have any Φ satisfying $\mathbf{S}\Phi\mathbf{R} = I$, then clearly, $\tilde{\mathbf{S}} \equiv \mathbf{R}^L = \mathbf{S}\Phi$ is a valid synthesis operator for $\mathbf{R}$, and $\tilde{\mathbf{R}} \equiv \mathbf{S}^R = \Phi\mathbf{R}$ is a valid synthesis operator for $\mathbf{S}$: in this sense, the two frames are interlocking. Thus if we start from $\mathbf{S}\Phi\mathbf{R} = I$, we can derive $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{S}}$; conversely if we start from $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{R}}$, we can obtain a Born matrix $\Phi = \tilde{\mathbf{R}}\tilde{\mathbf{S}}$. Of course, this is not the only choice of Φ. As we have seen, satisfying $\mathbf{S}\Phi\mathbf{R} = I$ is equivalent to $P\Phi P = P$, where $P = \mathbf{R}\mathbf{S}$: that is, Φ must be a $\{1\}$-inverse of P, and unless P is invertible, there is no unique choice. But however the twin synthesis operators are obtained, we may define dual effects and dual states as the rows and columns respectively of

$$\tilde{\mathbf{S}} = \sum_i |\tilde{S}_i)(i| \qquad\qquad \tilde{\mathbf{R}} = \sum_i |i)(\tilde{R}_i|. \tag{3.80}$$

Our two resolutions of the identity

$$\tilde{\mathbf{S}}\mathbf{R} = I = \mathbf{S}\tilde{\mathbf{R}} \tag{3.81}$$

then yield

$$|\rho) = \tilde{\mathbf{S}}\mathbf{R}|\rho) = \sum_i |\tilde{S}_i)(R_i|\rho) = \sum_i P(R_i|\rho)\ |\tilde{S}_i) \tag{3.82}$$

$$|\rho) = \mathbf{S}\tilde{\mathbf{R}}|\rho) = \sum_i |S_i)(\tilde{R}_i|\rho) = \sum_i W(R_i|\rho)\ |S_i) = \sum_{ij} \Phi_{ij} P(R_j|\rho)|S_i), \tag{3.83}$$

where we write $W(R_i|\rho)$ to signify that these quantities need not be probabilities.

As usual, the Moore-Penrose pseudoinverse plays a special role in picking out a Φ. Let F_S and F_R be the frame operators for our two interlocking frames. In the canonical case, we have $\mathbf{S}^{\dagger} = \mathbf{S}^{\dagger} F_S^{-1} = \Phi\mathbf{R}$ and $\mathbf{R}^{L} = F_R^{-1}\mathbf{R}^{\dagger} = \mathbf{S}\Phi$, and the choice of $\Phi = \mathbf{S}^{R}\mathbf{R}^{L} = \mathbf{S}^{\dagger} F_S^{-1} F_R^{-1}\mathbf{R}^{\dagger}$ tells us that in fact $\Phi = P^{+}$.

Theorem 3.6.5. *If $M = AB$ is a rank factorization, $M^{+} = B^{+}A^{+} = B^{\dagger}(BB^{\dagger})^{-1}(A^{\dagger}A)^{-1}A^{\dagger}$ [71].*

Finally, just as $P = \mathbf{RS} = \sum_{ij}(R_i|S_j)|i)(j|$, we may define the channel operator

$$C = \mathbf{SR} = \sum_i |S_i)(R_i|, \tag{3.84}$$

so that $(E|C|\rho) = \sum_i (E|S_i)(R_i|\rho) = P(E|R)P(R|\rho) = P(E|R,\rho)$. If P is the double sided frame analogue of a Gram matrix, then C is the double sided frame analogue of the frame operator. In fact, C picks out yet another choice of Born matrix, namely $\Phi = \mathbf{R}C^{-2}\mathbf{S}$.

Theorem 3.6.6. *Let $M = AB$ be a rank factorization and BA is invertible. Then $M^{\sharp} = B(BA)^{-2}A$ is a $\{1,2,5\}$-inverse [71], or* group inverse.

3.7 Informational completeness

3.7.1 Separation

The essence of a GPT is a state space and effect space which separate each other. This means that states are uniquely determined by the probabilities they induce on all the effects, and the effects are uniquely determined by the probabilities they induce on all the states. In other words, a GPT is a model which is entirely grounded in probability assignments: it contains no extraneous structure beyond the probabilities an agent might assign to all possible measurements on a system. This property of "separation" may be checked in a purely linear algebraic way, as the following theorem shows.

Theorem 3.7.1 (Adapted from [81])**.** *Let $\mathcal{S}$ and $\mathcal{E}$ be state and effect spaces. The following two statements are equivalent:*

1. *$\mathcal{S}$ and $\mathcal{E}$ separate each other.*
2. *$\mathcal{E}$ separates vectors in span$(\mathcal{S})$ and $\mathcal{S}$ separates vectors in span$(\mathcal{E})$.*

Proof. (2) $\implies$ (1) is immediate since $\mathcal{S} \subseteq \text{span}(\mathcal{S})$ and $\mathcal{E} \subseteq \text{span}(\mathcal{E})$. As for (1) $\implies$ (2): by the assumption,

$$\forall \pi \in \mathcal{S} : (E|\pi) = (E'|\pi) \iff (E| = (E'| \tag{3.85}$$

$$\forall E \in \mathcal{E} : (E|\pi) = (E|\pi') \iff |\pi) = |\pi'). \tag{3.86}$$

Now suppose on the contrary that $\mathcal{E}$ did *not* separate $\mathrm{span}(\mathcal{S})$. This would mean that

$$\exists |x), |y) \in \mathrm{span}(\mathcal{S}) \text{ such that } \forall E \in \mathcal{E} : (E|x) = (E|y) \text{ but } |x) \neq |y). \tag{3.87}$$

On the one hand, $\forall E \in \mathcal{E} : (E|\big(|x) - |y)\big) = 0$ tells us in particular that $(1|\big(|x) - |y)\big) = 0$, where $\forall \pi \in \mathcal{S} : (1|\pi) = 1$. On the other hand, $|x) - |y) \in \mathrm{span}(\mathcal{S})$ so we can express it as a linear combination of states $|x) - |y) = \sum_i \alpha_i |\pi_i)$ yielding

$$(1|\big(|x) - |y)\big) = \sum_i \alpha_i (1|\pi_i) = \sum_i \alpha_i = 0. \tag{3.88}$$

Separating out the nonnegative from the negative components in the linear combination, we have $\sum_{\alpha_i \geq 0} \alpha_i = \sum_{\alpha_i < 0} |\alpha_i| = N$ for some N. Thus

$$|\sigma^+) = \sum_{\alpha_i \geq 0} \frac{\alpha_i}{N} |\pi_i) \qquad\qquad |\sigma^-) = \sum_{\alpha_i < 0} \frac{|\alpha_i|}{N} |\pi_i) \tag{3.89}$$

are both convex combinations of states and so states themselves. Therefore

$$\forall E \in \mathcal{E} : \frac{1}{N}(E|\big(|x) - |y)\big) = \frac{1}{N}(E| \sum_i \alpha_i |\pi_i) = (E| \left\{ \sum_{\alpha_i \geq 0} \frac{\alpha_i}{N} |\pi_i) - \sum_{\alpha_i < 0} \frac{|\alpha_i|}{N} |\pi_i) \right\} \tag{3.90}$$

$$= \big(E|(|\sigma^+) - |\sigma^-\big) = 0. \tag{3.91}$$

But since the effects separate the states, $\forall E \in \mathcal{E} : (E|\sigma^+) = (E|\sigma^-)$ implies that $|\sigma^+) = |\sigma^-)$. But $|\sigma^+) - |\sigma^-) = \frac{1}{N}\big(|x) - |y)\big)$ this implies that $|x) = |y)$, contradicting our original assumption. We conclude $\mathcal{E}$ separates $\mathrm{span}(\mathcal{S})$ after all.

We use a different proof strategy to show that $\mathcal{S}$ must separate $\mathrm{span}(\mathcal{E})$. Again assuming the contrary means that

$$\exists (x|, (y| \in \mathrm{span}(\mathcal{E}) \text{ such that } \forall \pi \in \mathcal{S} : (x|\pi) = (y|\pi) \text{ but } (x| \neq (y|. \tag{3.92}$$

In other words, $\forall \pi \in \mathcal{S} : \big((x| - (y|)|\pi) = 0$. Now suppose we pick an effect $(e_0|$ in the *relative interior* [106] of the effect space. Recall that a point is in the relative interior of a set S just when there exists an $\epsilon > 0$ such that the intersection of a ball of radius ϵ centered on x with the affine hull of S is contained in the set: the relative interior is appropriate for talking about the interior of a set which is confined e.g., to a subspace. Since the effect space contains

the 0 vector, in fact, its affine hull is the span$(\mathcal{E})$[3]. Then for sufficiently small $\epsilon > 0$, let

$$(e| = (e_0| + \epsilon\big((x| - (y|\big). \tag{3.93}$$

Since $(e_0|$ is in the relative interior, then $(e|$ will be a valid effect. But

$$\forall \pi : (e|\pi) = (e_0|\pi) + \epsilon\big((x| - (y|\big)|\pi) = (e_0|\pi), \tag{3.94}$$

showing that the π's can't distinguish $(e_0|$ from $(e|$. But this contradicts our assumption: we conclude that $\mathcal{S}$ can separate span$(\mathcal{E})$ after all. □

Corollary 3.7.1.1. *Let $\{(R_i|\}$ be a set of effects whose span is span$(\mathcal{E})$ and $\{|R_i)\}$ be a set of states whose span is span$(\mathcal{S})$. Then the probabilities $P(R|\rho)$ suffice to characterize any state ρ and the probabilities $P(E|R)$ suffice to characterize any effect E. We call $\{(R_i|\}$ and $\{|R_i)\}$* informationally complete*, and we call a preparatory measurement characterized by the conditional probability matrix $P(R|R)$ a* reference measurement*.*

In our earlier development, we took the existence of a reference measurement as an assumption. If we start instead with an arbitrary GPT, is there guaranteed to be a reference measurement? We reproduce here a proof given in [41, 107] that informationally complete measurements exist in any GPT which allows all mathematically possible effects, that is, which satisfies the no-restriction hypothesis. In fact, it shows that minimal IC (MIC) measurements always exist, where $n = r$: the number of outcomes is precisely the dimension of the GPT. The effects of a MIC therefore form a basis, and so $\Phi = P(R|R)^{-1}$.

Theorem 3.7.2. *There exists a MIC in any GPT satisfying the no-restriction hypothesis.*

Proof. Pick an arbitrary basis $\{(b_i|\}_{i=1}^r$. We have $\sum_i (b_i| = (s|$, whatever $(s|$ may be, but we can always find some invertible matrix V which takes $(s|V = (1|$ so that the covectors $(b'_i| = (b_i|V$ sum to $(1|$. Now let m be the minimal value attained by any of the $(b'_i|$'s on any of the states ρ of the GPT, that is, $m = \min_{i,\rho\in\mathcal{S}}(b'_i|\rho)$. Because the state space is assumed compact, the minimum can be attained. In finite dimensions, every linear functional is continuous, and so the image of $f(\rho) = (E|\rho) : \mathcal{S} \to \mathbb{R}$ must be itself compact. Thus the minimum is some finite value. Clearly, $(b''_i| = (b'_i| - m(1|$ satisfies

$$\forall i, \rho \in \mathcal{S} : (b''_i|\rho) = (b'_i|\rho) - m \geq \min_{i,\rho\in\mathcal{S}}(b'_i|\rho) - m \geq 0. \tag{3.95}$$

[3]Recall that aff$(S) = \{\sum_i \alpha_i x_i, x_i \in S, \sum_i \alpha_i = 1\}$. If $0 \in S$, then we can extend $y = \sum_i \alpha_i x_i + (1 - \overline{\alpha})0$ where $\overline{\alpha} = \sum_i \alpha_i$. Then y is written as an affine combination. In the other direction, an affine combination is already a linear combination.

Moreover, $\sum_i (b_i''| = (1 - rm)(1|$ so that $\sum_i (b_i''|\rho) = 1 - rm > 0$, since each term is individually nonnegative, and they are assumed to not all be identically 0. Thus let

$$(R_i| = \frac{1}{1 - rm}\Big((b_i|V - m(1|\Big). \tag{3.96}$$

$(R_i|$ must still span the space, and since there are r elements, it must form a basis. Moreover, the $(R_i|$ are nonnegative on all states, and sum to $(1|$, and so represent a logically possible measurement. □

Remark. Any *measurement is informationally complete for* some *set of states and effects.*

Remark. *In particular, one may always construct an equiangular measurement in no-restriction GPT. One simply begins with a regular* $(r-1)$*-simplex with* r *vertices. Then as in the theorem, one shrinks the simplex until the effects corresponding to the vertices (obtained by adding in a component proportional to* $(1|$*) are nonnegative on all states in the GPT. Such effects must form a linearly independent basis.*

Remark. *In* any *GPT, not just no-restriction GPTs, reference measurements exist. At worst, one may simply consider all allowed measurements: the effects of each individually sum to* $(1|$ *so that they may be rescaled to form one single measurement. Since GPT effects must separate the states, this will be a reference measurement.*

3.8 de Finetti for GPT's

Using informationally complete measurements, [107] proves a de Finetti representation theorem for a wide class of GPT's[4]. Analogous to the classical de Finetti representation theorem, this theorem gives a subjective Bayesian justification for when one can act *as if* repeated reference measurements are estimating "unknown states."

Theorem 3.8.1. *Suppose* $|\rho^n) \in V^{\otimes n}$ *is exchangeable, meaning that it is*

(a) invariant under permutations of the n systems;

(b) n-fold nonsignalling, that is, nonsignalling across every bipartite split: this means that the marginal probability of obtaining an outcome on system A does not depend on what measurement is done on B;

(c) extendable, so that $|\rho^n)$ *may be obtained as the marginal of a state* $|\rho^N)$ *for any* $N > n$*,*

then $|\rho^n)$ *may be uniquely expressed*

$$|\rho^n) = \int_{\mathcal{S}} d\mu(\rho)|\rho)^{\otimes n}, \tag{3.97}$$

[4] In fact, they work in the essentially equivalent framework of "test spaces."

where μ is a unique probability measure on the single system state space $\mathcal{S}$ which is independent of n.

The authors adopt the proof strategy of [73] in the setting of quantum mechanics: they introduce an informationally complete reference measurement to express $|\rho^n)$ as a joint probability distribution on n copies of the reference measurement, and then appeal to the classical de Finetti theorem. A crucial assumption, then, is that performing reference measurements separately on each part of the collective is equivalent to performing a reference measurement on the whole: this assumption is usually called *local tomography*. A famous example where local tomography fails is quantum theory over the real numbers. Finally, we note that [108] provides an analogous de Finetti theorem justifying the "estimation of unknown quantum processes": the proof strategy makes use of the Choi representation theorem and so may be extended to GPTs which admit a similar duality between states and transformations [109].

3.9 Conclusion

In this chapter, we have explored the properties of the Born matrix Φ, seeing that it must form a $\{1\}$-inverse of $P(R|R)$, the conditional probability matrix that characterizes the reference measurement. Further pushing the notion of nonclassical coherence allowed us to derive fundamental subspace restrictions for states, effects, and valuations. Using the technique of rank factorization, we then showed that the QBist framework is essentially equivalent to that of the framework of generalized probabilistic theories, allowing results of the latter to be imported into the former and vice versa. We now turn to the question of what is the simplest form the fundamental nonclassical coherence condition $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ can take, showing that in GPT-land, it may almost always be expressed as a particularly gentle modification of the law of total probability.

CHAPTER 4

DIE URGLEICHUNGEN

4.1 Introduction

As we have seen, on very general principles, we may derive a nonclassical coherence condition

$$P(E|\rho) = P(E|R)\Phi P(R|\rho), \tag{4.1}$$

where Φ is any $\{1\}$-inverse of $P(R|R)$, the conditional probability matrix that characterizes the reference measurement. In particular, in quantum theory, if the reference measurement is taken to be a SIC-POVM [58], we arrive at the QBist *Urgleichung*, or fundamental equation,

$$P(E|\rho) = \sum_i P(E|R_i)\left\{(d+1)P(R_i|\rho) - \frac{1}{d}\right\}, \tag{4.2}$$

which looks quite close to the law of total probability: the only difference is that one rescales and shifts the reference probabilities $P(R|\rho)$. Now this precise expression is unique to a SIC, but a SIC is an example of a so-called quantum state 2-design, and any measurement forming a 2-design will enjoy a similar expression [6]. More generally, we can ask: in any GPT, when it is possible to choose a reference measurement which yields a *Protourgleichung*,

$$P(E|\rho) = \sum_i P(E|R_i)\Big\{\alpha P(R_i|\rho) + (1-\alpha)P(R_i|\mu)\Big\}, \tag{4.3}$$

where μ represents a state of complete uncertainty? If $\alpha = d+1$ and $P(R|\mu) = 1/d^2$, as it does for a SIC, we recover the Urgleichung itself. We now show that in fact such expressions are ubiquitous in the landscape of GPTs: in a GPT which allows all mathematically possible states and effects, a set of reference states may be chosen so that any reference measurement gives rise to a *Protourgleichung*. An equivalent condition can be stated in terms of the channel operator associated with the reference measurement: it must be depolarizing, and we show that this places restrictions on the action of $P(R|R)$. We then explore in our terms the recently introduced notion of morphophoricity [6, 76], the idea that a reference measurement maps a state space into the probability simplex in a shape preserving way, providing an alternative proof that such measurements correspond to the tight IC-POVMs pioneered by Andrew Scott [77]. Finally, we examine critically a measure of nonclassicality found in the QBist literature,

$\|I - \Phi\|$ with respect to any unitarily invariant norm [78], showing that in the overcomplete case, this measure is unable to capture the essence of the *Protourgleichung* in a univocal way.

4.2 Bloch form

In what follows, it will prove useful to bring our states and effects into a standard *Bloch (block) form*. First we would like the normalization functional, corresponding to $P(1|R)$, to be represented as $(1| = (1, 0, \ldots, 0)$ so that states can be represented as

$$|\rho) = \begin{pmatrix} 1 \\ \rho_0 \end{pmatrix}. \tag{4.4}$$

We would also like to distinguish a particular state $|\mu)$ to represent the "state of complete uncertainty." In some GPTs, there is a unique choice of such a state. For example, in classical probability theory, such a state would correspond to $(1/n, \ldots, 1/n)$. Notice that this is the unique state which is invariant under all the symmetries of the state space which in this case are permutations of the simplex vertices. In quantum theory, such a state would correspond to I/d: similarly, this is the unique state invariant under the *unitary* symmetries of the state space. More generally, given a GPT state space, we may consider a state $|\mu)$ which is invariant under the reversible transformations which preserve the state space, and adopt this as the "state of complete uncertainty" [37, 76]. But in general, there may not be a single unique such state [110]. Moreover, in classical and quantum theory, several ideas coincide: a state which is invariant under state space symmetries, the geometric center of the state space, the maximal entropy state. In terms of our reference measurement formalism, we might be inclined to pick a state corresponding to an eigenvector with eigenvalue 1 of $P(R|R)$: indeed, since $P(R|R)$ is column stochastic, it must have at least one such eigenvector which is a probability distribution [111]: but in general, it will not have a unique such eigenvector. Thus the choice of $|\mu)$ depends essentially on the nature of the GPT. Of course, the simplest solution is to take the uniform average of all the extremal states. For our purposes, however, it suffices merely to distinguish *some* state $|\mu)$ in the interior of the state space, and we leave open the question of how to choose it. We will then work in a representation such that $|\mu) = (1, 0, \ldots, 0)^\dagger$.

Lemma 4.2.1. *Any finite-dimensional GPT can be brought, non-uniquely, into Bloch form: for a choice of normalized state $|\mu)$, one may choose coordinates such that*

$$(1| = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix} \qquad |\mu) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{4.5}$$

where $(1|$ is the normalization functional.

Proof. Let V be the r dimensional real vector space spanned by the states of the GPT, and let $(1| \in V^*$ be the unit effect such that $\forall |\rho) \in \mathcal{S} : (1|\rho) = 1$. Choose a normalized state $|\mu)$. Every vector $|v) \in V$ decomposes uniquely as

$$|v) = (1|v)|\mu) + |v_0), \tag{4.6}$$

where $|v_0) = |v) - (1|v)|\mu)$ and $(1|v_0) = 0$. We can thus split $V = \mathbb{R}|\mu) \oplus V_0$ where we shall call $V_0 = \{|x) \in V : (1|x) = 0\}$ the "traceless subspace." Picking a basis $\{|b_1), \ldots, |b_{r-1})\}$ for V_0, we may define an invertible matrix

$$B = \begin{pmatrix} |\mu) & |b_1) & \cdots & |b_{r-1}) \end{pmatrix}. \tag{4.7}$$

Then $(E|\rho) = (E|BB^{-1}|\rho) = (E'|\rho')$ shifts us into the new coordinate system in a probability preserving way. In particular, since $B(1, 0, \ldots)^\dagger = |\mu)$,

$$|\mu') = B^{-1}|\mu) = \begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix}. \tag{4.8}$$

At the same time, since $(1|\mu) = 1$ and $(1|b_i) = 0$,

$$(1'| = (1|B = (1, 0, \ldots), \tag{4.9}$$

as desired. The choice of block representation is not unique, since it depends on the chosen normalized state $|\mu)$ and on the choice of basis for V_0. □

In what follows therefore without loss of generality, we will assume that $(1| = (1, 0, \ldots)$ and $|\mu) = (1, 0, \ldots)^\dagger$ so that effects and states may be expressed

$$(E| = \left(\begin{array}{c|c} w & E_0 \end{array}\right) \qquad\qquad |\rho) = \begin{pmatrix} 1 \\ \hline \rho_0 \end{pmatrix}, \tag{4.10}$$

where $w = P(E|\mu)$. Moreover, a measurement may be expressed as a block matrix whose rows are effects,

$$\mathbf{E} = \left(\begin{array}{c|c} w & \mathbf{E}_0 \end{array}\right), \tag{4.11}$$

where w is a column vector of weights satisfying $\sum_{i=1}^n w_i = 1$ and the columns of $\mathbf{E}_0 \in \mathbb{R}^{n\times(r-1)}$ sum to 0. Thus $\sum_{i=1}^n (E_i| = (1|$, or more compactly, $u^\dagger \mathbf{E} = (1|$ where $u^\dagger = (1, \ldots, 1)$. Interpreting $|\mu)$ as the state of complete ignorance, the weights w_i can be understood as the *biases* of the measurement $P(E_i|\mu) = w_i$: even in the state of complete ignorance, the

gambler might nevertheless assign nonuniform probabilities to the outcomes of a measurement signifying its bias.

4.3 Ubiquitous Urgleichung

An operator $\mathbf{R}$ representing a measurement can be thought of as a linear map from the state space into the probability simplex, $P(R|\rho) = \mathbf{R}|\rho)$. Expressing it in block form makes clear that it can alternatively be understood as an *affine transformation* from the traceless subspace instead,

$$P(R|\rho) = \mathbf{R}|\rho) = \left(\begin{array}{c|c} w & \mathbf{R}_0 \end{array} \right) \left(\frac{1}{\rho_0} \right) = w + \mathbf{R}_0 \rho_0. \tag{4.12}$$

Assuming that $\{R_i\}$ is an informationally complete measurement, if we were to demand that the quasiprobabilities $W(R|\rho) = \tilde{\mathbf{R}}|\rho)$ be close as possible to $P(R|\rho) = \mathbf{R}|\rho)$, we could do no better than to assume that

$$W(R|\rho) = w + \alpha \mathbf{R}_0 \rho_0, \tag{4.13}$$

for some choice of a single parameter α. Comparing Eq. (4.12) and Eq. (4.13), we have

$$W(R|\rho) = \alpha(P(R|\rho) - w) + w = \alpha P(R|\rho) + (1 - \alpha)P(R|\mu), \tag{4.14}$$

since $w = P(R|\mu)$. Recalling that $P(E|\rho) = P(E|R)\Phi P(R|\rho) = P(E|R)W(R|\rho)$, we obtain what we shall call the *Protourgleichung*,

$$P(E|\rho) = \sum_{jk} P(E|R_j)\Phi_{jk}P(R_k|\rho) \tag{4.15}$$

$$= \sum_{j} P(E|R_j)\Big\{\alpha P(R_j|\rho) + (1 - \alpha)P(R_j|\mu)\Big\}, \tag{4.16}$$

that is, we may take $\Phi = \alpha I + (1 - \alpha)wu^\dagger$. Can this always be achieved? We now establish the following theorem:

Theorem 4.3.1. *Given an informationally complete measurement with n outcomes in a no-restriction GPT of dimension r with a distinguished interior state $|\mu)$, one can always choose a set of reference states which furnish a quasiprobability representation $W(R|\rho) = \alpha P(R|\rho) + (1 - \alpha)P(R|\mu)$, for some parameter $|\alpha| \geq 1$.*

Proof. Let

$$\mathbf{S} = \left(\begin{array}{c|c} 1 \quad \ldots \quad 1 & \\ \hline & \\ \mathbf{S}_0 & \\ & \end{array}\right) \qquad \tilde{\mathbf{R}} = \left(\begin{array}{c|c} w & \alpha\mathbf{R}_0 \end{array}\right). \tag{4.17}$$

The Born identity requires that $\mathbf{S}\tilde{\mathbf{R}} = I$. We have

$$\left(\begin{array}{c|c} 1 \quad \ldots \quad 1 & \\ \hline & \\ \mathbf{S}_0 & \\ & \end{array}\right)\left(\begin{array}{c|c} w & \alpha\mathbf{R}_0 \end{array}\right) = \begin{pmatrix} u^\dagger w & \alpha u^\dagger \mathbf{R}_0 \\ \mathbf{S}_0 w & \alpha \mathbf{S}_0 \mathbf{R}_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathbf{S}_0 w & \alpha \mathbf{S}_0 \mathbf{R}_0 \end{pmatrix}, \tag{4.18}$$

where the last follows from $u^\dagger w = 1$ and $u^\dagger \mathbf{R}_0 = 0$. To satisfy the Born identity, then, our states $\mathbf{S}$ must satisfy

$$\mathbf{S}_0 w = 0 \qquad \alpha \mathbf{S}_0 \mathbf{R}_0 = I_{r-1}. \tag{4.19}$$

The measurement matrix itself is

$$\mathbf{R} = \left(\begin{array}{c|c} w & \mathbf{R}_0 \end{array}\right). \tag{4.20}$$

By informational completeness, it must have (not uniquely) a left inverse $\mathbf{R}^L$, which we may express in block form as

$$\mathbf{R}^L = \begin{pmatrix} \ell \\ \mathbf{R}_0^L \end{pmatrix}. \tag{4.21}$$

Indeed,

$$\mathbf{R}^L\mathbf{R} = \begin{pmatrix} \ell \\ \mathbf{R}_0^L \end{pmatrix}\begin{pmatrix} w & \mathbf{R}_0 \end{pmatrix} = \begin{pmatrix} \ell w & \ell \mathbf{R}_0 \\ \mathbf{R}_0^L w & \mathbf{R}_0^L \mathbf{R}_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & I_{r-1} \end{pmatrix}, \tag{4.22}$$

so that $\mathbf{R}_0^L w = 0$ and $\mathbf{R}_0^L \mathbf{R}_0 = I_{r-1}$. We conclude that if we take $\mathbf{S}_0 \propto \mathbf{R}_0^L$, we can satisfy the two constraints in Eq. 4.19. The key is to choose the constant α so that $|S_i)$, the columns of $\mathbf{S}$, are valid states. To this end, let $\mathbf{S}_0 = \alpha^{-1}\mathbf{R}_0^L$ where α has been chosen so that

$$\forall i, E \in \mathcal{E} : (E|S_i) = \left(\begin{array}{c|c} E_1 & E_0 \end{array}\right)\left(\begin{array}{c} 1 \\ \hline \frac{1}{\alpha}(R_0^L)_i \end{array}\right) = E_1 + \frac{1}{\alpha} E_0 \cdot (R_0^L)_i \geq 0, \tag{4.23}$$

where $(R_0^L)_i$ denotes the i'th column of $\mathbf{R}_0^L$. Now $E_1 = P(E|\mu) > 0$ since we chose $|\mu)$ to be an interior state. Thus we can ensure $\forall i, E \in \mathcal{E} : (E|S_i) \geq 0$ by choosing $|\alpha| \geq 1$ sufficiently large. That such an α can always be found follows from the same considerations as in Theorem 3.7.2: the set of effects is assumed to be compact, and in finite dimensions, linear functionals are all continuous, so that $f(E) = (E|S_i)$ sends the effect space to a compact subset of $\mathbb{R}$ on which the function attains its minimum and maximum. Moreover, because we work in a no-restriction GPT, $\mathcal{S} = \{\rho \mid \forall E \in \mathcal{E} : (E|\rho) \geq 0\}$.

In short, we select some left inverse $\mathbf{R}^L$ of the measurement matrix $\mathbf{R}$. The latter $r-1$ rows of $\mathbf{R}^L$ form a left inverse $\mathbf{R}_0^L$ of $\mathbf{R}_0$ which additionally satisfies $\mathbf{R}_0^L w = 0$. We then take $\mathbf{S}_0 = \frac{1}{\alpha}\mathbf{R}_0^L$ where $|\alpha|$ is chosen sufficiently large that $|S_i)$ are valid states. Thus $\mathbf{S}_0 w = 0$ and $\alpha \mathbf{S}_0 \mathbf{R}_0 = I_{r-1}$, as desired, and since $\mathbf{S}\tilde{\mathbf{R}} = I$, $\mathbf{S}$ has a right inverse, and so r linearly independent rows: the states $|S_i)$ are informationally complete. Finally, since $w = P(R|\mu)$, we have $\tilde{\mathbf{R}}|\rho) = \alpha P(R|\rho) + (1-\alpha)P(R|\mu)$ as desired. □

Remark. *The condition $\boldsymbol{S}_0 w = 0$ amounts to $\boldsymbol{S}w = |1)$: the weighted average of the reference states is the state of complete uncertainty. In a self-dual theory, and for an unbiased measurement, this implies that the reference states themselves correspond to a measurement.*

Another perspective on this construction is provided by considering the channel operator corresponding to reference measurement $C = \sum_{i=1}^n |S_i)(R_i|$. Notice that $(E|C|\rho) = P(E|R)P(R|\rho)$. We can understand this in reflection terms: the gambler is now contemplating their future where they expect first to perform the reference measurement followed by an arbitrary final measurement $\{E_i\}$. What state should they use now to assign probabilities to the final measurement: precisely, $C|\rho) = \sum_i P(R_i|\rho)|S_i)$.

Corollary 4.3.1.1. *A reference measurement furnishes a Protourgleichung with quasiprobabilities $W(R|\rho) = \alpha P(R|\rho) + (1-\alpha)P(R|\mu)$ iff its channel operator is depolarizing with parameter $1/\alpha$.*

Proof. Working out the form of the channel operator under the assumptions of Theorem (4.3.1) yields

$$\mathbf{SR} = \left(\begin{array}{c|c} 1 \quad \dots \quad 1 & \\ \hline \mathbf{S}_0 & \end{array}\right)\left(\begin{array}{c|c} w & \mathbf{R}_0 \end{array}\right) \tag{4.24}$$

$$= \begin{pmatrix} u^\dagger w & u^\dagger \mathbf{R}_0 \\ \mathbf{S}_0 w & \mathbf{S}_0 \mathbf{R}_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\alpha} I_{r-1} \end{pmatrix} \tag{4.25}$$

$$= \frac{1}{\alpha} I + \left(1 - \frac{1}{\alpha}\right) |1)(\mu|, \tag{4.26}$$

implying the reference measurement acts as a depolarizing channel with parameter $1/\alpha$,

$$|\rho) \mapsto \frac{1}{\alpha}|\rho) + \left(1 - \frac{1}{\alpha}\right)|\mu). \tag{4.27}$$

The argument can be read in reverse so that the assumption that the channel is depolarizing and informationally complete implies a quasiprobability representation of the desired form. □

Remark. *It may be useful to see step by step "why" the Protourgleichung works. We begin by inserting the Born identity into the expression for $P(E|\rho)$,*

$$P(E|\rho) = \boldsymbol{E}|\rho) = \boldsymbol{E}\left[\boldsymbol{S}\Phi\boldsymbol{R}\right]|\rho). \tag{4.28}$$

If Φ acts as a Protourgleichung, we have

$$= \boldsymbol{E}\boldsymbol{S}\Big[\alpha\boldsymbol{R}|\rho) + (1-\alpha)P(R|\mu)\Big] = \boldsymbol{E}\Big[\alpha\boldsymbol{S}\boldsymbol{R}|\rho) + (1-\alpha)\boldsymbol{S}P(R|\mu)\Big]. \tag{4.29}$$

But we know that $\boldsymbol{SR}$ is just the channel operator C, which is depolarizing with parameter $1/\alpha$, and $\boldsymbol{S}P(R|\mu) = |\mu)$. The effect, then, is simply to balance depolarization with repolarization,

$$= \boldsymbol{E}\left[\alpha\left[\frac{1}{\alpha}|\rho) + \left(1 - \frac{1}{\alpha}\right)|\mu)\right] + (1-\alpha)|\mu)\right] \tag{4.30}$$

$$= \boldsymbol{E}\Big[|\rho) - (1-\alpha)|\mu) + (1-\alpha)|\mu)\Big] \tag{4.31}$$

$$= \boldsymbol{E}|\rho). \tag{4.32}$$

Remark. *In the case that $\alpha = 1$, $W(R|\rho) = P(R|\rho)$: the Protourgleichung collapses back to the law of total probability. Since $\tilde{\boldsymbol{R}} = \boldsymbol{R}$, the channel operator is simply the identity $C = \boldsymbol{SR} = I$, which implies that $P^2 = \boldsymbol{R}\big[\boldsymbol{SR}\big]\boldsymbol{S} = \boldsymbol{RS} = P$: the conditional probability matrix which characterizes the reference measurement is a projector. Moreover, notice that*

$$\forall E, \rho : (E|\rho) = \Big[(E|\boldsymbol{S}\Big]\Big[\boldsymbol{R}|\rho)\Big] = P(E|R)P(R|\rho). \tag{4.33}$$

Thus $\boldsymbol{S}$ provides a linear map which embeds the effect space into the hypercube dual to the probability simplex, and $\boldsymbol{R}$ provides a linear map which embeds the state space into the probability simplex itself in such a way that all probabilities $P(E|\rho)$ are preserved. We have therefore constructed a simplex embedding*, or equivalently a noncontextual ontological model [79, 82]: in short, if we can take $\alpha = 1$, then our GPT must be equivalent to a classical theory. In the sequel, we shall discuss the possibility of classical representations more thoroughly.*

For a depolarizing reference measurement constructed according to the above prescription, the conditional probability matrix $P(R|R)$ which characterizes the reference

measurement takes the general form,

$$P(R|R) = \mathbf{RS} = \left(\begin{array}{c|c} w & \mathbf{R}_0 \end{array}\right)\left(\begin{array}{c} 1 \quad \ldots \quad 1 \\ \hline \mathbf{S}_0 \end{array}\right) \tag{4.34}$$

$$= wu^\dagger + \mathbf{R}_0\mathbf{S}_0. \tag{4.35}$$

We then have the following theorem:

Theorem 4.3.2. *The conditional probability matrix $P \equiv P(R|R)$ which characterizes a depolarizing reference measurement acts on vectors $y \in col(P)$ in its column space as*

$$Py = \frac{1}{\alpha}y + \left(1 - \frac{1}{\alpha}\right)\overline{y}w, \tag{4.36}$$

where $w = P(R|\mu)$ and $\overline{y}$ denotes the sum of the components of y. If the reference measurement is a MIC, then in fact

$$P(R|R) = \frac{1}{\alpha}I + \left(1 - \frac{1}{\alpha}\right)wu^\dagger, \tag{4.37}$$

and if moreover the reference measurement is unbiased, then $P(R|R) = aI + bJ$, where J is the matrix of all 1's with $a = 1/\alpha$ and $b = (1 - 1/\alpha)/r$, that is, $P(R|R)$ is compound symmetric.

Proof. From $P = wu^\dagger + \mathbf{R}_0\mathbf{S}_0$ and using the fact that $\mathbf{S}_0 w = 0$, we see that $Pw = w$. Moreover, since P is column stochastic, $u^\dagger P = u^\dagger$. By construction $P\Phi P = P$ where $\Phi = \alpha I + (1 - \alpha)wu^\dagger$, or $\alpha P^2 + (1 - \alpha)wu^\dagger = P$. Now suppose $y = Px \in \mathrm{col}(P)$. Noticing that $wu^\dagger Px = wu^\dagger x$, we have for $y \in \mathrm{col}(P)$,

$$\alpha Py + (1 - \alpha)wu^\dagger y = y, \tag{4.38}$$

or

$$Py = \frac{1}{\alpha}y + \left(1 - \frac{1}{\alpha}\right)\overline{y}w. \tag{4.39}$$

Now if the reference measurement is a MIC then $P(R|R)$ is an $r \times r$ invertible matrix and $\mathrm{col}(P) = V$, the entire vector space so that

$$P(R|R) = \frac{1}{\alpha}I + \left(1 - \frac{1}{\alpha}\right)wu^\dagger. \tag{4.40}$$

If the measurement is unbiased, then $w = u/n$, and $P(R|R) = aI + bJ$ where $a = 1/\alpha$ and $b = (1-a)/r$. □

Remark. *Supposing $y \in col(P)$ is a probability vector $P(R|\rho)$, we have*

$$P(R|R)P(R|\rho) = \frac{1}{\alpha}P(R|\rho) + \left(1 - \frac{1}{\alpha}\right)P(R|\mu). \tag{4.41}$$

We could have proven this equally well starting from the structure of the channel operator.

Lemma 4.3.3. *Let $\{R_i\}$ be a depolarizing reference measurement such that $\Phi = \alpha I + (1-\alpha)wu^\dagger$. Then $P(R|R)$, the conditional probability matrix which characterizes it must have 1 eigenvalue equal to 1, $r-1$ eigenvalues equal to $1/\alpha$, and the rest 0.*

Proof. Recall $P(R|R) = wu^\dagger + \mathbf{R}_0\mathbf{S}_0$, and that: $u^\dagger w = 1$, $u^\dagger\mathbf{R}_0 = 0$, $\mathbf{S}_0 w = 0$, and $\mathbf{S}_0\mathbf{R}_0 = \alpha^{-1}I_{r-1}$. We first note that

$$P(R|R)w = wu^\dagger w + \mathbf{R}_0\mathbf{S}_0 w = w, \tag{4.42}$$

so that w is an eigenvector with eigenvalue 1. Meanwhile, for any vector x,

$$P(R|R)\mathbf{R}_0 x = wu^\dagger\mathbf{R}_0 x + \mathbf{R}_0\mathbf{S}_0\mathbf{R}_0 x = \alpha^{-1}\mathbf{R}_0 x, \tag{4.43}$$

so that every vector in $\mathrm{col}(\mathbf{R}_0)$ is an eigenvector with eigenvalue α^{-1}. Since $u^\dagger w = 1$, but $u^\dagger\mathbf{R}_0 = 0$, we conclude that $w \notin \mathrm{col}(\mathbf{R}_0)$. We have so far accounted for r linearly independent eigenvectors. Finally,

$$\mathrm{rank}(P(R|R)) \leq \mathrm{rank}(wu^\dagger) + \mathrm{rank}(\mathbf{R}_0\mathbf{S}_0) \leq 1 + (r-1) = r, \tag{4.44}$$

so that the remaining eigenvalues are zero. □

4.4 Morphophoricity

The key to the construction of a depolarizing reference measurement is the demand that

$$\alpha\mathbf{S}_0\mathbf{R}_0 = I_{r-1}, \tag{4.45}$$

so that $\mathbf{S}_0 = \frac{1}{\alpha}\mathbf{R}_0^L$ where α is chosen so that the resulting states are nonnegative on all effects in the GPT. Suppose however that we further demand that the reference states take the particular

form

$$\mathbf{S} = \left(\begin{array}{ccc} 1 & \dots & 1 \\ \hline & c\mathbf{R}_0^\dagger W^{-1} & \end{array} \right), \tag{4.46}$$

where W^{-1} is the diagonal matrix with the reciprocals of the weights $1/w_i$ along the diagonal. If the constant c were equal to 1, this would amount to taking $|S_i) = (1/w_i)|E_i)$, that is, to taking the reference states proportional to effects. This is always possible in a self-dual GPT: otherwise, employing the same trick as before, one may choose the constant c to sufficiently depolarize the prospective states until they are nonnegative on all effects. Either way, we'll call such a reference measurement a *parallel update* reference measurement. By the same reasoning as in Corollary (4.3.1.1), if such a measurement is to furnish a depolarizing channel, we must have

$$\alpha c \mathbf{R}_0^\dagger W^{-1} \mathbf{R}_0 = I_{r-1}, \tag{4.47}$$

or $\mathbf{S}_0 = c\mathbf{R}_0^\dagger W^{-1} = \frac{1}{\alpha}\mathbf{R}_0^L$, and since c is fixed by the demand that $c\mathbf{R}_0^\dagger W^{-1}$ furnish valid states, this fixes the value of α. Finally, yet another way of saying it is that we require that the columns of $\mathbf{R}_0$ be orthonormal with respect to the metric provided by $\mathcal{M} = \alpha c W^{-1}$.

If this is possible, we have the following interesting consequence. Recalling that $P(R|\rho) = \mathbf{R}|\rho) = w + \mathbf{R}_0\rho_0$, we have

$$\begin{aligned} \big\|P(R|\rho) - P(R|\sigma)\big\|_{\mathcal{M}}^2 &= \big\|w + \mathbf{R}_0\rho_0 - w - \mathbf{R}_0\sigma_0\big\|_{\mathcal{M}}^2 && (4.48)\\ &= \big\|\mathbf{R}_0(\rho_0 - \sigma_0)\big\|_{\mathcal{M}}^2 && (4.49)\\ &= (\rho_0 - \sigma_0)^\dagger \mathbf{R}_0^\dagger \mathcal{M} \mathbf{R}_0 (\rho_0 - \sigma_0) && (4.50)\\ &= (\rho_0 - \sigma_0)^\dagger \alpha c \mathbf{R}_0^\dagger W^{-1} \mathbf{R}_0 (\rho_0 - \sigma_0) && (4.51)\\ &= \big\|\rho_0 - \sigma_0\big\|^2. && (4.52) \end{aligned}$$

In other words, up to the diagonal metric $\mathcal{M}$, the measurement map from the traceless subspace into the probability simplex preserves Euclidean distances. We'll call such a measurement *weighted morphophoric*. In particular, for an unbiased measurement, we have $\mathbf{R}_0^\dagger \mathbf{R}_0 = \frac{1}{\alpha c n} I$ so that the traceless part of the effects form a tight frame for the traceless subspace (the subspace orthogonal to $|\mu)$), and

$$\big\|P(R)_\rho - P(R)_\sigma\big\|^2 = \frac{1}{\alpha c n}\big\|\rho_0 - \sigma_0\big\|^2, \tag{4.53}$$

that is, the measurement map is a *similarity* with respect to the Euclidean metric on both the probability simplex and the traceless subspace. Such a measurement is properly *morphophoric*.

Remark. *"Morphophoric" in Greek means* form-bearing*: indeed, a morphophoric measurement embeds a state space into the probability simplex in a shape preserving way. The terminology was introduced in the quantum mechanical setting in [6]. A rigorous mathematical treatment of the subject in the more general setting of "geometric GPTs" equipped with inner products can be found in [76]. Morphophoric measurements in general yield Urgleichung-like expressions which are somewhat more sophisticated; we confine our interest here to the very simplest deformations of the law of total probability which correspond to quasiprobability representations of the form* $W(R|\rho) = \alpha P(R|\rho) + (1-\alpha)P(R|\mu)$*. This motivates our introduction of the notion of weighted morphophoricity, that is, the demand that* $\boldsymbol{R}_0^\dagger W^{-1}\boldsymbol{R}_0 \propto I$ *in order to handle biased measurements on equal footing with unbiased.*

Lemma 4.4.1. *Weighted morphophoric measurements exist in any no-restriction GPT of dimension* r *with a distinguished interior state* $|\mu)$*, for any number of outcomes* $n \geq r$ *and arbitrary choice of biases* $w_i > 0$ *and* $\sum_i w_i = 1$*.*

Proof. We seek $\mathbf{R}, \mathbf{S}$ of the form

$$\mathbf{R} = \left(\begin{array}{c|c} w & \mathbf{R}_0 \end{array}\right) \qquad \mathbf{S} = \left(\begin{array}{c} 1 \quad \dots \quad 1 \\ \hline c\mathbf{R}_0^\dagger W^{-1} \end{array}\right). \tag{4.54}$$

Pick $r-1$ orthonormal vectors in the subspace orthogonal to $W^{1/2}u$, and use them to build an $n \times (r-1)$ matrix $\mathbf{K}_0$ such that $\mathbf{K}_0^\dagger W^{1/2}u = 0$. Then let $\mathbf{R}_0 = (\alpha c)^{-1/2} W^{1/2}\mathbf{K}_0$. Clearly, $u^\dagger \mathbf{R}_0 = 0$ and moreover, $\alpha c \mathbf{R}_0^\dagger W^{-1}\mathbf{R}_0 = \mathbf{K}_0^\dagger \mathbf{K}_0 = I_{r-1}$. As in Theorem 4.3.1, by taking $|\alpha|$ sufficiently large with $c = 1$, the traceless part of $(R_i|$ can be made arbitrarily small and so since $w_i > 0$, the $(R_i|$'s can be made positive on the compact state space. This is enough to ensure they are valid effects since we work in a no-restriction GPT: $\mathcal{E} = \{E \mid \forall \rho \in \mathcal{S} : (E|\rho) \in [0,1]\}$. Similarly, since $|\mu)$ is interior, vectors sufficiently close to $|\mu)$ will be valid states: thus we can take $|c|$ sufficiently large that $|S_i)$ are positive on the compact effect space. Since $\mathbf{R}_0$ has rank $r-1$, the effects (and states) span the full r-dimensional space, giving informational completeness. □

Suppose we have a weighted morphophoric measurement. We have shown that if we consider the (properly depolarized) states proportional to the reference effects, they furnish a depolarizing channel. But we are not limited to that particular choice of reference states: as we have also seen, any $\mathbf{S}$ built from a left inverse of $\mathbf{R}_0$ will suffice. There is a special case, however.

Theorem 4.4.2. *For a weighted morphophoric MIC reference measurement, the choice of reference states which furnish a depolarizing channel is fixed up to sign.*

Proof. From Theorem (4.3.1), given a choice of reference effects, we require of the reference states that they satisfy $\beta \mathbf{S}_0 \mathbf{R}_0 = I_{r-1}$ in order for the resulting channel to be depolarizing. At the same time, morphophoricity demands Eq. (4.47), that is, $\alpha c \mathbf{R}_0^\dagger W^{-1} \mathbf{R}_0 = I_{r-1}$. Hence

$$(\alpha c \mathbf{R}_0^\dagger W^{-1} - \beta \mathbf{S}_0)\mathbf{R}_0 = M\mathbf{R}_0 = 0. \tag{4.55}$$

For a MIC, $\mathbf{R}_0^\dagger \in \mathbb{C}_{r-1}^{(r-1)\times r}$: by the rank nullity theorem, therefore, $\dim(\text{null}(\mathbf{R}_0^\dagger)) = 1$. Because the reference effects sum to (1|, we have $\mathbf{R}_0^\dagger u = 0$, and so $\text{null}(\mathbf{R}_0^\dagger) = \text{span}(u)$. Meanwhile, denoting the ith row of M by $m_i^\dagger$, we have $\mathbf{R}_0^\dagger m_i = 0$: that is, each of the m_i must lie in $\text{null}(\mathbf{R}_0^\dagger) = \text{span}(u)$. Thus every row of M must be proportional to $u^\dagger$ so that we can write $M = xu^\dagger$ for some $x \in \mathbb{F}^{r-1}$. From Theorem (4.3.1), we also have $\mathbf{S}_0 w = 0$ and $\mathbf{R}_0^\dagger W^{-1} w = \mathbf{R}_0^\dagger u = 0$. Thus $Mw = 0$. But also $Mw = xu^\dagger w = 0$. But since the weights must sum to 1, we conclude that $x = 0$, and so $M = 0$, and that $\alpha c \mathbf{R}_0^\dagger W^{-1} = \beta \mathbf{S}_0$. For a weighted morphophoric MIC, the choice of depolarizing reference states is essentially fixed up to c: we assume one takes $|c|$ as small as possible. Note, however, that we can take $\mathbf{S}_0 = -c\mathbf{R}_0^\dagger W^{-1}$ and $\beta = -\alpha$: hence the reference states are fixed up to sign. □

Corollary 4.4.2.1. *Suppose we have an unbiased morphophoric MIC reference measurement. Its depolarizing reference states are fixed up to sign to be proportional to effects (modulo c). By Theorem 4.3.2, the conditional probability matrix which characterizes the reference measurement must be compound symmetric, that is, of the form $P = aI + bJ$. Thus any unbiased morphophoric MIC measurement is a SIC, that is, a* symmetric informationally complete *measurement, forming a regular simplex in the effect space. Note, however, that such measurements need not be extremal, that is, lying on the boundary of the effect space. As we have observed, it is easy to see that nonextremal SICs exist in any GPT: one may simply embed a sufficiently small $(r-1)$-simplex (with r vertices) in the effect space such that the effects are non-negative on all states.*

4.4.1 An example from quantum theory

Consider the following extremal symmetric informationally complete measurement (SIC-POVM) in $d = 2$ quantum theory. The effects can be written in terms of Pauli matrices,

$$R_0 = \frac{1}{4}\left(I + \frac{1}{\sqrt{3}}(\sigma_x + \sigma_y + \sigma_z)\right) \qquad R_2 = \frac{1}{4}\left(I + \frac{1}{\sqrt{3}}(-\sigma_x - \sigma_y + \sigma_z)\right) \tag{4.56}$$

$$R_1 = \frac{1}{4}\left(I + \frac{1}{\sqrt{3}}(\sigma_x - \sigma_y - \sigma_z)\right) \qquad R_3 = \frac{1}{4}\left(I + \frac{1}{\sqrt{3}}(-\sigma_x + \sigma_y - \sigma_z)\right), \tag{4.57}$$

or more compactly in the Bloch representation,

$$\mathbf{R} = \left(\begin{array}{c|c} \frac{1}{4}u & \mathbf{R}_0 \end{array} \right) \qquad \mathbf{R}_0 = \frac{1}{4\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{pmatrix}. \tag{4.58}$$

Since quantum theory is self-dual, we can choose states proportional to effects $\mathbf{S} = \begin{pmatrix} u^\dagger \\ 4\mathbf{R}_0^{\mp\dagger} \end{pmatrix}$, in which case $\mathbf{R}_0^\dagger W^{-1} \mathbf{R}_0 = 4\mathbf{R}_0^\dagger \mathbf{R}_0 = (1/3)I$ so that $\alpha = 3$ (and $c = 1$). Then $\mathbf{SR}$ corresponds to a depolarizing channel with parameter $1/3$,

$$|\rho) \mapsto \frac{1}{3}|\rho) + \frac{2}{3}|\mu), \tag{4.59}$$

and we have the Urgleichung

$$P(E|\rho) = \sum_i P(E|R_i) \left[3P(R_i|\rho) - \frac{1}{2} \right]. \tag{4.60}$$

Alternatively, however, we could choose the states *antipodal* to the aforementioned, which also form a SIC, that is, we could take $\mathbf{S}' = \begin{pmatrix} u^\dagger \\ -4\mathbf{R}_0^{\mp\dagger} \end{pmatrix}$, in which case $\mathbf{S}_0 = -\mathbf{R}_0^\dagger W^{-1} = -4\mathbf{R}_0^\dagger$ and $\alpha = -3$. Then

$$|\rho) \mapsto -\frac{1}{3}|\rho) + \frac{4}{3}|\mu) \tag{4.61}$$

with $P(E|\rho) = \sum_i P(E|R_i)\,[-3P(R_i|\rho) + 1]$. In quantum theory, complete positivity for a depolarizing channel demands that $-1/(d^2 - 1) \leq 1/\alpha \leq 1$ [112]. Indeed, when $d = 2$, the lower bound becomes $-1/3$, which this latter channel saturates. Thus for both choices of reference states, we have a valid depolarizing channel. It is worth remarking, however, that the latter choice also leads to an inversion of the state space. For an arbitrary GPT, it may or may not be the case that the channel with an inverted depolarization parameter is a valid channel.

4.4.2 Tight IC-POVMs

Thanks to the pioneering work of Andrew Scott, it has long been known that in the case of quantum theory, measurements which give rise to tight frames for the traceless subspace (called tight IC-POVMs) correspond to so-called complex projective 2-designs [77]. We will have more to say about t-designs in the sequel: here we confine ourselves to giving an alternative proof using the formalism we have developed in this chapter.

Theorem 4.4.3. *In quantum theory over $\mathbb{C}$, weighted morphophoric reference measurements whose effects are extremal correspond to weighted complex projective 2-designs.*

Proof. For a depolarizing reference measurement, we have

$$C = \mathbf{SR} = \sum_{i=1}^{n} |S_i)(R_i| = \frac{1}{\alpha} I + \left(1 - \frac{1}{\alpha}\right) |\mu)(1| \tag{4.62}$$

Given that we are working with quantum theory over $\mathbb{C}$, we can interpret this expression in terms of vectorized operators, which will be elements of $\mathbb{C}^{d^2}$ [113]. Let $|A) = (A \otimes I) \sum_{i=1}^{d} |i, i\rangle$ be the (row) vectorization of A. It consists of the rows of A laid end to end to form a column vector. Then

$$(A|B) = \left[\sum_{i=1}^{d} \langle i, i|(A^\dagger \otimes I) \right] \left[(B \otimes I) \sum_{j=1}^{d} |j, j\rangle \right] \tag{4.63}$$

$$= \sum_{i,j} \langle i|A^\dagger B|j\rangle \otimes \langle i|j\rangle = \sum_{i} \langle i|A^\dagger B|i\rangle = \mathrm{tr}(A^\dagger B). \tag{4.64}$$

In other words, $(A|B)$ reproduces the Hilbert-Schmidt inner product $\mathrm{tr}(A^\dagger B)$ on operators. Note that in this case $(A|$ really does mean the conjugate transpose of $|A)$.

In this formalism, the vectorized identity operator $(I| \equiv (1|$ is the normalization functional, while $\frac{1}{d}|I) = |\mu)$ corresponds to the maximally mixed state. Thus

$$\sum_{i=1}^{n} |S_i)(R_i| = \frac{1}{\alpha} I_{d^2} + \left(1 - \frac{1}{\alpha}\right) \frac{1}{d} |I_d)(I_d|. \tag{4.65}$$

Taking the trace of the right-hand side of Eq. (4.65), we find

$$\sum_{i=1}^{n} \mathrm{tr}(R_i S_i) = \frac{1}{\alpha}(d^2) + \left(1 - \frac{1}{\alpha}\right) \frac{1}{d}(d) = \frac{1}{\alpha}(d^2 - 1) + 1. \tag{4.66}$$

Quantum theory over $\mathbb{C}$ is a self-dual theory, and so we can always take reference states directly proportional to effects: thus let $R_i = \mathrm{tr}(R_i) S_i$—this along with the depolarizing assumption amounts to weighted morphophoricity. Moreover, let us assume that our states are pure so that $\mathrm{tr}(S_i^2) = 1$. Then taking the trace of the left hand side of Eq. (4.65), we find

$$\sum_{i=1}^{n} \mathrm{tr}(R_i S_i) = \sum_{i=1}^{n} \mathrm{tr}(R_i)\mathrm{tr}(S_i S_i) = \sum_{i=1}^{n} \mathrm{tr}(R_i) = \mathrm{tr}\left(\sum_{i=1}^{n} R_i\right) = \mathrm{tr}(I) = d. \tag{4.67}$$

Equating these two expressions fixes the value of α:

$$\frac{1}{\alpha}(d^2-1)+1=d \Longrightarrow \alpha = d+1. \tag{4.68}$$

Thus our channel operator becomes

$$C=\sum_{i=1}^{n}\mathrm{tr}(R_i)|S_i)(S_i| = \frac{1}{d+1}I_{d^2}+\left(1-\frac{1}{d+1}\right)\frac{1}{d}|I_d)(I_d| \tag{4.69}$$

$$= \frac{1}{d+1}\Big(I_{d^2}+|I_d)(I_d|\Big). \tag{4.70}$$

Normalizing so that the trace of both sides is 1, we find

$$\sum_{i=1}^{n} w_i|S_i)(S_i| = \frac{1}{d(d+1)}\Big(I_{d^2}+|I_d)(I_d|\Big), \tag{4.71}$$

where $w_i = \mathrm{tr}(R_i)/d$ so that $\sum_i w_i = 1$. Considering the right hand side of Eq. (4.71), notice that we can rewrite it as

$$I_{d^2}+|I_d)(I_d| = I_d\otimes I_d + \sum_{jk}|j,j\rangle\langle k,k|. \tag{4.72}$$

Taking the partial transpose of the second tensor factor and multiplying by $\frac{1}{2}$, we find:

$$\frac{1}{2}\Big(I_d\otimes I_d+\sum_{jk}|j,k\rangle\langle k,j|\Big) = \frac{1}{2}\Big(I_d\otimes I_d+\mathrm{SWAP}\Big) = \Pi_{\mathrm{sym}^2}, \tag{4.73}$$

where Π_{sym^2} is the projector onto the permutation symmetric subspace of two tensor factors. Considering the left hand side of Eq. (4.71), we must also take the transpose of the second tensor factor. It is a useful lemma that for a pure state $S=|\psi\rangle\langle\psi|$, we have $|S)(S| = S\otimes S^T$. We conclude

$$\sum_{i=1}^{n} w_i S_i\otimes S_i = \frac{2}{d(d+1)}\Pi_{\mathrm{sym}^2}. \tag{4.74}$$

Meanwhile, Schur-Weyl duality tells us that the integral over tensor powers of pure states is

$$\int |\psi\rangle\langle\psi|^{\otimes t}d\psi = \frac{\Pi_{\mathrm{sym}^t}}{\mathrm{tr}(\Pi_{\mathrm{sym}^t})} = \frac{1}{\binom{t+d-1}{d-1}}\Pi_{\mathrm{sym}^t}, \tag{4.75}$$

which in the case of $t = 2$, delivers us

$$\int |\psi\rangle\langle\psi|^{\otimes 2} d\psi = \frac{2}{d(d+1)} \Pi_{\text{sym}^2} = \sum_{i=1}^{n} w_i S_i \otimes S_i. \tag{4.76}$$

Indeed, this is the defining equation for a weighted complex projective 2-design, showing that such designs coincide precisely with pure weighted morphophoric reference measurements [77] in quantum theory. □

Corollary 4.4.3.1. *For a pure weighted morphophoric reference measurement in quantum theory, we have $\alpha = d + 1$ so that $W(R|\rho) = \alpha P(R|\rho) + (1 - \alpha)P(R|\mu)$ becomes*

$$W(R|\rho) = (d+1)P(R|\rho) - dP(R|\mu). \tag{4.77}$$

For an unbiased measurement:

$$W(R|\rho) = (d+1)P(R|\rho) - \frac{d}{n}u. \tag{4.78}$$

For a MIC, with $n = d^2$:

$$W(R|\rho) = (d+1)P(R|\rho) - \frac{1}{d}u. \tag{4.79}$$

It has long been known that the unbiased complex projective 2-designs with the minimal number of elements correspond to pure SICs [114] with $P(R_i|R_j) = \frac{d\delta_{ij}+1}{d(d+1)}$, if indeed SICs exist in any dimension [115, 116]. We then arrive at the Urgleichung [2],

$$P(E|\rho) = \sum_{j=1}^{d^2} P(E|R_j) \left[(d+1)P(R_j|\rho) - \frac{1}{d} \right]. \tag{4.80}$$

Remark. *In appendix 4.7, we give the analogous result for quantum mechanics over $\mathbb{R}$.*

4.5 The LTP Deformation

As we have seen, the only difference between the law of total probability and our nonclassical coherence condition is the presence of the Born matrix Φ, a $\{1\}$-inverse of $P(R|R)$, the conditional probability matrix which characterizes the reference measurement itself. Inspired by this, [78] introduced the distance between the identity and Φ with respect to any unitarily invariant matrix norm[1] as a measure of the nonclassicality of the reference

[1] For example, a Schatten p-norm $||A||_p = \left(\sum_i \sigma_i^p\right)^{\frac{1}{p}}$ where the σ_i's are the singular values of A [111].

measurement,

$$\mathcal{D}(\Phi) = \|I - \Phi\|. \tag{4.81}$$

We shall call it here the *LTP deformation*. By minimizing this quantity over all allowed reference measurements with a certain number of outcomes, one obtains the *irreducible LTP deformation*, a property of the underlying theory itself (e.g., d dimensional quantum theory over $\mathbb{C}$), and which is meant to be a measure of nonclassicality on grounds that classically one may always pick a reference measurement for which $\Phi = I$, simply reading off an objective property of the system without disturbing it. The conclusion of [78] was among all minimal informationally complete (MIC) measurements (with d^2 outcomes), parallel update SIC reference measurements achieve the irreducible LTP deformation in all dimensions—if they exist.

Now for a MIC, the reference states and effects must form bases, and we have $\Phi = P(R|R)^{-1}$ univocally. But if we want to calculate $\mathcal{D}(\Phi)$ for "overcomplete" reference measurements, we must make a choice of Born matrix. We therefore introduce the *minimal LTP deformation*

$$\mathcal{D}_{\min} = \min_{\Phi} ||I - \Phi|| \quad \text{s.t.} \quad P\Phi P = P, \tag{4.82}$$

which is a constrained convex optimization problem. In particular, the Frobenius norm squared is strictly convex, and since the feasible set is itself convex, the problem in this case has a unique solution [106]. One can further obtain the irreducible LTP deformation by minimizing over both reference measurements with a certain number of outcomes as well as over Φ itself.

But this quantity may be misleading. On the one hand, we have seen that for any depolarizing reference measurement we may take $\Phi_1 = \alpha I + (1 - \alpha)wu^\dagger$. This is arguably the simplest Born matrix in algebraic terms. It gives an LTP deformation of e.g.,

$$\|I - \Phi_1\|_F = |1 - \alpha|\sqrt{n(1 + \|w\|^2) - 2}. \tag{4.83}$$

which in the unbiased case becomes $\|I - \Phi_1\|_F = |1 - \alpha|\sqrt{n - 1}$. But in general this will *not* be the Φ matrix which gives the smallest possible LTP deformation for the given measurement. In other words, in the non-MIC case, there may be a divergence between algebraic simplicity and matrix distance. For example, there are other Φ matrices that act as a Protourgleichung, that is, which act just like $\Phi = \alpha I + (1 - \alpha)wu^\dagger$ on the relevant subspace, but which may have different $\mathcal{D}(\Phi)$ values. For example, putting two forms of the Born identity together

$\mathbf{S}\tilde{\mathbf{R}} = \tilde{\mathbf{S}}\mathbf{R} = I$ yields $\mathbf{S}\tilde{\mathbf{R}}\tilde{\mathbf{S}}\mathbf{R} = I$ so that we may take

$$\Phi_2 = \tilde{\mathbf{R}}\tilde{\mathbf{S}} = \left(\begin{array}{c|c} w & \alpha\mathbf{R}_0 \end{array}\right)\left(\begin{array}{c} 1 \quad \dots \quad 1 \\ \hline \alpha\mathbf{S}_0 \end{array}\right) \tag{4.84}$$

$$= wu^\dagger + \alpha^2\mathbf{R}_0\mathbf{S}_0. \tag{4.85}$$

This Born matrix indeed acts as a Protourgleichung for a depolarizing reference measurement. Since $P(R|\rho) = \mathbf{R}|\rho) = w + \mathbf{R}_0\rho_0$, we have

$$\Phi_2 P(R|\rho) = \left[wu^\dagger + \alpha^2\mathbf{R}_0\mathbf{S}_0\right]\left[w + \mathbf{R}_0\rho_0\right] \tag{4.86}$$

$$= w + \alpha^2\mathbf{R}_0\mathbf{S}_0\mathbf{R}_0\rho_0 \tag{4.87}$$

$$= w + \alpha\mathbf{R}_0\rho_0 \tag{4.88}$$

$$= \alpha P(R|\rho) + (1-\alpha)P(R|\mu), \tag{4.89}$$

where we have used the fact that $\alpha\mathbf{S}_0\mathbf{R}_0 = I$. If we suppose that this Born matrix is unitarily diagonalizable, that is, a normal matrix, then we can easily calculate its LTP deformation. By the same argument as in Lemma 4.3.3, the Born matrix $\Phi = wu^\dagger + \alpha^2\mathbf{R}_0\mathbf{S}_0$ will have 1 eigenvalue equal to 1, $r-1$ eigenvalues equal to α, and the rest 0. Then $\|I - \Phi_2\| = \|I - VDV^\dagger\| = \|V(I-D)V^\dagger\| = \|I - D\|$, where we have used the unitary invariance of the norm, and D is the diagonal matrix with the aforementioned eigenvalues along its diagonal. We conclude

$$\|I - \Phi_2\|_F = \left\|\mathrm{diag}\left(0, \overbrace{1-\alpha, \dots, 1-\alpha}^{r-1}, \overbrace{-1, \dots, -1}^{n-r}\right)\right\|_F \tag{4.90}$$

$$= \sqrt{(r-1)(1-\alpha)^2 + (n-r)}. \tag{4.91}$$

But suppose we add $\Pi_\perp$, the projector onto the complement of $\mathrm{col}(P)$, so that $\Phi_3 = wu^\dagger + \alpha^2\mathbf{R}_0\mathbf{S}_0 + \Pi_\perp$. This will make no difference in the action of Φ_3 since $\Pi_\perp$ gives 0 on any vector in $\mathrm{col}(P)$. But crucially, now

$$\|I - \Phi_3\|_F = \left\|\mathrm{diag}\left(0, \overbrace{1-\alpha, \dots, 1-\alpha}^{r-1}, \overbrace{0, \dots, 0}^{n-r}\right)\right\|_F \tag{4.92}$$

$$= |1-\alpha|\sqrt{r-1}. \tag{4.93}$$

In fact, we can get a more explicit form for this Born matrix. Using $P = wu^\dagger + \mathbf{R}_0\mathbf{S}_0$ and $\Phi_1 = \alpha I + (1-\alpha)wu^\dagger$, we have for the projector onto col(P), $\Pi = P\Phi = wu^\dagger + \alpha\mathbf{R}_0\mathbf{S}_0$. Then

$$\Phi_3 = wu^\dagger + \alpha^2\mathbf{R}_0\mathbf{S}_0 + \Pi_\perp = I + \alpha(\alpha-1)\mathbf{R}_0\mathbf{S}_0. \tag{4.94}$$

The point, of course, is that all three choices of Φ matrices behave the same on the subspace col$(P(R|R))$, even if they differ elsewhere. The Φ_1 matrix we considered is arguably the simplest: it depends only on the fact that the reference measurement is depolarizing without being sensitive to the details of the measurement itself. Φ_2 and Φ_3 act exactly the same as the first on the relevant subspace: the second is confined entirely to that subspace, while the third is able to achieve the smallest LTP deformation of the three by adding support judiciously onto the complement.

Now since $P(R|R)\Phi$ is a projector onto col$(P(R|R))$, we could define an alternative measure,

$$\mathcal{D}_{\text{sub}}(\Phi) = \|(I-\Phi)\Pi\|, \tag{4.95}$$

where Π is the orthogonal projector onto col(P). Any Φ matrices which act identically on the subspace to which probability vectors are in any case constrained would be given identical scores by this measure: thus this measure would judge the three Φ matrices we considered above to have the same LTP deformation. But then again, is this what we really want? If we can make a Φ matrix simpler in an algebraic sense by adding support on the complement of col(P), why shouldn't we? We conclude that measure one uses depends on the goal: for non-MIC reference measurements, there is no one univocal measure of LTP deformation.

We further dramatize this issue by proving the following theorem. Using the method of Lagrange multipliers, we can solve for the Φ matrix which achieves $\min_\Phi ||I-\Phi||$ s.t. $P\Phi P = P$ with respect to the Schatten 2-norm (the Frobenius norm) analytically, at least for unbiased reference measurements.

Theorem 4.5.1. *The unique Born matrix which minimizes $||I-\Phi||_F$ for an unbiased depolarizing reference measurement is*

$$\Phi = I + \left(1 - \frac{1}{\alpha}\right) \boldsymbol{S}_0^\dagger(\boldsymbol{S}_0\boldsymbol{S}_0^\dagger)^{-1}(\boldsymbol{R}_0^\dagger\boldsymbol{R}_0)^{-1}\boldsymbol{R}_0^\dagger. \tag{4.96}$$

Proof. Define a Lagrangian

$$\mathcal{L}(\Phi,\Lambda) = ||I-\Phi||_2^F + \text{tr}\Big(\Lambda^T(P\Phi P - P)\Big) \tag{4.97}$$

$$= n - 2\text{tr}(\Phi) + ||\Phi||^2 + \text{tr}(\Lambda^T P\Phi P) - \text{tr}(\Lambda^T P) \tag{4.98}$$

where Λ is a matrix of Lagrange multipliers. Recalling that

$$\frac{\partial \mathrm{tr}(X)}{\partial X} = I \qquad \frac{\partial ||X||^2}{\partial X} = 2X \qquad \frac{\partial \mathrm{tr}(AX)}{\partial X} = A^T, \tag{4.99}$$

we differentiate with respect to Φ and demand that the result vanishes.

$$\frac{\partial \mathcal{L}(\Phi, \Lambda)}{\partial \Phi} = -2I + 2\Phi + (P\Lambda^T P)^T = 0 \tag{4.100}$$

yields

$$\Phi = I - \frac{1}{2} P^T \Lambda P^T, \tag{4.101}$$

where, considering that $P\Phi P = P$, Λ must satisfy

$$P^2 - P = \frac{1}{2} P P^T \Lambda P^T P. \tag{4.102}$$

Considering the LHS of Eq. (4.102), and recalling that $\mathbf{S}_0\mathbf{R}_0 = \mathbf{R}_0^\dagger \mathbf{S}_0^\dagger = \frac{1}{\alpha} I$ and $u^\dagger \mathbf{R}_0^\dagger = \mathbf{S}_0 w = 0$, we find

$$P^2 = \left[wu^\dagger + \mathbf{R}_0\mathbf{S}_0\right]\left[wu^\dagger + \mathbf{R}_0\mathbf{S}_0\right] = wu^\dagger + \frac{1}{\alpha}\mathbf{R}_0\mathbf{S}_0, \tag{4.103}$$

so that

$$P^2 - P = wu^\dagger + \frac{1}{\alpha}\mathbf{R}_0\mathbf{S}_0 - wu^\dagger - \mathbf{R}_0\mathbf{S}_0 = \frac{1-\alpha}{\alpha}\mathbf{R}_0\mathbf{S}_0. \tag{4.104}$$

As for the RHS of Eq. (4.102), we first note that

$$PP^T = \left[wu^\dagger + \mathbf{R}_0\mathbf{S}_0\right]\left[uw^\dagger + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\right] \tag{4.105}$$

$$= wu^\dagger uw^\dagger + wu^\dagger \mathbf{S}_0^\dagger \mathbf{R}_0^\dagger + \mathbf{R}_0\mathbf{S}_0 uw^\dagger + \mathbf{R}_0\mathbf{S}_0\mathbf{S}_0^\dagger\mathbf{R}_0^\dagger \tag{4.106}$$

$$P^T P = \left[uw^\dagger + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\right]\left[wu^\dagger + \mathbf{R}_0\mathbf{S}_0\right] \tag{4.107}$$

$$= uw^\dagger wu^\dagger + uw^\dagger \mathbf{R}_0\mathbf{S}_0 + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger wu^\dagger + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\mathbf{R}_0\mathbf{S}_0. \tag{4.108}$$

Let us assume that $\Lambda = \mathbf{S}_0^\dagger \Lambda' \mathbf{R}_0^\dagger$. We will justify this after the fact, since we will show that it leads to the unique solution to the optimization problem. With this ansatz, the terms in PP^T

which end in $w^\dagger$ and the terms in $P^T P$ which begin with u all vanish, leaving

$$= \frac{1}{2}\left[wu^\dagger \mathbf{S}_0^\dagger \mathbf{R}_0^\dagger + \mathbf{R}_0\mathbf{S}_0\mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\right]\mathbf{S}_0^\dagger\Lambda'\mathbf{R}_0^\dagger\left[\mathbf{S}_0^\dagger\mathbf{R}_0^\dagger wu^\dagger + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\mathbf{R}_0\mathbf{S}_0\right] \tag{4.109}$$

$$= \frac{1}{2\alpha^2}\left[wu^\dagger\mathbf{S}_0^\dagger + \mathbf{R}_0\mathbf{S}_0\mathbf{S}_0^\dagger\right]\Lambda'\left[\mathbf{R}_0^\dagger wu^\dagger + \mathbf{R}_0^\dagger\mathbf{R}_0\mathbf{S}_0\right]. \tag{4.110}$$

If we further assume that the measurement is unbiased, so that $w = \frac{1}{n}u$, then this becomes simply $\frac{1}{2\alpha^2}\left[\mathbf{R}_0\mathbf{S}_0\mathbf{S}_0^\dagger\right]\Lambda'\left[\mathbf{R}_0^\dagger\mathbf{R}_0\mathbf{S}_0\right]$. Recall that informational completeness implies that $\mathbf{S}_0\mathbf{S}_0^\dagger$ and $\mathbf{R}_0^\dagger\mathbf{R}_0$ are invertible. So let $\Lambda' = \beta(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}$. As for the constant β, we require

$$\frac{1}{2\alpha^2}\left[\mathbf{R}_0\mathbf{S}_0\mathbf{S}_0^\dagger\right]\beta(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\left[\mathbf{R}_0^\dagger\mathbf{R}_0\mathbf{S}_0\right] = \frac{\beta}{2\alpha^2}\mathbf{R}_0\mathbf{S}_0 \stackrel{!}{=} \frac{1-\alpha}{\alpha}\mathbf{R}_0\mathbf{S}_0. \tag{4.111}$$

Thus $\beta = 2\alpha(1-\alpha)$, and

$$\Lambda = 2\alpha(1-\alpha)\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger. \tag{4.112}$$

Putting it all together, the unique Born matrix which minimizes the LTP deformation with respect to the Frobenius norm for an unbiased depolarizing reference measurement is

$$\Phi = I - \frac{1}{2}P^T\Lambda P^T \tag{4.113}$$

$$= I - \frac{1}{2}\left[\frac{1}{n}uu^\dagger + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\right]\left[2\alpha(1-\alpha)\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger\right]\left[\frac{1}{n}uu^\dagger + \mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\right] \tag{4.114}$$

$$= I + \alpha(\alpha-1)\mathbf{S}_0^\dagger\mathbf{R}_0^\dagger\left[\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger\right]\mathbf{S}_0^\dagger\mathbf{R}_0^\dagger \tag{4.115}$$

$$= I + \frac{\alpha-1}{\alpha}\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger. \tag{4.116}$$

Finally, we show this this gives us the unique solution. The Frobenius norm is strictly convex: letting $f(\Phi) = \|I-\Phi\|^2$, for $\Phi_1 \neq \Phi_2$ and $0 < t < 1$, we have $f(t\Phi_1 + (1-t)\Phi_2) < tf(\Phi_1) + (1-t)f(\Phi_2)$. Moreover, the feasible set is $\{\Phi : P\Phi P = P\}$. This is affine since it is the set of solutions of a linear equation: indeed, if Φ_1, Φ_2 are in the set, one may easily check that so is $t\Phi_1 + (1-t)\Phi_2$. Suppose that Φ_1 and Φ_2 were distinct minimizers. But then

$$f\left(\frac{1}{2}(\Phi_1+\Phi_2)\right) < \frac{1}{2}f(\Phi_1) + \frac{1}{2}f(\Phi_2), \tag{4.117}$$

which contradicts minimality, and so $\Phi_1 = \Phi_2$. Thus the solution to the optimization problem is unique. □

But how does this Born matrix act on probability vectors? Recalling that $P(R|\rho) = \mathbf{R}|\rho) = w + \mathbf{R}_0\rho_0 = \frac{1}{n}u + \mathbf{R}_0\rho_0$ in the unbiased case, we have

$$\Phi P(R|\rho) = \left[I + \frac{\alpha-1}{\alpha}\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger\right]\left[\frac{1}{n}u + \mathbf{R}_0\rho_0\right] \tag{4.118}$$

$$= \frac{1}{n}u + \mathbf{R}_0\rho_0 + \frac{\alpha-1}{\alpha}\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger\mathbf{R}_0\rho_0 \tag{4.119}$$

$$= \frac{1}{n}u + \left(\mathbf{R}_0 + \frac{\alpha-1}{\alpha}\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}\right)\rho_0. \tag{4.120}$$

In other words, this Born matrix does *not* act as an Protourgleichung! That said, there is a special case. We can always choose $\mathbf{S}_0 = \frac{1}{\alpha}(\mathbf{R}_0)^+ = \frac{1}{\alpha}(\mathbf{R}_0^\dagger\mathbf{R}_0)^{-1}\mathbf{R}_0^\dagger$—that is, $\mathbf{S}_0$ can be taken proportional to the pseudoinverse of $\mathbf{R}_0$, which is a left inverse since $\mathbf{R}_0$ has linearly independent columns. Symmetrically, we have $\mathbf{R}_0 = \frac{1}{\alpha}\mathbf{S}_0^\dagger(\mathbf{S}_0\mathbf{S}_0^\dagger)^{-1}$, and so

$$\Phi = I + \alpha(\alpha-1)\mathbf{R}_0\mathbf{S}_0. \tag{4.121}$$

and

$$\Phi P(R|\rho) = \frac{1}{n}u + \left(\mathbf{R}_0 + (\alpha-1)\mathbf{R}_0\right)\rho_0 = \frac{1}{n}u + \alpha\mathbf{R}_0\rho_0. \tag{4.122}$$

In this special case, when the reference states derive from the pseudoinverse, the Φ which minimizes $\|I - \Phi\|_F$ acts as a Protourgleichung. In fact, even if the measurement is biased and the states were not obtained from the pseudoinverse, we can nevertheless always use a Born matrix of the form $\Phi = I + \alpha(\alpha-1)\mathbf{R}_0\mathbf{S}_0$. Indeed,

$$P\Phi P = \left[wu^\dagger + \mathbf{R}_0\mathbf{S}_0\right]\left[I + \alpha(\alpha-1)\mathbf{R}_0\mathbf{S}_0\right]\left[wu^\dagger + \mathbf{R}_0\mathbf{S}_0\right] \tag{4.123}$$

$$= wu^\dagger + \frac{1}{\alpha}\mathbf{R}_0\mathbf{S}_0 + \alpha(\alpha-1)\left(\frac{1}{\alpha^2}\right)\mathbf{R}_0\mathbf{S}_0 \tag{4.124}$$

$$= wu^\dagger + \mathbf{R}_0\mathbf{S}_0 = P. \tag{4.125}$$

Indeed, notice that $\Phi = I + \alpha(\alpha-1)\mathbf{R}_0\mathbf{S}_0$ is precisely the Born matrix we derived in Eq. (4.94), by finding a Φ supported entirely on $\mathrm{col}(P)$, and then adding in the projector onto its complement.

4.5.1 A choice of measures

Recall that for a MIC, we have

$$P(R|R) = \frac{1}{\alpha}I + \left(1 - \frac{1}{\alpha}\right)wu^\dagger. \tag{4.126}$$

It is not hard to confirm that the unique Φ matrix is

$$\Phi = \alpha I + (1-\alpha) w u^\dagger, \tag{4.127}$$

which clearly acts as a Protourgleichung. For overcomplete reference measurements, however, we must make a choice of Φ. We've seen many examples,

$$\Phi_1 = \alpha I + (1-\alpha) w u^\dagger, \tag{4.128}$$

which is algebraically simplest,

$$\Phi_2 = w u^\dagger + \alpha^2 \mathbf{R}_0 \mathbf{S}_0, \tag{4.129}$$

which is supported entirely on $\mathrm{col}(P)$,

$$\Phi_3 = I + \alpha(\alpha-1) \mathbf{R}_0 \mathbf{S}_0 \tag{4.130}$$

which achieves smaller LTP deformation by adding to the former the projector onto the complement of $\mathrm{col}(P)$, and which in certain circumstances (unbiasedness, and the reference states derive from the pseudoinverse) coincides with

$$\Phi_{\min} = I + \frac{\alpha-1}{\alpha} \mathbf{S}_0^\dagger (\mathbf{S}_0 \mathbf{S}_0^\dagger)^{-1} (\mathbf{R}_0^\dagger \mathbf{R}_0)^{-1} \mathbf{R}_0^\dagger, \tag{4.131}$$

which we have shown achieves the smallest LTP deformation with respect to the Frobenius norm. The first three act as a Protourgleichung on $\mathrm{col}(P)$, and yet achieve different values for the LTP deformation. The final one, the one which explicitly minimizes $\|I - \Phi\|_F$, need not. These examples show that the two concepts (minimal deformation with respect to a matrix norm, and the form of the Protourgleichung itself) are in fact distinct. We could amend the measure, e.g., to consider the distance not from the identity per se, but from the identity on the relevant subspace, that is, $\mathrm{col}(P)$: then e.g., the first three Φ matrices would be scored the same. But if we can achieve algebraic simplicity by adding support on the complement, why shouldn't we? We conclude simply that care must be taken to choose the measure proper to the circumstance at hand.

Happily, however, the exercise has allowed us to identify a distinguished set of reference states for any unbiased informationally complete reference measurement, namely those given by

$$\mathbf{S} = \left(\begin{array}{ccc} 1 & \dots & 1 \\ \hline & \frac{1}{\alpha} (\mathbf{R}_0^\dagger \mathbf{R}_0)^{-1} \mathbf{R}_0^\dagger & \end{array} \right) \tag{4.132}$$

for which the minimal Φ (with respect to the 2-norm) does in fact act as a Protourgleichung. In fact, if the measurement is weighted morphophoric, then this just gives us a parallel update reference measurement.

4.6 Conclusion

In this chapter, we have established that the Protourgleichung

$$P(E|\rho) = \sum_{j=1}^{n} P(E|R_j)\Big[\alpha P(R_j|\rho) + (1-\alpha)P(R_j|\mu)\Big] \tag{4.133}$$

is ubiquitous. In any no-restriction GPT (which allows all mathematically possible states and effects), and for any reference measurement, one can always find a set of reference states which yield an expression of this form. The key requirement is that the channel associated with the reference measurement must be depolarizing—and this requirement can always be satisfied. Considering the case where the reference states can be chosen proportional to the effects (modulo a depolarization parameter), we related our construction to so-called "morphophoric measurements," and ultimately to complex projective 2-designs, from which we recovered the standard Urgleichung. In light of these results, we reconsidered a matrix norm based measure of LTP deformation, showing by example that it fails to adequately capture the essence of the Protourgleichung, even as it still holds interesting lessons about the nature of reference measurements to teach us.

The Protourgleichung is the simplest nonclassical normative constraint between probability assignments $P(E|\rho)$, $P(E|R)$, and $P(R|\rho)$: it is universally available whether one works in quantum theory, classical probability theory, or any of the more exotic members of the GPT community. Nevertheless, the Urgleichung itself is a child of quantum theory over $\mathbb{C}$, its particular parameter ($\alpha = d + 1$) and the cardinality of its outcomes (d^2) hinting at quantum theory's special structure, and whispering of the possibility of *pure* SIC-POVMs. Indeed, these latter arguably represent the simplest nonclassical generalization of the trivial classical reference measurement, which simply reads off an objective property of a system without disturbing it.

Consider that, in the terms we've developed here, the trivial classical reference measurement has states and effects which are simply the r-dimensional basis vectors. It is a parallel update measurement, with reference states proportional to reference effects. The corresponding channel is simply the identity channel, which is trivially depolarizing, and the Protourgleichung collapses to the LTP. The measurement is trivially morphophoric, simply mapping the probability simplex exactly to itself. It is unbiased. It is a MIC. It is equiangular. Its reference states are extremal, corresponding to the vertices of the probability simplex itself. The reference effects are extremal too. Each reference state can be perfectly

distinguished by some effect[2], and each reference effect can be perfectly distinguished by some state. Moreover, the reference states and reference effects perfectly distinguish each other. Finally, such a measurement exists in any dimension, where the maximal number of pairwise perfectly distinguishable states scales with the dimension.

Compare this to the properties of a SIC reference measurement in quantum theory over $\mathbb{C}$, which is a parallel update measurement, whose associated channel is depolarizing with parameter $\frac{1}{\alpha} = \frac{1}{d+1}$. It is morphophoric, unbiased, a MIC with d^2 elements. It is equiangular. Its reference states and reference effects are extremal, corresponding to the vertices of a simplex. Each reference state can be perfectly distinguished by some effect, and each reference effect can be perfectly distinguished by some state. And finally, such a measurement (apparently) exists in any dimension, where the maximal number of pairwise perfectly distinguishable states scales with the dimension[3].

The only differences, then, are the values of the depolarizing parameter, the fact that the state space is mapped to a subset of the probability simplex, and the fact that the reference states and reference effects don't perfectly distinguish each other. Thus pure SIC reference measurements are arguably the closest analogue to the trivial classical reference measurement, and this is true not just in one theory, but in a whole infinite sequence of them, for each Hilbert space dimension d. From these considerations, if one were trying to generalize classical probability theory as minimally as possible, incorporating the idea that nature forgoes hidden variables[4], one could imagine nothing simpler than positing the existence of pure SIC reference measurements, with their attendant, the Urgleichung—and indeed, this has been the basis of QBist reconstructions of quantum mechanics hitherto.

4.7 Appendix: Real Vector Space QM

A close cousin of standard quantum theory is quantum mechanics over $\mathbb{R}$. Here there is a subtlety if one works with vectorized operators. For quantum mechanics over $\mathbb{C}$, if we take $P(R|R) = \mathbf{RS}$ where the rows of $\mathbf{R}$ are vectorized effects $(R_i|$ and the columns of $\mathbf{S}$ are vectorized states $|R_i)$, then $\mathbf{RS}$ will be a full rank factorization: this is because the real space of Hermitian matrices is d^2, but also the vectorized operators are d^2 dimensional. On the

[2]That is, there is some effect which yields probability 1 on that state.

[3]The point of emphasizing that the maximal number of pairwise perfectly distinguishable states scales with the dimension is to exclude, for example, so-called *spin-factor* theories, whose state spaces are n-dimensional balls, and for which SICs exists trivially. In such theories, regardless of the dimension, the maximal number of pairwise perfectly distinguishable states is always 2. Such a theory is an example of a *Euclidean Jordan Algebra* [24, 22] of which quantum theory over $\mathbb{C}$ is also an example, as well as quantum theory over $\mathbb{R}$, $\mathbb{H}$, and $d = 3$ quantum theory over $\mathbb{O}$. For quantum theory over $\mathbb{R}$, pure SICs are only known to exist in $d = 2, 3, 7, 23$, and are known *not* to exist in most dimensions [117]. For quantum theory over $\mathbb{H}$, pure SICs are only known to exist in $d = 2, 3$, and there is numerical evidence suggesting that they don't exist already in $d = 4$ [118]. There is a pure SIC in octonionic $d = 3$, but this is the only dimension in which such a theory can be constructed. We shall hear more about Euclidean Jordan algebras in subsequent chapters.

[4]Or more positively, as Blake Stacey would put it, respecting nature's *vitality* [24].

other hand, if one works over $\mathbb{R}$, the state space is the space of real symmetric matrices which has dimension $d(d+1)/2$ even as the vectorized operators remain d^2 dimensional. In this latter case, **RS** built from vectorized operators will *not* be a full rank factorization. One could instead work with the states and effects expanded in an operator basis of $d(d+1)/2$ elements; otherwise, one will have to modify the Born identity to $\mathbf{S\Phi R} = \Pi_{\text{sym}^2}$ where Π_{sym^2} is the projector onto symmetric subspace on two factors, i.e. onto the space of vectorized symmetric matrices. This issue is quite general if one doesn't work with full rank factorizations.

Thus for a depolarizing channel in real vector space quantum mechanics, we have (in terms of vectorized operators)

$$\sum_{i=1}^{n} |S_i)(R_i| = \frac{1}{\alpha}\Pi_{\text{sym}^2} + \left(1 - \frac{1}{\alpha}\right)\frac{1}{d}|I_d)(I_d|. \tag{4.134}$$

This theory is also self dual, so taking the trace of both sides, and assuming our states are pure, we find

$$d = \frac{1}{\alpha}\frac{d(d+1)}{2} + 1 - \frac{1}{\alpha} \Longrightarrow \alpha = \frac{d+2}{2}, \tag{4.135}$$

and

$$\sum_{i=1}^{n} |S_i)(R_i| = \frac{2}{d+2}\Pi_{\text{sym}^2} + \frac{1}{d+2}|I_d)(I_d| \tag{4.136}$$

$$= \frac{1}{d+2}\Big(I_{d^2} + \text{SWAP} + |I_d)(I_d|\Big). \tag{4.137}$$

Again dividing by $1/d$ so the trace of both sides is 1, and taking the partial transpose, which actually leaves both sides of the equation invariant, we obtain

$$\sum_{i=1}^{n} w_i S_i \otimes S_i = \frac{1}{d(d+2)}\Big(I_{d^2} + \text{SWAP} + |I_d)(I_d|\Big), \tag{4.138}$$

which according to [119] coincides precisely with $\int |\psi\rangle\langle\psi|^{\otimes 2} d\psi$ where the integral is taken over pure states in the real vector space theory. Thus weighted morphophoric references measurements for quantum theory over $\mathbb{R}$ coincide with real projective 2-designs.

On the subject of real vector space quantum mechanics, [120] conducted a numerical search for MIC reference measurements minimizing $\|I - \Phi\|$ in $d = 4$. It is known that in this dimension one cannot construct a SIC-POVM out of pure effects: the maximum of 6 equiangular lines is not enough to furnish the 10 elements required for a MIC. In fact, in $d = 4$, the minimum number of elements in a pure biased 2-design is 11, and the minimum number of elements in a pure unbiased 2-design is 12 [121]. These efforts were unable to yield

minimal reference measurements whose Born matrix acts as a Protourgleichung, a fact which is now contextualized by the present work. Indeed, such reference measurements *can* be found, although they will not minimize $\|I - \Phi\|$. It is worth noting that subsequent numerical searches suggest that in this case, if the reference effects are pure, then the depolarizing reference states must be mixed. It is also worth noting that unlike in [78], the minimal reference measurement does depend on the choice of matrix norm, which is likely to be true in the general overcomplete case.

CHAPTER 5

(NO) RETURN TO CLASSICALITY

5.1 Introduction

Having supplemented standard probability with a set of coherence conditions rich enough to encompass the brand of nonclassicality exemplified by quantum mechanics, we may then ask: when is a gambler justified in using vanilla probability theory after all? When is the gambler justified in acting *as if* objects have definite properties before they are measured and which are revealed by measurement, and which do not depend on the context in which they are measured? To put it more precisely, when is the gambler justified in supposing that there is a reference measurement $\{\lambda_i\}$ satisfying $\Phi_\lambda = I$ so that regardless of whether the reference measurement is actually performed, the gambler ought to appeal to the law of total probability $P(E|\rho) = P(E|\lambda)P(\lambda|\rho)$ for guidance. To answer this, we assimilate Spekkens's notion of *noncontextual ontological models* to our general philosophy. Following [81], we show how an equirank nonnegative matrix factorization of the probability table characterizing a scenario corresponds to so-called simplex embedding [82], which furnishes a noncontextual ontological model of the scenario. In particular, we show that if the probability table characterizing the scenario captures a reference measurement $\{R_i\}$, then the existence of a noncontextual ontological model can be understood as the existence of a classical reference measurement $\{\lambda_i\}$ for which $\{R_i\}$ itself is a reference. Thus we can understand noncontextual ontological models in terms more suitable to the QBist: indeed, this provides a motivation for the idea of a noncontextual ontological model to begin with. We also show how simplex embeddings implicate a variant of the Born matrix whose nonnegativity implies classicality, and discuss the general conditions under which Φ may be taken to be a stochastic matrix. Finally, we show how violations of Bell inequalities and noncontextuality inequalities more generally can be interpreted as the cost of abandoning the fundamental nonclassical coherence condition.

5.2 Noncontextuality

We have developed the notion of a *state* as an equivalence class of consequences conditional upon any of which the gambler would assign the same probabilities to the reference measurement. In short, for us the concept of a state is fully grounded in a gambler's state of expectation about reference outcomes regardless of the peculiarities of prior circumstances. If the gambler would assign probabilities $P(R|A)$ given circumstance A and probabilities

$P(R|B)$ given circumstance B, and $P(R|A) = P(R|B)$, we would identify the circumstances A and B, saying: they amount to the same state. Calling A or B the *context*, we see that our definition of a state is precisely *noncontextual*. The state is what for the gambler is invariant across contexts. And of course, by coherence, assigning the same state of expectation to the reference measurement conditional on either circumstances means that the gambler ought to assign the same probabilities to *any* measurement—and the same could be said of the effects.

We also showed how we may form a GPT-style representation using a rank decomposition, and we will now see how this preserves noncontextuality. To this end, let us consider an $m \times n$ table of conditional probabilities where the columns correspond to outcomes of preparatory measurements and the rows correspond to outcomes of subsequent measurements,

$$P(E|\pi) = \begin{pmatrix} P(E_1|\pi_1) & \cdots & P(E_1|\pi_n) \\ \vdots & \ddots & \vdots \\ P(E_m|\pi_1) & \cdots & P(E_m|\pi_n) \end{pmatrix}.$$

This matrix $P(E|\pi)$ will have some rank r, and we may form a rank decomposition $P(E|\pi) = \mathbf{EP}$ for $\mathbf{E} \in \mathbb{R}_r^{m\times r}, \mathbf{\Pi} \in \mathbb{R}_r^{r\times n}$. Now suppose two rows of $P(E|\pi)$ are identical: $P(E_i|\pi) = P(E_j|\pi)$. Then in the rank decomposition $(E_i|\mathbf{\Pi} = (E_j|\mathbf{\Pi}$, or $((E_i| - (E_j|)\mathbf{\Pi} = 0$. But $\mathbf{\Pi} \in \mathbb{R}_r^{r\times n}$: it has r linearly independent rows. Thus if a linear combination of them yields the zero vector, each component in the linear combination must be identically 0, and so $(E_i| = (E_j|$. Similarly, if we assign identical probabilities to two outcomes E_i and E_j given any π, then we assign the same row vector $(E_i| = (E_j|$. In this way, the representation of context becomes simply the label i or j which remembers which consequence in the equivalence class actually occurred. More generally, suppose that a row is linearly dependent on a subset of other rows, $P(E_i|\pi) = \sum_{j\in J} \alpha_j P(E_j|\pi)$ for J some subset of rows. Then $(E_i|\mathbf{\Pi} = \sum_{j\in J} \alpha_j (E_j|\mathbf{\Pi}$ where the same argument implies $(E_i| = \sum_{j\in J} \alpha_j (E_j|$. In this way, a rank factorization preserves linear equivalences in probability assignments, guaranteeing the new representation is noncontextual. In particular, if e.g., two rows of $P(E_i|\pi)$ and $P(E_j|\pi)$ are identical, we ought to identify the two outcomes even though they may appear in two different measurements: in a contextual representation, however, the vector representatives $(E_i|$ and $(E_j|$ might depend on which measurement the outcome appears in.

As we will see, the feature that guarantees that $P(E|\pi) = \mathbf{EP}$ is a noncontextual representation is that $\text{rank}(P(E|\pi)) = \text{rank}(\mathbf{E}) = \text{rank}(\mathbf{\Pi})$. Otherwise, the argument from linear independence would not hold: we might have $P(E_i|\pi) = P(E_j|\pi)$, but it would not necessarily follow that $(E_i| = (E_j|$. Similarly, we might have $P(E|\pi_i) = P(E|\pi_j)$, but it would not necessarily follow that $|\pi_i) = |\pi_j)$. We would not want to call $|\pi_i)$ or $|\pi_j)$ themselves states, since according to the gambler's own lights, assigning $|\pi_i)$ or $|\pi_j)$ would lead precisely to the same probabilities for all measurement outcomes. The gambler would be completely indifferent to the choice of representation $|\pi_i)$ or $|\pi_j)$ as far as their mesh of beliefs is concerned: the only difference between them is in the irrelevant context, that it was

in fact outcome π_i as opposed to π_j which occurred. In contrast, a rank decomposition yields one vector representative for equivalence class.

Thus if we forget the labels, in the representation furnished by a rank decomposition, each effect is assigned an $(E|$ and each state a $|\pi)$. In the language of Spekkens [80], the table $P(E|\pi)$ gives an *operational theory*, assigning a probability to each pairing of E_i and π_j. Quotienting by "operational equivalence" yields a set of states and effects which *separate* each other: two separate states must yield different probabilities on some effect, and two separate effects must yield different probabilities on some state. This mutual separation of states and effects is a prerequisite for passing to a GPT, or generalized probabilistic theory. (One must also complete the theory by taking conic combinations of states and effects, and adding in opposite effects $(\neg E| = (1| - (E|$ for each effect, if they are not in the table.) In the language of Spekkens, a GPT is then a noncontextual model of an operational theory[1].

If instead we use an arbitrary decomposition $P(E|\pi) = \mathbf{EP}$, however, even if we "forget the labels," we may end up with vector representatives $(E_i|$ and $|\pi_i)$ that remember their context even though that context makes no "operational difference" in terms of the gambler's own probability assignments. Indeed, even if $P(E|\pi)$ has no repeated rows or columns, if we use an arbitrary decomposition, because of the linear dependence, there will be many different choices of vector representatives that make no difference at all, thus introducing context dependence where there was none before. So which decompositions guarantee a noncontextual representation? We have the following lemma:

Lemma 5.2.1. *Let $M = AB$ and let $\{(A_i|\}$ denote the rows of A and $\{|B_i)\}$ denote the columns of B. Then $\{(A_i|\}$ separates points in span$(\{|B_i\})$ and $\{|B_i)\}$ separates points in span$(\{(A_i|\})$ iff rank(M) = rank(A) = rank(B).*

Proof. We first note that $\text{span}(\{(A_i|\}) = \text{row}(A)$ and $\text{span}(\{|B_i\}) = \text{col}(B)$. Moreover, $\dim \text{row}(A) = \text{rank}(A)$ and $\dim \text{col}(B) = \text{rank}(B)$. Now to say that $\{(A_i|\}$ separates points in $\text{span}(\{|B_i)\}$ is to say that for

$$\forall |x), |y) \in \text{col}(B), A|x) = A|y) \iff |x) = |y). \tag{5.1}$$

Thus A is 1-to-1 on $\text{col}(B)$. 1-to-1 linear maps must preserve the dimensionality of the subspace on which they act. Thus $\dim A\text{col}(B) = \dim \text{col}(B) = \text{rank}(B)$. But since $M = AB$, $\text{col}(M) = A\text{col}(B)$ and so $\text{rank}(M) = \dim A\text{col}(B)$. We conclude that $\text{rank}(B) = \text{rank}(M)$. Similarly, $\text{row}(M) = \text{row}(A)B$ so $\text{rank}(M) = \dim \text{row}(A)B$. To say that $\{|B_i)\}$ separates points in $\text{span}(\{(A_i|\})$ is to say that

$$\forall (x|, (y| \in \text{row}(A) : (x|B = (y|B \iff (x| = (y|. \tag{5.2}$$

[1]For us, of course, we begin from states as distributions $P(R|\pi)$ and effects as distributions $P(E|R)$. This is a noncontextual representation even though the dimension n of the probability vectors might be larger than the rank of $P(R|R)$: as we have seen the coherence conditions $P(R|\pi) = P(R|R)\Phi P(R|\rho)$ and $P(E|R) = P(E|R)\Phi P(R|R)$ ensure that any $P(R|\pi)$ and $P(E|R)$ live in the column and row spaces of $P(R|R)$.

So $\dim \mathrm{row}(A)B = \dim \mathrm{row}(A) = \mathrm{rank}(A)$ and since $\mathrm{row}(M) = \mathrm{row}(A)B$, $\mathrm{rank}(M) = \dim \mathrm{row}(A)B$ and $\mathrm{rank}(A) = \mathrm{rank}(M)$.

Conversely, suppose $\mathrm{rank}(M) = \mathrm{rank}(A) = \mathrm{rank}(B)$. We have $\mathrm{rank}(M) = \dim A\mathrm{col}(B) = \mathrm{rank}(B) = \dim \mathrm{col}(B)$ so that A is 1-to-1 on $\mathrm{col}(B)$, and so the rows of A separate vectors in $\mathrm{col}(B)$. Similarly, $\mathrm{rank}(M) = \dim \mathrm{row}(A)B = \mathrm{rank}(A) = \dim \mathrm{row}(A)$ so that B is 1-to-1 on $\mathrm{row}(A)$, and the columns of B separate vectors in $\mathrm{row}(A)$. □

Thus to preserve noncontextuality we ought to confine ourselves to decompositions $P(E|\pi) = \mathbf{EP}$ where $\mathrm{rank}(P(E|\pi)) = \mathrm{rank}(\mathbf{E}) = \mathrm{rank}(\mathbf{\Pi})$. Indeed, there would not seem to be any good reason for exploring contextual representations, unless of course they had some other perhaps desirable structure. With this in mind, we now turn to the possibility of constructing a classical model of $P(E|\pi)$. This would require us to be able to write for each state π and effect E,

$$P(E|\pi) = \sum_i P(E|\lambda_i)P(\lambda_i|\pi), \tag{5.3}$$

that is, we factorize $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$. In the language of Spekkens, we thereby construct an *ontological model*: we introduce a set of "hidden variables" which take values $\{\lambda_i\}$, we map effects to response functions $P(E|\lambda)$, and we map states to distributions $P(\lambda|\rho)$. In our terms, this is equivalent to assuming there exists a reference measurement with outcomes $\{\lambda_i\}$ whose performance the gambler is indifferent to so that $\Phi_\lambda = I$. It is this indifference that licenses the gambler to act as if the system is characterized by some underlying properties which the reference measurement $\{\lambda_i\}$ reads off. Those without a commitment to expressing everything in terms of coherence between the beliefs of a gambler betting on the consequences of their own actions, would simply want to say: if I can give an ontological model, then the system must really "have" such properties.

But in fact, constructing such ontological models is trivial. Indeed, there is an ontological model for every such table $P(E|\pi)$, namely, $P(E|\pi) = P(E|\pi)I_n$. Here we introduce as many reference outcomes as π's. Each π_j is associated with a state of complete certainty about the outcome of the reference measurement: a column vector of all 0's and a 1 in the jth place. We take the response function $P(E_i|\lambda)$ to simply be the ith row of $P(E|\pi)$ so that indeed $P(E_i|\lambda)P(\lambda|\pi_j) = P(E_i|\pi_j)$ as desired. In fact, in the language of Spekkens, we have in this model measurement noncontextuality but preparation contextuality.

Now it may be that $P(E|\pi)$ is invertible. Then this trivial decompositions will be equirank, and by Lemma 5.2.1 the construction will be noncontextual. In the language of Spekkens, we would have a *noncontextual ontological model*. But in general the trivial decomposition will yield a contextual representation. For example, suppose the third column of $P(E|\pi)$, that is, $P(E|\pi_3)$ were a convex combination of the first two: $P(E|\pi_3) = pP(E|\pi_1) + (1-p)P(E|\pi_2)$. This is a linear dependency, and so $P(E|\pi)$ cannot have full rank. In the decomposition $P(E|\pi) = P(E|\lambda)P(\lambda|\pi) = P(E|\pi)I_n$, we map $P(\lambda|\pi_1) = (1, 0, \dots)^\dagger$, $P(\lambda|\pi_2) = (0, 1, 0, \dots)^\dagger$, and $P(\lambda|\pi_3) = (0, 0, 1, \dots)^\dagger$ so that each $P(\lambda|\pi_j)$ picks out

the jth column $P(E_i|\pi_j)$. But since $P(E|\pi_3) = pP(E|\pi_1) + (1-p)P(E|\pi_2)$, it follows that $P(\lambda|\pi_3) = (0,0,1,0,\dots)^\dagger$ is indistinguishable from $pP(\lambda|\pi_1) + (1-p)P(\lambda|\pi_2) = (p, 1-p, 0, \dots)^\dagger$. In other words, there are two completely different states of uncertainty about the reference outcomes $\{\lambda_i\}$ which by the gambler's own lights are consistent with assigning identical probabilities to all measurements. The difference between the two we call *context*. Now this particular example could be avoided by identifying which columns of $P(E|\pi)$ are extremal (in the sense of convex independence) and which rows are extremal (in the sense of conic independence) and restricting attention to those: but the underlying problem is quite general.

This trivial example shows that a *nonnegative matrix factorization* (NMF) is always possible, that is, a decomposition $M = AB$ where M, A, B are elementwise nonnegative. But it is generally desirable to have more compact such factorizations.

Definition 5.2.1. *Let $M \in \mathbb{R}^{m\times n}_{\geq 0}$ be a nonnegative matrix. The* nonnegative rank *$rank_+(M) = r_+$ is the smallest inner dimension such that M can be factored as $M = AB$ where $A \in \mathbb{R}^{m\times r_+}$ and $B \in \mathbb{R}^{r_+\times n}$. It satisfies $rank(M) \leq rank_+(M) \leq \min(m,n)$.*

In particular, we would like to find an NMF $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ where the columns $P(\lambda|\pi_i)$ are proper probability distributions summing to 1. The nonnegative matrix rank puts a lower bound on the number of the outcomes of the classical reference measurement.

Lemma 5.2.2. *Let M be column stochastic and $M = AB$ a nonnegative matrix factorization. After deleting any zero columns of A, there is an equivalent factorization $M = A'B'$ with A' and B' column stochastic.*

Proof. Let a_i denote the columns of A and let $\overline{a}_i$ denote ith column sum. Now if $\overline{a}_i = 0$, delete that column as well as the ith row of B: since A is nonnegative, the whole column must be 0, and since $M = a_1 b_1^\dagger + \dots$, neither will contribute to M. Now let $K = \mathrm{diag}(\overline{a}_1, \dots)$: then $A' = AK^{-1}$ will be column stochastic. $B' = KB$ will ensure $M = A'B'$. Meanwhile, let $u = (1, \dots, 1)^\dagger$. Clearly, $u^\dagger B' = u^\dagger KB$. But $u^\dagger K = (\overline{a}_1, \dots) = u^\dagger A$, so $u^\dagger B' = u^\dagger AB = u^\dagger M = u^\dagger$ since M is assumed column stochastic. □

Corollary 5.2.2.1. *We may always construct an NMF of $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ so that $P(\lambda|\pi)$ is column stochastic.*

Proof. We may always organize the rows of $P(E|\pi)$ in terms of measurements, i.e., $E_1^{(1)}, E_2^{(1)}, \dots, E_1^{(2)}, E_2^{(2)}, \dots$ where the superscript refers to which measurement the outcome is a part of. If a measurement has a full set of outcomes, then $\forall j : \sum_i P(E_i^{(k)}|\pi_j) = 1$: that is, the column sums of the rectangular block $P(E^{(k)}|\pi)$ are all 1. If a measurement doesn't have a full set of outcomes, we can always add a row namely $P(1|\pi) - \sum_i P(E_i^{(k)}|\pi)$ which completes the measurement: we may always consider the outcome that none of the other outcomes occurs. Suppose there are h measurements. In any column, there are h blocks which sum to 1: the column sums of $P(E|\pi)$ are therefore h, so that $h^{-1}P(E|\pi)$

is column stochastic and so by the previous lemma, $h^{-1}P(E|\pi) = P'(E|\lambda)P(\lambda|\pi)$ are all column stochastic. Thus $P(E|\pi) = [hP(E'|\lambda)]P(\lambda|\pi)$ is a decomposition of the original matrix, and $P(\lambda|\pi)$ is column stochastic, so its columns can be interpreted as distributions of the outcomes of the reference measurement $\{\lambda_i\}$. □

Invoking Lemma 5.2.1, we may now characterize when a noncontextual ontological model exists of a probability table $P(E|\pi)$.

Corollary 5.2.2.2 (Theorem 1 of [81])**.** *$P(E|\pi)$ admits a noncontextual ontological model iff there exists a nonnegative matrix factorization $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ satisfying $rank(P(E|\pi)) = rank(P(E|\lambda)) = rank(P(\lambda|\pi))$.*

Finally, it may be that $P(E|\pi)$ contains within it probabilities $P(R|R)$ characterizing a reference measurement $\{R_i\}$. We shall see that the existence of a noncontextual ontological model amounts to the assumption that a classical reference measurement $\{\lambda_i\}$ exists for which $\{R_i\}$ is a reference.

5.2.1 An example

An example might be helpful here. Let us consider a probability table with three measurements each with three outcomes, and four preparatory outcomes.

$$P(E|\pi) = \left(\begin{array}{c|cccc} & \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \hline A_{AB} & 1/2 & 1/2 & 1/2 & 1/2 \\ B_{AB} & 1/2 & 0 & 1/2 & 0 \\ N_{AB} & 0 & 1/2 & 0 & 1/2 \\ \hline B_{BC} & 1/2 & 0 & 1/2 & 0 \\ C_{BC} & 1/2 & 1/2 & 0 & 0 \\ N_{BC} & 0 & 1/2 & 1/2 & 1 \\ \hline C_{AC} & 1/2 & 1/2 & 0 & 0 \\ A_{AC} & 1/2 & 1/2 & 1/2 & 1/2 \\ N_{AC} & 0 & 0 & 1/2 & 1/2 \end{array}\right). \tag{5.4}$$

We will label the three measurements AB, BC and AC. Indeed, notice that $P(A_{AB}|\pi) = P(A_{AC}|\pi)$: thus we are justified in identifying an effect A whose equivalence class is $\{A_{AB}, A_{AC}\}$ since we assign the same probabilities regardless of which π occurs. Similarly, we may identify effects B and C. The measurement labels then serve to remind us which pair of effects appear together in a single measurement, along with a third effect to ensure probabilities sum to 1. This table has the following contextual factorization with a classical

reference measurement of four outcomes,

$$P(E|\pi) = P(E|\lambda)P(\lambda|\pi) \tag{5.5}$$

$$= \left(\begin{array}{c|cccc} & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \\ \hline A_{AB} & 1 & 0 & 0 & 1 \\ B_{AB} & 0 & 1 & 0 & 0 \\ N_{AB} & 0 & 0 & 1 & 0 \\ \hline B_{BC} & 0 & 1 & 0 & 0 \\ C_{BC} & 1 & 0 & 0 & 0 \\ N_{BC} & 0 & 0 & 1 & 1 \\ \hline C_{AC} & 1 & 0 & 0 & 0 \\ A_{AC} & 0 & 1 & 1 & 0 \\ N_{AC} & 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c|cccc} & \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \hline \lambda_1 & 1/2 & 1/2 & 0 & 0 \\ \lambda_2 & 1/2 & 0 & 1/2 & 0 \\ \lambda_3 & 0 & 1/2 & 0 & 1/2 \\ \lambda_4 & 0 & 0 & 1/2 & 1/2 \end{array}\right). \tag{5.6}$$

Indeed, notice for example that

$$P(A_{AB}|\lambda) = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}, \tag{5.7}$$

but

$$P(A_{AC}|\lambda) = \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}. \tag{5.8}$$

The response function assigned to the effect A depends on context: whether A appears in the measurement AB or the measurement AC. By construction $P(A_{AB}|\lambda)P(\lambda|\pi) = P(A_{AC}|\lambda)P(\lambda|\pi)$: this is necessary in order to reproduce the table. But if the gambler were able to perform the classical reference measurement and obtain outcome λ_1, notice that $P(A_{AB}|\lambda_1) = 1$ but $P(A_{AC}|\lambda_1) = 0$: they would be certain that A would occur in the measurement AB but certain that A would *not* occur in the measurement AC. Thus the gambler would not regard "A" as a property of the system independent of the choice of measurement: whether "A" depends on the context in which it is measured.

Moreover, $\text{rank}(P(E|\pi)) = 3, \text{rank}(P(E|\lambda)) = 4, \text{rank}(P(\lambda|\pi)) = 3$. Thus indeed the equirank condition is violated. Denote the four columns of $P(E|\pi)$ as p_1, p_2, p_3, p_4, and note that $p_1 - p_2 - p_3 + p_4 = 0$. We can therefore write an arbitrary vector in the column space of $P(E|\pi)$ as e.g., $x = ap_1 + bp_2 + cp_3$. Forming the explicit expression for x and demanding that it be elementwise nonnegative leads to several inequalities which a, b, c must satisfy, not all of which are independent: they can be reduced down to four, namely $b \geq 0, c \geq 0, a+b \geq 0, a+c \geq 0$. In three dimensions, extreme rays are where two independent inequalities are both saturated [122]. Solving for these explicitly and discarding solutions where $(a, b, c) = (0, 0, 0)$, we find four extreme rays: $(1, 0, 0), (0, 1, 0), (0, 0, 1), (-1, 1, 1)$. But assigning these values to a, b, c yields just p_1, p_2, p_3, and p_4 again. Suppose that $P(E|\lambda)$ only had three columns. Since $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$, these three columns would have to

generate a cone with four extreme rays, which is contradictory. Thus $\text{rank}_+(P(E|\pi)) \geq 4$, and since we have an example with inner dimension 4, it follows that $\text{rank}_+(P(E|\pi)) = 4$: we cannot find an NMF with fewer hidden variables, and so any classical model must be contextual.

5.2.2 Simplex embedding

We now show that we can understand the construction of a noncontextual classical model in terms of a linear embedding of states and effects into the simplex and its dual hypercube in such a way that all probabilities are preserved.

Lemma 5.2.3. *Let $M = AB$ be a rank decomposition so that $M \in \mathbb{F}_r^{m\times n}$, $A \in \mathbb{F}_r^{m\times r}$, and $B \in \mathbb{F}_r^{r\times n}$. At the same time, let $M = CD$ be an equirank decomposition with arbitrary inner dimension: $C \in \mathbb{F}^{m\times k}$, and $D \in \mathbb{F}^{k\times n}$ with $r = \text{rank}(M) = \text{rank}(C) = \text{rank}(D)$. We may express this decomposition equivalently as*

$$C = AT_A \qquad\qquad D = T_B B \tag{5.9}$$

for T_A, T_B satisfying $T_A T_B = I_r$. Moreover, $T_B T_A$ projects column vectors into $\text{col}(D)$ and row vectors into $\text{row}(C)$.

Proof. Since $M = CD$, every column of M is a linear combination of the columns of C. Hence $\text{col}(M) \subseteq \text{col}(C)$. But since $\text{rank}(M) = \text{rank}(C)$, the column spaces have the same dimension, and so in fact $\text{col}(M) = \text{col}(C)$. Meanwhile, because $M = AB$ is a rank decomposition, $\text{col}(M) = \text{col}(A)$ and thus $\text{col}(A) = \text{col}(C)$. Since A and C share the same column space, and the columns of A form a basis for $\text{col}(A)$, we can write $C = AT_A$ for some operator T_A which specifies the (unique) expansion coefficients. Similarly, $\text{row}(M) \subseteq \text{row}(D)$. But since $\text{rank}(M) = \text{rank}(D)$, we have in fact $\text{row}(M) = \text{row}(D)$. Meanwhile, because $M = AB$ is a rank decomposition, $\text{row}(M) = \text{row}(B)$ and thus $\text{row}(B) = \text{row}(D)$ and we can write $D = T_B B$ for some operator T_B. Finally, since $M = AB$ is a rank decomposition, A has a left inverse and B has a right inverse. Thus explicitly $T_A = A^L C$ and $T_B = DB^R$ and so $T_A T_B = A^L CDB^R = A^L MB^R = A^L ABB^R = I$. Conversely, suppose $C = AT_A$ and $D = T_B B$, then

$$\text{rank}(C) = \text{rank}(AT_A) \leq \text{rank}(A) = \text{rank}(M) = \text{rank}(CD) \leq \text{rank}(C), \tag{5.10}$$

and similarly for $\text{rank}(D)$, so that $M = CD$ is an equirank decomposition.

Finally, we note $T_B T_A = (T_B T_A)^2$ is a projector. Clearly, $\text{col}(T_B T_A) \subseteq \text{col}(T_B)$. Conversely, if $y \in \text{col}(T_B)$, then $y = T_B z$ for some z. But $(T_B T_A)y = T_B T_A T_B z = T_B z = y$ so every vector in $\text{col}(T_B)$ is fixed by $T_B T_A$. Thus $\text{col}(T_B T_A) = \text{col}(T_B)$. Meanwhile since $D = T_B B$ and B has full row rank, its columns span all of $\mathbb{F}^r$. So $\text{col}(D) = T_B \text{col}(B) = T_B(\mathbb{F}^r) = \text{col}(T_B)$. Thus $\text{col}(T_B T_A) = \text{col}(D)$ and so $T_B T_A$ projects onto $\text{col}(D)$. At the same time, $\text{row}(T_B T_A) \subseteq \text{row}(T_A)$. Conversely, if $y \in \text{row}(T_A)$, then $y = zT_A$. But

$y(T_B T_A) = zT_A T_B T_A = zT_A = y$ so every vector in $\text{row}(T_A)$ is fixed by $T_B T_A$. Thus $\text{row}(T_B T_A) = \text{row}(T_A)$. Meanwhile since $C = AT_A$ and A has full column rank, its rows span all of $\mathbb{F}^r$. So $\text{row}(C) = \text{row}(A)T_A = (\mathbb{F}^r)T_A = \text{row}(T_A)$. Thus $\text{row}(T_B T_A) = \text{row}(C)$ and so $T_B T_A$ projects onto $\text{row}(C)$.

□

Corollary 5.2.3.1. *Let $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ be an equirank stochastic NMF and let $P(E|\pi) = \boldsymbol{E}\boldsymbol{\Pi}$ be a rank decomposition. Then $P(E|\lambda) = \boldsymbol{E}T_E$ and $P(\lambda|\pi) = T_\Pi \boldsymbol{\Pi}$ where $T_E = \boldsymbol{E}^L P(E|\lambda)$ and $T_\Pi = P(\lambda|\pi)\boldsymbol{\Pi}^R$ satisfy $T_E T_\Pi = I_r$. Moreover, $\Pi_\lambda = T_\Pi T_E$ projects column vectors into $col(P(\lambda|\pi))$ and row vectors into $row(P(E|\lambda))$.*

Remark. *This formulation in terms of embedding maps T_E and T_Π is more appropriate if one wishes to consider whether one can embed an entire GPT (and not just a probability table $P(E|\pi)$) into the simplex and its dual.*

Remark. *We note that the more general idea of embedding one GPT into another GPT has been treated systematically in [123, 124] although finding such embeddings, if they exist, will not in general be as computationally tractable as finding a simplex embedding. In particular, one may ask when a GPT can be embedded into quantum theory itself [123] so that the state and effect spaces can be mapped into $d \times d$ Hermitian matrices, where $d = \sqrt{r}$. One then may test quantitatively whether $P(E|\pi)$, encompassing one's best judgement of the behavior of a system, is compatible with the quantum formalism.*

5.2.3 QBist noncontextuality

While so far we have hewn closely to the standard account of noncontextual ontological models, we now show that if among the scenarios captured by the table $P(E|\pi)$ there is a reference measurement $\{R_i\}$, then we can understand noncontextuality in terms more amenable to the QBist mode of thought. For a QBist, the ultimate classical reference measurement $\{\lambda_i\}$ is one for which $\Phi_\lambda = I$: the fundamental nonclassical coherence relation collapses to the law of total probability, signifying the gambler's indifference to whether the classical reference measurement is performed or not. Let $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ be a decomposition of $P(E|\pi)$. We will show that this representation is noncontextual if and only if $\{R_i\}$ forms a reference for $\{\lambda_i\}$. This makes sense conceptually: since the the states $P(E|R)$ and effects $P(R|\lambda)$ are by definition noncontextual, if $\{R_i\}$ forms a reference for $\{\lambda_i\}$ itself, then $P(E|\lambda)$ and $P(\lambda|\pi)$ must inherit this noncontextuality. For QBists, then, this gives a motivation for the idea of a noncontextual ontological model to begin with.

To see this, suppose that the gambler relies upon a reference measurement with outcomes $\{R_i\}$, but that they further suppose there is an ultimate classical reference measurement $\{\lambda_i\}$ (with c outcomes, taken to be a preparatory measurement) whose actual performance the gambler is indifferent to. Specifically, we suppose that $P(R|R)$ is a submatrix of $P(E|\pi)$, the probability table that characterizes the scenarios the gambler is invested in.

Let $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ be an equirank decomposition $P(E|\pi)$ into response functions and probability vectors. As we have seen, there then exist simplex embedding matrices such that $P(E|\lambda) = \mathbf{E}T_E$ and $P(\lambda|\pi) = T_\Pi \mathbf{\Pi}$, where $P(E|\pi) = \mathbf{E\Pi}$ is a rank decomposition. We then have $P(R|R) = \mathbf{RS}$, where $\mathbf{R}$ is constructed from the relevant rows of $\mathbf{E}$ and $\mathbf{S}$ is constructed from the relevant colums of $\mathbf{\Pi}$. Since $\{R_i\}$ forms a reference for the scenarios summarized by $P(E|\pi)$, we have $\text{rank}(P(R|R)) = \text{rank}(P(E|\pi))$, so that $P(R|R) = \mathbf{RS}$ is a rank decomposition and $\mathbf{S\Phi R} = I$. Thus in particular,

$$P(R|\lambda) = \mathbf{R}T_E \qquad\qquad P(\lambda|R) = T_\Pi \mathbf{S}, \tag{5.11}$$

so that

$$P(E|\lambda) = \mathbf{E}T_E = \mathbf{ES\Phi R}T_E = P(E|R)\Phi P(R|\lambda) \tag{5.12}$$

and

$$P(\lambda|\pi) = T_\Pi \mathbf{\Pi} = T_\Pi \mathbf{S\Phi R\Pi} = P(\lambda|R)\Phi P(R|\pi). \tag{5.13}$$

We conclude that if $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ is an equirank decomposition, furnishing a noncontextual ontological model, then

$$P(E|\lambda) = P(E|R)\Phi P(R|\lambda) \qquad\qquad P(\lambda|\pi) = P(\lambda|R)\Phi P(R|\pi), \tag{5.14}$$

which shows that $\{R_i\}$ is a reference for $\{\lambda_i\}$: we can treat its states and effects like any other in the fundamental nonclassical coherence relation.

There is, however, a catch. Notice that

$$\text{“}P(\lambda|\lambda)\text{”} = P(\lambda|R)\Phi P(R|\lambda) = T_\pi \mathbf{R\Phi S}T_E = T_\pi T_E = \Pi_\lambda, \tag{5.15}$$

where Π_λ projects column vectors onto $\text{col}(P(\lambda|\pi))$ and row vectors onto $\text{row}(P(E|\lambda))$. But if $\{\lambda_i\}$ is an ultimate classical reference measurement, we would expect that $P(\lambda|\lambda) = I_c$, where c is the number of outcomes of the classical reference measurement! In a special case, this is possible. Suppose $c = r = \text{rank}(P(E|\pi))$. Then $P(E|\pi) = P(E|\lambda)P(\lambda|R)$ would be not just an equirank decomposition, but a rank decomposition. In this case, T_Π and T_E would be $r \times r$ matrices and since $T_E T_\Pi = I_r$, we must have $T_E = T_\Pi^{-1}$ so that $\Pi_\lambda = T_\Pi T_E = I_r$, as desired. In other words, if $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ is in fact a rank decomposition, $P(\lambda|\lambda) = I$, as expected. Otherwise if $c > r$, although “$P(\lambda|\lambda)$” $= \Pi_\lambda$, the matrix nevertheless behaves exactly as one would want,

$$P(E|\lambda)P(\lambda|\lambda)P(\lambda|\pi) = P(E|\pi), \tag{5.16}$$

since Π_λ projects column vectors into $\text{col}(P(\lambda|\pi))$ and row vectors into $\text{row}(P(E|\lambda))$. Thus in the relevant sense, Π_λ acts just like the identity. In fact, it couldn't be otherwise. Since

$P(R|\lambda)\Pi_\lambda = P(R|\lambda)$ and $\Pi_\lambda P(\lambda|R) = P(\lambda|R)$, we have

$$P(R|\lambda) = P(R|\lambda)P(\lambda|R)\Phi P(R|\lambda) = P(R|R)\Phi P(R|\lambda) \tag{5.17}$$

$$P(\lambda|R) = P(\lambda|R)\Phi P(R|\lambda)P(\lambda|R) = P(\lambda|R)\Phi P(R|R), \tag{5.18}$$

so that $P(R|\lambda) \in \text{col}(P(R|R))$ and $P(\lambda|R) \in \text{row}(P(R|R))$. If $\text{rank}(P(R|R)) = r < c$, there is no way that $P(\lambda|R)\Phi P(R|\lambda) = I_c$ since this would imply that $c \leq \text{rank}(P(\lambda|R)) \leq \text{rank}(P(R|R)) = r$, a contradiction. If we like, however, we could extend $P(\lambda|\lambda)$ to the full identity matrix: clearly it will make no difference since Π_λ already acts on the identity on our response functions and probability vectors. Moreover, we may adopt $\Phi_\lambda = I$, since

$$\text{“}P(\lambda|\lambda)\text{”}\Phi_\lambda\text{“}P(\lambda|\lambda)\text{”} = \Pi_\lambda^2 = \text{“}P(\lambda|\lambda)\text{”}. \tag{5.19}$$

With this technicality aside, we now show the converse. Let $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$. If $P(R|R)$ is a submatrix of $P(E|\pi)$ which characterizes a reference measurement, and $\{R_i\}$ forms a reference for $\{\lambda_i\}$, in other words, $P(\lambda|R)\Phi P(R|\lambda) = \Pi_\lambda$, then $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ must be an equirank decomposition, constituting a noncontextual ontological model.

Theorem 5.2.4. *Let $P(E|\pi) = P(E|\lambda)P(\lambda|E)$ and assume that a conditional probability matrix $P(R|R)$ characterizing a reference measurement is a submatrix of $P(E|\pi)$. Thus in particular, $P(R|R) = P(R|\lambda)P(\lambda|R)$. Let Φ be a Born matrix for $P(R|R)$. Then $P(\lambda|R)\Phi P(R|\lambda) = \Pi_\lambda$ where Π_λ projects column vectors onto col$(P(\lambda|\pi))$ and row vectors onto row$(P(E|\lambda))$ iff $P(E|\pi) = P(E|\lambda)P(\lambda|E)$ is a noncontextual ontological model.*

Proof. On the one hand, we have shown that the existence of a noncontextual ontological model is equivalent to $P(E|\pi) = P(E|\lambda)P(\lambda|E)$ being an equirank decomposition. Then $P(E|\lambda) = ET_{\mathbf{E}}$, $P(\lambda|\pi) = T_\Pi \mathbf{\Pi}$, and $T_\Pi T_E = \Pi_\lambda$ for simplex embedding matrices T_E and T_Π (Corollary 5.2.3.1). Here $P(E|\pi) = \mathbf{E\Pi}$ is a rank decomposition. Since $\{R_i\}$ is a reference measurement, $\text{rank}(P(R|R)) = \text{rank}(P(E|\pi))$. Thus $P(R|R) = \mathbf{RS}$ is a rank decomposition, where $\mathbf{R}$ is constructed from the corresponding rows of $\mathbf{E}$ and $\mathbf{S}$ is constructed from the corresponding columns of $\mathbf{\Pi}$. Since Φ is a $\{1\}$-inverse of $P(R|R)$, we have $\mathbf{S}\Phi\mathbf{R} = I$ (Lemma 3.4.4). Thus

$$\Pi_\lambda = T_\Pi T_E = T_\Pi \mathbf{S}\Phi\mathbf{R}T_E = P(\lambda|R)\Phi P(R|\lambda). \tag{5.20}$$

Conversely, let $P(\lambda|R)\Phi P(R|\lambda) = \Pi_\lambda$. Then since $\{P(\lambda|R)\} \subseteq \{P(\lambda|\pi)\}$,

$$P(\lambda|R) = \Pi_\lambda P(\lambda|R) = P(\lambda|R)\Phi P(R|\lambda)P(\lambda|R) = P(\lambda|R)\Phi P(R|R). \tag{5.21}$$

Thus $\text{rank}(P(\lambda|R)) \leq \text{rank}(P(R|R))$. But from $P(R|R) = P(R|\lambda)P(\lambda|R)$, we have $\text{rank}(P(R|R)) \leq \text{rank}(P(\lambda|R))$. We conclude that $\text{rank}(P(\lambda|R)) = \text{rank}(P(R|R)) =$

rank$(P(E|\pi))$. Similarly, since $\{P(R|\lambda)\} \subseteq \{P(E|\lambda)\}$,

$$P(R|\lambda) = P(R|\lambda)\Pi_\lambda = P(R|\lambda)P(\lambda|R)\Phi P(R|\lambda) = P(R|R)\Phi P(R|\lambda). \tag{5.22}$$

Thus rank$(P(R|\lambda)) \leq$ rank$(P(R|R))$. But from $P(R|R) = P(R|\lambda)P(\lambda|R)$, we have rank$(P(R|R)) \leq$ rank$(P(\lambda|R))$. We conclude that rank$(P(R|\lambda))$ = rank$(P(R|R))$ = rank$(P(E|\pi))$ and so the decomposition $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ is equirank, corresponding to a noncontextual ontological model. □

5.3 Computational interlude

5.3.1 Alternating projections

We now discuss some of the practicalities of computing such matrix factorizations. We have already seen that one may calculate a rank decomposition using the singular value decomposition,

$$P(E|\pi) = U\Sigma V^\dagger = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\dagger \\ V_2^\dagger \end{bmatrix} = U_1(\Sigma_r V_1^\dagger) = \mathbf{ES}, \tag{5.23}$$

where Σ_r is a diagonal matrix of the r nonzero singular values. But in practice, even as one refines one's probabilities, one will generically find $P(E|\pi)$ is in fact full rank, perhaps with a few large singular values, and many more smaller singular values. One could set these small singular values to 0 (and which singular values to zero out is a matter of judgement), and from the SVD decomposition obtain a low rank approximation to $P(E|\pi)$. But then it is possible that the low rank approximation will contain numbers that lie outside [0,1]. To deal with this, the literature [125] suggests using an alternating series of optimization. Fixing a choice of r, one initializes $\mathbf{E}$ at random, and then solves numerically the following convex program (with e.g. python's `cvxpy` [126]),

$$\min_{\mathbf{\Pi}} ||P(E|\pi) - \mathbf{EP}||^2 \text{ such that } \forall i, j : 0 \leq [\mathbf{EP}]_{ij} \leq 1, \tag{5.24}$$

which gives the closest elementwise nonnegative approximation to $P(E|\pi)$ holding $\mathbf{E}$ fixed and varying $\mathbf{\Pi}$. Then holding $\mathbf{\Pi}$ fixed, one solves

$$\min_{\mathbf{E}} ||P(E|\pi) - \mathbf{EP}||^2 \text{ such that } \forall i, j : 0 \leq [\mathbf{EP}]_{ij} \leq 1, \tag{5.25}$$

finding the closest approximation over choices of $\mathbf{E}$. One alternates between these two programs until one finds a reasonable enough rank r approximation of $P(E|\pi)$ which is elementwise nonnegative.

Each individual step is a convex quadratic program and can be efficiently computed, large dimensionality notwithstanding; the optimal product may be unique even when the factor

itself is not. The same method may be extended to find a nonnegative matrix factorization if one further demands that the elements of $\mathbf{E}$ and $\mathbf{\Pi}$ lie in $[0, 1]$.

5.3.2 *A linear program for simplex embedding*

[82] offers an alternative, principled way of calculating a simplex embedding, if it exists: here we give it our own spin, showing that it amounts to the nonnegativity of a variant of the Born matrix. Let $P(E|\pi) = \mathbf{EP}$ be a rank decomposition yielding effects $\{(E_i|\}$ and states $\{|\pi_i)\}$. On the one hand, we can consider the cone generated by the states

$$\text{cone}(\mathbf{\Pi}) = \left\{ |\rho) \;\middle|\; |\rho) = \sum_i \alpha_i |\pi_i), \alpha_i \geq 0 \right\}; \tag{5.26}$$

on the other hand, we can consider its dual cone, the set of linear functionals nonnegative on the states,

$$\text{cone}^*(\mathbf{\Pi}) = \left\{ (\eta| \;\middle|\; (\eta|\mathbf{\Pi} \geq_e 0 \right\}. \tag{5.27}$$

Let $\mathbf{\Pi}^*$ be the matrix whose n_π^* rows correspond to representatives of the dual cone's extreme rays. Then $\text{cone}^*(\mathbf{\Pi}) = \text{cone}(\mathbf{\Pi}^*)$. Clearly, then, we can write any row vector nonnegative on every column of $\mathbf{\Pi}$ as some conic combination of the rows of $\mathbf{\Pi}^*$, i.e. $x^\dagger \mathbf{\Pi}^*$ for $x \geq_e 0$.

For a geometric intuition: consider the convex hull of several points. Such a convex set may be described in two alternative ways [127]. The first is called the *vertex representation*: it specifies the set by giving its extremal points, those points which cannot be written as convex combinations of any other points in the set. The second is called the *halfspace representation*: here each facet of the convex set corresponds to an inequality $a_i \cdot x + b_i \leq 0$ defining a halfspace, and the convex set is expressed as the intersection of all these halfspaces. For example, suppose we have the following 4 states given in Bloch representation so that $(1| = (1, 0, 0)$,

$$|S_1) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad |S_2) = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \qquad |S_3) = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \qquad |S_4) = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}: \tag{5.28}$$

the traceless part of the states form the vertices of a square. These are clearly the extreme points of the square: any point in the interior of the square can be written as a convex combination of these four states. We can consider more generally the cone they generate, whose base is cut by $(1| = (1, 0, 0)$. The four states then lie on extremal rays of the cone which terminate at $|0)$. At the same time, we could describe the same square as an intersection of halfspaces: all the points (x, y) satisfying $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$. Putting these in

standard form,

$$y - 1 \leq 0 \qquad -y - 1 \leq 0 \qquad z - 1 \leq 0 \qquad -z - 1 \leq 0. \tag{5.29}$$

The square is the intersection of all points satisfying these four inequalities. Introducing an auxiliary dimension, we can write these affine inequalities as linear inequalities, indeed, as a single matrix inequality,

$$\begin{pmatrix} -1 & 1 & 0 \\ -1 & -1 & 0 \\ -1 & 0 & 1 \\ -1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix} \leq 0, \tag{5.30}$$

which, incidentally, also allows one to appreciate the role of the "traceful" part of the state in a new light. Multiplying by -1 gives

$$\begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix} \geq 0, \tag{5.31}$$

which shows that any conic combination of the rows of the left hand matrix will be nonnegative on the square. In fact, these rows represent extremal rays of the dual cone $\text{cone}^*(\mathbf{\Pi})$. Geometrically, they correspond also to a square, albeit rotated 45 degrees: by duality, extremal points have been exchanged with facets. In GPT-language, this dual square correspond to all of the logically possible effects in the theory.

Returning to the main thread, we can go in the opposite direction, considering the cone generated by the effects $\mathbf{E}$ in the rank decomposition $P(E|\pi) = \mathbf{EP}$,

$$\text{cone}(\mathbf{E}) = \left\{ (\eta| \;\middle|\; (\eta| = \sum_i \beta_i (E_i|, \beta_i \geq 0 \right\}, \tag{5.32}$$

as well as its dual cone

$$\text{cone}^*(\mathbf{E}) = \left\{ |\sigma) \;\middle|\; \mathbf{E}|\sigma) \geq_e 0 \right\}, \tag{5.33}$$

for which $\text{cone}^*(\mathbf{E}) = \text{cone}(\mathbf{E}^*)$, the n_E^* columns of $\mathbf{E}^*$ being representatives of the extremal rays. Then we can write any column vector nonnegative on every row of $\mathbf{E}$ as some conic combination of the columns of $\mathbf{E}^*$, i.e., $\mathbf{E}^* y$ for $y \geq_e 0$. Suppose now that we can find a matrix Φ^* of shape $n_E^* \times n_\pi^*$ which is elementwise nonnegative and which satisfies

$$\mathbf{E}^* \Phi^* \mathbf{\Pi}^* = I. \tag{5.34}$$

Then since by construction $\mathbf{E}\mathbf{E}^* \geq_e 0$ and $\mathbf{\Pi}^*\mathbf{\Pi} \geq_e 0$, we have

$$(\mathbf{E}\mathbf{E}^*)\Phi^*(\mathbf{\Pi}^*\mathbf{\Pi}) = P(E|\pi), \tag{5.35}$$

where each parenthesized matrix is nonnegative. Consolidating this into the product of two matrices, we therefore obtain a nonnegative matrix factorization, and after possibly renormalizing it, a decomposition $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$. Indeed, letting e.g., $T_E = \mathbf{E}^*\Phi^*$ and $T_\Pi = \mathbf{\Pi}^*$, we have therefore constructed by Lemma 5.2.3 a simplex embedding, and thus a noncontextual ontological model.

Given $\mathbf{E}^*$ and $\mathbf{\Pi}^*$, which may be computed using standard software packages (e.g., `cdd` [128]), it is a simple linear program to find an elementwise nonnegative Φ^* satisfying $\mathbf{E}^*\Phi^*\mathbf{\Pi}^* = I$ if one exists. If it does not exist, following [82], we may relax the problem and seek a Φ^* satisfying the following program:

$$\begin{aligned}
&\text{minimize } p \\
&\quad \text{over } p, \Phi^* \\
&\text{subject to } \mathbf{E}^*\Phi^*\mathbf{\Pi}^* = (1-p)I + p|\mu)(1| \\
&\quad \Phi^* \geq_e 0, 0 \leq p \leq 1,
\end{aligned}$$

where $|\mu)(1|$ is the completely depolarizing channel. Given sufficient noise, any $P(E|\pi)$ has a noncontextual ontological model: depolarizing noise will eventually shrink the (traceless part of the) state and effect spaces sufficiently that they can fit simultaneously into the simplex and the dual hypercube of response functions. The minimum amount of depolarizing noise that makes this possible is a measure of the nonclassicality of the scenario.

Finally, we note that $\mathbf{E}^*\Phi^*\mathbf{\Pi}^* = I$ is a kind of cousin of the Born identity $\mathbf{\Pi}\Phi\mathbf{E} = I$ where instead of actual states $\mathbf{\Pi}$, we have representatives of the extremal rays of the cone dual to the effects $\mathbf{E}^*$ and instead of actual effects $\mathbf{E}$, we have representatives of the extremal rays of the cone dual to the states $\mathbf{\Pi}^*$. In the special case that $\mathbf{E}^* = \mathbf{\Pi}$ and $\mathbf{\Pi}^* = \mathbf{E}$, then $\Phi^* = \Phi$ is a nonnegative $\{1\}$-inverse $P(E|\pi)$, if it exists. In the next section, we will pursue this idea further.

5.4 Stochastic Φ

We showed in Lemma 3.2.2 that Φ may always be taken to be quasistochastic. When might it be possible to go further and take Φ to be stochastic? Now if $P(R|R)$ is invertible, the answer is very simple: the only stochastic matrices with stochastic inverses are permutation matrices [129]. To handle the more general case, we adapt an argument given in [83].

Theorem 5.4.1. *Let $P \equiv P(X|Y)$ be a column stochastic matrix. Then there exists a column stochastic $\{1\}$-inverse Φ, satisfying $P\Phi P = P$ iff*

1. *$P(X|Y)$ factors into the product of two stochastic matrices $P(X|Y) = P(X|\lambda)P(\lambda|Y)$.*

2. *The columns of $P(X|\lambda)$ have pairwise disjoint support:* $\forall i \neq j : \operatorname{supp} P(X|\lambda_i) \cap \operatorname{supp} P(X|\lambda_j) = \emptyset$.

3. *Every column of $P(X|\lambda)$ is also a column of $P(X|Y)$.*

Proof. First, we assume a column stochastic Φ exists satisfying $P\Phi P = P$. Recall that $\Pi = P\Phi$ is a projector onto the column space of P. The jth column of Π is $\pi_j = P\Phi e_j = P\phi_j$ where ϕ_j is the jth column of Φ. We assumed Φ is column stochastic so π_j is a convex combination of columns of P. Now denote by $P(X|\lambda_i)$ the extreme columns of $P(X|Y)$, that is, the distinct extreme points of the convex hull of the columns of $P(X|Y)$. Now in general, $\Pi P(X|\lambda_i) = P(X|\lambda_i)$ and so

$$P(X|\lambda_i) = \sum_j P(X_j|\lambda_i)\pi_j = \sum_j P(X_j|\lambda_i)[P\phi_j], \tag{5.36}$$

that is, $P(X|\lambda_i)$ must be a convex combination of the π_j which are themselves convex combinations of columns of P. But we assumed that $P(X|\lambda_i)$ is extremal. Thus if $P(X_j|\lambda_i)$ is nonzero, $\pi_j = P(X|\lambda_i)$ since $P(X|\lambda_i)$ can't be expressed as a convex combination of any other columns. Now suppose $P(X|\lambda_i)$ and $P(X|\lambda_k)$ share support at index j. Then by the above argument, $\pi_j = P(X|\lambda_i) = P(X|\lambda_k)$: they must be equal. We conclude that if two extreme columns are distinct, they must have disjoint supports, which gives us (2). We already constructed $P(X|\lambda)$ from the extreme columns of $P(X|Y)$, giving us (3). And since every column of $P(X|Y)$ is a convex combination of its extreme columns, there must be some column stochastic matrix $P(\lambda|Y)$ such that $P(X|Y) = P(X|\lambda)P(\lambda|Y)$, which gives us (1).

Conversely, assume (1), (2), and (3). By (1), $P(X|Y) = P(X|\lambda)P(\lambda|Y)$. By (3), for each λ_i there is some column index $f(i)$ such that $P(X|\lambda_i) = P(X|Y_{f(i)})$. Let $P(Y|\lambda)$ satisfy $P(Y_i|\lambda_j) := \delta_{i,f(j)}$ so that $P(X|\lambda) = P(X|Y)P(Y|\lambda)$. Meanwhile, define $P(\lambda|X)$ by first setting

$$P(\lambda_i|X_j) = \begin{cases} 1, & j \in \operatorname{supp} P(X|\lambda_i), \\ 0, & \text{otherwise}, \end{cases} \tag{5.37}$$

and then, if there is some index j which doesn't lie in the support of any column $P(X|\lambda_i)$, put a single 1 anywhere in that column. In the end, because the supports are disjoint, each column of $P(\lambda|X)$ will have exactly one nonzero entry, and $P(\lambda|X)P(X|\lambda) = I$. To see this, first assume $i \neq k$. Then

$$\sum_j P(\lambda_i|X_j)P(X_j|\lambda_k) = 0, \tag{5.38}$$

since by construction $P(\lambda_i|X)$ and $P(X|\lambda_k)$ have disjoint supports for $i \neq k$. If $i = k$,

$$\sum_j P(\lambda_i|X_j)P(X_j|\lambda_i) = \sum_{j \in \operatorname{supp} P(X|\lambda_i)} P(X_j|\lambda_i) = 1, \tag{5.39}$$

since $P(X|\lambda)$ is column stochastic, and we are summing just the nonzero entries in a column. Finally, let $\Phi = P(Y|\lambda)P(\lambda|X)$. Clearly, Φ is column stochastic and

$$\begin{aligned} P\Phi P &= P(X|Y)\big\{P(Y|\lambda)P(\lambda|X)\big\}P(X|Y) && (5.40)\\ &= \big\{P(X|Y)P(Y|\lambda)\big\}P(\lambda|X)\big\{P(X|\lambda)P(\lambda|Y)\big\} && (5.41)\\ &= P(X|\lambda)P(\lambda|Y) = P. && (5.42) \end{aligned}$$

□

Corollary 5.4.1.1. *In the above construction, $P(X|Y) = P(X|\lambda)P(\lambda|Y)$ is a rank decomposition.*

Proof. By (2) the columns of $P(X|\lambda)$ have pairwise disjoint supports, and so they must be linearly independent. Thus $P(X|\lambda) \in \mathbb{R}_r^{n\times r}$ and $\operatorname{rank}(P) = \operatorname{rank}\big(P(X|\lambda)P(\lambda|Y)\big) \leq \operatorname{rank}\big(P(X|\lambda)\big) = r$. But by (3) every column of $P(X|\lambda)$ is also a column of P, and so $\operatorname{rank}\big(P(X|\lambda)\big) \leq \operatorname{rank}(P)$, from which we conclude $\operatorname{rank}(P) = r$. Finally, since $P(\lambda|Y) \in \mathbb{R}^{r\times m}$, $\operatorname{rank}\big(P(\lambda|Y)\big) \leq \min(r, n) \leq r$. But $\operatorname{rank}(P) \leq \operatorname{rank}\big(P(\lambda|Y)\big) \leq r = \operatorname{rank}(P)$ so that $\operatorname{rank}\big(P(\lambda|Y)\big) = r$ as well. □

Remark. *Notice that just as $P(X|Y) = P(X|\lambda)P(\lambda|Y)$, $\Phi = P(Y|\lambda)P(\lambda|X) = P(Y|X)$, which we could call a stochastic retrodiction map. Moreover, the key step is that $P(\lambda|X)P(X|\lambda) = I$. In other words, the events $\{X_i\}$ allow one to completely retrodict the λ_i's.*

Corollary 5.4.1.2. *In the above construction, $P(\lambda|X)P(X|\lambda) = I$: so $P(X|\lambda)$ has a stochastic left inverse. Similarly, $P(Y|\lambda)$ is a stochastic right inverse of $P(\lambda|Y)$.*

Proof. We defined $P(Y_i|\lambda_j) = \delta_{i,f(j)}$ so that $P(X|\lambda) = P(X|Y)P(Y|\lambda)$. But that means $P(X|\lambda) = P(X|\lambda)P(\lambda|Y)P(Y|\lambda)$. Acting from the left with $P(\lambda|X)$, we have $I = P(\lambda|Y)P(Y|\lambda)$ as desired. □

Corollary 5.4.1.3. *Let $P(R|R)$ be the conditional probability matrix that characterizes the reference measurement. If $P(R|R)$ has a stochastic Born matrix, then the corresponding GPT is classical.*

Proof. From Theorem 5.4.1 and its corollaries, we know that $P(R|R)$ has a stochastic rank decomposition $P(R|R) = P(R|\lambda)P(\lambda|R)$. Then $P(\lambda|\pi) = P(\lambda|R)\Phi P(R|\pi)$ is a probability

distribution and $P(E|\lambda) = P(E|R)\Phi P(R|\lambda)$ is a response function which satisfy

$$P(E|\lambda)P(\lambda|\pi) = P(E|R)\Phi P(R|\lambda)P(\lambda|R)\Phi P(R|\pi) \tag{5.43}$$
$$= P(E|R)\Phi P(R|R)\Phi P(R|\pi) \tag{5.44}$$
$$= P(E|R)\Phi P(R|\pi) \tag{5.45}$$
$$= P(E|\pi). \tag{5.46}$$

Now suppose that the effects $\{E_i\}$ and the states $\{\pi_i\}$ are informationally complete for the GPT. Then $\text{rank}(P(E|R)) = \text{rank}(P(R|\pi)) = \text{rank}(P(R|R)) = r$. Now from $P(E|\lambda) = P(E|R)\Phi P(R|\lambda)$, we have $\text{rank}(P(E|\lambda)) \leq \text{rank}(P(E|R))$. But from $P(E|R) = P(E|\lambda)P(\lambda|R)$, we have $\text{rank}(P(E|R)) \leq \text{rank}(P(E|\lambda))$. Thus $\text{rank}(P(E|\lambda)) = r$. Similarly, comparing $P(\lambda|\pi) = P(\lambda|R)\Phi P(R|\pi)$ and $P(R|\pi) = P(R|\lambda)P(\lambda|\pi)$, we have $\text{rank}(P(\lambda|\pi)) = r$. Since $P(R|R) = P(R|\lambda)P(\lambda|R)$ is a rank decomposition, there must be r λ's. Thus $P(E|\lambda)$ has full column rank and $P(\lambda|\pi)$ has full row rank; from $P(E|\pi) = P(E|\lambda)P(\lambda|\pi)$ we may conclude that $\text{rank}(P(E|\pi)) = r$. The factorization is therefore equirank. In particular, this will hold if we consider *all* the states and effects of the GPT: the GPT is therefore entirely classical. □

5.4.1 An example

For concreteness, we give an example of a situation when a stochastic Φ matrix is possible. Let

$$P(X|Y) = P(X|\lambda)P(\lambda|Y) = \begin{pmatrix} \frac{2}{3} & 0 \\ \frac{1}{3} & 0 \\ 0 & \frac{1}{4} \\ 0 & \frac{3}{4} \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{2} & \frac{3}{4} \\ 0 & 1 & \frac{1}{2} & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & 0 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{6} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & \frac{3}{4} & \frac{3}{8} & \frac{3}{8} \end{pmatrix}. \tag{5.47}$$

Then

$$\Phi = P(Y|\lambda)P(\lambda|X) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5.48}$$

Slightly more generally, suppose that

$$P(X|\lambda) = \begin{pmatrix} p & 0 \\ (1-p) & 0 \\ 0 & q \\ 0 & 1-q \end{pmatrix} \qquad P(\lambda) = \begin{pmatrix} a \\ b \end{pmatrix} \qquad P(\lambda|X) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \tag{5.49}$$

which yields

$$P(X) = P(X|\lambda)P(\lambda) = \begin{pmatrix} ap \\ a(1-p) \\ bq \\ b(1-q) \end{pmatrix}. \tag{5.50}$$

Clearly, $P(\lambda) = P(\lambda|X)P(X)$. Thinking of $P(\lambda)$ as a state and $P(X|\lambda)$ as a set of effects, we see that we can tell a very classical story. First λ_1 or λ_2 is chosen with some probability. Conditional on λ_1, a coin is flipped with bias p, and if it lands heads, the measurement yields X_1 and if it lands tails, the measurement yields X_2. Similarly, conditional on λ_2, a coin is flipped with bias q, and this gives either X_3 or X_4. If you like, first the ultimate classical reference measurement is performed, reading off the property of the system (which λ is the case), and then conditional on that value, one of several outcomes associated to that property is chosen at random. Meanwhile, the n distributions over the r hidden variables that give rise to P, that is, $P(\lambda|Y_i)$, are constrained: all the states of complete certainty about which λ is the case must be among them.

5.5 Negativity in quantum theory

For completeness we reproduce here the nice proof of [130] that negativity is necessary in any frame representation of quantum mechanics, and so no noncontextual ontological model is possible for full quantum theory. See also [131, 132, 133] for a general discussion of the relationship between quantum theory, frames and their duals, negativity, and contextuality.

Theorem 5.5.1 (Theorem 2 of [130])**.** *There does not exist a dual frame of positive semidefinite operators for a frame of positive semidefinite operators for $d > 1$.*

Proof. Let $\{E_i\}$ be a PSD operator frame, and $\{\tilde{E}_i\}$ its dual. Consider the map

$$M(\rho) = \sum_i \tilde{E}_i \mathrm{tr}(E_i \rho) = \rho. \tag{5.51}$$

Forming its Choi state,

$$C_M = \frac{1}{d}\sum_{ij} M(|i\rangle\langle j|) \otimes |i\rangle\langle j| \tag{5.52}$$

$$= \frac{1}{d}\sum_{ijk} \tilde{E}_k \mathrm{tr}(E_k|i\rangle\langle j|) \otimes |i\rangle\langle j| \tag{5.53}$$

$$= \frac{1}{d}\sum_{k} \tilde{E}_k \otimes \sum_{ij}\langle j|E_k|i\rangle|i\rangle\langle j| \tag{5.54}$$

$$= \frac{1}{d}\sum_{k} \tilde{E}_k \otimes E_k^T, \tag{5.55}$$

we see that it is a sum of tensor products of positive semidefinite operators: in other words, C_M is a separable state. But M is supposed to be the identity map, whose Choi state is maximally entangled. Thus the dual frame of a positive semidefinite operator frame cannot itself be composed entirely of positive semidefinite operators. □

Corollary 5.5.1.1. *Since the dual frame contains operators with negative eigenvalues, $W(E|\rho) = \boldsymbol{E}|\rho)$ will in general take negative values. Thus in quantum theory, there must be negativity in any frame representation, and so no noncontextual ontological model is possible [80].*

5.6 QBist Bell Inequalities

As we noted in the previous chapter, the authors of [78] explored $\|I - \Phi\|$ with respect to any unitarily invariant norm as a measure of nonclassicality. The intuition is that classically it is always possible to pick a reference measurement such that $\Phi = I$: but this is precisely *forbidden* by quantum theory. This is suggestive of the idea that quantum systems do not have underlying properties which are read off by measurements, even reference measurements, but that, as QBism holds, the outcomes of measurement are better regarded as the possible consequences for an individual agent gambling upon a nature which is being continually cocreated. But for a single system, this is a heuristic argument and not at all a proof: after all, the outcomes of any single system quantum experiment can be spoofed by a laptop computer which is generally regarded to operate on definite properties by perfectly definite rules. From this point of view, there is nothing to exclude the hypothesis that a single quantum system has some definite, yet unknown properties which are however disturbed by measurement. The situation is different, however, when we consider multipartite scenarios. There the violation of Bell's inequalities force us into a quandary: either quantum systems have underlying properties that condition the results of measurement, but these properties may be disturbed *without regard for locality*; or else the outcomes of measurements are created in the act of measurement for the individual observer. (We exclude here the evasion of the quandary by

the hypothesis of a multiverse.) For QBists, the point is that when one formulates a Bell inequality, one in fact appeals to the law of total probability in defiance of the fact that performing a reference measurement makes a difference. We now show that violations of Bell inequalities, and noncontextuality inequalities more generally, may be understood as quantifying the cost of abandoning the fundamental nonclassical coherence condition.

5.6.1 Local realism?

The usual way of introducing Bell inequalities is to consider a multipartite scenario, where n distant parties share parts of a system, and each party agrees to choose from a menu of potential measurements on the part local to them. In the simplest case, suppose Alice has a choice of measurements $\{A_i\}$ each of which has outcomes $\{a_i\}$, and Bob has a choice of measurements $\{B_i\}$ each of which has outcomes $\{b_i\}$. We can then consider the joint probability distribution $P(A_k = a_i, B_l = b_j)$. In particular, following Bell, we may consider the subset of all joint probability distributions which can be expressed

$$P(A_k = a_i, B_l = b_j) = \sum_m P(A_k = a_i|\lambda_m)P(B_l = b_j|\lambda_m)P(\lambda_m), \tag{5.56}$$

where $P(\lambda_m)$ is a distribution over some underlying "hidden variables" $\{\lambda_m\}$, while $P(A_k = a_i|\lambda_m)$ and $P(B_l = b_j|\lambda_m)$ are response functions which depend only upon Alice (or Bob's, respectively) local choice of measurement—and the hidden variables. The point is that, for Bell, any correlations in Alice and Bob's measurement outcomes ought to be explained by correlations between the hidden variables which their systems are supposed to have agreed upon before they were separated. The set of all such "classical" joint probability distributions for a given multipartite scenario forms a so-called *Bell polytope* $\mathcal{L}$ [134, 135].

For us, we take the point of view of a gambler who expects to act upon Alice and upon Bob, and who believes them when they promise that they will try to communicate the results of their measurements. But for the gambler, there is no measurement until they themselves act upon Alice and upon Bob [136]. Introducing reference measurements $\{R_i^A\}$ and $\{R_i^B\}$ for each of them, the gambler ought in general to appeal to the coherence condition,

$$P(A_k = a_i, B_l = b_j) = \Big\{P(A_k = a_i|R^A) \otimes P(B_l = b_j|R^B)\Big\}\Big\{\Phi^A \otimes \Phi^B\Big\}P(R^A, R^B|\rho). \tag{5.57}$$

If the gambler is indifferent to whether the reference measurements are performed or not, then this reduces to

$$P(A_k = a_i, B_l = b_j) = \Big\{P(A_k = a_i|R^A) \otimes P(B_l = b_j|R^B)\Big\}P(R^A, R^B|\rho) \tag{5.58}$$

$$= \sum_{ru} P(A_k = a_i|R_r^A)P(B_l = b_j|R_u^B)P(R_r^A, R_u^B|\rho), \tag{5.59}$$

which indeed has Bell form, where the outcomes of the reference measurements play the role of the so-called hidden variables[2]. In this way, the assumption of "Bell locality" amounts for us to the assumption that the gambler can introduce a reference measurement for Alice and for Bob, and that they are indifferent to whether these reference measurements are performed or not. The key point is, as always, that the gambler may very well *not* be indifferent to whether the reference measurements are performed. Turning it around, and as we will see, we can bound $\|I - \Phi\|$ in terms of the magnitude of a Bell inequality violation.

5.6.2 CHSH

For sake of intuition, consider for example the CHSH scenario. This is a bipartite scenario in which the gambler can ask Alice and Bob for the outcomes of two possible measurements they made. The outcomes are valued in $\{\pm 1\}$. Supposing that $P(A_k = a_i, B_l = b_j)$ takes Bell form, one may derive the famous inequality

$$-2 \leq \langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2. \tag{5.60}$$

Let us consider A_1, A_2 first. Suppose they take opposite values. In that case, $(-1) + (+1) = (+1) + (-1) = 0$ so that $A_1 + A_2 = 0$, and since either we have $(+1) - (-1) = +2$ or $(-1) - (+1) = -2$, we have $A_1 - A_2 = \pm 2$. On the other hand, suppose they take the same values. In that case, $(+1) - (+1) = (-1) - (-1) = 0$ so that $A_1 - A_2 = 0$, and since either we have $(+1) + (+1) = +2$, or $(-1) + (-1) = -2$, we conclude $A_1 + A_2 = \pm 2$. Let us now consider the quantity

$$C = (A_1 + A_2)B_1 + (A_1 - A_2)B_2. \tag{5.61}$$

If A_1, A_2 take opposite values, then $A_1 + A_2 = 0$ and $A_1 - A_2 = \pm 2$ so that $C = 0 + (\pm 2)(\pm 1) = \pm 2$. On the other hand, if A_1, A_2 take the same values, then $A_1 - A_2 = 0$ and $A_1 + A_2 = \pm 2$ so that $C = (\pm 2)(\pm 1) + 0 = \pm 2$. Either way, $C = \pm 2$. Assuming a joint probability distribution over the outcomes of the four measurements, we may consider the expectation value of C:

$$\langle C \rangle = \langle (A_1 + A_2)B_1 + (A_1 - A_2)B_2 \rangle \tag{5.62}$$

$$= \langle A_1 B_1 \rangle + \langle A_2 B_1 \rangle + \langle A_1 B_2 \rangle - \langle A_2 B_2 \rangle. \tag{5.63}$$

[2]This is perfectly sensible if we analogize to classical physics. Classically, the ultimate reference device is simply reading off the positions and momenta of all the particles: these are the "hidden variables" upon which any other observable supervenes—and in principle they need not be hidden at all! And this is as it should be: what is the point of hypothesizing some underlying properties which could never in principle be measured? But to consider something measurable according to e.g., quantum theory is quite restrictive: distributions $P(R^A_r, R^B_s|\rho)$ are not as general as $P(\lambda_m)$. The former are constrained e.g., to correspond to quantum states, which rules out large swathes of the probability simplex, whereas the latter are otherwise unconstrained. Thus joint probability distributions in the form of Eq. 5.58 are a subset of those defined by Eq. 5.56. Luckily, this is all we will need.

Since $C = \pm 2$, we must have $-2 \leq \langle C \rangle \leq 2$, and so

$$-2 \leq \langle A_1 B_1 \rangle + \langle A_2 B_1 \rangle + \langle A_1 B_2 \rangle - \langle A_2 B_2 \rangle \leq 2, \tag{5.64}$$

as desired. We derived the inequality solely by assuming that A_1, A_2, B_1, B_2 all take preexisting values in ± 1, and it holds *whatever* joint probability distribution we assign to the outcomes of the four measurements.

Let us take another point of view, however. Since $\langle A_k B_l \rangle = \sum_{ij} a_i b_j P(A_k = a_i, B_l = b_j)$, letting

$$s = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad\qquad c_{ijkl} = a_i b_j s_{kl}, \tag{5.65}$$

and denoting by $\mathbf{p}$ and $\mathbf{c}$ the tensors $P(A_k = a_i, B_l = b_j)$ and c_{ijkl} flattened into vectors, we may rewrite the CHSH inequality as

$$-2 \leq \mathbf{c}^\dagger \mathbf{p} \leq 2. \tag{5.66}$$

Here $\mathbf{p}$ is often called a *behavior*. This is quite general: a Bell inequality always takes the form

$$\mathbf{c}^\dagger \mathbf{p} \leq \beta, \tag{5.67}$$

where $\mathbf{c}$ represents a particular observable and β is a classical bound derived by supposing that $\mathbf{p}$ takes Bell form. Geometrically speaking, each Bell inequality $(\mathbf{c}, \beta)$ defines a halfspace: their collective intersection is the Bell polytope $\mathcal{L}$, the set of all behaviors which can be written in Bell form. In fact, the CHSH inequality corresponds to a facet of this polytope.

Of course, quantum mechanics generally violates Bell inequalities. In the CHSH scenario, let $A_1 = Z, A_2 = X$, and $B_1 = (Z + X)/\sqrt{2}, B_2 = (Z - X)/\sqrt{2}$, where Z, X are Pauli observables and let the initial state be $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. Then by the Born rule

$$P(A_k = a_i, B_l = b_j) = \mathrm{tr}\Big((\Pi_{a_i|A_k} \otimes \Pi_{b_j|B_l})\rho\Big), \tag{5.68}$$

where $\Pi_{a_i|A_k}$ is the rank-1 projector onto the a_i outcome of observable A_k. Famously, this choice of state and observables leads to $|\mathbf{c}^\dagger \mathbf{p}| = 2\sqrt{2}$, the maximal violation compatible with quantum mechanics.

5.6.3 *Bounding* $\|I - \Phi\|$

In order to show how a Bell inequality violation can bound $\|I - \Phi\|$, we first rewrite Eq. 5.57 as

$$\begin{aligned} &P(A_k = a_i, B_l = b_j) \\ &= \sum_{rs,uv} P(A_k = a_i|R_r^A)\Phi_{rs}^A P(B_l = b_j|R_u^B)\Phi_{uv}^B P(R_s^A, R_v^B|\rho) &(5.69)\\ &= \sum_{rs,uv} \Big[P(A_k = a_i|R_r^A)P(B_l = b_j|R_u^B)P(R_s^A, R_v^B|\rho)\Big]\Big[\Phi_{rs}^A\Phi_{uv}^B\Big] &(5.70)\\ &= \sum_{rs,uv} T_{ijkl,rsuv}\phi_{rsuv}, &(5.71)\end{aligned}$$

so that denoting by $\mathbf{p}$ the tensor $P(A_k = a_i, B_l = b_j)$ flattened into a vector, reshaping the tensor T into a rectangular matrix $\mathbf{T}$, and finally flattening $\Phi^A \otimes \Phi^B$ into a vector $\phi = |\Phi^A \otimes \Phi^B)$, we can achieve a yet more compact expression $\mathbf{p} = \mathbf{T}\phi$. In other words, the reference probabilities which specify the Bell scenario are packaged into the matrix $\mathbf{T}$, separate from the choice of Born matrices. Consequently, we can write any Bell inequality as

$$\mathbf{c}^\dagger \mathbf{T}\phi \leq \beta. \tag{5.72}$$

Now let $\phi_Q = |\Phi^A \otimes \Phi^B)$ and $\phi_C = |I \otimes I)$, where the soft brackets denote the vectorization of the matrices. On the one hand, $\mathbf{p}_Q = \mathbf{T}\phi_Q$ gives us our nonclassical distribution, while $\mathbf{p}_C = \mathbf{T}\phi_C$ lies in the Bell polytope $\mathcal{L}$. We now recall that the operator norm associated with a vector norm $\|\cdot\|$ is

$$\|\mathbf{A}\|_{\text{op}} = \sup_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{v}\|, \tag{5.73}$$

so that immediately we have $\|\mathbf{A}\mathbf{v}\| \leq \|\mathbf{A}\|_{\text{op}}\|\mathbf{v}\|$. Putting this together, we have

$$\|\mathbf{p}_Q - \mathbf{p}_C\| = \|\mathbf{T}\phi_Q - \mathbf{T}\phi_C\| = \|\mathbf{T}(\phi_Q - \phi_C)\| \leq \|\mathbf{T}\|_{\text{op}}\|\phi_Q - \phi_C\|, \tag{5.74}$$

or

$$\|\phi_Q - \phi_C\| \geq \frac{\|\mathbf{p}_Q - \mathbf{p}_C\|}{\|\mathbf{T}\|_{\text{op}}}. \tag{5.75}$$

But $\|\mathbf{p}_Q - \mathbf{p}_C\| \geq \text{dist}(\mathbf{p}_Q, \mathcal{L})$ since $\mathbf{p}_C$ lies in $\mathcal{L}$. We then have the following theorem.

Theorem 5.6.1. *Let $\boldsymbol{p}$ be a behavior which violates $\boldsymbol{c}^\dagger \boldsymbol{p} \leq \beta$. Then $\boldsymbol{c}^\dagger \boldsymbol{p} = \beta + \Delta$ for $\Delta \geq 0$, and*

$$dist(\boldsymbol{p}, \mathcal{L}) \geq \frac{\Delta}{\|\boldsymbol{c}\|_*}, \tag{5.76}$$

where $\|\cdot\|_*$ is the dual norm[3] *of $\|\cdot\|$.*

Proof. We first reprove a standard result. Let $H = \{\mathbf{x} \mid \langle \mathbf{y}, \mathbf{x} \rangle = \beta\}$ be a hyperplane. The distance of a point $\mathbf{x}_0$ to H (with respect to a choice of norm $\|\cdot\|$) is

$$\text{dist}(\mathbf{x}_0, H) = \frac{|\Delta|}{\|\mathbf{y}\|_*}, \tag{5.77}$$

where $\|\cdot\|_*$ is the dual norm and $\Delta = \langle \mathbf{y}, \mathbf{x}_0 \rangle - \beta$. To see this, we observe that for a norm $\|\cdot\|$, its *dual norm* [106] is defined as

$$\|\mathbf{x}\|_* = \sup_{\|\mathbf{y}\| \leq 1} \langle \mathbf{x}, \mathbf{y} \rangle. \tag{5.78}$$

Immediately, we have the tight inequality

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{y}\| \left\langle \mathbf{x}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle \leq \|\mathbf{y}\| \|\mathbf{x}\|_*. \tag{5.79}$$

Let us now consider the distance from a point $\mathbf{x}_0 \in \mathbb{R}^n$ to a hyperplane $H = \{\mathbf{x} \mid \langle \mathbf{y}, \mathbf{x} \rangle = \beta\}$. In other words, we seek the point in H closest to $\mathbf{x}_0$,

$$\text{dist}(\mathbf{x}_0, H) = \min_{\mathbf{x} \in H} \|\mathbf{x}_0 - \mathbf{x}\|. \tag{5.80}$$

By definition, for $\mathbf{x} \in H$, $\langle \mathbf{y}, \mathbf{x} \rangle = \beta$; by the same token, for $\mathbf{x}_0 \notin H$, $\langle \mathbf{y}, \mathbf{x}_0 \rangle = \beta + \Delta$. Subtracting these two expressions yields

$$\langle \mathbf{y}, \mathbf{d} \rangle = \Delta \qquad\qquad \mathbf{d} = \mathbf{x}_0 - \mathbf{x}, \tag{5.81}$$

so that we can reformulate our problem as $\min_{\langle \mathbf{y}, \mathbf{d} \rangle = \Delta} \|\mathbf{d}\|$. But $|\Delta| = |\langle \mathbf{y}, \mathbf{d} \rangle| \leq \|\mathbf{y}\|_* \|\mathbf{d}\|$, or

$$\|\mathbf{d}\| \geq \frac{|\Delta|}{\|\mathbf{y}\|_*}. \tag{5.82}$$

[3] In particular, for a p-norm, $\|\mathbf{x}\|_p = \left(\sum_i x_i^p\right)^{1/p}$, the dual norm is the Hölder dual $\|\mathbf{x}\|_q$ satisfying $1/p + 1/q = 1$ [106]. Notice that the Euclidean norm is self-dual in this sense.

Since the inequality is tight, we can achieve the minimum, and thus

$$\mathrm{dist}(\mathbf{x}_0, H) = \frac{|\Delta|}{\|\mathbf{y}\|_*}. \tag{5.83}$$

Finally, since every such inequality defines a halfspace $\{\mathbf{x}|\mathbf{c}^\dagger\mathbf{x} \leq \beta\}$ bounded by the corresponding hyperplane $\{\mathbf{x}|\mathbf{c}^\dagger\mathbf{x} = \beta\}$, and since the polytope $\mathcal{L}$ lies entirely in the halfspace, the distance to the hyperplane gives a lower bound on the distance to the whole polytope. In other

$$\mathrm{dist}(\mathbf{p}, \mathcal{L}) \geq \mathrm{dist}(\mathbf{p}, H) = \frac{\Delta}{\|\mathbf{c}\|_*}. \tag{5.84}$$

□

With this result in hand, we conclude that

$$\|\phi_Q - \phi_C\| \geq \frac{\|\mathbf{p}_Q - \mathbf{p}_C\|}{\|\mathbf{T}\|_{\mathrm{op}}} \geq \frac{\mathrm{dist}(\mathbf{p}_Q, \mathcal{L})}{\|\mathbf{T}\|_{\mathrm{op}}} \geq \frac{\Delta}{\|\mathbf{T}\|_{\mathrm{op}}\|\mathbf{c}\|^*} = \frac{\mathbf{c}^\dagger\mathbf{p} - \beta}{\|\mathbf{T}\|_{\mathrm{op}}\|\mathbf{c}\|_*}, \tag{5.85}$$

where Δ is the Bell inequality violation. Specializing to the Euclidean norm $\|\cdot\|_2$, we have that the associated operator norm is $\|\mathbf{T}\|_{\mathrm{op}} = \sigma_{\max}(\mathbf{T})$, the largest singular value [111]; that the Euclidean norm is self-dual; and finally, that $\|\mathrm{vec}(\mathbf{A})\|_2 = \|\mathbf{A}\|_F$, the Frobenius norm $\sqrt{\mathrm{tr}(\mathbf{A}^\dagger\mathbf{A})}$. Letting $\Phi = \Phi^A \otimes \Phi^B$, we arrive at the following theorem:

Theorem 5.6.2. *Consider a two party Bell scenario. Introduce a reference measurement for each party and consider the joint reference measurement with Born matrix $\Phi = \Phi_A \otimes \Phi_B$. Let $\boldsymbol{c}$ be the Bell observable, Δ be the Bell inequality violation, and $\sigma_{\max}$ denote the largest singular value. Finally, let $\boldsymbol{T}$ be the matrix with entries*

$$T_{ijkl,rsuv} = P(A_k = a_i|R^A_r)P(B_l = b_j|R^B_u)P(R^A_s, R^B_v|\rho), \tag{5.86}$$

where $P(A_k = a_i|R^A_r)$ and $P(B_l = b_j|R^B_u)$ are the response functions characterizing Alice and Bob's effects and $P(R^A_s, R^B_v|\rho)$ is the probability distribution characterizing the initial state. Then

$$\|I - \Phi\|_F \geq \frac{\Delta}{\sigma_{\max}(\boldsymbol{T})\|\boldsymbol{c}\|_2}. \tag{5.87}$$

The generalization to m parties is straightforward.

Returning to the CHSH example, assigning a qubit SIC reference measurement $\{R_i\}$ to Alice and Bob each, we can characterize the quantum state of their two qubits with the

joint probability distribution

$$P(R_i, R_j|\rho) = \mathrm{tr}\Big((R_i \otimes R_j)\rho\Big), \tag{5.88}$$

and characterize their measurements with

$$P(A_k = a_i|R_u) = \mathrm{tr}(\Pi_{a_i|A_k}\sigma_u) \qquad P(B_l = b_j|R_v) = \mathrm{tr}(\Pi_{b_j|B_l}\sigma_v), \tag{5.89}$$

where $\sigma_u = R_u/\mathrm{tr}(R_u)$ are reference states. Maximal violation of the CHSH inequality gives $\Delta = 2\sqrt{2} - 2$. Meanwhile, for two qubit SIC reference devices, $\sigma_{\max}(\mathbf{T}) = \sqrt{4/3}$, yielding $\|I - \Phi\|_F \geq 0.1794$. Now in fact, $\|I - \Phi\|_F \approx 24.4949$, so the bound is not particularly informative as to its exact value. The point is rather that *it cannot be 0*. Indeed, we may take

$$\mathcal{B} = \frac{\Delta}{\sigma_{\max}(\mathbf{T})\|\mathbf{c}\|_2} \tag{5.90}$$

itself to be a measure of the nonclassicality of the scenario as captured by the reference measurement. This result opens up a number of interesting further directions for research. In particular, one may ask: optimizing over number of parties, reference measurements, initial states, choices of measurements, and Bell inequalities, which scenario gives the largest bound? Is it possible to find a bound which is equal to the actual value of $\|I - \Phi\|$? Fixing a reference device (e.g., a SIC), which Bell scenario gives the largest bound? Conversely, fixing a Bell scenario, which reference measurement gives the largest bound?

5.6.4 Noncontextuality inequalities

The wonderful review article [137] gives an explicit algorithm for constructing a Bell inequality which witnesses the nonclassicality of a behavior $\mathbf{p}$ or else affirms its classicality. The first observation is that any uncertainty in the response functions (e.g., $P(A_j = a_i|\lambda_k)$) can be shunted instead into the hidden variables λ: thus it suffices to consider models where the local response functions take values in $\{0, 1\}$, that is to say, it suffices to consider deterministic local hidden variable models. In such a model, the hidden variables simply assign outcomes to measurements in a particular run. For example, let there be two parties, Alice and Bob, and let $\lambda = (\lambda_{A_2}, \ldots, \lambda_{A_m}; \lambda_{B_2}, \ldots, \lambda_{B_n})$ be an assignment of outcomes to each of m measurements for Alice and each of n measurements for Bob in a given run. To each possible λ corresponds a deterministic behavior

$$P_\lambda(A_k = a_i, B_l = b_j) = \delta_{a_i,\lambda_{A_k}}\delta_{b_j,\lambda_{B_l}}. \tag{5.91}$$

A behavior $\mathbf{p}$ then has Bell form iff it can be written as a convex combination of these deterministic behaviors,

$$\mathbf{p} = \sum_{\lambda} \mathbf{p}_\lambda p(\lambda) \qquad\qquad p(\lambda) \geq 0, \;\; \sum_{\lambda} p(\lambda) = 1. \tag{5.92}$$

Given a behavior $\mathbf{p}$, consider the following linear (and so efficiently solvable) program

$$\max_{\mathbf{c},\beta} \; S = \mathbf{c}^\dagger \mathbf{p} - \beta \;\text{ such that }\; \forall \lambda : \mathbf{c}^T \mathbf{p}_\lambda - \beta \leq 0 \text{ and } \mathbf{c}^T \mathbf{p} - \beta \leq 1. \tag{5.93}$$

Now if $\mathbf{p}$ has Bell form, from $\forall \lambda : \mathbf{c}^\dagger \mathbf{p}_\lambda - \beta \leq 0$, we have, summing over all λ's,

$$\sum_{\lambda} p(\lambda) \mathbf{c}^\dagger \mathbf{p}_\lambda - \beta \leq 0 \Longrightarrow \mathbf{c}^\dagger \mathbf{p} \leq \beta. \tag{5.94}$$

Thus $S \leq 0$. If $\mathbf{p}$ is not of Bell form, then by the second constraint $\mathbf{c}^\dagger \mathbf{p} - \beta = S \leq 1$. Since $\mathbf{p} \notin \mathcal{L}$, a separating hyperplane can be rescaled so that the optimum achieves $S = 1$. The point is that any deterministic behavior satisfies $\mathbf{c}^\dagger \mathbf{p}_\lambda \leq \beta$, and thus so does *any* behavior in Bell form. At the same time $\mathbf{c}^\dagger \mathbf{p} \leq \beta + 1$: the classical bound is violated by $\mathbf{p}$ itself. Thus the Bell functional $\mathbf{c}$ witnesses the nonclassicality of $\mathbf{p}$. We note that the Bell polytope $\mathcal{L}$ is the convex hull of the deterministic behaviors, which are its extreme points: switching from the vertex representation to the halfspace representation of the polytope, we see that $\mathcal{L}$ is characterized by a finite set of Bell inequalities. If $\mathbf{c}^\dagger \mathbf{p} \leq \beta$ is true for any point $\mathbf{p}$ in the polytope, then the set $\{\mathbf{p} \in \mathcal{L} : \mathbf{c}^\dagger \mathbf{p} = \beta\}$ is a *face*. Faces of dimension $\dim \mathcal{L} - 1$ are called *facets*, and the corresponding Bell inequalities are called *tight*: any other Bell inequality can be written as a non-negative combination of facet inequalities.

In fact, exactly the same technique works for the more general class of *noncontextuality inequalities*. The simplest example of a contextuality inequality is provided by the Specker triangle [138]. Suppose we have three measurements A, B, C each of which has two outcomes to which we assign valuations in $\{-1, +1\}$. Each pair of measurements can be made jointly, but not all three together. Suppose that these measurements reveal some preexisting properties of the system: $a, b, c \in \{-1, +1\}$, that is, the A measurement reads off whether a is ± 1, and so forth. Clearly, $(ab)(ac)(bc) = a^2 b^2 c^2 = 1$. In particular, this means that the products of each pair cannot all be -1: we cannot have $ab = ac = bc = -1$. Either none of them or two of them can be -1, and so at most two of the outcomes of the A, B, C measurements must disagree. Consequently,

$$\begin{Bmatrix} 1, & \text{if } A \neq B \\ 0, & \text{if } A = B \end{Bmatrix} + \begin{Bmatrix} 1, & \text{if } A \neq C \\ 0, & \text{if } A = C \end{Bmatrix} + \begin{Bmatrix} 1, & \text{if } B \neq C \\ 0, & \text{if } B = C \end{Bmatrix} \leq 2. \tag{5.95}$$

Taking expectations, this implies

$$P(A \neq B) + P(A \neq C) + P(B \neq C) \leq 2, \tag{5.96}$$

and using the fact that $P(A \neq B) = P(A = +1, B = -1) + P(A = -1, B = +1)$ as well as the fact that outcomes are valued in $\{-1, +1\}$ so that $\langle AB \rangle = 1 - 2P(A \neq B)$, we may rewrite the inequality as

$$\langle AB \rangle + \langle BC \rangle + \langle CA \rangle \geq -1. \tag{5.97}$$

Finally, letting

$$\mathbf{p} = \begin{pmatrix} P(A = +1, B = +1) \\ P(A = +1, B = -1) \\ P(A = -1, B = +1) \\ P(A = -1, B = -1) \\ P(A = +1, C = +1) \\ P(A = +1, C = -1) \\ P(A = -1, C = +1) \\ P(A = -1, C = -1) \\ P(B = +1, C = +1) \\ P(B = +1, C = -1) \\ P(B = -1, C = +1) \\ P(B = -1, C = -1) \end{pmatrix} \tag{5.98}$$

and

$$\mathbf{c} = \begin{pmatrix} 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \end{pmatrix}, \tag{5.99}$$

we see that we can write the inequality as

$$\mathbf{c} \cdot \mathbf{p} \geq -1. \tag{5.100}$$

More generally, we may define response functions

$$P(A = a_0, B = b_0 | a, b, c) = \delta_{a_0,a} \delta_{b_0,b} \tag{5.101}$$
$$P(A = a_0, C = c_0 | a, b, c) = \delta_{a_0,a} \delta_{c_0,c} \tag{5.102}$$
$$P(B = b_0, C = c_0 | a, b, c) = \delta_{b_0,b} \delta_{c_0,c}, \tag{5.103}$$

which formalize the idea that the measurements A, B, C simply read off the values of the properties a, b, c, so that given a distribution over their values $P(a, b, c)$, we can write

$$P(A = a_0, B = b_0) = \sum_{abc} P(A = a_0, B = b_0|a, b, c)P(a, b, c) \tag{5.104}$$

$$P(A = a_0, C = c_0) = \sum_{abc} P(A = a_0, C = c_0|a, b, c)P(a, b, c) \tag{5.105}$$

$$P(B = b_0, C = c_0) = \sum_{abc} P(B = b_0, C = c_0|a, b, c)P(a, b, c). \tag{5.106}$$

Gathering up these three distributions into a behavior $\mathbf{p}$, we may then consider the geometry of the allowed behaviors implied by the law of total probability. As before, they form a polytope $\mathcal{L}$ of allowed classical behaviors, and the inequality $\mathbf{c} \cdot \mathbf{p} \geq -1$ corresponds to a facet of this polytope. We could have considered nondeterministic response functions, and more general types of hidden variables, but as long as the pairwise distributions may be obtained by marginalizing over a distribution $P(a, b, c)$, the resulting behavior will lie in the abovementioned polytope whose vertices are precisely states of certainty about which values a, b, c take. Converting to the halfspace representation yields all the noncontextuality inequalities, of the form $\mathbf{c} \cdot \mathbf{p} \leq \beta$.

As we discussed earlier, if we introduce a reference measurement for the system, to say that A, B, C are pairwise jointly measurable is to say that there exist three measurements $\{J^{AB}\}, \{J^{AC}\}, \{J^{BC}\}$ such that

$$P(A, B|\rho) = P(A, B|J^{AB})P(J^{AB}|R)\Phi P(R|\rho) \tag{5.107}$$

$$P(A, C|\rho) = P(A, C|J^{AC})P(J^{AC}|R)\Phi P(R|\rho) \tag{5.108}$$

$$P(B, C|\rho) = P(B, C|J^{BC})P(J^{BC}|R)\Phi P(R|\rho) \tag{5.109}$$

satisfying

$$P(A|\rho) = \sum_i P(A, B_i|\rho) = \sum_i P(A, C_i|\rho) \tag{5.110}$$

$$P(B|\rho) = \sum_i P(A_i, B|\rho) = \sum_i P(B, C_i|\rho) \tag{5.111}$$

$$P(C|\rho) = \sum_i P(A_i, C|\rho) = \sum_i P(B_i, C|\rho). \tag{5.112}$$

Gathering up $P(A, B|\rho), P(A, C|\rho), P(B, C|\rho)$ into a behavior $\mathbf{p}_Q$, we can compare it to the behavior $\mathbf{p}_C$ obtained by setting $\Phi = I$, and relate the violation of any noncontextuality inequality for the scenario to $\|I - \Phi\|$ by the same argument we gave for Bell inequalities.

5.7 Conclusion

In this chapter, we have given a QBist-friendly account of classicality, adapting work on so-called noncontextual ontological models to the QBist framework. We have shown how such models correspond to nonnegative matrix factorizations, or equivalently simplex embeddings, relating such constructions to the existence (or not) of a nonnegative cousin of the Born matrix, emphasizing along the way the QBist idea that classicality involves indifference to whether or not a reference measurement is performed. Finally, we have shown that for QBists, violations of Bell inequalities and the like may be understood as quantifying the cost of the failure to adopt the fundamental nonclassical coherence condition.

CHAPTER 6

CHARACTERIZING QUANTUM STATE SPACE WITH A SINGLE QUANTUM MEASUREMENT

6.1 Introduction

In 1927, Niels Bohr introduced the notion of *complementarity*, that not all aspects of a physical system may be simultaneously definite, as the distinctive feature of the new quantum mechanics [139, 140]. For example, the Heisenberg uncertainty principle, $\sigma_x \sigma_p \geq \frac{1}{2}\hbar$, tells us that if we experiment upon an ensemble of identically prepared particles, then a small variance in the measured position of the particles implies a large variance in their momentum, and vice versa. In particular, regardless of the choice of ensemble, the variance of the two quantities cannot be made arbitrarily small together while remaining consistent with quantum theory. The uncertainty principle may be generalized e.g., to arbitrary pairs of observables, and even to collections of observables [141]. If Bohr was right that complementarity is the defining feature of quantum theory, then it ought to be possible to characterize "quantum states" as nothing other than probability assignments which satisfy appropriate uncertainty relations for all possible observables.

From this point of view, what is essential is not the traditional Hilbert space formalism, but instead the constraints quantum theory urges on probability assignments: indeed, the former can be seen as a convenient mathematical technique for imposing those very constraints. As we have discussed, this is the central contention of QBism [1], a subjective Bayesian interpretation of quantum mechanics, which holds that quantum theory should not be viewed as a description of physical reality but rather as a set of consistency constraints on probability assignments motivated by nature's lack of hidden variables. In particular, QBists side with Schrödinger in viewing the quantum state as nothing more than a "catalogue of expectations" [53]. Indeed, while Bohr and Heisenberg focused on relations between observables like position and momentum, contemporary quantum information theory contemplates a more general class of measurements, so-called *informationally complete* (IC) measurements [73]. Remarkably, assigning appropriate probabilities to the outcomes of a single IC measurement is equivalent to assigning a quantum state. Thus one may take Schrödinger's "catalogue of expectations" a step further, and identify quantum states with probability distributions directly.

The caveat is that while all quantum states correspond to probability distributions, not all probability distributions correspond to quantum states [142]. On the Hilbert space side, such invalid distributions correspond to self adjoint matrices which are not positive

semidefinite, and thus cannot be regarded as density matrices. A picture thereby emerges of quantum state space as a privileged *subset* of the probability simplex, corresponding to just those probability distributions which map back to valid states. The shape of the subset depends on the choice of informationally complete measurement used as a "reference," and can be derived by calculating certain quantities associated with the Hilbert space representation of the measurement. From a foundational perspective, however, one might wonder whether it is possible to characterize the shape of such a privileged subset without reference to Hilbert space as such. Related questions have recently been asked in the context of GPT tomography [143, 144], which provides a theory agnostic approach to reconstructing state and effect spaces from the experimental data collected from a wide range of measurements. Certainly, a state is invalid if it implies a negative probability for some effect. From a QBist point of view, the Born rule is just an example of a general nonclassical coherence constraint, expressed entirely in terms of probabilities with respect to a reference measurement. One may thus rule out a probability distribution if the Born rule, formulated in terms of probabilities, yields a negative number for some measurement outcome. In this way, the shape of state space would emerge out of the demand for probabilistic consistency.

Nevertheless, just as a single IC reference measurement may characterize a quantum system entirely in terms of reference probabilities, one might wonder whether a single specially chosen reference measurement could perform "state space tomography" by appealing only to probabilities assigned to that single measurement. Probabilistic consistency with this one measurement would imply consistency with every possible measurement, and provide an important proof of principle for the QBist approach to quantum mechanics. In this chapter, we demonstrate that this can in fact be done. Taking one's reference measurement to be a *complex projective 3-design* allows the set of valid probability distributions—that is, the shape of quantum state space—to be elegantly characterized by an uncertainty principle[1], precisely in the spirit of Bohr. Crucially, the only building blocks needed to formulate this principle are reference measurement probabilities.

In particular, we will show that probability assignments to the outcomes of the reference measurement cannot be too sharp in a prescribed way: they must satisfy a lower bound on the variance with respect to any of a natural class of observables. This is so even as those same probability assignments may imply a sharp distribution on the outcomes of some alternative measurement via the Born rule. This is a dramatic reversal of the situation classically, where any alternative measurement may be regarded as a coarse graining of a reference measurement e.g., of the positions and momenta of a set of particles. In the classical case, on the one hand, certainty is achievable for the reference; on the other hand, one cannot achieve more certainty about alternative measurements than about the reference. Thus our result underscores in a novel and perspicacious way the degree to which quantum mechanics resists hidden variable interpretations. Moreover, the notion of defining the set of valid probability distributions in terms of an uncertainty principle is a promising approach to constructing generalizations of

[1]We note that whereas the usual uncertainty principle relates probability distributions with respect to different measurements, our generalization is formulated with respect to a single measurement.

quantum theory in the spirit of *generalized probabilistic theories* [4, 145] or in the QBist literature, the *qplex* research program [5]. Independently motivating the features of a 3-design representation, and demonstrating how a Hilbert space representation arises from them (rather than the other way around), would represent a significant advance in the ongoing quest to derive the quantum formalism from satisfying quantum information theoretic principles, and at the same time likely lead to a novel characterization of 3-designs themselves. Indeed, it is to this subject we turn in the next and final chapter. Moreover, we note that 3-designs are of great contemporary interest due to their special role in the theory of classical shadow estimation, where employing a 3-design allows one for example to achieve constant sample complexity in fidelity estimation independent of system size [7, 8, 9, 10]. We hope the present work places these developments in a broader context.

In the spirit of completeness, we begin by reviewing the probability-first formalism for quantum mechanics furnished by an *informationally complete reference measurement*, and then lay out the basic features of complex projective 3-designs. These mathematical preliminaries aside, we derive scalar constraints on probability assignments corresponding to pure states: pure state probability assignments turn out to live in the intersection of 2-norm and 3-norm spheres of specified radii restricted to a natural subspace. Operationally, these constraints can be interpreted as bounds on the agreement between several copies of the reference measurement; conceptually, they can be understood as entropic uncertainty principles [146]. We then establish a vector constraint on pure probability assignments, which draws our attention to a particular 3-index tensor built out of the probabilities which characterize the reference measurement itself. To handle the case of mixed states, we report the key insight that for a 3-design measurement, one can relate the variance of an observable as measured directly to the variance of the observable as measured by the reference. This is what allows quantum state space to be characterized in its entirety by a single uncertainty principle, which lower bounds the variance of any of a natural class of observables. In closing, we observe that the reason 3-designs play such a privileged role is that for just these measurements, the structure coefficients for the Jordan algebra of observables [27] can be expressed solely in terms of the probabilities which characterize the reference measurement. The essence of this observation was already made in [63, 147], but its significance for probabilistic representations of quantum mechanics was left unexplored. In fact, we show that just as a 2-design allows the Born rule to be expressed as a gentle modification of the law of total probability, 3-designs allow the quantum Jordan product on observables to be expressed as a gentle modification of the classical Jordan product, the elementwise product of valuations on reference outcomes.

6.2 The reference measurement formalism

In quantum mechanics, the most general form of a measurement with a finite number of outcomes consists of a set $\{E_i\}_{i=1}^n$ of positive semidefinite matrices called *effects*, acting on a Hilbert space $\mathcal{H}_d$ satisfying $\sum_i E_i = I$. If a set of effects $\{R_i\}$ span the d^2-dimensional operator

space, we call the measurement *informationally complete*: the probabilities $P(R_i|\rho) = \mathrm{tr}(R_i\rho)$ fully characterize the density matrix ρ representing a quantum state, which must itself be positive semidefinite with $\mathrm{tr}(\rho) = 1$. We suppose that upon obtaining an outcome R_i, an agent would assign a corresponding state σ_i to a subsequent reference measurement[2]. We suppose that $\{\sigma_i\}$ also span the d^2-dimensional operator space, and in fact we will always assume that for our reference measurement $\sigma_i = \mathrm{tr}(R_i)^{-1}R_i$, that $\forall i : \mathrm{tr}(R_i)^{-1} = n/d$, and that each σ_i is a pure state. In other words, we specialize to the case of unbiased rank one measurements $R_i = (d/n)|\psi_i\rangle\langle\psi_i|$.

The minimal number of effects in an IC measurement is d^2: moreover, at best an IC-measurement may furnish a linearly independent, but not orthonormal basis [148, 149, 150], and more generally an informationally *over*complete set. We must therefore take up the matter of its dual representation—with a probabilistic twist. Let $|\sigma_i) = \mathrm{vec}(\sigma_i) = (I \otimes \sigma_i)\sum_j |j, j\rangle$ be the vectorization of a reference state σ_i, and similarly let $(R_i| = \sum_i \langle i, i|(I \otimes R_i)$ be the vectorization of a reference effect. In general, $(A|B) = \mathrm{tr}(A^\dagger B)$, and so arranging $(R_i|$ into the rows of a matrix $\mathbf{R}$, and $|\sigma_i)$ into the columns of a matrix $\mathbf{S}$, we can write the conditional probability matrix with elements $P(R_i|R_j)$, for the probability of a subsequent reference outcome given an initial reference outcome, $P \equiv \mathbf{RS}$. By informational completeness, these probabilities fully characterize the reference measurement itself, and they will play a fundamental role in the sequel.

We call a *Born matrix* any matrix Φ which satisfies $P\Phi P = P$, the defining equation of a $\{1\}$-inverse of P [71]. As we have seen, it follows from informational completeness that $P\Phi P = P \iff \mathbf{S}\Phi\mathbf{R} = I$ [75], which provides a resolution of the identity, and thus a dual representation $|\rho) = \mathbf{S}\Phi\mathbf{R}|\rho)$ or,

$$\rho = \sum_{ij} \Phi_{ij} P(R_j|\rho)\sigma_i. \tag{6.1}$$

If the measurement operators (and states) are linearly independent, then P will be invertible, and thus $\Phi = P^{-1}$. Otherwise, there will be a variety of choices for the Born matrix[3], and different assignments of probabilities will lead to the same ascription of a density matrix.

[2] In the language of the previous chapters, we take the reference measurement to be a preparatory measurement.

[3] Recall that the $\{1\}$-inverses of a matrix $P = U\Sigma V^\dagger$ may all be calculated from its singular value decomposition via

$$\Phi = V \begin{pmatrix} \sigma^{-1} & A \\ B & C \end{pmatrix} U^\dagger,$$

where A, B, C are completely arbitrary matrices, and σ is the diagonal matrix of nonzero singular values. A typical example is the Moore-Penrose pseudoinverse, for which $A = B = C = 0$ [71].

Consider now some alternative measurement $\{E_i\}_{i=1}^m$. We can write the Born rule probability $P(E_i|\rho) = \mathrm{tr}(E_i\rho)$ in terms of reference probabilities,

$$\begin{aligned} P(E_i|\rho) = \mathrm{tr}(E_i\rho) &= (E_i|\mathbf{S}\Phi\mathbf{R}|\rho) \\ &= \sum_{jk} P(E_i|R_j)\Phi_{jk}P(R_k|\rho), \end{aligned} \tag{6.2}$$

and appreciate that the Born rule has become a simple *deformation* [133] of the law of total probability $P(E_i) = \sum_j P(E_i|R_j)P(R_j)$.

In particular, taking $\mathbf{E} = \mathbf{R}$, we have that $P(R_i|\rho) = \sum_{jk} P(R_i|R_j)\Phi_{jk}P(R_k|\rho)$. This is a fundamental consistency criterion in an overcomplete probability representation. As we saw in Chapter 3, it follows from the defining equation of a $\{1\}$-inverse that $P\Phi$ is a projector onto $\mathrm{col}(P)$: recall the column space of P (or equivalently, its range) is the span of its columns. Now $P = \mathbf{RS}$ is in fact a full rank factorization of P, and hence the columns of $\mathbf{R}$ form a basis for $\mathrm{col}(P)$ [103]. Thus any vector with components $x_i = \mathrm{tr}(R_i X)$, that is, $x = \mathbf{R}|X)$, must lie in $\mathrm{col}(P)$. Conversely, if $x \in \mathrm{col}(P)$, there must exist a vector $\tilde{x}$ such that $x_i = \sum_j P(R_i|R_j)\tilde{x}_j = \mathrm{tr}\left(R_i \sum_j \tilde{x}_j\sigma_j\right) = \mathrm{tr}(R_i\tilde{X})$ for some operator $\tilde{X}$. By the same token, just as distributions $P(R|\rho)$ must live in $\mathrm{col}(P)$, response functions $P(E|R)$ ought to live in $\mathrm{row}(P)$. Moreover without loss of generality, valuations on reference outcomes may be taken to be vectors in $\mathrm{col}(P)$. Finally, taking $\sigma_i = \mathrm{tr}(R_i)^{-1}R_i = (n/d)R_i$ means that $P = P^\dagger$ and so $\mathrm{col}(P) = \mathrm{row}(P)$.

6.3 Making designs

It follows from the representation theory of the unitary and symmetric groups [119, 151] that the *tth moment of quantum state space*, that is, the tth tensor power of a pure state averaged over all pure states is

$$\int |\psi\rangle\langle\psi|^{\otimes t} d\psi = \binom{d+t-1}{t}^{-1} \Pi_{\mathrm{sym}^t}, \tag{6.3}$$

where $d\psi$ denotes the Haar measure on pure states, and Π_{sym^t} is the projector onto the permutation symmetric subspace on t tensor factors [62]. Note $\mathrm{tr}(\Pi_{\mathrm{sym}^t}) = \binom{d+t-1}{t}$ is just the dimension of that subspace. As Π_{sym^t} can be expressed as a sum over all permutation operators, we have in fact

$$\int |\psi\rangle\langle\psi|^{\otimes t} d\psi = \binom{d+t-1}{t}^{-1} \frac{1}{t!} \sum_{\pi\in S_t} T_\pi \tag{6.4}$$

where $T_\pi = \sum_{a_1,\ldots,a_t} |a_{\pi^{-1}(1)}, \ldots, a_{\pi^{-1}(t)}\rangle\langle a_1, \ldots, a_t|$.

A *quantum state t-design*, also called a *complex projective t-design* [62], is an ensemble of pure states $\{p_i, |\psi_i\rangle\}_{i=1}^n$ which satisfy

$$\sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|^{\otimes t} = \int |\psi\rangle\langle\psi|^{\otimes t} d\psi, \tag{6.5}$$

that is, the average over the design ensemble mimics the average over all pure states up to the t-th moment. We will call a design *unbiased* or *unweighted* if $\forall i : p_i = \frac{1}{n}$. The number of elements in a t-design satisfies

$$n \geq \binom{d-1+\lfloor t/2 \rfloor}{\lfloor t/2 \rfloor} \binom{d-1+\lceil t/2 \rceil}{\lceil t/2 \rceil}.$$

We note that a t-design of any order always exists for sufficiently large n [152], and a t-design is also a $(t-1)$-design. For $t = 1$, we have $\sum_i p_i |\psi_i\rangle\langle\psi_i| = \frac{1}{d} I$ which shows that a 1-design furnishes a set of rank-1 projectors which, rescaled, sum to the identity: thus a 1-design is a quantum measurement. For a 2-design

$$\sum_i p_i |\psi_i\rangle\langle\psi_i|^{\otimes 2} = \frac{1}{d(d+1)} (I \otimes I + T_{21}), \tag{6.6}$$

where T_{21} is the swap operator: two typical examples are symmetric informationally complete (SIC) states and states corresponding to a complete set of mutually unbiased bases (MUBs) [153]. In fact, it follows from Eq. 6.6 that for any unbiased 2-design reference device, we can take $\Phi = (d+1)I - \frac{d}{n}J$, where J is the Hadamard identity, the matrix of all 1's (Appendix 6.8)[4].

The Born rule then takes the profoundly elegant form

$$P(E|\rho) = \sum_i P(E|R_i) \left[(d+1) P(R_i|\rho) - \frac{d}{n} \right], \tag{6.7}$$

which was given an independent motivation in the case that $n = d^2$ in [78], and is related to the fact that 2-designs are optimal for linear quantum state tomography [77, 154]. The terrain of 3-designs is an area of active investigation: several examples were presented in [6], and more general constructions were given in [155, 156]. In particular, 3-designs have been studied for their use in the theory of classical shadows [7, 8, 9] since they lead to well-controlled variance in the estimation of expectation values. In $d = 2$, the complete set of MUBs forms a 3-design as does the union of the tetrahedral SIC and the antipodal SIC formed by Bloch

[4]Moreover, viewing $\mathbf{R}$ as a frame analysis operator and $\tilde{\mathbf{S}} = \mathbf{S}\Phi$ as an frame synthesis operator, we see that the dual analysis operator $\tilde{\mathbf{S}}^\dagger$ has the same range as $\mathbf{R}$ itself. Thus by Lemma 3.6.3, the columns of $\tilde{\mathbf{S}}$ furnish canonical dual elements

sphere inversion. Notably, the set of n-qubit stabilizer states [157] also forms an unbiased 3-design. When $t = 3$, $n \geq \frac{1}{2}d^2(d+1)$, a bound which is not tight [62].

Finally, we note that the *order-t frame potential* $\mathcal{F}_t$ provides a variational characterization of t-designs [62, 114].

Theorem 6.3.1.

$$\mathcal{F}_t[\{p_i, |\psi_i\rangle\}] = \sum_{ij} p_i p_j |\langle\psi_i|\psi_j\rangle|^{2t} \geq \binom{d+t-1}{t}^{-1},$$

with equality if and only if $\{p_i, |\psi_i\rangle\}$ *forms a complex-projective t-design.*

Proof. Theorem 3.6.2, we showed that the frame potential $\mathrm{tr}(F^2) \geq \mathrm{tr}(F)/d$ achieves its lower bound only for a tight frame. Here F is the frame operator. From Eq. (6.3) and Eq. (6.5), we have

$$F_t = \sum_i \sqrt{p_i}|\psi_i\rangle^{\otimes t}\sqrt{p_i}\langle\psi_i|^{\otimes t} = \binom{d+t-1}{t}^{-1}\Pi_{\mathrm{sym}^t}. \tag{6.8}$$

This implies that $\{\sqrt{p_i}|\psi_i\rangle^{\otimes t}\}$ forms a tight frame for the symmetric subspace on which Π_{sym^t} is the identity operator. Applying Theorem 3.6.2 to $\mathrm{tr}(F_t^2)$ gives the result. □

6.4 The shape of quantum state space

6.4.1 Bounding agreement

We begin by providing a characterization of pure state probability distributions in terms of a set of entropic uncertainty principles. They can be motivated by considering the following scenario. Suppose an agent performs t preparatory measurements $\rho_1, \ldots, \rho_t$, and afterwards performs a reference measurement on each of the t systems. What is the probability that all t reference measurements give the same outcome? In other words, we are interested in the agreement probability

$$P(\mathrm{agree}|\rho_1, \ldots, \rho_t) = \sum_{i=1}^{n}\prod_{j=1}^{t} P(R_i|\rho_j) = \mathrm{tr}\left(\sum_{i=1}^{n} R_i^{\otimes t} \otimes_{j=1}^{t} \rho_j\right).$$

Assuming the reference measurement is unbiased and that the reference states are proportional to effects $R_i = \frac{d}{n}\sigma_i$, we have $P(\mathrm{agree}|\rho_1, \ldots, \rho_t) = \frac{d^t}{n^{t-1}}\mathrm{tr}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^{\otimes t} \otimes_{j=1}^{t}\rho_j\right)$. If we further assume that the reference measurement forms a t-design, $\frac{1}{n}\sum_i \sigma_i^{\otimes t} = \binom{d+t-1}{t}^{-1}\Pi_{\mathrm{sym}^t}$, and so

the agreement probability can be written

$$P(\text{agree}|\rho_1,\ldots,\rho_t) = \frac{d^t}{n^{t-1}}\binom{d+t-1}{t}^{-1}\frac{1}{t!}\sum_{\pi\in S_t}\text{tr}(T_\pi\otimes_{j=1}^t\rho_j).$$

To evaluate expressions like this, it suffices to consider traces with cyclic permutations. For $t=2$, the swap operator $T_{21}=\sum_{ab}|b,a\rangle\langle a,b|$ yields

$$\text{tr}\left((X\otimes Y)\sum_{ab}|b,a\rangle\langle a,b|\right)=\sum_{ab}\langle a|X|b\rangle\langle b|Y|a\rangle=\text{tr}(XY),$$

while similarly, for $t=3$, a cyclic permutation of three elements delivers

$$\text{tr}\left((X\otimes Y\otimes Z)\sum_{abc}|b,c,a\rangle\langle a,b,c|\right)=\sum_{abc}\langle a|X|b\rangle\langle b|Y|c\rangle\langle c|Z|a\rangle=\text{tr}(XYZ).$$

In light of this, we have for the order-2 agreement probability,

$$P(\text{agree}|\rho_1,\rho_2)=\frac{1}{d+1}\left(\frac{d}{n}\right)\Big[\text{tr}(\rho_1)\text{tr}(\rho_2)+\text{tr}(\rho_1\rho_2)\Big]. \tag{6.9}$$

Now $\text{tr}(\rho_1\rho_2)\leq\sqrt{\text{tr}(\rho_1^2)\text{tr}(\rho_2^2)}$, while the purity satisfies $\text{tr}(\rho^2)\leq 1$. Thus when $\rho_1=\rho_2=\rho$ pure, we saturate the upper bound of this quantity. For the lower bound, we note $\text{tr}(\rho_1\rho_2)\geq 0$ with equality if and only if ρ_1,ρ_2 are orthogonal, which leads to bounds

$$\left(\frac{d}{n}\right)\frac{1}{d+1}\leq\sum_i P(R_i|\rho_1)P(R_i|\rho_2)\leq\left(\frac{d}{n}\right)\frac{2}{d+1}. \tag{6.10}$$

Similarly, for $t=3$, we find

$$\begin{aligned}P(\text{agree}|\rho_1,\rho_2,\rho_3)=\frac{1}{(d+1)(d+2)}\left(\frac{d}{n}\right)^2\Big[&\text{tr}(\rho_1)\text{tr}(\rho_2)\text{tr}(\rho_3)+\text{tr}(\rho_1)\text{tr}(\rho_2\rho_3)\\&+\text{tr}(\rho_2)\text{tr}(\rho_1\rho_3)+\text{tr}(\rho_3)\text{tr}(\rho_1\rho_2)+\text{tr}(\rho_1\rho_2\rho_3)+\text{tr}(\rho_1\rho_3\rho_2)\Big],\end{aligned}$$

which is maximized when $\rho_1=\rho_2=\rho_3=\rho$ pure, so that $\text{tr}(\rho^2)=\text{tr}(\rho^3)=1$, and minimized when ρ_1,ρ_2,ρ_3 are mutually orthogonal (when the dimension permits), delivering bounds

$$\left(\frac{d}{n}\right)^2\frac{1}{(d+1)(d+2)}\leq\sum_i P(R_i|\rho_1)P(R_i|\rho_2)P(R_i|\rho_3)\leq\left(\frac{d}{n}\right)^2\frac{6}{(d+1)(d+2)}. \tag{6.11}$$

Since in both cases, the upper bounds are saturated by identical pure states, we conclude that pure state probability assignments with respect to a 3-design lie in the nonnegative orthant, in the intersection of three kinds of spheres. From $\sum_i P(R_i|\rho)$, they live on a 1-norm sphere of radius 1; from $\sum_i P(R_i|\rho)^2$, they live on a 2-norm sphere of radius $\sqrt{\left(\frac{d}{n}\right)\frac{2}{d+1}}$; and from $\sum_i P(R_i|\rho)^3$, they live on a 3-norm sphere of radius $\sqrt[3]{\left(\frac{d}{n}\right)^2\frac{6}{(d+1)(d+2)}}$. Finally, since we derived all probabilities from the trace of a reference effect on a state, our probability distributions live in $\mathrm{col}(P)$.

We could continue on, contemplating the t-fold agreement probability for an unbiased t-design reference device for any t. Since a pure state satisfies $\forall t : \mathrm{tr}(\rho^t) = 1$, we would find that pure probability vectors live on t-norm spheres with fixed radii determined by the agreement probability

$$P(\text{agree}|\rho^{\otimes t}) = \sum_i P(R_i|\rho)^t = \frac{d^t}{n^{t-1}}\binom{d+t-1}{t}^{-1}\frac{1}{t!}\sum_{\pi\in S_t}\mathrm{tr}(T_\pi\rho^{\otimes t}) = \frac{d^t}{n^{t-1}}\binom{d+t-1}{t}^{-1}, \tag{6.12}$$

where the last follows from the fact that since $\forall t : \mathrm{tr}(\rho^t) = 1$, all of the $t!$ terms in the sum will be 1.

The following lemma [158, 159], however, assures us that a 3-design is all we need.

Lemma 6.4.1. *A Hermitian operator A is a rank-1 projector if and only if $tr(A^2) = tr(A^3) = 1$.*

Proof. Let $\{\lambda_i\}$ be the eigenvalues of A. $\mathrm{tr}(A^2) = \mathrm{tr}(A^3) = 1$ means that $\sum_i \lambda_i^2 = \sum_i \lambda_i^3 = 1$. On the one hand, $\sum_i \lambda_i^2 = 1$ implies that $\forall i : -1 \leq \lambda_i \leq 1$. On the other hand, $\sum_i \lambda_i^3 \leq \sum_i \lambda_i^2$ with equality if and only if $\forall i : \lambda_i \in \{0, 1\}$. But since the whole sum must be 1, we must have exactly one $\lambda_i = 1$ and the rest 0. Thus A is a rank-1 projector, or equivalently a pure state ρ. □

In light of this, as long as $t \geq 3$, we can fully characterize the pure states of quantum theory with respect to an unbiased t-design reference device by $\forall i : P(R_i|\rho) \geq 0$ and

$$\sum_i P(R_i|\rho) = 1 \tag{6.13}$$

$$\sum_i P(R_i|\rho)^2 = \left(\frac{d}{n}\right)\frac{2}{d+1} \tag{6.14}$$

$$\sum_i P(R_i|\rho)^3 = \left(\frac{d}{n}\right)^2\frac{6}{(d+1)(d+2)}, \tag{6.15}$$

along with $P(R_i|\rho) = \sum_{jk} P(R_i|R_j)\Phi_{jk}P(R_k|\rho)$, that is, the probability distribution lives in $\mathrm{col}(P)$[5].

Thus in a sense we have already achieved our goal since quantum state space is the convex hull of all pure state probability assignments. The bounds on agreement probabilities, from which the constraints on pure states were derived, already suggest a kind of uncertainty principle: after all, classically, there is nothing in principle preventing such an agreement probability from being 1 when $\rho_1 = \cdots = \rho_t$, or 0 for perfectly distinguishable states. In fact, the upper bounds may be understood as implying a set of *entropic* uncertainty principles. Defining the order-t Rényi entropy of a probability-distribution $\{P(R_i|\rho)\}$ as

$$H_t\Big(\{P(R_i|\rho)\}\Big) = \frac{1}{1-t}\log\left(\sum_i P(R_i|\rho)^t\right) \tag{6.16}$$

for $0 < t < \infty$, it is clear from the above discussion that pure state probability distributions achieve the lower bound on the Rényi entropies of order $t = 2, 3$ over all states, and this, along with the restriction to $\mathrm{col}(P)$, is enough to characterize them completely.

6.4.2 *The contour of idempotents*

As an alternative approach, we can derive a single equation picking out pure state probability assignments by appealing to the fact that, for a normalized state, $\rho = \rho^2$ if and only if $\rho = \rho^\dagger$ is pure, that is, a rank one projector. Substituting the resolution of the identity $\rho = \sum_{ij}\Phi_{ij}P(R_j|\rho)\sigma_i$ into $P(R_i|\rho^2) = \mathrm{tr}(R_i\rho^2)$, we find that

$$P(R_i|\rho) = \sum_{lm} P(R_l|\rho)P(R_m|\rho)\sum_{jk}\Phi_{jl}\Phi_{km}\Re\big[\mathrm{tr}(R_i\sigma_j\sigma_k)\big].$$

We need only consider the real part since $\mathrm{tr}(R_i\sigma_k\sigma_j) = \mathrm{tr}\big(R_i(\sigma_j\sigma_k)^\dagger\big) = \mathrm{tr}\big(R_i(\sigma_j^*\sigma_k^*)^T\big) = \mathrm{tr}\big(R_i^T\sigma_j^*\sigma_k^*\big) = \mathrm{tr}\big(R_i^*\sigma_j^*\sigma_k^*\big) = \mathrm{tr}(R_i\sigma_j\sigma_k)^*$ as R_i, σ_j, σ_k are all positive semidefinite and so every term in Eq. 6.17 is added to its complex conjugate.

Let $\mathcal{M}_t = \int |\psi\rangle\langle\psi|^{\otimes t}d\psi$ be the tth moment of quantum state space. This is itself a valid state, and so we can consider its probability distribution $P(R_i, R_j, R_k, \ldots|\mathcal{M}_t)$ with respect to t copies of the reference measurement. If we assume an unbiased set of effects, by

[5]In fact, we can show more specifically that a probability distribution corresponds to a pure state if and only if its component that lies in $\mathrm{col}(P)$ satisfies the quadratic and cubic constraints. To see this, let $\rho = \sum_{ij}\Phi_{ij}P(R_j|\rho)\sigma_i$. Then $\mathrm{tr}(\rho^2) = \frac{n}{d}\sum_{ij}\Phi_{ij}P(R_j|\rho)\sum_{kl}P(R_i|R_k)\Phi_{kl}P(R_l|\rho)$. Decomposing $P(R|\rho) = x + y$, for $x \in \mathrm{col}(P)$ and $y \in \mathrm{col}(P)^\perp$, and using the form of Φ for an unbiased 2-design, we find $\mathrm{tr}(\rho^2) = (d+1)\left(\frac{n}{d}\right)\sum_i x_i^2 - 1$. Thus if and only if $\sum_i x_i^2 = \left(\frac{d}{n}\right)\frac{2}{d+1}$ does $\mathrm{tr}(\rho^2) = 1$. The result is analogous for $\mathrm{tr}(\rho^3) = 1$, using the expression for $\Re\big[\mathrm{tr}(R_i\sigma_j\sigma_k)\big]$ derived in section 6.4.2.

the same argument as in Section 6.4.1, it follows that

$$P(R_i|\mathcal{M}_1) = \frac{1}{n} \tag{6.17}$$

$$P(R_i, R_j|\mathcal{M}_2) = \frac{1}{d+1}\left(\frac{1}{n}\right)\left[\frac{d}{n} + P(R_i|R_j)\right] \tag{6.18}$$

$$P(R_i, R_j, R_k|\mathcal{M}_3) = \frac{1}{(d+1)(d+2)}\left(\frac{d}{n^2}\right)\times \tag{6.19}$$

$$\left[\frac{d}{n} + P(R_j|R_k) + P(R_i|R_j) + P(R_i|R_k) + 2\Re\left[\mathrm{tr}(R_i\sigma_j\sigma_k)\right]\right].$$

If we further assume that the reference states form a 3-design then $\mathcal{M}_3 = \frac{1}{n}\sum_i \sigma_i^{\otimes 3}$ and so $P(R_i, R_j, R_k|\mathcal{M}_3) = \frac{1}{n}\sum_m P(R_i|R_m)P(R_j|R_m)P(R_k|R_m)$. Equating these two expressions allows us to calculate $\Re\left[\mathrm{tr}(R_i\sigma_j\sigma_k)\right]$ directly from the conditional probability matrix $P(R_i|R_j)$ which characterizes the reference measurement itself,

$$\Re\left[\mathrm{tr}(R_i\sigma_j\sigma_k)\right] = \frac{1}{2}\Bigg[(d+1)(d+2)\left(\frac{n}{d}\right)\sum_m P(R_i|R_m)P(R_j|R_m)P(R_k|R_m)$$
$$- P(R_j|R_k) - P(R_i|R_j) - P(R_i|R_k) - \frac{d}{n}\Bigg].$$

Then Eq. 6.17, which expresses $\rho = \rho^2$ in terms of probability-assignments, simplifies to

$$P(R_i|\rho) = \frac{1}{2}\left[\frac{1}{2}(d+1)(d+2)\left(\frac{n}{d}\right)\sum_m P(R_i|R_m)P(R_m|\rho)^2 - \frac{d}{n}\right],$$

which depends only upon $P(R_i|R_j)$ and $P(R_i|\rho)$. A probability distribution satisfying Eq. 6.20 is clearly in $\mathrm{col}(P)$, and it is straightforward to check that Eq. 6.20 implies the scalar constraints proved in the previous section.

Finally, we note that Eq. 6.19 implies a state assignment $\mathcal{M}_3$ allows one to extract $\Re[\mathrm{tr}(E_i\sigma_j\sigma_k)]$ (where $\sigma_i = \mathrm{tr}(E_i)^{-1}E_i$) from the joint probability distribution $P(E_i, E_j, E_k|\mathcal{M}_3)$ for *any* measurement $\{E_i\}$. One may compare this method to the procedure described in [160], which exploits the backaction on a control qubit after a controlled cyclic permutation to extract the real part of a trace of an n-product of states from an expectation value. In particular, having estimated $\Re[\mathrm{tr}(E_i\sigma_j\sigma_k)]$ for any informationally complete measurement, we may characterize pure state probability distributions with respect to that measurement according to Eq. 6.17. What makes a 3-design distinctive is that $\Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]$

can be calculated from $P(R_i|R_j)$ alone, and so pure state probability distributions can be characterized by appealing to probabilities assigned to one single measurement.

6.4.3 A variance-based uncertainty principle

We now give a condition for the validity of *any* probability assignment, pure or mixed. For a 3-design measurement, the variance of an observable as measured by a standard von Neumann measurement can be directly related to the variance of the same observable as estimated by the reference measurement. From this consideration, we can characterize the validity of any probability distribution in terms of a lower bound on the variance of all observables of a natural class, that is, those that live in $\mathrm{col}(P)$.

We begin by noting that a Hermitian matrix ρ is positive semidefinite if and only if its second moment with respect to all Hermitian observables X is nonnegative:

$$\forall X : \mathrm{tr}(X^2\rho) \geq 0 \Longleftrightarrow \rho \geq 0. \tag{6.20}$$

This follows from the fact that $X^2 \geq 0$ and the fact that the cone of positive semidefinite matrices is self-dual [27]. More simply, one can observe that $\rho \geq 0$ is equivalent to $\forall\psi : \langle\psi|\rho|\psi\rangle = \mathrm{tr}(|\psi\rangle\langle\psi|\rho) \geq 0$, and any X^2 can be decomposed into a sum of rank-1 projectors weighted by nonnegative numbers via the spectral decomposition.

Substituting $X = \sum x_i R_i$ and $\rho = \sum_{ij} \Phi_{ij} P(R_j|\rho)\sigma_i$, we find that

$$\forall X : \mathrm{tr}(X^2\rho) = \left(\frac{d}{n}\right) \sum_{ijkl} x_i x_j \Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]\Phi_{kl}P(R_l|\rho) \geq 0. \tag{6.21}$$

If we assume that $x \in \mathrm{col}(P)$, and exploit the expression for $\Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]$ in terms of $P(R_i|R_j)$, we may simplify the expression for $\mathrm{tr}(X^2\rho)$.

Lemma 6.4.2. *For an unbiased complex-projective 3-design reference measurement,*

$$tr(X^2\rho) = \frac{1}{2}\left(\frac{d+2}{d+1}\right)\left[\langle x^2\rangle_\rho - \frac{d}{d+2}\Big(\langle x^2\rangle_\mu - 2\langle x\rangle_\mu\langle x\rangle_\rho\Big)\right], \tag{6.22}$$

where e.g. $\langle x^2\rangle_\rho = \sum_i x_i^2 P(R_i|\rho)$ *and* $\forall i : P(R_i|\mu) = \frac{1}{n}$ *are the probabilities for the maximally mixed state.*

Proof. Let x be a valuation on reference outcomes, that is, an assignment of real numerical values to the outcomes of the reference measurement. This is equivalent to the assignment of a self-adjoint operator X since $\langle x\rangle_\rho = \sum_i x_i P(R_i|\rho) = \mathrm{tr}\left(\sum_i x_i R_i\rho\right) = \mathrm{tr}(X\rho) = \langle X\rangle_\rho^{(\mathrm{vn})}$, although in an overcomplete representation different choices of x will yield the same operator X. In other words, any real valuation x determines a self-adjoint operator $X = \sum_i x_i R_i$, though the representation is not unique. The variance with respect to a standard

von Neumann measurement of X is $\mathrm{Var}[X]_\rho^{(\mathrm{vn})} = \mathrm{tr}(X^2\rho) - \mathrm{tr}(X\rho)^2$, where $\mathrm{tr}(X^2\rho) = \left(\frac{d}{n}\right)\sum_{ijkl} x_i x_j \Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]\Phi_{kl}P(R_l|\rho)$. Substituting in the expression for $\Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]$ yields

$$\mathrm{tr}(X^2\rho) = \frac{1}{2}\Bigg[(d+1)(d+2)\sum_k\left(\sum_i P(R_k|R_i)x_i\right)^2 P(R_k|\rho) - \left(\frac{d}{n}\right)\sum_{ij} x_i P(R_i|R_j)x_j - 2d\langle x\rangle_\mu\langle x\rangle_\rho - d^2\langle x\rangle_\mu^2\Bigg]. \tag{6.23}$$

We assume that $x \in \mathrm{col}(P)$ so that $x_i = \mathrm{tr}(R_i\tilde{X})$ for some Hermitian $\tilde{X}$. By the 2-design property then,

$$\sum_j P(R_i|R_j)x_j = \sum_j P(R_i|R_j)\mathrm{tr}(R_j\tilde{X}) = d\mathrm{tr}\left(\frac{1}{n}\sum_j \sigma_j^{\otimes 2}(R_i\otimes\tilde{X})\right) \tag{6.24}$$

$$= \frac{1}{d+1}\mathrm{tr}\big((I+T_{21})(R_i\otimes\tilde{X})\big) = \frac{1}{d+1}\left(\frac{d}{n}\mathrm{tr}(\tilde{X}) + \mathrm{tr}(R_i\tilde{X})\right), \tag{6.25}$$

where $\mathrm{tr}(\tilde{X}) = \sum_i \mathrm{tr}(R_i\tilde{X}) = \sum_i x_i = n\langle x\rangle_\mu$, so that

$$\sum_j P(R_i|R_j)x_j = \frac{d\langle x\rangle_\mu + x_i}{d+1} \tag{6.26}$$

$$\left(\sum_j P(R_i|R_j)x_j\right)^2 = \frac{d^2\langle x\rangle_\mu^2 + 2d\langle x\rangle_\mu x_i + x_i^2}{(d+1)^2} \tag{6.27}$$

$$\sum_{ij} x_i P(R_i|R_j)x_j = \frac{n\big(d\langle x\rangle_\mu^2 + \langle x^2\rangle_\mu\big)}{d+1}. \tag{6.28}$$

Substituting these expressions into Eq. 6.23 yields Eq. 6.22. □

As we have seen, the restriction that $x \in \mathrm{col}(P)$ is natural since already any Hermitian observable X can be expressed with respect to the reference device in this form. Remarkably, this expression relates the second moment of X with respect to a standard von Neumann measurement, whose outcomes are eigenvalues of X, to the second moment of X *with respect to the reference measurement*, where now the x_i's are interpreted as numerical values assigned to reference measurement outcomes. Only for a 3-design is such a simple relationship possible, and this is what will allow us to characterize valid probability distributions in terms of a lower bound on the variance with respect to the reference measurement.

The restriction $x \in \mathrm{col}(P)$ (for real x) is equivalent to the assumption that $x_i = \mathrm{tr}(R_i \tilde{X})$ for some Hermitian $\tilde{X}$. By the 2-design property, if $X = \sum_i x_i R_i$ for $x_i = \mathrm{tr}(R_i \tilde{X})$, then $\tilde{X} = \left(\frac{n}{d}\right)\left[(d+1)X - \mathrm{tr}(X)I\right]$. Consequently, if the RHS of Eq. 6.22 is nonnegative for some real $x \in \mathrm{col}(P)$, then $\mathrm{tr}(X^2\rho) \geq 0$ for $X = \frac{1}{d+1}\left(\frac{d}{n}\right)\left[\tilde{X} + \mathrm{tr}(\tilde{X})I\right]$; and if the RHS is nonnegative for *all* real $x \in \mathrm{col}(P)$, $\mathrm{tr}(X^2\rho) \geq 0$ for all Hermitian X. We conclude that probability assignments $P(R_i|\rho)$ are valid if and only if

$$\forall x \in \mathrm{col}(P) : \mathrm{Var}[x]_\rho \geq \frac{d}{d+2}\Big(\langle x^2\rangle_\mu - 2\langle x\rangle_\mu \langle x\rangle_\rho\Big) - \langle x\rangle_\rho^2, \tag{6.29}$$

where $\mathrm{Var}[x]_\rho = \sum_i x_i^2 P(R_i|\rho) - \left(\sum_i x_i P(R_i|\rho)\right)^2$. In this way, the shape of quantum state space can be understood in terms of a variance based uncertainty principle: valid probability assignments on reference outcomes cannot be too sharp lest they violate a lower bound on the variance for any observable in $\mathrm{col}(P)$. We note that a related inequality was derived recently in [10] in the context of bounding the variance of expectation values in shadow estimation.

6.5 The Jordan product

We can shed further light on the special role that 3-designs play in encoding the shape of quantum state space by considering their relationship to the *Jordan algebra* of observables. To see this, let us return to the inequality in Eq. 6.21, which we may reshape into $\forall x : \sum_i x_i [\mathcal{L}_\rho]_{ij} x_j \geq 0$, where

$$[\mathcal{L}_\rho]_{ij} = \sum_{kl} \Re[\mathrm{tr}(R_i \sigma_j \sigma_k)] \Phi_{kl} P(R_l|\rho). \tag{6.30}$$

From this we conclude that probability assignments $P(R_i|\rho)$ are valid if and only if $\mathcal{L}_\rho$ is positive semidefinite. In particular, for an unbiased 3-design reference device, Eq. 6.30 simplifies to

$$[\mathcal{L}_\rho]_{ij} = \frac{1}{2}\Bigg[(d+1)(d+2)\left(\frac{n}{d}\right)\sum_m P(R_m|R_i)P(R_m|R_j)P(R_m|\rho) - P(R_i|R_j) - P(R_i|\rho) - P(R_j|\rho) - \frac{d}{n}\Bigg],$$

which has the virtue of depending only upon reference measurement probabilities, and which provides a straightforward way to check whether the variance bound is satisfied.

But there is another interpretation of the operator $\mathcal{L}_\rho$. Recall that under the Jordan product $A \odot B = \frac{1}{2}(AB + BA)$, Hermitian matrices over $\mathbb{C}$ form a *Euclidean Jordan algebra*

[27, 24, 22]. A Jordan algebra is a nonassociative algebra which satisfies commutativity and the Jordan identity,

$$A \odot B = B \odot A$$
$$A^2 \odot (B \odot A) = A \odot (B \odot A^2).$$

If we define L_A to be the linear operator which takes the Jordan product with A, that is, $L_A(B) = A \odot B$, the Jordan identity is equivalent to $\left[L_A, L_{A^2}\right] = 0$. A *Euclidean* Jordan algebra enjoys the additional property that there exists an inner product on the underlying vector space $\mathcal{V}$ such that $\forall A, B, C \in \mathcal{V} : \langle L_A(B), C\rangle = \langle B, L_A(C)\rangle$.

By introducing an informationally complete reference measurement, we identify quantum states with probability distributions. We may then ask: how can we represent the Jordan product in terms of probabilities? Treating states ρ and τ as observables of the Jordan algebra, if $|\rho \odot \tau) = L_\rho|\tau)$, using the resolution of the identity $\mathbf{S\Phi R} = I$, we have

$$\mathbf{R}L_\rho|\tau) = \mathbf{R}L_\rho\mathbf{S\Phi R}|\tau) = \mathcal{L}_\rho \Phi P(R|\tau), \tag{6.31}$$

where $\mathcal{L}_\rho = \mathbf{R}L_\rho\mathbf{S}$, whose matrix elements are

$$\begin{aligned}[\mathcal{L}_\rho]_{ij} &= \mathrm{tr}(R_i L_\rho(\sigma_j)) = \frac{1}{2}\big(\mathrm{tr}(R_i\rho\sigma_j) + \mathrm{tr}(R_i\sigma_j\rho)\big) \\ &= \sum_{kl} \Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]\Phi_{kl}P(R_l|\rho).\end{aligned} \tag{6.32}$$

Indeed, this is precisely the matrix we developed earlier, whose positive semidefiniteness diagnoses the validity of probability assignments $P(R_i|\rho)$.

Significantly, Eq. 6.32 reveals that the three-index tensor $\Re[\mathrm{tr}(R_i\sigma_j\sigma_k)]$ encodes the structure coefficients for the Jordan product on $d \times d$ Hermitian matrices over $\mathbb{C}$, and thus fully defines it by its action on the reference states and effects. Since pure states are idempotents of the Jordan algebra with trace 1, the structure coefficients implicitly determine the geometry of the state space. At the same time, we have shown that the components of this tensor can be extracted from the joint probability distribution $P(R_i, R_j, R_k|\mathcal{M}_3)$. Finally, taking our reference measurement to be an unbiased 3-design means that $P(R_i, R_j, R_k|\mathcal{M}_3) = \frac{1}{n}\sum_m P(R_i|R_m)P(R_j|R_m)P(R_k|R_m)$ so that $P(R_i|R_j)$ *alone* is sufficient to characterize the Jordan product, and through this algebraic structure, the entire state space.

6.6 A gentle modification

As we have seen for a 2-design, the Born rule appears as a remarkably gentle modification of the classical law of total probability,

$$P(E|\rho) = \sum_i P(E|R_i)\Big\{(d+1)P(R_i|\rho) - \frac{d}{n}\Big\}. \tag{6.33}$$

We will now see that our results imply that for a 3-design, the quantum Jordan product appears as a similarly gentle modification of the classical Jordan product. In the previous section, we worked out the formula for the quantum Jordan product on probability vectors: expressing it instead directly in terms of valuations on the reference measurement makes the correspondence with the classical rule most manifest.

But first let us remind ourselves what the classical Jordan product is in this case. As we discussed in Chapter 2, classically, real valued random variables, which are valuations $\Omega \to \mathbb{R}$ on a sample space Ω, form an algebra. Given valuations $x(\omega)$ and $y(\omega)$, we may form their product $(x \circ y)(\omega) = x(\omega)y(\omega)$. Treating x and y as vectors, this amounts to the elementwise product $x \circ y$. Quantum mechanically, the role of the classical sample space is played by the outcomes of the reference measurement, and if x and y are two valuations on reference outcomes, we may form their classical product $x \circ y$: in fact, since $\circ$ is commutative and associative, this is nothing other than the Jordan product on $\mathbb{R} \oplus \cdots \oplus \mathbb{R} \simeq \mathbb{R}^n$. As we have seen, valuations on the reference measurement are equivalent to the assignment of a quantum mechanical observable. But of course, quantum mechanically there is another way of multiplying observables: the Jordan product $X \odot Y = (XY + YX)/2$. Expressing this latter directly in terms of valuations yields the classical formula with a gentle quantum correction.

Theorem 6.6.1. *For $x, y \in col(P)$, the quantum Jordan product $X \odot Y = (XY + YX)/2$ for $X = \sum_i x_i R_i$ and $Y = \sum_i y_i R_i$ may be expressed with respect to an unbiased 3-design reference measurement as*

$$x \odot y = \gamma(x \circ y) + (1-\gamma)\Big(\overline{x}y + \overline{y}x - (x \cdot y)u\Big)/n, \tag{6.34}$$

where $x \circ y$ is the Hadamard or elementwise vector product, $\overline{x}$ denotes the sum of the vector x, and $\gamma = \frac{1}{2}\frac{d+2}{d+1}$.

Proof. See appendix 6.8. □

Remark. *This formula is valid for $x, y \in col(P)$: $x \odot y$ may not be in $col(P)$, and hence ought to be projected into it. Even if $x \odot y \notin col(P)$, however, since $P(R|\rho) \in col(P)$, we have $\langle X \odot Y\rangle = \sum_i (x \odot y)_i P(R_i|\rho)$, that is, the formula properly reproduces the expectation value since contracting with the probability vector will kill any components in $col(P)^\perp$.*

Remark. *Continuing to count, a* 4*-design will make the symmetric product of three operators look as close as possible to* $x \circ y \circ z$*, and so on. But for characterizing quantum mechanics, counting to 3 is enough.*

Remark. *In appendix 6.9, we show how unitary maps may be expressed in terms of this formula.*

6.7 Conclusion

We have thus shown that the shape of quantum state space can be understood in terms of an uncertainty principle which constrains the probabilities one ought to assign to the outcomes of a 3-design reference measurement. Compatibility with this uncertainty principle can be diagnosed through the positive semidefiniteness of a particular operator $\mathcal{L}_\rho$ constructed from reference probabilities. We have also provided a set of scalar constraints that pick out pure state probability distributions, which can alternatively be summarized by a single vector constraint. Conceptually, these constraints can be understood as entropic uncertainty principles, and operationally they relate to the agreement probability on multiple copies of the reference measurement.

Crucially, each term that appears in our equations is grounded in a probability assignment, and even better, these probabilities refer to the behavior of a single reference measurement. The possibility of achieving this rests on the delicate interplay between unitary symmetry and the Jordan algebra of observables. The algebraic structure of quantum theory implies that the 3rd moment of quantum state space determines them all, and so does a reference measurement furnished by a 3-design. This further vindicates the centrality of 3-designs already suggested by their optimality in classical shadow estimation tasks.

Our result holds particular significance for the QBist research program in the foundations of quantum mechanics. As we have explained, QBism argues that quantum theory is best understood not as a description of physical reality, but rather as a set of normative guidelines for gambling on the consequences of one's actions in a world undergoing ceaseless creation [1, 2, 89]. Consequently, QBist "reconstructions" of quantum mechanics proceed [57, 5] by motivating the constraints on probability assignments implied by quantum theory in the same spirit in which de Finetti derived the usual rules of probability theory by contemplating what constraints a gambler ought to place on their different bets in order to prevent a sure loss.

The simplicity of our result is very much in the spirit of [24], which suggests that the constraints implied by quantum theory are in some sense the "most symmetrical" compatible with what the author calls "vitality," i.e. the nonexistence of a hidden variable model. Indeed, our result shows that these constraints may be understood as a fundamental expression of complementarity. That said, our derivation presumes a prior knowledge of traditional quantum theory: the significance of our work here is that it exposes the structure that must be aimed for in any future reconstructive effort, and it is to the latter that we turn in Chapter 7.

Here we observe that previous efforts at QBist reconstruction [5] took as their starting place the generalization of the image of quantum state space within the probability simplex

induced by a symmetric informationally complete (SIC) reference measurement. SIC measurements have a host of virtues: the corresponding states form a simplex in quantum state space whose vertices are pure states; the conjecture of their existence in any Hilbert space dimension has led to a fruitful and unexpected interplay between physics and algebraic number theory [5]. SIC states, however, only form 2-designs, and thus the Jordan structure coefficients cannot be extracted directly from probability assignments assigned to a single reference device. Breaking this barrier is the central innovation of the present work.

More specifically, the SIC based reconstructive effort began from the observation that the 2-design condition implies the inner product between any two probability vectors must lie between certain upper and lower bounds. Inspired by this, [5] defined a *qplex* to be set of probability vectors which mutually satisfy these bounds, to which no more elements can be added without inconsistency, and which contains a simplex of pure states (corresponding to a SIC). The goal of the program was to motivate these bounds on independent grounds, situate quantum theory in the vaster landscape of qplexes, and provide a principle by which quantum theory could be identified within this landscape. This approach is consonant with developments in quantum foundations over the last 25 years where the study of so-called *generalized probabilistic theories* [4, 145] has played a central role in providing new perspectives on quantum theory.

The authors of [5] demonstrated that any qplex whose symmetry group is a stochastic subgroup of the orthogonal group isomorphic to the unitary group must correspond to quantum theory, and vice versa: moreover, the existence of such a subgroup is equivalent to the existence of a particular SIC. At the same time, the authors left open the possibility of there being a simpler principle which could pick out quantum theory among the qplexes. The present work shows that demanding the fundamental reference measurement to have the properties of a 3-design, rather than a 2-design, means that a single finite set of probability distributions is sufficient to characterize the theory, which is much more tractable, analytically and computationally, as well as more conceptually satisfying.

Just as a qplex generalizes the representation of quantum mechanics according to a SIC, it is natural to consider analogous generalizations of a 3-design representation, what we might call *3-qplexes*. For instance, one could explore the landscape of all state spaces defined by an uncertainty principle as in Eq. 6.29, without at first restricting $P(R_i|R_j)$, the probabilities that characterize the reference measurement, to correspond to an actual quantum 3-design. Which properties of quantum theory are preserved in such generalized theories, and which fall by the wayside? Just as identifying quantum theory among the qplexes led to an alternative characterization of SICs themselves, picking out quantum theory among the 3-qplexes would lead to an alternative characterization of 3-designs. A new approach to identifying and constructing 3-designs, in particular, of self-testing them, would have significant practical application in quantum computing and beyond—it is to this subject that we now turn.

6.8 Appendix: The first three moments

For reference, we first prove in one place some basic properties of the first three moments of an unbiased quantum 3-design. In general, for an unbiased t-design $\{\sigma_i\}$, we have

$$\mathcal{M}_t = \frac{1}{n}\sum_i \sigma_i^{\otimes t} = \int |\psi\rangle\langle\psi|^{\otimes t} d\psi = \binom{d+t-1}{t}^{-1} \Pi_{\mathrm{sym}^t} = \binom{d+t-1}{t}^{-1} \frac{1}{t!}\sum_{\pi\in S_t} T_\pi.$$

We will again and again use the generalized swap trick. Let T_π be the unitary operator which performs a cyclic permutation of m subsystems. Then

$$\mathrm{tr}\big(T_\pi(A_1\otimes\cdots\otimes A_m)\big) = \mathrm{tr}(A_1\ldots A_m). \tag{6.35}$$

We now examine the implications for unbiased t-designs that follow from the structure of the moment operators $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$.

1. $\mathcal{M}_1 = \frac{1}{d}I$. For an unbiased 1-design, therefore, $\frac{1}{n}\sum_i \sigma_i = \frac{1}{d}I$. Letting $R_i = \frac{d}{n}\sigma_i$ yields $\sum_i R_i = I$, from we conclude that $\{R_i\}$ may be viewed as POVM elements corresponding to a measurement.

2. $\mathcal{M}_2 = \frac{1}{d(d+1)}(I + T_{21})$. This yields

$$\mathrm{tr}(\mathcal{M}_2(X\otimes I)) = \frac{1}{n}\sum_i \mathrm{tr}(\sigma_i X)\sigma_i = \frac{1}{d(d+1)}\Big(\mathrm{tr}(X)I + X\Big), \tag{6.36}$$

   which we may rearrange into a resolution of the identity,

$$X = (d+1)\sum_i \mathrm{tr}(R_i X)\sigma_i - \mathrm{tr}(X)I \tag{6.37}$$

$$= \sum_i \mathrm{tr}(R_i X)\Big[(d+1)\sigma_i - I\Big] \tag{6.38}$$

$$= \sum_i \mathrm{tr}(R_i X)\tilde{R}_i, \tag{6.39}$$

   where $\{\tilde{R}_i\}$ are elements of the frame dual to the frame provided by the POVM. Alternatively,

$$X = \sum_i \left[(d+1)\mathrm{tr}(R_i X) - \frac{d}{n}\mathrm{tr}(X)\right]\sigma_i = \sum_{ij}\Phi_{ij}\mathrm{tr}(R_j X)\sigma_i, \tag{6.40}$$

where $\Phi = (d+1)I - \frac{d}{n}J$: here J is the matrix of all 1's. In this way, the form of the Born matrix may be derived from the second moment operator $\mathcal{M}_2$. Indeed,

$$P(E_i|\rho) = \mathrm{tr}(E_i\rho) = \mathrm{tr}\left(E_i \sum_{jk} \Phi_{jk} P(R_k|\rho)\sigma_j\right) \tag{6.41}$$

$$= \sum_{jk} P(E_i|R_j)\Phi_{jk}P(R_k|\rho) \tag{6.42}$$

$$= \sum_{j} P(E_i|R_j)\left[(d+1)P(R_j|\rho) - \frac{d}{n}\right], \tag{6.43}$$

reproduces the Born rule. At the same time,

$$\mathrm{tr}(\mathcal{M}_2(X\otimes Y)) = \frac{1}{n}\sum_i \mathrm{tr}(\sigma_i X)\mathrm{tr}(\sigma_i Y) = \frac{1}{d(d+1)}\Big(\mathrm{tr}(X)\mathrm{tr}(Y) + \mathrm{tr}(XY)\Big). \tag{6.44}$$

Let $x_i = \mathrm{tr}(\sigma_i \tilde{X})$. Noticing that $\sum_i x_i = (n/d)\mathrm{tr}(\tilde{X})$, the action of $P(R_i|R_j) = \mathrm{tr}(R_i\sigma_j)$ on any such vector x can be expressed

$$\sum_j P(R_i|R_j)x_j = \frac{d}{n}\sum_j \mathrm{tr}(\sigma_i\sigma_j)\mathrm{tr}(\sigma_j\tilde{X}) = \frac{1}{(d+1)}\Big(\mathrm{tr}(\tilde{X}) + \mathrm{tr}(\sigma_i\tilde{X})\Big) \tag{6.45}$$

$$= \frac{1}{(d+1)}\Big((d/n)\sum_j x_j + x_i\Big). \tag{6.46}$$

This holds for any vector $x \in \mathrm{col}(P(R|R))$ since if $x \in \mathrm{col}(P(R|R))$, then $x_i = \sum_j P(R_i|R_j)y_j = \mathrm{tr}\big(\sigma_i(d/n)\sum_j y_j\sigma_j\big) = \mathrm{tr}(\sigma_i\tilde{X})$.

3. $\mathcal{M}_3 = \frac{1}{d(d+1)(d+2)}(I + T_{132} + T_{321} + T_{213} + T_{312} + T_{231})$. On the one hand,

$$\mathrm{tr}_1(\mathcal{M}_3(X\otimes I\otimes I)) = \frac{1}{n}\sum_i \mathrm{tr}(\sigma_i X)\sigma_i\otimes\sigma_i$$

$$= \frac{1}{d(d+1)(d+2)}\Big(\mathrm{tr}(X)I\otimes I + \mathrm{tr}(X)T_{21} + I\otimes X + X\otimes I + (I\otimes X)T_{21} + (X\otimes I)T_{21}\Big) \tag{6.47}$$

$$= \frac{2}{d(d+1)(d+2)}\Big(\big(\mathrm{tr}(X)I\otimes I + I\otimes X + X\otimes I\big)\Pi_{\mathrm{sym}^2}\Big). \tag{6.48}$$

On the other hand,

$$\begin{aligned}\mathrm{tr}_{12}(\mathcal{M}_3(X\otimes Y\otimes I)) &= \frac{1}{n}\sum_i \mathrm{tr}(\sigma_i X)\mathrm{tr}(\sigma_i Y)\sigma_i \\ &= \frac{1}{d(d+1)(d+2)}\Big((\mathrm{tr}(X)\mathrm{tr}(Y)+\mathrm{tr}(XY))I+\mathrm{tr}(X)Y+\mathrm{tr}(Y)X+2X\odot Y\Big), \quad (6.49)\end{aligned}$$

where $X\odot Y=(XY+YX)/2$ is the Jordan product on $d\times d$ Hermitian matrices over $\mathbb{C}$.

We now turn to the proof of Theorem 6.6.1.

Proof. Let $X=\sum_i x_i R_i$ and $Y=\sum_i y_i R_i$ for $x, y\in \mathrm{col}(P)$. We will rewrite

$$\begin{aligned}&\frac{1}{n}\sum_i \mathrm{tr}(\sigma_i X)\mathrm{tr}(\sigma_i Y)\sigma_i \\ &= \frac{1}{d(d+1)(d+2)}\Big((\mathrm{tr}(X)\mathrm{tr}(Y)+\mathrm{tr}(XY))I+\mathrm{tr}(X)Y+\mathrm{tr}(Y)X+2X\odot Y\Big) \quad (6.50)\end{aligned}$$

entirely in terms of x and y. Let $u=(1,\ldots,1)^\dagger$ so that

$$\mathrm{tr}\left(\sigma_i X\right)=\sum_j \mathrm{tr}\left(\sigma_i R_j\right)x_j=\frac{1}{d+1}\Big(\frac{d}{n}u^\dagger x+x_i\Big), \quad (6.51)$$

and

$$\begin{aligned}\mathrm{tr}(XY)=\sum_{ij}x_i\mathrm{tr}(R_iR_j)y_j &= \frac{d}{n}\sum_i x_i\left[\frac{1}{d+1}\left(\frac{d}{n}u^\dagger y+y_i\right)\right] \quad (6.52)\\ &= \frac{1}{d+1}\frac{d}{n}\left(\frac{d}{n}(u^\dagger x)(u^\dagger y)+x^\dagger y\right). \quad (6.53)\end{aligned}$$

Substituting these identities into Eq. (6.50) and expressing each term as linear combination of the POVM elements $\{R_i\}$ yields

$$\begin{aligned}&\sum_i \frac{1}{d+1}\Big(\frac{d}{n}u^\dagger x+x_i\Big)\frac{1}{d+1}\Big(\frac{d}{n}u^\dagger y+y_i\Big)R_i \\ &= \frac{1}{(d+1)(d+2)}\left[\left(\frac{d^2}{n^2}(u^\dagger x)(u^\dagger y)+\frac{1}{d+1}\frac{d}{n}\left(\frac{d}{n}(u^\dagger x)(u^\dagger y)+x^\dagger y\right)\right)\sum_i R_i\right. \\ &\left.+\frac{d}{n}(u^\dagger x)\sum_i y_iR_i+\frac{d}{n}(u^\dagger y)\sum_i x_iR_i+2\sum_i(x\odot y)_iR_i\right]. \quad (6.54)\end{aligned}$$

Matching coefficients of R_i yields

$$\frac{1}{(d+1)^2}\left(\frac{d^2}{n^2}(u^\dagger x)(u^\dagger y)+\frac{d}{n}(u^\dagger x)y_i+\frac{d}{n}(u^\dagger y)x_i+x_iy_i\right)$$
$$=\frac{1}{(d+1)(d+2)}\left(\frac{d^2}{n^2}(u^\dagger x)(u^\dagger y)+\frac{d^2}{n^2}\frac{1}{d+1}(u^\dagger x)(u^\dagger y)\right.$$
$$\left.+\frac{1}{d+1}\frac{d}{n}x^\dagger y+\frac{d}{n}(u^\dagger x)y_i+\frac{d}{n}(u^\dagger y)x_i+2(x\odot y)_i\right), \tag{6.55}$$

which simplifies to

$$x\odot y=\frac{1}{2}\left(\frac{d+2}{d+1}(x\circ y)+\frac{1}{d+1}\frac{d}{n}\Big((u^\dagger x)y+(u^\dagger y)x-(x^\dagger y)u\Big)\right) \tag{6.56}$$
$$=\gamma(x\circ y)+(1-\gamma)\Big(\overline{x}y+\overline{y}x-(x\cdot y)u\Big)/n, \tag{6.57}$$

where $x\circ y$ is the Hadamard or elementwise vector product and $\gamma=\frac{1}{2}\frac{d+2}{d+1}$. Since a 3-design furnishes an overcomplete representation, however, this expression for the quantum Jordan product is not unique. Letting $\Pi=P\Phi$ be the projector onto $\mathrm{col}(P)$, then $\Pi(x\odot y)$ gives the canonical representation of the product on $\mathrm{col}(P)$. □

6.9 Appendix: The associator and the von Neumann equation

Unitary state updates in quantum mechanics are usually expressed in terms of the commutator $[X,Y]=XY-YX$ which measures how noncommutative the product of two Jordan elements is. Given an operator O and a Hamiltonian H, one may express the time derivative of O in terms of the von Neumann equation,

$$\frac{d}{dt}O(t)=-i[H,O(t)], \tag{6.58}$$

so that $O(t)=U(t)OU(t)^\dagger$ where $U=e^{-iHt}$ is a unitary operator, and we note that in the Heisenberg picture, we ought to flip the sign so that $O(t)=U(t)^\dagger OU(t)$. In this appendix, using the results of [161, 162, 163, 27, 164], we show how unitary state updates may be alternatively expressed in entirely Jordan algebraic terms using the *associator*,

$$(X,Y,Z)=(X\odot Y)\odot Z-X\odot(Y\odot Z), \tag{6.59}$$

which measures how nonassociative is the product of three Jordan elements.

First, let us give a little context. Suppose we are working with an arbitrary Euclidean Jordan algebra, and we consider a one parameter family of automorphisms G_t. We can then consider a continuous reversible "evolution" of an element x, $x(t) = G_t x$. By definition, $G_0 = u$, where u is the identity element of the EJA; $G_{s+t} = G_s G_t$; and $G_t(x \odot y) = G_t x \odot G_t y$. Now for small t, $G_t = I + tD + O(t^2)$ for some linear map D. Suppose G_t acts on the Jordan product $x \odot y$. On the one hand,

$$G_t(x \odot y) = x \odot y + tD(x \odot y) + O(t^2). \tag{6.60}$$

On the other hand,

$$G_t x \odot G_t y = (x + tDx) \odot (y + tDy) + O(t^2) \tag{6.61}$$

$$= x \odot y + t(Dx \odot y + x \odot Dy) + O(t^2). \tag{6.62}$$

Equating coefficients of t gives

$$D(x \odot y) = Dx \odot y + x \odot Dy, \tag{6.63}$$

that is, D satisfies the *Leibniz rule*, and so is called a *derivation*. It represents an infinitesimal change that preserves a product, in this case, the Jordan product. Indeed, if we consider the differential equation

$$\frac{d}{dt}x(t) = Dx(t), \tag{6.64}$$

since D is a derivation, we have

$$\frac{d}{dt}(x \odot y)(t) = \frac{d}{dt}x(t) \odot y + x \odot \frac{d}{dt}y(t). \tag{6.65}$$

Conversely, every derivation integrates to a one-parameter family of automorphisms. The differential equation $\frac{d}{dt}x(t) = Dx(t)$ has the solution $x(t) = e^{tD}x$. We must show that this preserves the Jordan product. On the one hand, let $f(t) = e^{tD}(x \odot y)$. Then $\frac{d}{dt}f(t) = Df(t)$. On the other hand, let $g(t) = e^{tD}x \odot e^{tD}y$. Differentiating gives $\frac{d}{dt}g(t) = De^{tD}x \odot e^{tD}y + e^{tD}x \odot De^{tD}y = Dg(t)$ by the Leibniz rule. Since f and g satisfy the same linear ODE ($\frac{d}{dt}z(t) = Dz(t)$) with the same initial conditions $f(0) = g(0) = x \odot y$, they must be equal. Thus $e^{tD}(x \odot y) = e^{tD}x \odot e^{tD}y$, as desired.

Now let $L_a x = a \odot x$ be the linear operator that performs the Jordan product. It is a remarkable theorem of Jacobson and Koecker that any derivation of an EJA D lies in $\mathrm{span}\{[L_a, L_b]\}$ [164, 27, 165]: in other words, the derivations of an EJA are precisely given

by linear combinations of commutators of Jordan product operators. But

$$[L_a, L_b]c = a \odot (b \odot c) - b \odot (a \odot c) = (a, b, c), \tag{6.66}$$

which is precisely the associator we defined above.

For associative Jordan algebras, for example, $\mathbb{R}^n$ equipped with the elementwise product $\circ$, corresponding to classical probability theory, the associator always vanishes. While one can formulate continuous "evolution" in classical theory using the formalism of continuous Markov chains and their rate matrices, this evolution does not originate intrinsically from the theory itself. In contrast, the geometric meaning of derivations is particularly transparent for the so-called spin-factor Jordan algebras which are defined on $V_n = \mathbb{R} \oplus \mathbb{R}^n$ and whose Jordan product is

$$(x_1, x_0) \odot (y_1, y_0) = (x_1 y_1 + x_0 \cdot y_0,\ x_1 y_0 + y_1 x_0). \tag{6.67}$$

For simplicity, let us pick elements $a = (0, a_0)$ and $b = (0, b_0)$. Then

$$L_a(x_1, x_0) = (a_0 \cdot x_0, x_1 a_0), \qquad L_b(x_1, x_0) = (b_0 \cdot x_0, x_1 b_0). \tag{6.68}$$

A straightforward calculation gives

$$[L_a, L_b](x_1, x_0) = (0,\ (b_0 \cdot x_0)u - (a_0 \cdot x_0)b_0)\,. \tag{6.69}$$

Notice that the scalar component is unchanged. As for the vector part, suppose that $n = 3$, and recall that in three dimensions $(b_0 \cdot x_0)a_0 - (a_0 \cdot x_0)b_0 = (a_0 \times b_0) \times x_0$, so that if $\frac{d}{dt}(x_1, x_0)(t) = [L_a, L_b](x_1, x_0)$, we have

$$\frac{d}{dt}x(t) = \omega \times x(t), \tag{6.70}$$

where $\omega = a_0 \times b_0$. This is precisely the equation for a rigid-body rotation around the axis ω.

The further remarkable fact about quantum mechanics over $\mathbb{C}$ is that all these equations can be rewritten not in terms of an associator with two observables, but in terms of commutators with a *single* observable. In fact, the $n = 3$ spin-factor case is just such an example as it coincides with a qubit. Quantum mechanics over $\mathbb{C}$ thus generalizes in a surprising way the fact that in three dimensions, any vector representing an axis of rotation can be expressed as the cross product of two vectors. We first observe that the associator can be rewritten as a nested commutator. Letting $X \odot Y = (XY + YX)/2$ be the Jordan product on Hermitian

matrices, we have

$$(X, O, Y) = \frac{1}{2}(XO + OX) \odot Y - X \odot \frac{1}{2}(OY + YO) \tag{6.71}$$

$$= \frac{1}{4}\Big(XOY + YXO + OXY + YOX - XOY - OYX - XYO - YOX\Big) \tag{6.72}$$

$$= \frac{1}{4}\Big(YXO + OXY - OYX - XYO\Big) \tag{6.73}$$

$$= \frac{1}{4}\Big(O[X, Y] - [X, Y]O\Big) \tag{6.74}$$

$$= -\frac{1}{4}[[X, Y], O]. \tag{6.75}$$

This holds for any EJA built out of Hermitian matrices. For quantum mechanics over $\mathbb{C}$ in particular, however, we have the von Neumann equation $\frac{d}{dt}O(t) = -i[H, O(t)]$. Writing $H = \mathrm{tr}(H)I/d + H_0$ for $\mathrm{tr}(H_0) = 0$, we see that the von Neumann equation cares only about the traceless part of the Hamiltonian,

$$\frac{d}{dt}O(t) = -i[\mathrm{tr}(H)I/d + H_0, O] = -i[H_0, O]. \tag{6.76}$$

What is not trivial is that over $\mathbb{C}$, *any* traceless Hermitian matrix H_0 can be written as i times the commutator between two Hermitian matrices X, Y: $H_0 = i[X, Y]$ [162]. Thus let $H_0 = -\frac{1}{4}i[X, Y]$ so that

$$\frac{d}{dt}O(t) = -i\left[-\frac{1}{4}i[X, Y], O(t)\right] = -\frac{1}{4}[[X, Y], O(t)] = (X, O(t), Y). \tag{6.77}$$

As desired, we have reexpressed the von Neumann equation in terms of the associator instead of the commutator, which in fact is the form that generalizes to any EJA.

What to make of this in the light of 3-designs? We have for $x, y \in \mathrm{col}(P)$,

$$x \odot y = \gamma(x \circ y) + (1 - \gamma)\Big(\overline{x}y + \overline{y}x - (x \cdot y)u\Big)/n. \tag{6.78}$$

Let $\Pi = P\Phi$ be the projector onto $\mathrm{col}(P)$. Making sure to project onto $\mathrm{col}(P)$ after an application of $\odot$, we find after some algebra,

$$\begin{aligned}&(x, y, z) \\ &= \frac{1}{4(d+1)^2}\Bigg\{(d+2)^2\Big(\Pi(x \circ y) \circ z - x \circ \Pi(y \circ z)\Big) \\ &+ \left(\frac{d}{n}\right)^2 \Big(\big(n(y \cdot z) - \overline{yz}\big)x - \big(n(x \cdot y) - \overline{xy}\big)z + \big(\overline{z}(x \cdot y) - \overline{x}(y \cdot z)\big)u\Big)\Bigg\}.\end{aligned} \tag{6.79}$$

Meanwhile, let $L_x y = x \odot y$ so that

$$L_x = \frac{1}{2}\frac{1}{d+1}\Big((d+2)D_x + \frac{d}{n}\big(\overline{x}I + xu^\dagger - ux^\dagger\big)\Big). \tag{6.80}$$

Similarly, we can consider the linear operator $A_{x,y}$ such that $A_{x,y}o = (x, o, y)$,

$$\begin{aligned}A_{x,y} = \frac{1}{4(d+1)^2}\Bigg\{&(d+2)^2\big(D_y \Pi D_x - D_x \Pi D_y\big) \\ &+ \left(\frac{d}{n}\right)^2 \Big(n\,(xy^\dagger - yx^\dagger) + \overline{x}\,(yu^\dagger - uy^\dagger) + \overline{y}\,(ux^\dagger - xu^\dagger)\Big)\Bigg\}.\end{aligned} \tag{6.81}$$

For compactness, writing $x \wedge y = xy^\dagger - yx^\dagger$, we have

$$\Pi A_{x,y} = \frac{1}{4(d+1)^2}\left\{(d+2)^2[\Pi D_y, \Pi D_x] + \left(\frac{d}{n}\right)^2 \Big(n\,x \wedge y + \overline{x}\,y \wedge u + \overline{y}\,u \wedge x\Big)\right\}, \tag{6.82}$$

so that

$$O(t) = U(t)OU(t)^\dagger = e^{-iHt}Oe^{iHt} = \sum_i o(t)_i R_i, \tag{6.83}$$

where $o(t) = \exp(\Pi A_{x,y}t)o$, $O = \sum_i o_i R_i$ for $o \in \mathrm{col}(P)$, and $H = -\frac{1}{4}i[X, Y]$ for $X = \sum_i x_i R_i$ and $Y = \sum_i y_i R_i$. We may connect this to the general reference measurement representation

of a unitary map. Let $P(R_i|R_j^{(U)}) = \mathrm{tr}(R_i U \sigma_j U^\dagger)$. Then

$$\sum_{jk} P(R_i|R_j^{(U)})\Phi_{jk}P(R_k|\rho) = \sum_j \mathrm{tr}\left(R_i U \sum_{jk} \Phi_{jk}\mathrm{tr}(R_k\rho)\sigma_j U^\dagger\right) \tag{6.84}$$

$$= \mathrm{tr}(R_i U \rho U^\dagger) \tag{6.85}$$

$$= P(R_i|\rho^{(U)}). \tag{6.86}$$

At the same time, from the above considerations,

$$P(R|R^{(U)})\Phi = \Pi \exp(\Pi A_{x,y} t)\Pi, \tag{6.87}$$

where we sandwich the operator with Π's since the expression $P(R_i|R_j^{(U)}) = \mathrm{tr}(R_i U \sigma_j U^\dagger)$ does not assume that $P(R_i|R_j^{(U)})\Phi$ acts solely on vectors in $\mathrm{col}(P)$ although it produces only vectors in $\mathrm{col}(P)$. In this way, unitary maps can be represented entirely in Jordan algebraic terms.

CHAPTER 7

RECONSTRUCTION

7.1 Introduction

In the last chapter, we began within quantum mechanics, assumed our reference measurement was constructed from a complex projective 3-design, and then presented an image of quantum state space within the probability simplex, showing that it can be defined in terms of an uncertainty principle. In particular, we showed that the geometry of quantum mechanics as a whole is encoded in the single matrix $P(R|R)$ which characterizes the reference measurement itself. The keystone in the argument was the connection between the third tensor moment of quantum state space and the Jordan algebra of observables. In this chapter, we go in the opposite direction. We begin from almost nothing: the gambler equipped with the nonclassical coherence criterion $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ which we have seen can be derived on very general grounds. We then proceed to place increasingly strong constraints on the matrix $P(R|R)$ until we arrive back at quantum mechanics, and can conclude that the reference measurement in fact forms a 3-design. The conceptual innovation is that just as we may consider the gentlest possible modification of the law of total probability, the Protourgleichung, in the same way we may consider the gentlest possible modification of the classical rule for multiplying valuations on the reference outcomes. We make a series of assumptions which then allow us to identify this rule with the quantum Jordan product. On the one hand, this reconstruction of quantum mechanics allows us to see precisely how it is, from the inside as it were, that 3-designs make the quantum Jordan product look as close as possible to the classical Jordan product. In the QBist spirit, it is grounded entirely in constraints on the reference measurement. On the other hand, our result may be interpreted in practical terms as a way of *self-testing* [166] 3-design measurements. If $P(R|R)$ satisfies our constraints, we show it must have a Hilbert space representation, and in fact correspond to an unbiased complex-projective 3-design: other methods, e.g., calculating the frame potential, do not in themselves guarantee that such a representation exists. Moreover, once our conditions are met, if $P(R|R)$ minimizes the t'th order frame potential, the same framework certifies t-designs for $t \geq 3$.

7.2 The Urgleichung

We begin as always with the gambler trying to make better decisions. We assume they have identified a reference measurement $\{R_i\}_{i=1}^n$ for a domain they are interested in so that

they adopt the nonclassical coherence condition

$$P(E|\rho) = P(E|R)\Phi P(R|\rho), \tag{7.1}$$

where Φ is a chosen $\{1\}$-inverse of $P(R|R)$, the conditional probability matrix which characterizes the reference measurement. Guided by simplicity, we assume that $P \equiv P(R|R)$ is in fact symmetric ($P = P^T$), and so bistochastic. Thus $\mathrm{col}(P) = \mathrm{row}(P)$, and so probability vectors $P(R|\rho)$, response functions $P(E|R)$, and valuations all live in the same subspace: $\mathrm{col}(P)$. We denote the projector onto this subspace $\Pi = P\Phi$. Finally, we also assume that P is constant along its diagonal: $\forall i : P(R_i|R_i) = \mathrm{const}$.

Next, we assume that $\Phi = \alpha I + \beta J$, where J is the matrix of all 1's, is a Born matrix for P: thus the nonclassical coherence condition takes Protourgleichung form. If we require that Φ is quasistochastic, having negative entries, but with columns summing to 1, so that it preserves the normalization of probability vectors, then we must have $\alpha + n\beta = 1$, so that $\beta = (1 - \alpha)/n$.

Lemma 7.2.1. *Suppose $P\Phi P = P$ for $\Phi = \alpha I + \beta J$ where $\beta = (1 - \alpha)/n$. Then*

$$\forall y \in col(P) : Py = \frac{1}{\alpha} y + \left(1 - \frac{1}{\alpha}\right) \overline{y} u/n, \tag{7.2}$$

for $u = (1, \ldots, 1)^\dagger$.

Proof. We have $P\Phi P = \alpha P^2 + \beta J = P$ since $PJP = J$. Moreover, since $JPx = Jx$, it follows that if we let $y = Px \in \mathrm{col}(P)$, then $\alpha Py + \beta Jy = y$. Thus $Py = \frac{1}{\alpha}(y - \beta u u^\dagger y)$, from which the result follows. □

Lemma 7.2.2. *Suppose P is stochastic, symmetric, and $P\Phi P = P$ for $\Phi = \alpha I + \beta J$ where $\beta = (1 - \alpha)/n$. Then the eigenvalues of P are $\in \{1, 1/\alpha, 0\}$.*

Proof. Since P is bistochastic, $Pu = u$, showing that 1 is an eigenvalue. Suppose that $v \neq 0$ is an eigenvector with eigenvalue λ satisfying $u^\dagger v = 0$. Applying v to $\alpha P^2 + \beta J = P$ gives $\alpha\lambda^2 v = \lambda v$ so that $\lambda(\alpha\lambda - 1) = 0$. We conclude that $\lambda = 0$ or $\lambda = 1/\alpha$. □

Corollary 7.2.2.1. *If we further assume that P is constant along its diagonal, then $\forall i$: $P(R_i|R_i) = \big(1 + (r - 1)/\alpha\big)/n$.*

Proof. $\mathrm{tr}(P) = nP(R_i|R_i) = \sum_i \lambda_i = 1 + (r - 1)/\alpha$, where $r = \mathrm{rank}(P)$, from which the result follows. □

Lemma 7.2.3. *$\Pi = P\Phi$ is an orthogonal projector onto $col(P)$.*

Proof. Since $P\Phi P = P$, we have $\Pi^2 = P\Phi P\Phi = P\Phi = \Pi$, so Π is a projector. Then since $P = P^\dagger$ (so that P is bistochastic) and $\Phi = \alpha I + (1 - \alpha)J/n$, we have $P\Phi = \Phi P$. Moreover, $\Pi^\dagger = (P\Phi)^\dagger = \Phi^\dagger P^\dagger = \Phi P = P\Phi$. Finally, since $\Pi = P\Phi$, $\mathrm{range}(\Pi) \subseteq \mathrm{col}(P)$. Let $y \in \mathrm{col}(P)$ so that $y = Px$. Then $\Pi y = P\Phi Px = Px = y$, so that Π fixes $\mathrm{col}(P)$. □

7.3 A nonclassical product

As we discussed in Chapter 2, a real-valued random variable is a map from a sample space $\Omega \to \mathbb{R}$. Viewing the elements of the sample space as the outcomes of the finest grained classical reference measurement, it becomes clear that more generally we ought to define a real valued random variable as a map from reference outcomes to the reals: $\{R_i\} \to \mathbb{R}$. Indeed, if we have a valuation $\mathfrak{x}$ on any measurement $\{X_i\}$, then $x = \mathfrak{x}P(X|R)\Phi$ is an equivalent valuation on $\{R_i\}$, where $\mathfrak{x}$ and x are understood as row vectors[1]. By equivalent, we mean that $\sum_i \mathfrak{x}_i P(X_i|\rho) = \sum_i x_i P(R_i|\rho)$ so that the gambler ought to assign the same price to both.

Moreover, such random variables form an algebra: multiplication of random variables is achieved by the elementwise or Hadamard product $x \circ y$. Clearly, the product should be valued at $\sum_i x_i y_i P(R_i|\rho)$. What we would like, although we do not assume it yet[2], is that there exists a measurement $\{M_i^{(x\circ y)}\}$ and a valuation $\mathfrak{m}^{(x\circ y)}$ such that $x \circ y = \mathfrak{m}^{(x\circ y)} P(M^{(x\circ y)}|R)\Phi$ which we would value at the same price,

$$\sum_i \mathfrak{m}_i^{(x\circ y)} P(M^{(x\circ y)}|\rho) = \sum_i x_i y_i P(R_i|\rho). \tag{7.3}$$

Since x and y are valuations on $\{R_i\}$, whatever this measurement $\{M_i^{(x\circ y)}\}$ is, should it exist, it will depend on the particular choice of reference measurement. But we could imagine introducing an alternative product on valuations, $x \odot y$, such that there exists a measurement $\{M_i^{(x\odot y)}\}$ and a valuation $\mathfrak{m}^{(x\odot y)}$ such that $x \odot y = \mathfrak{m}^{(x\odot y)} P(M^{(x\odot y)}|R)\Phi$, so that

$$\sum_i \mathfrak{m}_i^{(x\odot y)} P(M^{(x\odot y)}|\rho) = \sum_i [x \odot y]_i P(R_i|\rho), \tag{7.4}$$

but where the measurement $\{M_i^{(x\odot y)}\}$ *would not depend in any way on the reference measurement* used in the definition of the product $\odot$.

For motivation, recall that we ought to use $P(E|R,\rho) = P(E|R)P(R|\rho)$ in the case that we perform an intermediate reference measurement: in this way, $P(E|R,\rho)$ carries a reference measurement dependence. In contrast, we ought to use $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ in the case that the reference measurement remains hypothetical: here $P(E|\rho)$ does *not* depend on the choice of reference measurement, even as it can be expressed in terms of reference probabilities. At the same time, treating $P(E|R)\Phi$ as a valuation v on reference outcomes, we can interpret $P(E|\rho)$ as the price at which we value the random variable v on a reference measurement we actually perform.

Analogously, the random variable $x \circ y$ can be understood in terms of a reference measurement the gambler actually performs. At the same time, by hypothesis, it is equivalent

[1] In general we will write valuations x as row vectors. In order to simplify notation, however, we may sometimes treat x as a column vector when convenient.

[2] We introduce it as a guiding hypothesis, and later, by restricting the form of $P(R|R)$, will we justify it.

to a valuation on a measurement $\{M_i^{(x\circ y)}\}$, but the nature of the measurement depends on the choice of reference. In contrast, while $x \odot y$ can be understood as a valuation on the reference which is equivalent to a valuation on a measurement $\{M_i^{(x\odot y)}\}$, we would like the latter to not depend at all on the choice of reference. The reference measurement should remain in this sense "hypothetical," in that it merely provides a convenient way of expressing the observable, which is independent of it. In seeking such a product, we will be guided by this analogy: just as the Protourgleichung

$$P(E|\rho) = \sum_i P(E|R_i)\big\{\alpha P(R_i|\rho) + (1-\alpha)P(R_i|\mu)\big\}, \tag{7.5}$$

brings the fundamental nonclassical coherence relation as close as possible to the classical law of total probability, we will seek a nonclassical product on random variables $\odot$ which looks as close as possible to the classical product $\circ$. In other words, we have already assumed that $\Phi = \alpha I + (1-\alpha)J/n$, that is, the nonclassical coherence relation is a minimal deformation of the classical rule, the law of total probability. We now similarly attempt to deform the classical rule for the multiplication of random variables in as gentle a way as possible.

Notice that the Hadamard product is: *commutative*, that is, $x \circ y = y \circ x$; *permutation equivariant*, that is, for any permutation matrix P_π, it satisfies $(P_\pi x) \circ (P_\pi y) = P_\pi(x \circ y)$; and finally, it enjoys $u = (1, \ldots, 1)^\dagger$ as a *multiplicative identity*: $u \circ a = a$. We will now explore the space of products which enjoy these same properties, with a further demand: just as $P(E|R)\Phi P(R|\rho) = \alpha P(E|R)P(R|\rho) + (1-\alpha)P(E|R)P(R|\mu)$ where $P(R_i|\mu) = 1/n$, gives a one-parameter mixture with the classically expected probabilities, we want our product $\odot$ to be a one-parameter mixture with the classically expected product $\circ$ in the simplest possible way.

Theorem 7.3.1. *Let $\odot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be a bilinear map which*

1. *is commutative: $x \odot y = y \odot x$,*
2. *is permutation equivariant: $\forall \pi \in S_n : P_\pi(x \odot y) = (P_\pi x) \odot (P_\pi y)$,*
3. *is unital with unit u: $\forall x \in \mathbb{R}^n : u \odot x = x$, where $u = (1, \ldots, 1)^\dagger$,*
4. *is a one-parameter mixture with the classical product: $x \odot y = \gamma(x \circ y) + (1-\gamma)z$ where $z \neq 0$ contains as few terms as possible.*

Then any such product $\odot$ must take the form

$$x \odot y = \gamma(x \circ y) + (1-\gamma)(\overline{y}x + \overline{x}y - (x \cdot y)u)/n, \tag{7.6}$$

where $\circ$ is the entrywise or Hadamard product, $\overline{x} = \sum_i x_i$, $\cdot$ is the dot product, and γ is an arbitrary parameter.

Proof. Since $\odot$ is bilinear, there must be a rank-3 tensor T_{ijk} such that

$$(x \odot y)_i = \sum_{jk} T_{ijk} x_j y_k. \tag{7.7}$$

Commutativity is equivalent to invariance under swapping of the last two indices,

$$T_{ijk} = T_{ikj}. \tag{7.8}$$

Permutation equivariance means that for any permutation $\pi \in S_n$,

$$T_{\pi(i)jk} = T_{i\pi^{-1}(j)\pi^{-1}(k)} \Longleftrightarrow T_{ijk} = T_{\pi(i)\pi(j)\pi(k)}, \tag{7.9}$$

that is, T_{ijk} must be invariant under simultaneous permutations of its indices. Such permutations can shuffle the indices arbitrarily, but can't change the patterns of equality between the indices. There are five possibilities,

$$i = j = k, \qquad i = j \neq k, \qquad i = k \neq j, \qquad j = k \neq i, \qquad i, j, k \text{distinct} \tag{7.10}$$

and so the most general form for a permutation covariant tensor is

$$T_{ijk} = c_1 \delta_{ij} \delta_{ik} + c_2 \delta_{ij} + c_3 \delta_{ik} + c_4 \delta_{jk} + c_5, \tag{7.11}$$

which yields

$$[x \odot y]_i = \sum_{jk} T_{ijk} x_j y_k = c_1 x_i y_i + c_2 \overline{y} x_i + c_3 \overline{x} y_i + c_4 x \cdot y + c_5 \overline{xy}, \tag{7.12}$$

or

$$x \odot y = c_1 x \circ y + c_2 \overline{y} x + c_3 \overline{x} y + c_4 (x \cdot y) u + c_5 \overline{xy} u. \tag{7.13}$$

From commutativity, $T_{ijk} = T_{ikj}$, so that $c_2 = c_3$. Demanding that u is a multiplicative identity gives

$$\begin{aligned} u \odot y &= c_1 u \circ y + c_{23}(\overline{y} u + \overline{u} y) + c_4 (u \cdot y) u + c_5 \overline{uy} u && (7.14) \\ &= c_1 y + c_{23}(\overline{y} u + n y) + c_4 \overline{y} u + n c_5 \overline{y} u && (7.15) \\ &= (c_1 + n c_{23}) y + (c_{23} + c_4 + n c_5) \overline{y} u = y. && (7.16) \end{aligned}$$

We conclude

$$c_1 + n c_{23} = 1 \qquad\qquad c_{23} + c_4 + n c_5 = 0, \tag{7.17}$$

and

$$x \odot y = (1 - nc_{23})x \circ y + c_{23}(\overline{y}x + \overline{x}y) + c_4(x \cdot y)u - \frac{1}{n}(c_{23} + c_4)\overline{xy}u \tag{7.18}$$

$$= (1 - nc_{23})x \circ y + c_{23}(\overline{y}x + \overline{x}y - \overline{xy}u/n) + c_4(x \cdot y - \overline{xy}/n)u. \tag{7.19}$$

The final demand, that $x \odot y$ is a one-parameter mixture with $x \circ y$ forces $c_{23} \propto c_4$. If we take $c_4 = -c_{23}$, then $\overline{xy}u/n$ term drops out, leading to the simplest expression with the fewest terms,

$$x \odot y = (1 - nc_{23})x \circ y + c_{23}(\overline{y}x + \overline{x}y - (x \cdot y)u) \tag{7.20}$$

$$= \gamma(x \circ y) + (1 - \gamma)(\overline{y}x + \overline{x}y - (x \cdot y)u)/n. \tag{7.21}$$

□

Remark. *We defined $x \odot y$ on $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, but this isn't quite right for our purposes. On the one hand, we always assume that $x, y \in col(P)$; on the other hand, even if $x, y \in col(P)$, $x \odot y$ need not be in $col(P)$. Thus for consistency, we ought to project the result back into $col(P)$: therefore let $x \bar{\odot} y = \Pi(x \odot y)$ which takes $col(P) \times col(P) \to col(P)$. Through this column space restriction, the product $\bar{\odot}$ then depends on P.*

We may now fix the value of γ in terms of α. Notice that if we insert the projector $P\Phi$ into the nonclassical coherence relation,

$$\sum_i \mathfrak{x}_i P(X_i|\rho) = \mathfrak{x}P(X|R)\Phi P(R|\rho) = \Big\{\mathfrak{x}P(X|R)\Phi\Big\}P\Big\{\Phi P(R|\rho)\Big\} = x^\dagger P y = \langle x, y\rangle_P, \tag{7.22}$$

where $x^\dagger = \mathfrak{x}P(X|R)\Phi$ and $y = \Phi P(R|\rho)$, we can rewrite the relation as an inner product on vectors with metric P. By Lemma 7.2.2, P is positive semidefinite: and on $\mathrm{col}(P)$ it is positive definite. Thus $\langle\cdot,\cdot\rangle_P$ is a proper inner product on $\mathrm{col}(P)$. Clearly, x is a valuation: what about y? From

$$\sum_j y_j P(R_j|R_i) = \sum_{jk} \Phi_{jk} P(R_k|\rho)P(R_j|R_i) = P(R_i|\rho), \tag{7.23}$$

we see that we can interpret y as reference valuation such that $\langle y\rangle_{R_i} = P(R_i|\rho)$: the expectation of y on the reference measurement, conditional on the preparatory outcome R_i, is equivalent to the probability for the R_i'th reference outcome conditional on ρ. Finally, notice that $\langle (P(X_i|R)\Phi)^\dagger, \Phi P(R|\rho)\rangle_P = P(X_i|\rho)$ reproduces the fundamental nonclassical coherence rule itself.

Now we would like our product $\odot$ to be compatible with this inner product.

Lemma 7.3.2. *Let $x \odot y$ be a column vector and $u = (1, \ldots, 1)^\dagger$. Demanding*

$$u^\dagger (x \odot y) = \langle x, y \rangle_P. \tag{7.24}$$

fixes $\gamma = \frac{1}{2}(1 + 1/\alpha)$.

Proof. On the one hand,

$$u^\dagger (x \odot y) = \gamma (x \cdot y) + (1 - \gamma)(2\overline{xy}/n - (x \cdot y)) \tag{7.25}$$
$$= (2\gamma - 1)(x \cdot y) + 2(1 - \gamma)\overline{xy}/n. \tag{7.26}$$

On the other hand,

$$\langle x, y \rangle_P = x^\dagger P y = \frac{1}{\alpha}(x \cdot y) + \left(1 - \frac{1}{\alpha}\right)\overline{xy}/n. \tag{7.27}$$

Equating terms implies that $2\gamma - 1 = 1/\alpha$ from which the result follows. □

Corollary 7.3.2.1. *Fixing $\gamma = \frac{1}{2}(1 + 1/\alpha)$ implies that*

$$\forall x, y, z \in col(P) : \langle x, y \odot z \rangle_P = \langle x \odot y, z \rangle_P, \tag{7.28}$$

so that $\odot$ is self-adjoint with respect to this inner product.

Proof. Since for $x, y \in \mathrm{col}(P)$,

$$\langle x, y \rangle_P = x^\dagger P y = \frac{1}{\alpha}(x \cdot y) + \left(1 - \frac{1}{\alpha}\right)\overline{xy}/n, \tag{7.29}$$

we have

$$\langle x, y \odot z \rangle_P = \frac{1}{\alpha}\, x \cdot (y \odot z) + \left(1 - \frac{1}{\alpha}\right)\overline{x}\,\overline{y \odot z}/n, \tag{7.30}$$
$$\langle x \odot y, z \rangle_P = \frac{1}{\alpha}\, (x \odot y) \cdot z + \left(1 - \frac{1}{\alpha}\right)\overline{x \odot y}\,\overline{z}/n. \tag{7.31}$$

Now,

$$x \cdot (y \odot z) = \gamma \sum_i x_i y_i z_i + (1 - \gamma)\left(\overline{z}(x \cdot y) + \overline{y}(x \cdot z) - \overline{x}(y \cdot z)\right)/n, \tag{7.32}$$
$$(x \odot y) \cdot z = \gamma \sum_i x_i y_i z_i + (1 - \gamma)\left(\overline{y}(x \cdot z) + \overline{x}(y \cdot z) - \overline{z}(x \cdot y)\right)/n. \tag{7.33}$$

Thus

$$x \cdot (y \odot z) - (x \odot y) \cdot z = 2(1-\gamma)\left(\overline{z}(x \cdot y) - \overline{x}(y \cdot z)\right)/n \tag{7.34}$$

$$= \left(1 - \frac{1}{\alpha}\right)\left(\overline{z}(x \cdot y) - \overline{x}(y \cdot z)\right)/n. \tag{7.35}$$

Taking component sums gives

$$\overline{y \odot z} = \frac{1}{\alpha}(y \cdot z) + \left(1 - \frac{1}{\alpha}\right)\overline{y}\,\overline{z}/n, \qquad \overline{x \odot y} = \frac{1}{\alpha}(x \cdot y) + \left(1 - \frac{1}{\alpha}\right)\overline{x}\,\overline{y}/n, \tag{7.36}$$

so that

$$\overline{x}\,\overline{y \odot z} - \overline{x \odot y}\,\overline{z} = \frac{1}{\alpha}\left(\overline{x}(y \cdot z) - \overline{z}(x \cdot y)\right). \tag{7.37}$$

Putting this all together, we find

$$\begin{aligned}
&\langle x, y \odot z\rangle_P - \langle x \odot y, z\rangle_P \\
&= \frac{1}{\alpha}\left(1 - \frac{1}{\alpha}\right)\left(\overline{z}(x \cdot y) - \overline{x}(y \cdot z)\right)/n + \left(1 - \frac{1}{\alpha}\right)\frac{1}{\alpha}\left(\overline{x}(y \cdot z) - \overline{z}(x \cdot y)\right)/n && (7.38)\\
&= 0, && (7.39)
\end{aligned}$$

as desired. □

Remark. *The projected product $x\bar{\odot}y$ is also self-adjoint with respect to $\langle\cdot,\cdot\rangle_P$. Since $P\Pi = PP\Phi = P\Phi P = P$ and $\Pi P = P\Phi P = P$,*

$$\langle x, y\bar{\odot}z\rangle_P = x^\dagger P\Pi(y \odot z) = x^\dagger P(y \odot z) = \langle x, y \odot z\rangle_P \tag{7.40}$$

$$\langle x\bar{\odot}y, z\rangle_P = (x \odot y)^\dagger \Pi^\dagger P z = \langle x \odot y, z\rangle_P. \tag{7.41}$$

Lemma 7.3.3. *The algebra on $col(P)$ defined by $\bar{\odot}$ along with vector addition and scalar multiplication is* formally real*: if a sum of squares vanishes, each term must individually vanish.*

Proof. Let $\{v_i\}$ be a set of valuations. Suppose that $\sum_i v_i^{\odot 2} = 0$. Then

$$0 = \left\langle \sum_i v_i^{\odot 2}, u \right\rangle_P = \sum_i \langle v_i \odot v_i, u\rangle_P = \sum_i \langle v_i, v_i \odot u\rangle_P = \sum_i \langle v_i, v_i\rangle_P = \sum_i \|v_i\|_P^2, \tag{7.42}$$

which implies that $\forall i : v_i = 0$. □

7.4 Moment matching

Suppose we have a measurement $\{X_i\}$ with a valuation $\mathfrak{x}$. As we have seen, $x = \mathfrak{x}P(X|R)\Phi$ is an equivalent valuation on the reference measurement in the sense that

$$\sum_i \mathfrak{x}_i P(X_i|\rho) = \sum_i x_i P(R_i|\rho). \tag{7.43}$$

Similarly, higher moments of $\mathfrak{x}$ can be expressed as valuations on the reference measurement. Letting $x^{(m)} = \mathfrak{x}^{\circ m} P(X|R)\Phi$ where $\mathfrak{x}^{\circ m}$ denotes the m-th Hadamard power of the valuation vector, we have

$$\sum_i \mathfrak{x}_i^m P(X_i|\rho) = \sum_i x_i^{(m)} P(R_i|\rho). \tag{7.44}$$

At the same time, we supposed that the reference valuation $x\bar{\odot}y$ is equivalent to a valuation $\mathfrak{m}^{(x\bar{\odot}y)}$ on a measurement $\{M_i^{(x\bar{\odot}y)}\}$ independent of the choice of reference. In particular, we can consider $x^{\bar{\odot}m}$ for some power m: this too should correspond to a valuation on a measurement independent of the choice of reference. What meaning can we give to $\sum_i [x^{\bar{\odot}m}]_i P(R_i|\rho)$, the m-th moment of x with respect to $\bar{\odot}$? By assumption, there must exist a measurement $\{M_i^{(x^{\bar{\odot}m})}\}$, independent of the choice of reference, on which there exists an equivalent valuation. The most minimal supposition would be that this measurement is $\{X_i\}$ itself and further that $x^{\bar{\odot}m}$ is equivalent to $\mathfrak{x}^m$: this would certainly be independent of the choice of reference. But that would be too restrictive: after all, there may be multiple measurements and valuations which imply the same reference valuation $x = \mathfrak{x}'P(X'|R)\Phi$. We thus merely assume that there exists some class of measurements for which

$$\sum_i \mathfrak{x}_i^m P(X_i|\rho) = \sum_i x_i^{\bar{\odot}m} P(R_i|\rho), \tag{7.45}$$

that is, $x^{(m)} = \mathfrak{x}^{\circ m} P(X|R)\Phi = x^{\bar{\odot}m}$. For this class of measurements, then, $\langle \mathfrak{x}^m \rangle = \langle x^{\bar{\odot}m} \rangle$: we can calculate the higher moments of $\mathfrak{x}$ entirely in terms of the reference valuation x. This is in the spirit of $\bar{\odot}$ being a minimal modification of $\circ$: whereas the classical powers of a reference valuation $x^{\circ m}$ carry a reference measurement dependence, we assume that the nonclassical powers $x^{\bar{\odot}m}$ remove this dependence, and the simplest assumption is that they reflect the higher moments of the original valuation $\mathfrak{x}^m$.

This assumption, perhaps innocuous, will turn out to have profound consequences. The reason is that, as we have defined it, $x\bar{\odot}y$ is commutative, but it need not be *associative*: it may be that $(x\bar{\odot}y)\bar{\odot}z \neq x\bar{\odot}(y\bar{\odot}z)$. In order for $x^{\bar{\odot}m} = x\bar{\odot}x\bar{\odot}x\bar{\odot}\ldots$ to have an unambiguous meaning, we must therefore require that $\bar{\odot}$ powers of x associate with each other. Then even if the algebra so defined is in general nonassociative, $x\bar{\odot}x\bar{\odot}x\bar{\odot}\ldots$ will have an univocal meaning, independent of how it is parenthesized. An algebra with such a product is called

power associative. But how can we implement the constraint that $\bar{\odot}$ is power associative? Concretely, since $\bar{\odot}$ is defined in terms of P, what restrictions must be placed upon the structure of P to guarantee power associativity? There is a very elegant answer to this question: we may appeal to a remarkable theorem[3] of von Neumann, Jordan, and Wigner [167, 28] to show that our product $\bar{\odot}$ is power associative if and only if it satisfies an identity known as the *Jordan identity*.

Theorem 7.4.1 (von Neumann, Jordan, Wigner [28])**.** *Let V be a finite dimensional real vector space equipped with a commutative bilinear product $\bar{\odot}$, not assumed associative. Further assume that V is formally real with respect to this product. Then defining powers recursively by $x^{\bar{\odot}1} = x$ and $x^{\bar{\odot}a} = x\bar{\odot}x^{\bar{\odot}(a-1)}$, the following are equivalent:*

$$\forall x \in V, a, b \in \mathbb{N}_{\geq 1} : x^{\bar{\odot}a}\bar{\odot}x^{\bar{\odot}b} = x^{\bar{\odot}(a+b)} \tag{7.46}$$

$$\iff$$

$$\forall x, y \in V : ((x\bar{\odot}x)\bar{\odot}y)\bar{\odot}x = (x\bar{\odot}x)\bar{\odot}(y\bar{\odot}x). \tag{7.47}$$

We have already shown that $x\bar{\odot}y = \Pi(x \odot y)$ defines a commutative bilinear product on $\mathrm{col}(P)$, a finite dimensional real vector space. And moreover, because $\bar{\odot}$ is self-adjoint with respect to the inner product $\langle\cdot,\cdot\rangle_P$ on $\mathrm{col}(P)$, it defines a formally real algebra. The theorem then tells us that if this algebra is power associative (Eq. (7.46)), then $\bar{\odot}$ must satisfy the Jordan identity (Eq. (7.47)), and conversely, if $\bar{\odot}$ satisfies the Jordan identity, the algebra must be power associative. Therefore we can ensure that $x^{\bar{\odot}m}$ has a univocal meaning by imposing the Jordan identity on $\bar{\odot}$. Since $\bar{\odot}$ is fixed up to the choice of P, this ultimately means putting a constraint on P itself.

7.4.1 Polarization

We would like to massage the Jordan identity into a form which makes it easy to check whether a given P induces a product $\bar{\odot}$ which satisfies it. In particular, it would be convenient if one could introduce an orthonormal basis $\{e_i\}$ for $\mathrm{col}(P)$ and check the identity on just those basis vectors. This is the right idea: but in order to formulate it correctly, we will have to *polarize* the Jordan identity.

Let $L_x y = x\bar{\odot}y$ so that L_x is the linear operator which performs $\bar{\odot}$ with x. Explicitly,

$$L_x = \Pi\Big\{\gamma D_x + (1-\gamma)(xu^\dagger + \overline{x}I - ux^\dagger)/n\Big\}, \tag{7.48}$$

where D_x is the diagonal matrix with x along its diagonal. Rearranging the Jordan identity using commutativity, we may reexpress it in terms of L_x and $L_{x^{\bar{\odot}2}}$,

$$x\bar{\odot}((x\bar{\odot}x)\bar{\odot}y) = (x\bar{\odot}x)\bar{\odot}(x\bar{\odot}y) \iff L_x L_{x^{\bar{\odot}2}} y = L_{x^{\bar{\odot}2}} L_x y. \tag{7.49}$$

[3] One of many.

Since this must hold for all $x, y \in \mathrm{col}(P)$, we see that the Jordan identity is equivalent to

$$\forall x \in \mathrm{col}(P) : \mathcal{J}(x) = [L_x, L_{x^{\bar{\odot}2}}]\Pi = 0, \tag{7.50}$$

an expression which is cubic in x. By the bilinearity of $\bar{\odot}$, we have for any scalar λ, $L_{\lambda x} = \lambda L_x$ and $L_{(\lambda x)^{\bar{\odot}2}} = \lambda^2 L_{x^{\bar{\odot}2}}$ so that $[L_{\lambda x}, L_{(\lambda x)^{\bar{\odot}2}}]\Pi = \lambda^3 [L_x, L_{x^{\bar{\odot}2}}]\Pi$. Thus $\mathcal{J}(x) = 0$ is a homogeneous cubic identity in x. A real homogeneous polynomial identity of degree d is equivalent to its full d-linear polarization: since $d = 3$, it suffices to consider the third order polarization [164]. Indeed, checking only $\forall i : [L_{e_i}, L_{e_i^{\bar{\odot}2}}]\Pi = 0$ on an orthonormal basis $\{e_i\}$ for $\mathrm{col}(P)$ is not enough. To check the Jordan identity in terms of basis vectors $\{e_i\}$, we must polarize, or linearize, the Jordan identity so that it becomes a *trilinear identity* [164]: then it suffices to check it on all triples of basis vectors. The standard procedure for doing so is the following.

Let $a, b, c \in \mathrm{col}(P)$ be arbitrary, and let $x = ta + sb + rc$. Then $L_x = tL_a + sL_b + rL_c$, and because $\bar{\odot}$ is commutative and bilinear, we have

$$L_{x^{\bar{\odot}2}} = t^2 L_{a^{\bar{\odot}2}} + s^2 L_{b^{\bar{\odot}2}} + r^2 L_{c^{\bar{\odot}2}} + 2ts L_{a\bar{\odot}b} + 2sr L_{b\bar{\odot}c} + 2rt L_{c\bar{\odot}a}, \tag{7.51}$$

and so

$$\begin{aligned}
\mathcal{J}(x) = [L_x, L_{x^{\bar{\odot}2}}]\Pi = \Big\{ & t^3 [L_a, L_{a^{\bar{\odot}2}}] + s^3 [L_b, L_{b^{\bar{\odot}2}}] + r^3 [L_c, L_{c^{\bar{\odot}2}}] \\
& + t^2 s \Big(2[L_a, L_{a\bar{\odot}b}] + [L_b, L_{a^{\bar{\odot}2}}] \Big) \\
& + t s^2 \Big([L_a, L_{b^{\bar{\odot}2}}] + 2[L_b, L_{a\bar{\odot}b}] \Big) \\
& + t^2 r \Big(2[L_a, L_{a\bar{\odot}c}] + [L_c, L_{a^{\bar{\odot}2}}] \Big) \\
& + t r^2 \Big([L_a, L_{c^{\bar{\odot}2}}] + 2[L_c, L_{a\bar{\odot}c}] \Big) \\
& + s^2 r \Big(2[L_b, L_{b\bar{\odot}c}] + [L_c, L_{b^{\bar{\odot}2}}] \Big) \\
& + s r^2 \Big([L_b, L_{c^{\bar{\odot}2}}] + 2[L_c, L_{b\bar{\odot}c}] \Big) \\
& + 2tsr \Big([L_a, L_{b\bar{\odot}c}] + [L_b, L_{a\bar{\odot}c}] + [L_c, L_{a\bar{\odot}b}] \Big) \Big\} \Pi.
\end{aligned} \tag{7.52}$$

The full third order polarization is obtained from the tsr coefficient. Indeed, define the symmetric trilinear map

$$T(a, b, c) = \frac{1}{3} \Big\{ [L_a, L_{b\bar{\odot}c}] + [L_b, L_{c\bar{\odot}a}] + [L_c, L_{a\bar{\odot}b}] \Big\} \Pi. \tag{7.53}$$

Then $T(x,x,x) = [L_x, L_{x^{\bar{\odot}2}}]\Pi = \mathcal{J}(x)$, showing that T is the full third-order polarization of the homogeneous cubic map $\mathcal{J}(x)$. We now show that $\forall a,b,c \in \mathrm{col}(P) : T(a,b,c) = 0$ is equivalent to the original Jordan identity. On the one hand, suppose the Jordan identity holds. Then $\forall x \in \mathrm{col}(P) : [L_x, L_{x^{\bar{\odot}2}}]\Pi = 0$. Since the expression in Eq. (7.52) is a polynomial in t,s,r, and it must vanish for any choice of t,s,r, every coefficient in fact must be identically zero. In particular, this means that the tsr coefficient vanishes. Thus $T(a,b,c) = 0$. Since a,b,c are arbitrary, the Jordan identity therefore implies $\forall a,b,c \in \mathrm{col}(P) : T(a,b,c) = 0$. Conversely, suppose $\forall a,b,c \in \mathrm{col}(P) : T(a,b,c) = 0$. Then in particular, $\forall x \in \mathrm{col}(P) : T(x,x,x) = [L_x, L_{x^{\bar{\odot}2}}]\Pi = 0$, so the Jordan identity holds. We conclude that $\forall x \in \mathrm{col}(P) : \mathcal{J}(x) = 0$ is equivalent to $\forall a,b,c \in \mathrm{col}(P) : T(a,b,c) = 0$. The benefit of the latter trilinear identity is that if $\{e_i\}$ is an orthonormal basis and $a = \sum_i a_i e_i, b = \sum_i b_i e_i, c = \sum_i c_i e_i$, we have

$$T(a,b,c) = \sum_{ijk} a_i b_j c_k T(e_i, e_j, e_k). \tag{7.54}$$

Since T is trilinear, $\forall a,b,c \in \mathrm{col}(P) : T(a,b,c) = 0$ if and only if

$$\forall i,j,k : T(e_i, e_j, e_k) = 0. \tag{7.55}$$

Instead of using an orthonormal basis $\{e_i\}$ for $\mathrm{col}(P)$, we could simply use the columns of P itself (or just r linearly independent columns of P). We may then express the Jordan identity as a relation, while complicated, entirely among the matrix elements of $P(R|R)$.

Theorem 7.4.2. *Let $P \equiv P(R|R)$, $\Phi = \alpha I + (1-\alpha)J/n$, and let*

$$L_x = P\Phi\Big\{\gamma D_x + (1-\gamma)(xu^\dagger + \overline{x}I - ux^\dagger)/n\Big\}, \tag{7.56}$$

where $\gamma = (1 + 1/\alpha)/2$ so that $L_x y = x\bar{\odot}y$. Finally let

$$T(a,b,c) = \frac{1}{3}\Big\{[L_a, L_{b\bar{\odot}c}] + [L_b, L_{c\bar{\odot}a}] + [L_c, L_{a\bar{\odot}b}]\Big\}P\Phi. \tag{7.57}$$

Then $\bar{\odot}$ satisfies the Jordan identity iff

$$\forall i,j,k : T(P(R|R_i), P(R|R_j), P(R|R_k)) = 0. \tag{7.58}$$

7.4.2 Euclidean Jordan algebras

We arrived at the Jordan identity by first minimally modifying the classical product $x \circ y$ on reference valuations to the nonclassical product $x\bar{\odot}y$. Inspired by the relationship between the law of total probability and the fundamental nonclassical coherence rule, we conceived a hope: that both $x \circ y$ and $x\bar{\odot}y$ would correspond to equivalent valuations on

some measurements, and that while the former would depend on the choice of reference measurement, the latter would not. We then hoped that for some class of measurements,

$$\sum_i \mathfrak{x}_i^m P(X_i|\rho) = \sum_i x_i^{\bar{\odot} m} P(R_i|\rho), \tag{7.59}$$

that is, that $x^{\bar{\odot} m} = \mathfrak{x}^{\circ m} P(X|R)\Phi$. In assuming this, we were guided by simplicity: the reference-independent measurement with an equivalent valuation to $x^{\bar{\odot} m}$ ought to be $\{X_i\}$ itself, the equivalent valuation being $\mathfrak{x}^{\circ m}$. For this to have any hope of being realized, however, $\bar{\odot}$ would have to be power associative: otherwise $x^{\bar{\odot} m}$ would have no unambiguous meaning. But since the bilinear product $\bar{\odot}$ defines a real finite dimensional formally real commutative algebra, assuming power associativity is equivalent to assuming that $\bar{\odot}$ satisfies the Jordan identity. Finally, we were able to express this identity entirely in terms of constraints on the matrix $P(R|R)$ which characterized the reference measurement itself. Thus $\bar{\odot}$ defines a formally real or *Euclidean* Jordan algebra on $\mathrm{col}(P)$.

As we will see, the supposition that reference valuations form a Euclidean Jordan algebra places remarkably profound constraints on the state and effect spaces compatible with the reference measurement. Indeed, it means that our hopes are already realized: not only will there exist measurements with valuations equivalent to $x \circ y$ and $x \bar{\odot} y$, the latter being reference measurement independent, but also there will exist measurements whose moments satisfy Eq. (7.59). In fact, those measurements will turn out to be implied entirely by the assignment of the reference valuation x, and their effects will correspond to a set of *mutually orthogonal idempotents* of the Jordan algebra. States and effects will turn out to correspond to squares $x^{\bar{\odot} 2}$ of reference valuations, and in fact Euclidean Jordan algebras are self-dual, so that the state cone $\mathcal{S}$ will be equal to the effect cone $\mathcal{E}$. Finally, we avail ourselves of the seminal result of Jordan, von Neumann, and Wigner: the classification of the Euclidean Jordan algebras [28, 27, 164, 168]. Remarkably, all finite dimensional Euclidean Jordan algebras are isomorphic to direct sums of the so-called simple Euclidean Jordan algebras. These fundamental building blocks come in just a few shapes and sizes:

- $\mathrm{Herm}_d(\mathbb{R})$, the algebra of $d \times d$ self-adjoint real matrices with product $a \star b = \frac{1}{2}(ab + ba)$.
- $\mathrm{Herm}_d(\mathbb{C})$, the algebra of $d \times d$ self-adjoint complex matrices with product $a \star b = \frac{1}{2}(ab + ba)$.
- $\mathrm{Herm}_d(\mathbb{H})$, the algebra of $d \times d$ self-adjoint quaternionic matrices with product $a \star b = \frac{1}{2}(ab + ba)$.
- $\mathrm{Herm}_3(\mathbb{O})$, the algebra of 3×3 self-adjoint octonionic matrices with product $a \star b = \frac{1}{2}(ab + ba)$.
- The spin factors $\mathbb{R} \oplus \mathbb{R}^n$ with product $(t, x) \star (t', x') = (tt' + \langle x, x' \rangle, tx' + t'x)$.

We are thus well on our way along the royal road [22] to quantum mechanics. We now turn to unpacking the consequences of Jordan identity; then we will try to narrow our focus to $\mathrm{Herm}_d(\mathbb{C})$ specifically. But already we have learned something crucial to the QBist project: Euclidean Jordan algebras may be straightforwardly motivated by considering what is the most natural nonclassical product one can define on reference valuations. By centering the interplay between the reference measurement and the measurements for which they form a reference, we hope to have given new meaning to the formal algebra of observables defined by Jordan et al.

7.5 Lifting the product

Before we move on, however, it will be useful in the sequel to lift our product $\bar{\odot}$, which is defined on valuations, to a product on probability vectors and response functions. Recall that earlier we defined the inner product $\langle \cdot, \cdot \rangle_P$ so that

$$\sum_i \mathfrak{x}_i P(X_i|\rho) = \mathfrak{x} P(X|R)\Phi P(R|\rho) = \Big\{\mathfrak{x} P(X|R)\Phi\Big\} P\Big\{\Phi P(R|\rho)\Big\} = x^\dagger P y = \langle x, y\rangle_P, \tag{7.60}$$

where $x^\dagger = \mathfrak{x} P(X|R)\Phi$ and $y = \Phi P(R|\rho)$, and demanded that $u^\dagger(x \bar{\odot} y) = \langle x, y\rangle_P$. Moreover, $\langle (P(X_i|R)\Phi)^\dagger, \Phi P(R|\rho)\rangle_P = P(X_i|\rho)$ reproduces the fundamental nonclassical coherence relation. At the same time, however, we could define an inner product on $\mathrm{col}(P)$, $\langle \cdot, \cdot \rangle_\Phi$ such that

$$\sum_i \mathfrak{x}_i P(X_i|\rho) = \Big\{\mathfrak{x} P(X|R)\Big\}\Phi\Big\{P(R|\rho)\Big\} = g^\dagger \Phi h = \langle g, h\rangle_\Phi, \tag{7.61}$$

which reproduces $P(X_i|\rho) = \langle P(X_i|R)^\dagger, P(R|\rho)\rangle_\Phi$. One way of putting it is that from

$$P(X_i|\rho) = \langle (P(X_i|R)\Phi)^\dagger, \Phi P(R|\rho)\rangle_P, \tag{7.62}$$

$\langle \cdot, \cdot \rangle_P$ defines an inner product on *dual coordinates* while from

$$P(X_i|\rho) = \langle P(X_i|R)^\dagger, P(R|\rho)\rangle_\Phi, \tag{7.63}$$

$\langle \cdot, \cdot \rangle_\Phi$ defines an inner product on *regular coordinates*. Thus implicitly we defined $\bar{\odot}$ on dual coordinates: we would thus like to lift it to a product $\hat{\odot}$ on regular coordinates, and thus on response functions and probability vectors directly.

Now if we have dual coordinates $x = \Phi h$, then acting with P gives us regular coordinates $h = Px$ since $h = P\Phi h$ for $h \in \mathrm{col}(P)$. Similarly, if $y^\dagger = g^\dagger \Phi$, then $g^\dagger = y^\dagger P$.

We are thus led to consider

$$g\hat{\odot}h = P(\Phi g\bar{\odot}\Phi h) = P\Pi(\Phi g \odot \Phi h) = P(\Phi g \odot \Phi h), \tag{7.64}$$

where since $P\Pi = PP\Phi = P\Phi P = P$ it is inessential whether we use the original product $\odot$ or the projected product $\bar{\odot}$. Just as $\odot$ is compatible with $\langle\cdot,\cdot\rangle_P$, $\hat{\odot}$ is compatible with $\langle\cdot,\cdot\rangle_\Phi$. On the one hand,

$$u^\dagger(g\hat{\odot}h) = u^\dagger P(\Phi g \odot \Phi h) = u^\dagger(\Phi g \odot \Phi h) \tag{7.65}$$

$$= \langle \Phi g, \Phi h\rangle_P = g^\dagger \Phi P \Phi h = g^\dagger \Phi h = \langle g, h\rangle_\Phi, \tag{7.66}$$

and on the other hand,

$$\langle f, g\hat{\odot}h\rangle_\Phi = f^\dagger \Phi P(\Phi g \odot \Phi h) = \langle \Phi f, \Phi g \odot \Phi h\rangle_P \tag{7.67}$$

$$= \langle \Phi f \odot \Phi g, \Phi h\rangle_P = (\Phi f \odot \Phi g)^\dagger P \Phi h = \langle f\hat{\odot}g, h\rangle_\Phi, \tag{7.68}$$

so that $\hat{\odot}$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle_\Phi$. Working out the explicit form for $\hat{\odot}$, we find the following.

Lemma 7.5.1. *Lifting $\odot$ from a product on dual coordinates to a product $\hat{\odot}$ on regular coordinates gives*

$$g\hat{\odot}h = \frac{1}{2}\left\{\alpha(\alpha+1)P(g\circ h) - (\alpha-1)\left(\overline{h}g + \overline{g}h + \alpha(g\cdot h)u\right)/n.\right\} \tag{7.69}$$

Moreover, let $\hat{L}_g$ be the linear operator such that $\hat{L}_g h = g\hat{\odot}h$. It may be expressed

$$\hat{L}_g = PL_{\Phi g}\Phi = \frac{1}{2}\left\{\alpha(\alpha+1)PD_g - (\alpha-1)\left(gu^\dagger + \overline{g}\,I + \alpha u g^\dagger\right)/n\right\}. \tag{7.70}$$

Proof. We had $L_x y = x\bar{\odot}y$ where

$$L_x = \Pi\left\{\gamma D_x + (1-\gamma)(xu^\dagger + \overline{x}I - ux^\dagger)/n\right\}. \tag{7.71}$$

Let $x = \Phi g = \alpha g + (1-\alpha)\overline{g}\,u/n$ so that $D_x = D_{\Phi g} = \alpha D_g + (1-\alpha)\overline{g}\,I/n$. Moreover, notice that $\Phi g u^\dagger - u g^\dagger \Phi = \alpha(gu^\dagger - ug^\dagger)$. Using $\gamma = (1+1/\alpha)/2$, we arrive at

$$L_{\Phi g} = \frac{1}{2}\Pi\left\{(\alpha+1)D_g + (\alpha-1)\left(gu^\dagger - ug^\dagger - \overline{g}\,I\right)/n\right\}, \tag{7.72}$$

so that

$$\hat{L}_g = PL_{\Phi g}\Phi = \frac{1}{2}\left\{(\alpha+1)PD_g\Phi + (\alpha-1)P\left(gu^\dagger - ug^\dagger - \overline{g}\,I\right)\Phi/n\right\}. \tag{7.73}$$

Now $PD_g\Phi = \alpha PD_g + (1-\alpha)Pg\,u^\dagger/n$ and $Pg = (1/\alpha)g + (1-1/\alpha)\overline{g}\,u/n$ which yields

$$PD_g\Phi = \alpha PD_g - (\alpha-1)(1/\alpha)gu^\dagger/n - (\alpha-1)^2(1/\alpha)\overline{g}\,J/n^2. \tag{7.74}$$

Meanwhile

$$P\left(gu^\dagger - ug^\dagger - \overline{g}\,I\right)\Phi = Pg\,u^\dagger - ug^\dagger\Phi - \overline{g}\,P\Phi \tag{7.75}$$

$$= (1/\alpha)gu^\dagger - \alpha ug^\dagger - \overline{g}\,P\Phi + (\alpha^2-1)(1/\alpha)\overline{g}\,J/n. \tag{7.76}$$

We conclude that

$$\hat{L}_g = PL_{\Phi g}\Phi = \frac{1}{2}\left\{\alpha(\alpha+1)PD_g - (\alpha-1)\left(gu^\dagger + \alpha ug^\dagger + \overline{g}\,P\Phi\right)/n\right\}. \tag{7.77}$$

By assumption $\hat{L}_g$ acts on vectors in $\mathrm{col}(P)$ so we can substitute $P\Phi \to I$, from which the result follows. □

Finally, we may formulate the Jordan identity in terms of $\hat{\odot}$ instead of $\bar{\odot}$. First let $\mathcal{L}_g = \hat{L}_g P$ so that $\mathcal{L}_g\Phi h = \hat{L}_g P\Phi h = \hat{L}_g h = g\hat{\odot}h$ for $h \in \mathrm{col}(P)$. We have

$$\mathcal{L}_g = \frac{1}{2}\left\{\alpha(\alpha+1)PD_gP - (\alpha-1)\left(gu^\dagger P + \alpha ug^\dagger P + \overline{g}\,P\right)/n\right\} \tag{7.78}$$

$$= \frac{1}{2}\left\{\alpha(\alpha+1)PD_gP - (\alpha-1)\left(gu^\dagger + ug^\dagger + \overline{g}\,P + (\alpha-1)\overline{g}J/n\right)/n\right\}. \tag{7.79}$$

Notice that unlike L_g, $\mathcal{L}_g = \mathcal{L}_g^\dagger$ so that the latter is self-adjoint in the usual sense. Moreover, we note in passing that since $P^2 = \frac{1}{\alpha}P + (1-1/\alpha)J/n$, we have

$$\mathcal{L}_u = \frac{1}{2}\left\{\alpha(\alpha+1)P^2 - (\alpha-1)\left(uu^\dagger + uu^\dagger + nP + (\alpha-1)J/n\right)/n\right\} \tag{7.80}$$

$$= \frac{1}{2}\left\{\alpha(\alpha+1)\left[\frac{1}{\alpha}P + (1-1/\alpha)J/n\right] - (\alpha-1)\left(nP + (\alpha+1)J\right)/n\right\} \tag{7.81}$$

$$= \frac{1}{2}\left\{2P + 2(\alpha-1)J/n - 2(\alpha-1)J/n\right\} = P, \tag{7.82}$$

so that $\mathcal{L}_u\Phi = \Pi$, which indeed acts as the identity on $\mathrm{col}(P)$. Finally, consider in particular $\mathcal{L}_{P(R|R_k)}$. Treated as a three index tensor, this object fully defines $\hat{\odot}$. Explicitly,

$$[\mathcal{L}_P]_{ijk} = \frac{1}{2}\Bigg\{\alpha(\alpha+1)\sum_m P(R_m|R_i)P(R_m|R_j)P(R_m|R_k) \\ - (\alpha-1)\Big(P(R_i|R_k) + P(R_j|R_k) + P(R_i|R_j) + (\alpha-1)/n\Big)/n\Bigg\}. \tag{7.83}$$

Then $[\mathcal{L}_g]_{ij} = \sum_{kl}[\mathcal{L}_P]_{ijk}\Phi_{kl}g_l$ so that

$$[g\hat{\odot}h]_i = \sum_{jklm}[\mathcal{L}_{P(R|R)}]_{ijk}\Phi_{kl}\Phi_{jm}g_l h_m, \tag{7.84}$$

and

$$[P(R|R_j)\hat{\odot}P(R|R_k)]_i = \sum_{ab}[\mathcal{L}_P]_{ajk}\Phi_{ab}P(R_b|R_i), \tag{7.85}$$

or equivalently, $P(R|R_j)\hat{\odot}P(R|R_k) = \sum_i C_{ijk}P(R|R_i)$. Here $C_{ijk} = \sum_a[\mathcal{L}_P]_{ajk}\Phi_{ai}$ encodes the structure coefficients for $\hat{\odot}$, and we may formulate the Jordan identity in terms of them.

Lemma 7.5.2. *Let $C_{ijk} = \sum_a[\mathcal{L}_P]_{ajk}\Phi_{ai}$ be the structure coefficients for $\hat{\odot}$. Then $\hat{\odot}$ satisfies the Jordan identity iff*

$$\forall t,i,j,k,\ell : \sum_{rs} C_{rij}\left(C_{sk\ell}C_{trs} - C_{sr\ell}C_{tks}\right) \\ + \sum_{rs} C_{rjk}\left(C_{si\ell}C_{trs} - C_{sr\ell}C_{tis}\right) \\ + \sum_{rs} C_{rki}\left(C_{sj\ell}C_{trs} - C_{sr\ell}C_{tjs}\right) = 0. \tag{7.86}$$

Proof. Analogously to our earlier construction, using the third order polarization, $\hat{\odot}$ satisfies the Jordan identity iff

$$\forall a,b,c \in \mathrm{col}(P) : \frac{1}{3}\Big\{[\mathcal{L}_a\Phi, \mathcal{L}_{b\hat{\odot}c}\Phi] + [\mathcal{L}_b\Phi, \mathcal{L}_{c\hat{\odot}a}\Phi] + [\mathcal{L}_c\Phi, \mathcal{L}_{a\hat{\odot}b}\Phi]\Big\} = 0, \tag{7.87}$$

where we note that unlike in Eq. (7.53) there is no need to insert a trailing projector Π since $\mathcal{L}_g\Phi\Pi = (\hat{L}_g P)\Phi P\Phi = \hat{L}_g P\Phi = \mathcal{L}_g\Phi$. Using trilinearity, we can formulate the Jordan

identity entirely in terms of $P(R|R)$ as

$$\begin{aligned}\forall i,j,k : &[\mathcal{L}_{P(R|R_i)\hat{\odot}P(R|R_j)}\Phi, \mathcal{L}_{P(R|R_k)}\Phi] \\ &+ [\mathcal{L}_{P(R|R_j)\hat{\odot}P(R|R_k)}\Phi, \mathcal{L}_{P(R|R_i)}\Phi] \\ &+ [\mathcal{L}_{P(R|R_k)\hat{\odot}P(R|R_i)}\Phi, \mathcal{L}_{P(R|R_j)}\Phi] = 0. \end{aligned} \tag{7.88}$$

Since $P(R|R_j)\hat{\odot}P(R|R_k) = \sum_i C_{ijk}P(R|R_i)$, we have $\mathcal{L}_{P(R|R_j)\hat{\odot}P(R|R_k)} = \sum_i C_{ijk}\mathcal{L}_{P(R|R_i)}$ and so

$$\begin{aligned}\forall i,j,k : \sum_r \Big\{ &C_{rij}[\mathcal{L}_{P(R|R_r)}\Phi, \mathcal{L}_{P(R|R_k)}\Phi] \\ &+ C_{rjk}[\mathcal{L}_{P(R|R_r)}\Phi, \mathcal{L}_{P(R|R_i)}\Phi] \\ &+ C_{rik}[\mathcal{L}_{P(R|R_r)}\Phi, \mathcal{L}_{P(R|R_j)}\Phi] \Big\} = 0. \end{aligned} \tag{7.89}$$

To further resolve this expression, we will apply the LHS to some $P(R|R_\ell)$. First,

$$\mathcal{L}_{P(R|R_k)}\Phi P(R|R_\ell) = P(R|R_k)\hat{\odot}P(R|R_\ell) = \sum_s C_{sk\ell}P(R|R_s), \tag{7.90}$$

so that

$$\mathcal{L}_{P(R|R_r)}\Phi\mathcal{L}_{P(R|R_k)}\Phi P(R|R_\ell) = \sum_{st} C_{sk\ell}C_{trs}P(R|R_t), \tag{7.91}$$

$$\mathcal{L}_{P(R|R_k)}\Phi\mathcal{L}_{P(R|R_r)}\Phi P(R|R_\ell) = \sum_{st} C_{sr\ell}C_{tks}P(R|R_t), \tag{7.92}$$

which leads to

$$[\mathcal{L}_{P(R|R_r)}\Phi, \mathcal{L}_{P(R|R_k)}\Phi]P(R|R_\ell) = \sum_{st} \left(C_{sk\ell}C_{trs} - C_{sr\ell}C_{tks}\right) P(R|R_t). \tag{7.93}$$

Rewriting all the commutators in this way, we arrive at

$$\begin{aligned}\forall i,j,k,\ell : \sum_t \Bigg[&\sum_{rs} C_{rij}\left(C_{sk\ell}C_{trs} - C_{sr\ell}C_{tks}\right) \\ &+ \sum_{rs} C_{rjk}\left(C_{si\ell}C_{trs} - C_{sr\ell}C_{tis}\right) \\ &+ \sum_{rs} C_{rki}\left(C_{sj\ell}C_{trs} - C_{sr\ell}C_{tjs}\right) \Bigg] P(R|R_t) = 0. \end{aligned} \tag{7.94}$$

The expression in the brackets is a vector v indexed by t. Therefore we can reexpress the constraint simply as $Pv = 0$. But from the form of $C_{ijk} = \sum_a [\mathcal{L}_P]_{ajk} \Phi_{ai}$, and in particular the expression for $[\mathcal{L}_P]_{ijk}$, we see that $v \in \mathrm{col}(P)$ automatically. If $Pv = 0$ and $v \notin \mathrm{null}(P)$, we must have $v = 0$, from which the result follows. □

7.6 Self-duality

Now that we know that our nonclassical product defines a Euclidean Jordan algebra, we may appeal to many powerful results. For instance, the famed Koecher-Vinberg theorem [64, 27, 169] provides an alternative, geometric characterization of a Euclidean Jordan algebra: the squares of elements of an EJA form a *symmetric cone*, essentially a convex cone which is *homogeneous* and *self-dual*, and conversely any symmetric cone arises from the squares of the elements of some Euclidean Jordan algebra. Homogeneity tells us that there is an invertible transformation which takes any element in the interior of the cone to any other element while preserving the ordering of elements: in other words, the automorphism group of the cone acts transitively[4]. Meanwhile, let C be the cone of squares of elements of an EJA. Then self-duality means that if one considers C^*, the dual cone of C, that is, the space of linear functionals nonnegative on the cone, then $C = C^*$. For us, this means that we can interpret squares of elements of our EJA precisely as states or equivalently, effects. Indeed, if

$$P(E|R) = x \hat{\odot} x \qquad\qquad P(R|\rho) = y \hat{\odot} y, \tag{7.95}$$

then

$$P(E|\rho) = P(E|R)\Phi P(R|\rho) = \langle P(E|R), P(R|\rho) \rangle_\Phi = \langle x^{\hat{\odot} 2}, y^{\hat{\odot} 2} \rangle_\Phi \geq 0, \tag{7.96}$$

where the nonnegativity follows from self-duality. Indeed, the same principle guarantees that e.g., $P(R|\rho)$ is indeed a probability vector since by the same token,

$$P(R_i|\rho) = P(R_i|R)\Phi P(R|\rho) = \langle P(R_i|R), P(R|\rho) \rangle_\Phi = \langle r_i^{\hat{\odot} 2}, y^{\hat{\odot} 2} \rangle_\Phi \geq 0. \tag{7.97}$$

Similarly, in terms of $\bar{\odot}$, if

$$\Phi P(E|R)^\dagger = x \bar{\odot} x \qquad\qquad \Phi P(R|\rho) = y \bar{\odot} y, \tag{7.98}$$

then

$$P(E|\rho) = \big\{P(E|R)\Phi\big\} P\big\{\Phi P(R|\rho)\big\} = \langle \Phi P(E|R)^\dagger, \Phi P(R|\rho) \rangle_P = \langle x^{\bar{\odot} 2}, y^{\bar{\odot} 2} \rangle_P \geq 0. \tag{7.99}$$

[4] In quantum theory, this corresponds to the fact that we can map any state ρ to any state σ by a suitable Kraus operator: $K = \sigma^{1/2} \rho^{-1/2}$, where ρ, σ are full-rank.

7.6.1 The uncertainty principle

Letting $\mathfrak{x}$ be a valuation on a measurement $\{X_i\}$ and $x = \mathfrak{x}P(X|R)\Phi$ be the equivalent valuation on the reference measurement, we supposed that

$$\sum_i \mathfrak{x}_i^2 P(X_i|\rho) = \sum_i (x \odot x)_i P(R_i|\rho) \tag{7.100}$$

for some class of measurements. More concisely,

$$\langle \mathfrak{x}^2 \rangle_\rho = \langle x^{\odot 2} \rangle_\rho = \gamma \langle x^2 \rangle_\rho + (1-\gamma)\Big(2\langle x \rangle_\mu \langle x \rangle_\rho - \langle x^2 \rangle_\mu\Big), \tag{7.101}$$

where e.g., $\langle x^2 \rangle_\rho = \sum_i x_i^2 P(R_i|\rho)$ and $\forall i : P(R_i|\mu) = 1/n$[5]. We will soon show how the structure of a Euclidean Jordan algebra itself will pick out exactly for which class of measurements this relation holds. But at the moment, let us take it for granted, and notice that because the second moment of a random variable is always nonnegative $\langle \mathfrak{x}^2 \rangle_\rho \geq 0$, we immediately have a lower bound on the second moment of x with respect to the reference,

$$\langle x^2 \rangle_\rho \geq \frac{\alpha - 1}{\alpha + 1}\Big(\langle x^2 \rangle_\mu - 2\langle x \rangle_\mu \langle x \rangle_\rho\Big). \tag{7.102}$$

Since this should hold for any $x \in \mathrm{col}(P)$, we therefore have a restriction on probability assignments $P(R|\rho)$ on reference outcomes. In fact, this restriction follows entirely from the fact that the cone of squares of the EJA is self-dual without separately assuming $\langle \mathfrak{x}^2 \rangle_\rho = \langle x^{\odot 2} \rangle_\rho$, and in fact is a necessary and sufficient condition for $P(R|\rho)$ to be a valid state. Indeed, for $P(R|\rho) \in \mathrm{col}(P)$,

$$(x \odot x)^\dagger P(R|\rho) = (x \odot x)^\dagger P \Phi P(R|\rho) = \langle x^{\odot 2}, \Phi P(R|\rho) \rangle_P. \tag{7.103}$$

By self-duality, $\langle x^{\odot 2}, \Phi P(R|\rho) \rangle_P \geq 0$ iff $\Phi P(R|\rho) = y^{\bar{\odot} 2}$ for some y, that is, $P(R|\rho)$ is a valid state. (Notice we've used $\bar{\odot}$, the projected product, since by assumption $P(R|\rho) \in \mathrm{col}(P)$). Thus without even specifying anything about the nature of the measurement $\{X_i\}$, we have the constraint $\langle x^{\odot 2} \rangle_\rho \geq 0$. We can therefore fully characterize the state space via

$$P(R|\rho) \in \Delta_n \cap \mathrm{col}(P) \text{ valid} \iff \forall x \in \mathrm{col}(P) : \langle x^2 \rangle_\rho \geq \frac{\alpha - 1}{\alpha + 1}\Big(\langle x^2 \rangle_\mu - 2\langle x \rangle_\mu \langle x \rangle_\rho\Big), \tag{7.104}$$

Thinking of the lower bound as a bound on the variance of any observable, we see the shape of the state space is determined entirely by a kind of *uncertainty principle*: to respect this

[5] Indeed, a Euclidean Jordan algebra picks out a distinguished state $P(R|\mu)$ which is proportional to the unit of the algebra.

uncertainty principle, probability distributions on reference outcomes cannot be too sharp in a prescribed way.

But how can we check straightforwardly whether a given distribution $p \equiv P(R|\rho)$ is in fact valid? Notice that we can rewrite the lower bound in matrix form as

$$\forall x \in \mathrm{col}(P) : \sum_i (x \odot x)_i p_i = x^\dagger \Big\{ \gamma D_p + (1-\gamma)\big(up^\dagger + pu^\dagger - \overline{p} I\big)/n \Big\} x \tag{7.105}$$

$$= x^\dagger V_p x \geq 0, \tag{7.106}$$

so that the validity of $p \in \Delta_n \cap \mathrm{col}(P)$ amounts to the positive semidefiniteness of $V_p = V_p^\dagger$ on $\mathrm{col}(P)$. Thus if we let $\mathcal{L}_p = \Pi V_p \Pi$, where Π is the projector onto $\mathrm{col}(P)$, then the validity of p corresponds to the positive semidefiniteness of $\mathcal{L}_p$ in the full sense, so that

$$p \in \Delta_n \cap \mathrm{col}(P) \text{ valid} \iff \mathcal{L}_p \geq 0. \tag{7.107}$$

In this way, the positive semidefiniteness of $\mathcal{L}_p$ allows one to diagnose whether a given distribution p is consistent with the reference measurement.

The astute reader may have noticed that we are using the same symbol $\mathcal{L}$ as in the previous section where we defined $\mathcal{L}_x$ to act as $\mathcal{L}_x \Phi y = x \hat{\odot} y$. This is no accident! Indeed, let

$$V_x = \gamma D_x + (1-\gamma)\big(ux^\dagger + xu^\dagger - \overline{x} I\big)/n, \tag{7.108}$$

which coincides with the previous definition when $\overline{x} = 1$ as it does for probability vectors. Now on the one hand, we earlier defined

$$\mathcal{L}_x = \hat{L}_x P = P L_{\Phi x} \Phi P \tag{7.109}$$

$$= \frac{1}{2} P \left\{ (\alpha+1) D_x + (\alpha - 1) \left(xu^\dagger - ux^\dagger - \overline{x}\, I \right) / n \right\} \Phi P, \tag{7.110}$$

where we have used the expression we already worked out for $L_{\Phi x}$ in Eq. 7.72. On the other hand, we now propose that

$$\mathcal{L}_x = \Pi V_x \Pi = P\Phi \Big\{ \gamma D_x + (1-\gamma)(ux^\dagger + xu^\dagger - \overline{x} I)/n \Big\} P\Phi : \tag{7.111}$$

indeed, one can easily verify that the two expressions are the same by working out ΦV_x. Thus $\mathcal{L}_x$ has two interpretations. On the one hand, $\mathcal{L}_x \Phi y = x \hat{\odot} y$ in fact performs the Jordan product itself. On the other hand, the positive semidefiniteness of $\mathcal{L}_{P(R|\rho)}$ diagnoses the validity of a distribution $P(R|\rho)$ with respect to the reference measurement. This applies not just to states but also to effects. In fact, from

$$P(E|R)^\dagger y^{\odot 2} = P(E|R)^\dagger \Phi P y^{\odot 2} = \langle \Phi P(E|R), y^{\odot 2} \rangle_P \geq 0 \text{ iff } \Phi P(E|R) = x \bar{\odot} x, \tag{7.112}$$

we see that exactly the same argument applies. Indeed, by self-duality, the positive semidefiniteness of $\mathcal{L}_{P(E|R)}$ diagnoses the validity of $P(E|R)$.

7.6.2 The self-duality constant

Indeed, by self-duality, states can be rescaled into effects and vice versa. Defining the *self-duality constant* allows us to introduce a rescaled product $\check{\odot}$, which will prove particularly useful when we discuss idempotents of the Jordan algebra. To begin, let $P(R|\rho)$ be a state. Self-duality implies there must be a corresponding effect $P(\rho|R)$. By Bayes's theorem,

$$P(\rho|R_i) = \frac{P(R_i|\rho)P(\rho)}{P(R_i)} = \kappa P(R_i|\rho). \tag{7.113}$$

In particular, let us apply this to the reference states and effects. For clarity, we will write $P(R_i|R_j) \equiv P(R_i|\sigma_j)$ where σ_j denotes a reference state. Considering when $i = j$, we have

$$P(\sigma_i|R_i) = \kappa P(R_i|\sigma_i), \tag{7.114}$$

where we will call κ the *self-duality constant*. Now we supposed that $P(R_i|\sigma_i)$ is a constant, and in Corollary 7.2.2.1, we fixed its value to $P(R_i|\sigma_i) = (1 + (r-1)/\alpha)/n$. Let us further suppose that in fact $P(\sigma_i|R_i) = 1$: conditional on obtaining outcome $\{R_i\}$ on the reference measurement, if we then perform a measurement among whose outcomes is σ_i, we will be certain of getting that outcome. In fact, requiring that each reference state be certain for its corresponding effect fixes the value of κ,

$$\kappa = \frac{n}{1 + (r-1)/\alpha}, \tag{7.115}$$

where $r = \text{rank}(P)$. By linearity then we have in general, that

$$P(\rho|R) = \kappa P(R|\rho)^\dagger. \tag{7.116}$$

In fact, this places a constraint on all states $P(R|\rho)$ such that $P(\rho|\rho) = 1$.

$$P(\rho|\rho) = P(\rho|R)\Phi P(R|\rho) \tag{7.117}$$

$$= \kappa\Big\{\alpha P(R|\rho)^\dagger P(R|\rho) + (1-\alpha)/n\Big\} = 1, \tag{7.118}$$

from which it follows that

$$\sum_i P(R_i|\rho)^2 = \frac{\alpha^2 + r - 1}{\alpha^2 n}. \tag{7.119}$$

At the same time,

$$P(\rho|\rho) = \langle P(\rho|R)^\dagger, P(R|\rho)\rangle_\Phi \tag{7.120}$$
$$= u^\dagger(P(\rho|R)^\dagger \hat{\odot} P(R|\rho)) \tag{7.121}$$
$$= u^\dagger(\kappa P(R|\rho)\hat{\odot}P(R|\rho)). \tag{7.122}$$

We are thus motivated to introduce a rescaled product $\check{\odot} = \kappa\hat{\odot}$ and a rescaled inner product $\langle\cdot,\cdot\rangle_{\kappa\Phi}$ such that

$$P(\rho|\rho) = u^\dagger(P(R|\rho)\check{\odot}P(R|\rho)) \tag{7.123}$$
$$= \langle P(R|\rho), P(R|\rho)\rangle_{\kappa\Phi} = \|P(R|\rho)\|^2_{\kappa\Phi}, \tag{7.124}$$

which will turn out to be the right tool to use when discussing idempotents of the Jordan product.

7.7 Idempotents

In a Euclidean Jordan algebra, idempotents $x = x\bar{\odot}x$ of the Jordan product play a crucial role. Many key results from standard linear algebra may be ported to the Euclidean Jordan algebraic setting—in particular, the spectral theorem. We say two idempotents are orthogonal iff $x\bar{\odot}y = 0$. A *primitive idempotent* [27, 164, 168] is a nonzero element e of the algebra such that $e\bar{\odot}e = e$ and which cannot be expressed as a sum of two (or more) nonzero orthogonal idempotents. That is, an idempotent is primitive iff there does *not* exist any x, y such that $e = x + y$ and $x\bar{\odot}y = 0$. Euclidean Jordan algebras enjoy a *spectral theorem*: any element x of the algebra may be decomposed as a real linear combination of mutually orthogonal primitive idempotents,

$$x = \sum_i \lambda_i e_i \qquad\qquad e_i\bar{\odot}e_j = \delta_{ij}e_i \qquad\qquad \sum_i e_i = u \tag{7.125}$$

where we may call the λ_i's eigenvalues, and u is the unit of the algebra, such that $\forall x : u\bar{\odot}x = x$. Not only that, but every element can be expressed as a linear combination of no more than d mutually orthogonal primitive idempotents, where d is called the *rank* of the EJA. We may define the *trace* of an element $\mathrm{tr}(x)$, and indeed, $\mathrm{tr}(x) = \sum_i \lambda_i$ is the sum of the eigenvalues[6]. In particular, the only allowed eigenvalues for idempotents are 0 and 1, and so $\mathrm{tr}(e) = 1$ for any primitive idempotent. It follows also that any square $x\bar{\odot}x$ must have all nonnegative eigenvalues, and conversely any element with nonnegative eigenvalues can be expressed as a square. The squares form the *nonnegative cone*, $C = \{x \odot x : x \in V\}$ and this cone is self-dual with respect to the trace inner product, so that we have as well

[6]For us $\mathrm{tr}(x) = u^\dagger x$, and so e.g., $P(\rho|\rho) = \mathrm{tr}(P(R|\rho)\check{\odot}P(R|\rho))$.

that $C = \{y \in V : \forall x \in C : \mathrm{tr}(x \bar{\odot} y) \geq 0\}$. Clearly all idempotents are in the nonnegative cone: in particular, the primitive idempotents generate its extremal rays. Finally, we have observed a consequence of these considerations is that in a Euclidean Jordan algebraic theory, probabilities are given by the trace inner product of two nonnegative elements, one representing a state and the other an effect.

We briefly sketch how an element x of a Euclidean Jordan algebra can be spectrally decomposed [27, 168, 164]. Defining $x^{\bar{\odot}0} = u$, first form the powers of x: $u, x, x^{\bar{\odot}2}, x^{\bar{\odot}3}, \ldots$. Since the algebra is finite dimensional, these powers will eventually become linearly dependent. Concretely, one may arrange the powers as the columns of a matrix, adding column after column, until the matrix becomes singular. When this happens, we can write the final column as a linear combination of the earlier columns: let M be the matrix with the linearly independent powers of x and let y be the first linearly dependent power. Then $y = Mz$ for some z, and we can invert the equation by taking for example the Moore-Penrose pseudoinverse of M: $z = M^{+}y$. This first relation of linear dependence

$$x^{\bar{\odot}m} + a_{m-1}x^{\bar{\odot}(m-1)} + \cdots + a_1 x + a_0 u = 0 \tag{7.126}$$

gives the minimal polynomial of x,

$$p_x(t) = t^{\bar{\odot}m} + a_{m-1}t^{\bar{\odot}(m-1)} + \cdots + a_1 t + a_0 = \prod_i (t - \lambda_i), \tag{7.127}$$

where $\{\lambda_i\}$ are the distinct eigenvalues of x.

Lemma 7.7.1. *In a Euclidean Jordan algebra, the roots $\{\lambda_i\}$ will be real [27, 164, 168].*

Proof. Suppose $p_x(t)$ had a complex root $z = a + ib$. Since $p_x(t)$ has all real coefficients, then $z^* = a - ib$ must also be a root. Let $f(t) = (t - z)(t - z^*) = (t - a)^2 + b^2$. Since both $f(t)$ and $p_x(t)$ vanish at z and z^*, $f(t)$ divides $p_x(t)$ and we may write $p_x(t) = f(t)r(t)$ for some polynomial $r(t)$. Now

$$\begin{aligned} r(x) \bar{\odot} p_x(x) &= r(x) \bar{\odot} f(x) \bar{\odot} r(x) && (7.128)\\ &= r(x) \bar{\odot} \Big((x - au)^{\bar{\odot}2} + b^2 u \Big) \odot r(x) && (7.129)\\ &= \big((x - au) r(x) \big)^{\bar{\odot}2} + b^2 r(x)^{\bar{\odot}2} = 0, && (7.130) \end{aligned}$$

since $p_x(x) = 0$ and we are working in the associative subalgebra of powers of x. But this is a vanishing sum of squares and so by formal reality, $(x - au) \bar{\odot} r(x) = 0$ and $br(x) = 0$. Suppose $r(x) = 0$. Since $p_x(t) = f(t)r(t)$ where $f(t)$ has degree 2, the degree of $r(t)$ must be less than that of $p_x(t)$. But if $r(x) = 0$, then this is a polynomial that vanishes on x and which has smaller degree that the minimal polynomial—in other words, it must be the minimal polynomial itself! So $r(x) \neq 0$. Thus in fact, $b = 0$: the root was real to begin with. □

Now for each eigenvalue, using Lagrange interpolation, we may calculate the idempotents in the spectral decomposition in accordance with the following theorem [27, 164, 168].

Theorem 7.7.2. *Let x be an element of the Jordan algebra, and let $\{\lambda_j\}$ be its distinct eigenvalues. If we let*

$$e_i = \overline{\bigodot}_{i\neq j} \frac{x-\lambda_j u}{\lambda_i - \lambda_j}, \tag{7.131}$$

then $x = \sum_i \lambda_i e_i$ for $e_i \bar{\odot} e_j = \delta_{ij} e_i$ and $\sum_i e_i = u$.

Proof. Suppose that $x = \sum_{k=1}^{s} \lambda_k e_k$ for $e_k \bar{\odot} e_l = \delta_{kl} e_k$ and $\sum_k e_k = u$, for distinct eigenvalues λ_k. Let $q_i(t) = \prod_{j\neq i} \frac{t-\lambda_j}{\lambda_i - \lambda_j}$. Clearly, $q_i(\lambda_k) = \delta_{ik}$. By power associativity, we can consider

$$q_i(x) = \overline{\bigodot}_{i\neq j} \frac{x-\lambda_j u}{\lambda_i - \lambda_j}. \tag{7.132}$$

Since $x = \sum_k \lambda_k e_k$ and $u = \sum_k e_k$, we have $x - \lambda_j u = \sum_k \lambda_k e_k - \lambda_k \sum_k e_k = \sum_k (\lambda_k - \lambda_j) e_k$ so that

$$q_i(x) = \overline{\bigodot}_{i\neq j} \sum_k \frac{\lambda_k - \lambda_j}{\lambda_i - \lambda_j} e_k = \sum_k \left(\prod_{j\neq i} \frac{\lambda_k - \lambda_j}{\lambda_i - \lambda_j} \right) e_k = \sum_k \delta_{ik} e_k = e_i, \tag{7.133}$$

where we have used the fact that if $a = \sum_k a_k e_k$ and $b = \sum_l b_l e_l$, then $a \bar{\odot} b = \sum_k a_k b_k e_k$. Conversely, suppose

$$e_i = \overline{\bigodot}_{i\neq j} \frac{x-\lambda_j u}{\lambda_i - \lambda_j}. \tag{7.134}$$

Now $\sum_i q_i(\lambda_k) = \sum_i \delta_{ik} = 1$ so that $\sum_i q_i(t) - 1 = 0$ at the s eigenvalues $\lambda_1, \ldots, \lambda_s$. But since each q_i has degree $s-1$, the degree of this polynomial must be $\leq s-1$: at the same time, it vanishes at s points. Therefore it must be identically zero, and we conclude that $\sum_i q_i(t) = 1$, and therefore $\sum_i q_i(x) = u$. Similarly, consider $\sum_i \lambda_i q_i(\lambda_k) = \sum_i \lambda_i \delta_{ik} = \lambda_k$. We conclude that $\sum_i \lambda_i q_i(t) - t$ vanishes at the s eigenvalues, but since its degree cannot be more than $s-1$, it must vanish identically. We conclude $\sum_i \lambda_i q_i(t) = t$, or $\sum_i \lambda_i q_i(x) = \sum_i \lambda_i e_i = x$.

Finally, suppose that $i \neq j$, and consider that $q_i(\lambda_k) q_j(\lambda_k) = (q_i q_j)(\lambda_k) = \delta_{ik}\delta_{jk} = 0$. Thus $(q_i q_j)(t)$ vanishes at the s eigenvalues. The minimal polynomial of x is $p_x(t) = \prod_k (t - \lambda_k)$ which also vanishes at the s distinct eigenvalues. Thus $p_x(t)$ divides $(q_i q_j)(t)$, that is, we can write $(q_i q_j)(t) = p_x(t) h(t)$ for some polynomial $h(t)$. We conclude

$$e_i \bar{\odot} e_j = q_i(x) \bar{\odot} q_j(x) = (q_i q_j)(x) = p_x(x) \odot h(x) = 0, \tag{7.135}$$

since $p_x(x) = 0$. Now suppose $i = j$. Observe that $q_i(\lambda_k)^2 - q_i(\lambda_k) = (q_i^2 - q_i)(\lambda_k) = \delta_{ik}^2 - \delta_{ik} = 0$ so that the polynomial vanishes at the s eigenvalues. Thus the minimal polynomial $p_x(t)$ divides it, and so $q_i(t)^2 - q_i(t) = p_x(t)g(t)$ for some polynomial $g(t)$. We conclude

$$e_i \odot e_i - e_i = q_i(x) \odot q_i(x) - q_i(x) = (q_i^2 - q_i)(x) = p_x(x) \odot g(x) = 0, \tag{7.136}$$

which completes the proof. □

Finally, we note that the idempotents in the spectral decomposition will not be primitive if the eigenvalues are degenerate. To resolve this, let e be an idempotent, and let $V_1(e) = \{y : e\bar{\odot}y = y\}$: this is the eigenspace with eigenvalue 1 of the idempotent. If e is primitive, then $\dim V_1(e) = 1$. Letting $L_e y = e\bar{\odot}y$, it then suffices to check whether the +1 eigenspace of L_e is one dimensional. If not, then notice that $V_1(e)$ defines its own Euclidean Jordan algebra with unit e. Pick a random vector $y \in V_1(e)$, and consider its powers $e, y, y^{\bar{\odot}2}, \ldots$: by the same procedure as above, one may construct its spectral decomposition. One recursively repeats this recipe until one has (not uniquely) d mutually orthogonal primitive idempotents, where d is the rank of the EJA.

7.7.1 From valuations to measurements

From the beginning, we had tentatively supposed that if we have two measurements $\{X_i\}$ and $\{Y_i\}$ with valuations $\mathfrak{x}$ and $\mathfrak{y}$, and we assign equivalent valuations to the reference measurement $x = \mathfrak{x}P(X|R)\Phi$ and $y = \mathfrak{y}P(Y|R)\Phi$, then there ought to be a third measurement $\{M_i^{(x\bar{\odot}y)}\}$ and a valuation $\mathfrak{m}^{x\bar{\odot}y}$ such that $\mathfrak{m}^{x\bar{\odot}y}P(M^{(x\bar{\odot}y)}|R)\Phi = x\bar{\odot}y$. We promised that this is guaranteed by the fact that reference valuations form a Euclidean Jordan algebra. We will now see that the spectral decomposition implies something even stronger is true.

Take *any* valuation x with respect to the reference measurement, and decompose it into orthogonal idempotents $x = \sum_i \mathfrak{x}_i e_i$ with respect to $\bar{\odot}$. Since the e_i's are their own squares, they correspond to effects: we take the state and effect spaces to be subsets of the self-dual cone of squares of the EJA. We are working with $\bar{\odot}$, that is, in dual coordinates, where $P(E|\rho) = \langle P(E|R)\Phi, \Phi P(R|\rho)\rangle_P$. Thus we must have $e_i = P(X_i|R)\Phi$ for some $P(X_i|R)$. Switching to regular coordinates, where $P(E|\rho) = \langle P(E|R), P(R|\rho)\rangle_\Phi$, we have $P(X_i|R) = e_i P$. Indeed, $\sum_i P(X_i|R) = \sum_i e_i P = u^\dagger P = u^\dagger = P(1|R)$, so that these effects indeed form a measurement $\{X_i\}$ and the eigenvalues $\mathfrak{x}_i$ represent a valuation on its outcomes. Indeed, $\{X_i\}$ is not just any measurement. Because its effects are a set of mutually orthogonal idempotents,

$$P(X_i|X_j) = P(X_i|R)\Phi P(R|X_j) = u^\dagger(P(R|X_i)\check{\odot}P(R|X_j)) = \delta_{ij}, \tag{7.137}$$

so that if one assigns the state $P(R|X_j)$, then one will expect outcome X_j with certainty upon a measurement of $\{X_i\}$: the states and effects perfectly distinguish each other.

Moreover, because in the decomposition $x = \sum_i \mathfrak{x}_i P(X_i|R)\Phi$, the valuations $\mathfrak{x}_i$ are eigenvalues, we have $x^{\bar{\odot}m} = \sum_i \mathfrak{x}_i^m P(X_i|R)\Phi$, and so our moment matching condition is assured,

$$\sum_i \mathfrak{x}_i^m P(X_i|\rho) = \sum_i [x^{\bar{\odot}m}]_i P(R_i|\rho). \tag{7.138}$$

Thus the class of measurements for which $\langle \mathfrak{x}^m \rangle_\rho = \langle x^{\bar{\odot}m} \rangle_\rho$ are precisely those whose effects are mutually orthogonal idempotents, that is, those measurements which perfectly distinguish their corresponding states.

Finally, let us compare $x \circ y$ and $x \bar{\odot} y$. In either case, we may spectrally decompose the valuation to find an equivalent idempotent measurement. What is the difference between them? If one chose a different reference measurement, one could always express the product $x \circ y$ according to the original reference measurement with respect to the new one. But this would entirely depend on the relationship between the original reference and the new one. In contrast, for *any* reference measurement, the product $\bar{\odot}$ is picked out as the one which defines the state space itself. Indeed, if one chose a different reference measurement, the expression for $x \bar{\odot} y$ would change, but the product's relationship to the very consistency rules which ought to govern the gambler's probability assignments would not. In this sense, the valuation and measurement equivalent to $x \circ y$ depends on the choice of reference, while $x \bar{\odot} y$ is independent of that choice. Indeed, more generally, since any reference valuation x implies, through $\bar{\odot}$, an equivalent "spectral measurement," valuations have an autonomous existence, as it were, regardless of the choice of reference.

7.7.2 Characterizing idempotents

Let us now turn to an explicit characterization of the idempotents with respect to our product $\breve{\odot}$.

Lemma 7.7.3. *If $x \breve{\odot} x = x$, then*

$$x = \frac{1}{2}\left\{ \frac{\alpha(\alpha+1)}{(\kappa^{-1} + (\alpha-1)\overline{x}/n)} P(x \circ x) - (\alpha - 1)\overline{x}u/n \right\}. \tag{7.139}$$

Proof. From

$$g \breve{\odot} h = \frac{\kappa}{2}\left[\alpha(\alpha+1)P(g \circ h) - (\alpha-1)\left(\overline{h}g + \overline{g}h + \alpha(g \cdot h)u \right)/n \right], \tag{7.140}$$

we have for $x = x \breve{\odot} x$,

$$x = \frac{\kappa}{2}\left[\alpha(\alpha+1)P(x \circ x) - (\alpha-1)\left(2\overline{x}x + \alpha(x \cdot x)u \right)/n \right], \tag{7.141}$$

Taking the sum of both sides of the equation gives

$$\overline{x} = \frac{\kappa}{2}\left[\alpha(\alpha+1)(x\cdot x) - (\alpha-1)\left(2\overline{x}^2/n + \alpha(x\cdot x)\right)\right] \tag{7.142}$$

$$= \kappa\left[\alpha(x\cdot x) - (\alpha-1)\overline{x}^2/n\right], \tag{7.143}$$

so that

$$x\cdot x = \frac{1}{\kappa\alpha}\overline{x} + \left(1-\frac{1}{\alpha}\right)\overline{x}^2/n. \tag{7.144}$$

Substituting this in, we find

$$x = \frac{\kappa}{2}\left[\alpha(\alpha+1)P(x\circ x) - (\alpha-1)\left(2\overline{x}x + \alpha\left\{\frac{1}{\kappa\alpha}\overline{x} + \left(1-\frac{1}{\alpha}\right)\overline{x}^2/n\right\}u\right)/n\right], \tag{7.145}$$

so that collecting terms, we have

$$(1+\kappa(\alpha-1)\overline{x}/n)x = \frac{1}{2\kappa}\left[\alpha(\alpha+1)P(x\circ x) - \kappa^{-1}(\alpha-1)\overline{x}\left(1+\kappa\left(\alpha-1\right)\overline{x}/n\right)u/n\right], \tag{7.146}$$

or

$$x = \frac{1}{2}\left\{\frac{\alpha(\alpha+1)}{(\kappa^{-1}+(\alpha-1)\overline{x}/n)}P(x\circ x) - (\alpha-1)\overline{x}u/n\right\} \tag{7.147}$$

□

Remark. *Another way of interpreting this formula is that just as for $x \in col(P)$, the action of P on x is very simple, simply mixing x and u, the action of P on $x\circ x$ is also just as simple for idempotents $x = x\breve{\odot}x$: x is simply mixed with u.*

Corollary 7.7.3.1. *For $x\breve{\odot}x = x$,*

$$\sum_i x_i^2 = \frac{1}{\kappa\alpha}\overline{x} + \left(1-\frac{1}{\alpha}\right)\overline{x}^2/n. \tag{7.148}$$

Corollary 7.7.3.2. *For $x\breve{\odot}x = x$,*

$$\sum_i x_i^3 = \frac{\overline{x}\left(\kappa^{-1}+(\alpha-1)\overline{x}/n\right)\left(2\kappa^{-1}+(\alpha-1)\overline{x}/n\right)}{\alpha(\alpha+1)}. \tag{7.149}$$

Proof. Substituting our expression for an idempotent x into $x \cdot x$, we obtain

$$x \cdot x = \frac{1}{2}\left\{\frac{\alpha(\alpha+1)}{\kappa^{-1}+(\alpha-1)\overline{x}/n}x^\dagger P(x\circ x) - (\alpha-1)\overline{x}^2/n\right\}. \tag{7.150}$$

First,

$$x^\dagger P(x\circ x) = \frac{1}{\alpha}\sum_i x_i^3 + \left(1-\frac{1}{\alpha}\right)\overline{x}(x\cdot x)/n. \tag{7.151}$$

Then, solving for $\sum_i x_i^3$ yields

$$\sum_i x_i^3 = \frac{\kappa^{-1}+(\alpha-1)\overline{x}/n}{\alpha+1}\left[2(x\cdot x)+(\alpha-1)\overline{x}^2/n\right] - (\alpha-1)\overline{x}(x\cdot x)/n. \tag{7.152}$$

Using

$$x\cdot x = \frac{1}{\kappa\alpha}\overline{x} + \left(1-\frac{1}{\alpha}\right)\overline{x}^2/n = \frac{\overline{x}}{\alpha}\left(\kappa^{-1}+(\alpha-1)\overline{x}/n\right), \tag{7.153}$$

and simplifying leads to the result. □

In particular, a primitive idempotent satisfies $\mathrm{tr}(x) = u^\dagger x = \overline{x} = 1$, and since every state is represented by a probability distribution, primitive idempotents may be identified with distributions $P(R|\rho)$.

Corollary 7.7.3.3. *$P(R|\rho)$ is a primitive idempotent of $\hat{\odot}$ iff*

$$P(R_i|\rho) = \frac{1}{2}\left\{\frac{\alpha(\alpha+1)}{(\kappa^{-1}+(\alpha-1)/n)}\sum_j P(R_i|R_j)P(R_j|\rho)^2 - (\alpha-1)P(R_i|\mu)\right\}, \tag{7.154}$$

where $\kappa = \frac{n}{1+(r-1)/\alpha}$.

Moreover, the proof of Lemma 6.4.1 from the previous chapter is still valid in the broader Euclidean Jordan algebraic setting due to the existence of the spectral decomposition. Thus an element of an EJA is a primitive idempotent iff $\mathrm{tr}(x^{\check{\odot}2}) = \mathrm{tr}(x^{\check{\odot}3}) = 1$. But this means

$$\mathrm{tr}\left(P(R|\rho)^{\check{\odot}2}\right) = u^\dagger(P(R|\rho)\check{\odot}P(R|\rho)) = P(\rho|R)\Phi P(R|\rho) = P(\rho|\rho) = 1, \tag{7.155}$$

and we already saw that this implies

$$\sum_i P(R_i|\rho)^2 = \frac{1}{\kappa\alpha} + \left(1-\frac{1}{\alpha}\right)/n. \tag{7.156}$$

Remark. *In fact, because we took* $P(\sigma_i|\sigma_i) = tr\Big(P(R|\sigma_i)^{\check{\odot}2}\Big) = 1$ *for reference states, and the distributions* $P(R|\sigma_i)$ *are by construction valid states, we have that the reference states must be primitive idempotents already.* $tr\Big(P(R|\sigma_i)^{\check{\odot}2}\Big) = \sum_i \lambda_i^2 = 1$ *is only possible given the constraint that* $tr(P(R|\sigma_i)) = \sum_i \lambda_i = 1$ *and* $\lambda_i \geq 0$*, if just a single* λ_i *is 1 and the rest 0.*

Similarly,

$$\mathrm{tr}\Big(P(R|\rho)^{\check{\odot}3}\Big) = u^\dagger \Big\{P(R|\rho)\check{\odot}\big(P(R|\rho)\check{\odot}P(R|\rho)\big)\Big\} \tag{7.157}$$

$$= P(\rho|R)\Phi\big(P(R|\rho)\check{\odot}P(R|\rho)\big) = P(\rho|\rho) = 1, \tag{7.158}$$

so that substituting Eq. (7.154) for $P(R|\rho)\check{\odot}P(R|\rho)$, we find as in Corollary 7.7.3.2,

$$\sum_i P(R_i|\rho)^3 = \frac{\big(\kappa^{-1} + (\alpha-1)/n\big)\big(2\kappa^{-1} + (\alpha-1)/n\big)}{\alpha(\alpha+1)}. \tag{7.159}$$

Corollary 7.7.3.4. $P(R|\rho)$ *is a primitive idempotent of* $\hat{\odot}$ *iff* $P(R|\rho) \in \Delta_n \cap col(P)$ *and*

$$\sum_i P(R_i|\rho)^2 = \frac{\kappa^{-1} + (\alpha-1)/n}{\alpha} \tag{7.160}$$

$$\sum_i P(R_i|\rho)^3 = \frac{\big(\kappa^{-1} + (\alpha-1)/n\big)\big(2\kappa^{-1} + (\alpha-1)/n\big)}{\alpha(\alpha+1)}. \tag{7.161}$$

Along with the normalization $\sum_i P(R_i|\rho) = 1$ condition, we see that the primitive idempotents, or *pure states*, live in the intersection of three kinds of spheres: a 1-norm, 2-norm, and a 3-norm sphere of specified radii, intersected with $\mathrm{col}(P)$ and lying in the nonnegative orthant. Any valid state can be written as a convex combination of such pure states.

7.7.3 Orthogonality

In order to check whether primitive idempotents are mutually orthogonal, the following lemma will help.

Lemma 7.7.4. *Let* $P(\rho|R)\Phi P(R|\tau) = tr\big(P(R|\rho)\check{\odot}P(R|\tau)\big) = 0$*. Then*

$$P(R|\rho) \cdot P(R|\tau) = \frac{\alpha - 1}{\alpha n}. \tag{7.162}$$

Moreover if $P(R|\rho)\breve{\odot}P(R|\tau) = 0$*,*

$$\sum_j P(R_i|R_j)P(R_j|\rho)P(R_j|\tau) = \frac{\alpha-1}{\alpha(\alpha+1)}\Big(P(R_i|\rho) + P(R_i|\tau) + (\alpha-1)P(R_i|\mu)\Big)/n. \tag{7.163}$$

Proof. Let p, q be the probability distributions. Our product is

$$p\breve{\odot}q = \frac{\kappa}{2}\left[\alpha(\alpha+1)P(p\circ q) - (\alpha-1)\left(p+q+\alpha(p\cdot q)u\right)/n\right]. \tag{7.164}$$

The condition that $p\breve{\odot}q = 0$ is just that

$$P(p\circ q) = \frac{\alpha-1}{\alpha(\alpha+1)}\Big(p+q+\alpha(p\cdot q)u\Big)/n. \tag{7.165}$$

Summing over both sides gives

$$p\cdot q = \frac{\alpha-1}{\alpha(\alpha+1)}\Big(2/n+\alpha(p\cdot q)\Big) \Longrightarrow p\cdot q = \frac{\alpha-1}{\alpha n}. \tag{7.166}$$

Thus

$$P(p\circ q) = \frac{\alpha-1}{\alpha(\alpha+1)}\Big(p+q+(\alpha-1)u/n\Big)/n. \tag{7.167}$$

□

Remark. *In fact, for a Jordan algebra, if* $P(R|\rho)$ *and* $P(R|\tau)$ *are assumed to be idempotents, and* $tr(P(R|\rho)\breve{\odot}P(R|\tau)) = 0$*, then* $P(R|\rho)\breve{\odot}P(R|\tau) = 0$ *[164]. Thus the condition that* $P(R|\rho)$ *and* $P(R|\tau)$ *are mutually orthogonal primitive idempotents reduces down entirely to scalar constraints, along with the subspace restriction, and the nonnegativity of the entries of the probability vectors.*

Remark. *Just as* $P(p\circ p)$ *acts simply on an idempotent* p*, so too it acts simply on* $P(p\circ q)$ *for* $p\breve{\odot}q = 0$*.*

Corollary 7.7.4.1. *For any two reference distributions, since the state space is the convex hull of the primitive idempotents,*

$$\frac{\alpha-1}{\alpha n} \leq P(R|\rho)\cdot P(R|\tau) \leq \frac{\kappa^{-1}+(\alpha-1)/n}{\alpha} \tag{7.168}$$

Remark. *In a more general setting, these are the pairwise constraints defining a qplex [5].*

7.7.4 *Testing the rank of an EJA*

The *rank* of an EJA is the maximum number of mutually orthogonal primitive idempotents. One could determine this, for example, by searching directly for sets of distributions satisfying $P(\rho_i|\rho_j) = P(\rho_i|R)\Phi P(R|\rho_j) = \delta_{ij}$ subject to the primitive idempotent constraints we just derived. More simply, one may rely upon the spectral decomposition. As we have noted, every element of an EJA has a spectral decomposition $x = \sum_i \lambda_i e_i$ where e_i are mutually orthogonal idempotents and λ_i are the distinct eigenvalues of x [27]. For a generic element of the EJA, the eigenvalues will all be distinct, and so there will be d of them, where d is the rank of the EJA [170]. We can thus determine the rank of the EJA with a simple recipe: pick many elements of the algebra at random, take their spectral decompositions, and determine the maximum number of distinct eigenvalues. This will almost certainly be the rank of the Euclidean Jordan algebra.

7.8 A simple EJA

As we have said, Jordan–von Neumann–Wigner theorem of 1934 [28, 164] classifies the Euclidean Jordan algebras, showing they must be direct sums of the *simple Euclidean Jordan algebras*, which we now define. A simple EJA is one whose only ideals are $\{0\}$ and the entire vector space V on which the algebra is defined, in our case $\mathrm{col}(P)$. In this case, an *ideal* is a subspace $\mathcal{I} \subseteq V$ such that $y \in \mathcal{I}$ implies $\forall x \in V : x \bar{\odot} y \in \mathcal{I}$. In other words, if y is in the ideal, then under the Jordan product, it traps any other vector in that subspace with it. For us, then, an ideal satisfies e.g., $\forall x \in \mathrm{col}(P) : \mathcal{L}_x \mathcal{I} \subseteq \mathcal{I}$. A subspace $\mathcal{I}$ is left invariant by a matrix $M = M^\dagger$ iff the orthogonal projector onto that subspace commutes [111]: $[M, \Pi_\mathcal{I}] = 0$. Thus we require

$$\nexists \Pi_\mathcal{I} \text{ such that } \forall x \in \mathrm{col}(P) : [\mathcal{L}_x, \Pi_\mathcal{I}] = 0 \text{ unless } \Pi_\mathcal{I} = 0 \text{ or } \Pi_\mathcal{I} = \Pi. \tag{7.169}$$

where Π is the projector onto $\mathrm{col}(P)$ and where $\mathcal{L}_x = \mathcal{L}_x^\dagger$. We presume as well that $\Pi_\mathcal{I}\Pi = \Pi\Pi_\mathcal{I} = \Pi_\mathcal{I}$ so that $\Pi_\mathcal{I}$ projects inside $\mathrm{col}(P)$. Moreover, since $\mathcal{L}_x$ was constructed by sandwiching V_x with the projector Π, the commutator is already restricted to $\mathrm{col}(P)$ (otherwise, we might need to project again). By linearity, it suffices to check the above condition for each reference state $P(R|R_i)$. Practically speaking, what one ought to do is introduce a basis for $\mathrm{col}(P)$. Let B be the matrix whose columns form an orthonormal basis $\{b_i\}$ for $\mathrm{col}(P)$. We may then parameterize the space of symmetric matrices $Z = Z^\dagger$, and solve for

$$\forall i : [B^\dagger \mathcal{L}_{b_i} B, Z] = 0. \tag{7.170}$$

If such a Z exists, then its eigenspaces will be preserved by $B^\dagger \mathcal{L}_{b_i} B$: thus one can construct the projector onto those eigenspaces to get the projector onto the ideal. Simplicity of the EJA is equivalent to the only solution being $Z = \lambda I$, a multiple of the identity.

Another way of thinking about simplicity is in terms of the *center* of the EJA. Let $V = \mathrm{col}(P)$ and let L_x be the linear operator which performs the Jordan product with x. The center of a Euclidean Jordan algebra may be defined as $Z = \{z \in V : [L_z, L_x] = 0 \; \forall x \in V\}$ [164]. In other words, it is just those elements whose Jordan product operator commutes with all other Jordan product operators. If the EJA is composite $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, the center is spanned by the units of each summand $u_1, u_2, \ldots, u_k$. Thus $\dim Z = k$. The EJA is simple iff Z is spanned by just a single unit u [27]. To test this, let $\{b_i\}_{i=1}^{r}$ be a basis for $\mathrm{col}(P)$ and B be the matrix with the basis elements as its columns. Let $\tilde{L}_i = B^\dagger L_{b_i} B$ be the Jordan product operator for the b_ith element represented as an $r \times r$ matrix. We can write an arbitrary element $z = \sum_l z_l b_l$ so that $\tilde{L}_z = \sum_l z_l \tilde{L}_l$. Then z will be central iff

$$\forall i : [\tilde{L}_z, \tilde{L}_i] = \sum_l z_l [\tilde{L}_l, \tilde{L}_i] = 0. \tag{7.171}$$

Alternatively, let $Z_{ijkl} = [\tilde{L}_l, \tilde{L}_i]_{jk}$ so that we require $\forall i, j, k : \sum_l Z_{ijkl} z_l = 0$. Reshaping this into an $r^3 \times r$ matrix, this becomes $Zz = 0$. The EJA is simple, then, if $\dim \mathrm{null}(Z) = 1$: there is just one element $Z = u$ whose Jordan product operator commutes with all others.

While the above considerations yield a conceptually straightforward way of calculating whether the EJA is simple, when the rank of P is large, the tensor Z_{ijkl}, which has r^4 entries, will become intractable to work with. We therefore provide an algorithm which will certify simplicity with overwhelming probability. We begin from the observation that if an EJA is composite, a primitive idempotent must lie entirely in one of the factors [27].

Lemma 7.8.1. *Let V be the vector space on which a composite EJA is defined, and let $\odot$ denote the Jordan product. V may therefore be decomposed as $V = V_1 \oplus V_2 \oplus \ldots$. Primitive idempotents must live entirely in one of the summands.*

Proof. Without loss of generality, suppose $V = V_1 \oplus V_2$. Suppose e were a primitive idempotent supported in both factors. Thus $e = e_1 \oplus e_2$ satisfies $e \odot e = e$. Since $(x_1 \oplus x_2) \odot (y_1 \oplus y_2) = (x_1 \odot y_1) \oplus (x_2 \odot y_2)$, we have $e_1 \odot e_1 = e_1$ and $e_2 \odot e_2 = e_2$. In other words, e_1 and e_2 must themselves be idempotent. Thus $e = (e_1 \oplus 0) + (0 \oplus e_2)$ is a sum of idempotents. Since the summands are supported on different sectors, they are mutually orthogonal. Thus e can be expressed as a sum of mutually orthogonal idempotents, which contradicts primitivity. □

Suppose we pick a primitive idempotent e at random. By the above lemma, it must lie entirely in one sector of the EJA. Now select random elements $\{x_i\}$ of the algebra and take the Jordan product of each with e, obtaining new elements $\{x_i'\}$. Then repeat the process, selecting random elements of the algebra, and taking the Jordan product with each of the elements $\{x_i'\}$ to obtain $\{x_i''\}$, hitting those with random elements of the algebra, for some number of repetitions. Finally, consider the linear span of the elements so obtained: this is an approximation of the ideal of which e is a part. Indeed, if e were supported on a single sector

of the EJA, then the linear span of the elements obtained throughout this process should remain in that same sector. If the dimension of the linear span is in fact the dimension of V, the whole vector space, then the EJA must be simple.

To obtain a primitive idempotent e, one may find some element which satisfies the constraints we derived above: or one may obtain e by spectrally decompositing a random element of the Jordan algebra. Generically, such an element will have d distinct eigenvalues: thus its spectral decomposition will yield d mutually orthogonal primitive idempotents $\{e_i\}$. Once one has selected a starting primitive idempotent e, instead of hitting it with random elements of the algebra, one may more systematically hit e with elements of a basis $\{b_i\}$ for V. Then introduce a basis $\{b'_i\}$ for the linear span of $\{e \odot b_i\}$. Once this is done, hit each element of $\{b'_i\}$ with elements of the original basis $\{b_i\}$, and consider the linear span of $\{b_i \odot b'_j\}$. Again, introduce a basis $\{b''_i\}$ for this linear span, and then consider $\{b_i \odot b''_i\}$, and so on. Eventually the dimension of these subspaces will stabilize, and if it in fact is $\dim V$, then the EJA is simple.

Thus given $P(R|R)$ it is a matter of computation to determine whether the Jordan algebra it encodes is simple. And for simplicity's sake, we now suppose that it is. After all, if it were not, we could build it up from simpler components, considering a reference measurement for each part: the union of the reference measurement for each will be a reference measurement for the whole. (Later we will consider another type of motivation.)

7.9 Full circle

The 1934 paper of Jordan, Wigner, and von Neumann [28] not only shows that Euclidean Jordan algebras must be composites of simple EJAs, but also gives a direct way of characterizing these fundamental building blocks. Up to some special isomorphisms between the different classes of EJAs[7], a simple EJA is uniquely defined by two numbers. The first is its *rank*, that is, the maximum number of mutually orthogonal primitive idempotents, and the second is its *dimension*, the smallest vector space on which the algebra may be defined.

Euclidean Jordan algebra	**rank**	**dimension**
$\mathrm{Herm}_d(\mathbb{R})$	d	$d(d+1)/2$
$\mathrm{Herm}_d(\mathbb{C})$	d	d^2
$\mathrm{Herm}_d(\mathbb{H})$	d	$d(2d-1)$
$\mathrm{Herm}_3(\mathbb{O})$	3	27
spin factor	2	$n+1$

We thus make one final assumption about the conditional probability matrix $P(R|R)$ which characterizes the reference measurement, that $r = \mathrm{rank}(P) = d^2$ where d is the maximum number of mutually orthogonal idempotents. With this final piece in place, we conclude that reference valuations have a representation as $d \times d$ Hermitian matrices over $\mathbb{C}$, and that states

[7] For example, the $n = 3$ spin factor is isomorphic to $\mathrm{Herm}_2(\mathbb{C})$.

$P(R|\rho)$ and effects $P(E|R)$ have a representation as $d \times d$ positive semidefinite Hermitian matrices over $\mathbb{C}$. In other words, we have arrived at quantum mechanics.

We may now fix our constants. By self-duality, we had $P(\rho|R) = \kappa P(R|\rho)^\dagger$ where $\kappa = \frac{n}{1+(r-1)/\alpha}$. Thus in fact

$$\alpha = \frac{\kappa(r-1)}{n-\kappa}, \tag{7.172}$$

where we now know that $r = d^2$. We also now know that there is a representation of the reference measurement in terms of $d \times d$ positive semidefinite matrices satisfying $\sum_i R_i = I$. Denoting the corresponding reference states as σ_i, we have $\forall \rho : P(\rho|R_i) = \mathrm{tr}(\rho\sigma_i) = \kappa P(R_i|\rho) = \kappa \mathrm{tr}(R_i\rho)$, from which we conclude that $R_i = \kappa^{-1}\sigma_i$. Thus

$$\mathrm{tr}\left(\sum_i R_i\right) = \sum_i \mathrm{tr}(R_i) = n\kappa^{-1} = \mathrm{tr}(I) = d, \tag{7.173}$$

so that $\kappa = n/d$. Substituting $\kappa = n/d$ and $r = d^2$ into our expression for α, we find

$$\alpha = \frac{(n/d)(d^2-1)}{n - n/d} = \frac{d^2-1}{d-1} = d+1. \tag{7.174}$$

We may now express our formulas with all the constants determined. We have

$$g \breve{\odot} h = \frac{1}{2}\left\{(d+1)(d+2)\left(\frac{n}{d}\right)P(g \circ h) - \left(\overline{h}g + \overline{g}h + (d+1)(g \cdot h)u\right)\right\}, \tag{7.175}$$

so that the Jordan product tensor (rescaled to implement $\breve{\odot}$) is

$$\begin{aligned}\kappa[\mathcal{L}_P]_{ijk} = \frac{1}{2}\Bigg\{&(d+1)(d+2)\left(\frac{n}{d}\right)\sum_m P(R_m|R_i)P(R_m|R_j)P(R_m|R_k) \\ &- \Big(P(R_i|R_j) + P(R_i|R_k) + P(R_j|R_k) + d/n\Big)\Bigg\}.\end{aligned} \tag{7.176}$$

The pure state vector condition reads

$$P(R_i|\rho) = \frac{1}{2}\left\{\frac{1}{2}(d+1)(d+2)\left(\frac{n}{d}\right)\sum_j P(R_i|R_j)P(R_j|\rho)^2 - \frac{d}{n}\right\}, \tag{7.177}$$

and finally, the pure state scalar constraints become

$$\sum_i P(R_i|\rho)^2 = \frac{d}{n}\frac{2}{d+1} \tag{7.178}$$

$$\sum_i P(R_i|\rho)^3 = \frac{d^2}{n^2}\frac{6}{(d+1)(d+2)}, \tag{7.179}$$

all of which may be compared to the results of the last chapter. In particular, the cubic equation implies that

$$\sum_{ij} P(R_i|R_j)^3 = \sum_i \left[\sum_j P(R_i|R_j)^3\right] = \frac{d^2}{n}\frac{6}{(d+1)(d+2)}. \tag{7.180}$$

Since $P(R_i|R_j) = \mathrm{tr}(R_i\sigma_j) = \frac{d}{n}\mathrm{tr}(\sigma_i\sigma_j) = \frac{d}{n}|\langle\psi_i|\psi_j\rangle|^2$, we have

$$\mathcal{F}_3(\{\psi_i\}) = \frac{1}{n^2}\sum_{ij}|\langle\psi_i|\psi_j\rangle|^6 = \frac{6}{d(d+1)(d+2)}. \tag{7.181}$$

In other words, the pure states $\{\sigma_i\}$ attain the minimum of the 3rd order frame potential [62], which we introduced in Chapter 6. We conclude at last that our reference measurement must correspond to an unweighted complex projective 3-design.

7.10 Self-testing

While each of the conditions we placed on $P(R|R)$ were motivated by foundational considerations, in hopes of giving a reconstruction of quantum mechanics which puts the structure of the reference measurement itself in pride of place, we now observe that our considerations may be reinterpreted as providing a means of *self-testing* complex projective 3-designs in a theory-agnostic way [166]. A set of quantum states forms a complex projective t-design iff it minimizes the order-t frame potential: here the emphasis is on a set of *quantum* states. One can calculate this quantity from $P(R_i|R_j)$; alternatively, if one is not able to implement the corresponding measurement, one can calculate it from $P(\sigma_i|\sigma_j)$ obtained e.g., by performing a swap test on the states [171, 172]. But how can one be sure that the states in fact have a Hilbert space representation? The same problem arises even if one can measure the frame potential directly [173, 174]. Such certifications of t-designhood are conditional on the assumption that a Hilbert space representation is in fact possible. A theory-agnostic certification would guarantee that not only the frame potential is minimized, but also a Hilbert space representation can be constructed. In the course of this chapter, we have provided exactly such a means of doing this not just for 3-designs, but for t-designs for $t \geq 3$.

Theorem 7.10.1. *Let $P \equiv P(R|R)$ be a column stochastic matrix satisfying:*

- $r = rank(P) = d^2$ *for some* d.
- $P = P^T$.
- $\forall i : P(R_i|R_i) = d/n$
- $P\Phi P = P$ *for* $\Phi = \alpha I + (1-\alpha)J/n$ *where* $\alpha = d+1$.
- *Let* $L_x = P\Phi\left\{\gamma D_x + (1-\gamma)(xu^\dagger + \bar{x}I - ux^\dagger)/n\right\}$ *where* $\gamma = (1+1/\alpha)/2$ *so that* $L_x y = x\bar{\odot}y$. *Let* $\{b_i\}$ *be an orthonormal basis for* $col(P)$. *Then* $\forall i,j,k : T(b_i,b_j,b_k) = 0$ *where*

$$T(a,b,c) = \frac{1}{3}\Big\{[L_a, L_{b\bar{\odot}c}] + [L_b, L_{c\bar{\odot}a}] + [L_c, L_{a\bar{\odot}b}]\Big\}P\Phi.$$

- *The resulting EJA is simple, computed according to Section 7.8.*
- *The rank of the resulting EJA is* d, *that is, the maximum number of distributions* $\{P(R|\rho_j)\}$ *in* $col(P)$ *satisfying*

$$\forall j : \sum_i P(R_i|\rho_j)^2 = \frac{d}{n}\frac{2}{d+1} \qquad \text{and} \qquad \sum_i P(R_i|\rho_j)^3 = \frac{d^2}{n^2}\frac{6}{(d+1)(d+2)},$$

and for $j \neq k$,

$$\sum_i P(R_i|\rho_j)P(R_i|\rho_k) = \frac{d}{n}\frac{1}{d+1}$$

is d, *computed according to Section 7.7.4.*

Then $P(R_i|R_j) = \frac{d}{n}tr(\sigma_i\sigma_j)$ *where* $\{\sigma_i\}$ *is a set of rank-1 projectors on a complex Hilbert space of dimension* d *which forms an unbiased complex projective 3-design.*

Corollary 7.10.1.1. *Suppose the conditions in Theorem 7.10.1 are satisfied, and additionally that for* $t \geq 3$

$$\sum_{ij} P(R_i|R_j)^t = \frac{d^t}{n^{t-2}}\binom{d+t-1}{t}^{-1}. \tag{7.182}$$

Then $P(R|R)$ *corresponds to a unbiased complex projective* t*-design.*

Proof. The conditions in Theorem 7.10.1 guarantee that $P(R|R)$ has a Hilbert space representation as a complex projective 3-design. Recalling that a t-design is also a $(t-1)-$design, if $P(R|R)$ additionally minimizes the frame potential for $t > 3$, it must be a complex projective t-design. □

7.11 Assume a SIC?

We now show that if we make just one more assumption, we can rule out composite Jordan algebras, and moreover restrict the possible simple EJA's compatible with $P(R|R)$. This assumption is that the state space implied by $P(R|R)$ can host a *symmetric informationally complete* set of states [24]. If $\mathrm{rank}(P) = r$, then this means a set of r states, which are primitive idempotents, and which form a regular simplex.

Theorem 7.11.1. *Let $\{P(R|\rho_i)\}_{i=1}^r$ be a set of primitive idempotents such that $P(\rho_i|\rho_j) = c$ for $i \neq j$ and whose sum is proportional to u. Then $c = (\alpha - 1)/(r + \alpha - 1)$.*

Proof. We have

$$c = P(\rho_i|\rho_j) = P(\rho_i|R)\Phi P(R|\rho_j) = \kappa P(R|R_i)\Big[\alpha I + (1-\alpha)J/n\Big]P(R|R_j) \tag{7.183}$$

$$= \kappa\big\{\alpha p_i \cdot p_j + (1-\alpha)/n\big\}, \tag{7.184}$$

so that $p_i \cdot p_j = \frac{1}{\alpha}\{c/\kappa - (1-\alpha)/n\}$. Now on the one hand, using the formula for $\sum_i p_i^2$ for a primitive idempotent,

$$\left\|\sum_i p_i\right\|^2 = \sum_{ij} p_i \cdot p_j = \sum_i \|p_i\|^2 + \sum_{i\neq j} p_i \cdot p_j \tag{7.185}$$

$$= r\left(\frac{\kappa^{-1} + (\alpha-1)/n}{\alpha}\right) + (r^2 - r)\frac{1}{\alpha}\Big[c/\kappa - (1-\alpha)/n\Big]. \tag{7.186}$$

On the other hand, since $\sum_i p_i = \beta u$, we have $\sum_i \mathrm{tr}(p_i) = \beta \mathrm{tr}(u)$ which implies $r = \beta n$ or $\beta = r/n$. Thus

$$\left\|\sum_i p_i\right\|^2 = \|(r/n)u\|^2 = (r^2/n^2)n = r^2/n. \tag{7.187}$$

Equating these two expressions and using $\kappa = n/(1 + (r-1)/\alpha)$ yields the result. □

Corollary 7.11.1.1. *Such a set is informationally complete.*

Proof. Since $P(\rho_i|\rho_j) = \kappa p_i^\dagger \Phi p_j = \kappa\big(\alpha p_i \cdot p_j + (1-\alpha)/n\big)$, the $r \times r$ Gram matrix $G_{ij} = p_i \cdot p_j$ may be expressed

$$G = (A - B)I + BJ \qquad A = \frac{1}{\alpha}\left(\frac{1}{\kappa} - (1-\alpha)/n\right) \qquad B = \frac{1}{\alpha}\left(\frac{c}{\kappa} - \frac{1-\alpha}{n}\right). \tag{7.188}$$

The eigenvalues of G are therefore $A-B$ with multiplicity $r-1$ and $A+(r-1)B$ with multiplicity 1. $A-B = (1-c)/(\alpha\kappa)$. Since $c \neq 1$, $A-B \neq 0$. Meanwhile, $r(A+(r-1)B) = \|\sum_i p_i\|^2 > 0$. Thus G is full rank, and so the $\{p_i\}$ are linearly independent. □

Lemma 7.11.2. *If such a SIC set exists, the EJA must be simple or else classical.*

Proof. In Lemma 7.8.1, we showed that primitive idempotents must be supported entirely in a single factor of a composite EJA. Thus if $\{P(R|\rho_i)\}$ are primitive idempotents, they must be supported on one or the other sector. But if $P(R|\rho_i)$ and $P(R|\rho_j)$ are supported on different sectors, $P(\rho_i|\rho_j) = 0$, which contradicts the assumed equiangularity of the set unless $c = 0$. If $c = 0$, the primitive idempotents must be mutually orthogonal, $\alpha = 1$, $\Phi = I$, and $\odot = \circ$: the EJA is classical. Otherwise, the SIC must be supported entirely in one or the other summand: but this contradicts informational completeness. □

In light of this, just as we determined d, the cardinality of the set of mutually orthogonal primitive idempotents, by trying to find d distributions satisfying the scalar primitive idempotent constraints as well as the orthogonality constraint, we can try to find r distributions which satisfy the scalar primitive idempotent constraints as well as for $i \neq j$,

$$P(\rho_i|\rho_j) = P(\rho_i|R)\Phi P(R|\rho_j) = \frac{\alpha - 1}{r + \alpha - 1} \tag{7.189}$$

If this can be done, then a SIC set exists, and the EJA must be simple. Moreover, the existence of a SIC set narrows down the possible choices of simple EJA's as explained in [24]. Indeed, SICs are only known to exist for certain special ranks ($d = 2, 3, 7, 23$—although there may be more yet to be discovered) in the setting of quantum mechanics over $\mathbb{R}$ [120, 62], and they do not appear to exist for quantum mechanics over $\mathbb{H}$ for $d > 3$ [118]. The same paper, however, shows that a SIC *does* exist in the exceptional 3×3 octonionic case. Moreover, a SIC always exists in a spin-factor EJA: since the state spaces are spheres, it is trivial to inscribe a regular $(r-1)-$ simplex within them. But the rank of a spin-factor EJA—the maximum number of mutually orthogonal primitive idempotents—is always just 2. Thus if $d \neq 2, 3, 7$ or 23, unless one stumbles upon an exceptional SIC over $\mathbb{R}$ or $\mathbb{H}$, the EJA must correspond to quantum mechanics over $\mathbb{C}$—or it would, if SICs existed over $\mathbb{C}$ for all d. There is very strong numerical evidence that this is the case [117, 175], as well as a by now highly developed theory behind their exact construction. For the state of the art on the SIC existence question, see [176] which proves the existence of SICs over $\mathbb{C}$ in all dimensions conditional on the resolution of certain conjectures in algebraic number theory (related to Hilbert's twelfth problem) as well as the proof of a special function identity.

At the risk of specializing too much, we can go further and assume that among the reference states and effects themselves are not only a SIC set, but also $s \leq d$ mutually orthogonal primitive idempotents. Then we can rule out composite EJAs as well as incompatible simple EJAs from a simple inspection of $P(R|R)$. It remains to show that a

3-design with this structure can in fact always be constructed in quantum mechanics over $\mathbb{C}$, assuming SICs exist.

Theorem 7.11.3. *An unbiased 3-design can be constructed in quantum mechanics over $\mathbb{C}$ which contains a SIC as well as at least $s \leq d$ orthogonal states.*

Proof. Analogous to quantum state 3-designs, a unitary 3-design is an ensemble of unitaries which mimic the behavior of Haar distributed unitaries up to the third moment [177, 178, 179]. For sufficiently large cardinality, such sets always exist. Let $\mathcal{U}$ be an unweighted unitary 3-design. By definition,

$$\frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}}(U\Pi U^\dagger)^{\otimes 3} = \int dU (U\Pi U^\dagger)^{\otimes 3} = \binom{d+2}{3}^{-1}\Pi_{\text{sym}^3}, \tag{7.190}$$

so that $\{U\Pi U^\dagger\}$, the orbit of a state Π under the unweighted unitary 3-design, forms an unweighted state 3-design. Let $\{\Pi_i\}$ be a set of SIC states. Then

$$\frac{1}{d^2}\sum_{i=1}^{d^2}\frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}}(U\Pi_i U^\dagger)^{\otimes 3} = \binom{d+2}{3}^{-1}\Pi_{\text{sym}^3}, \tag{7.191}$$

so that we have constructed an unbiased 3-design which contains a SIC subset as the union of the orbit of each SIC state under the unitary 3-design (counting repeated unitaries with multiplicity). Now without loss of generality, we can take $I \in \mathcal{U}$: if not, we may take any element U_1 and consider the set $\{U_1^\dagger U\}$ which forms a 3-design which does contain I. Now take one of the SIC states Π_1 and $s-1$ unitaries V_i such that $\Pi_1, V_1\Pi_1 V_1^\dagger, V_2\Pi_1 V_2^\dagger, \ldots$ are s mutually orthogonal states: this can always be done. Then take the union (again preserving repeated unitaries) $\mathcal{U}' = \mathcal{U} \cup V_1\mathcal{U} \cup V_2\mathcal{U}\ldots$. Such an equal weight union of unitary 3-designs is again an unweighted 3-design. If we then consider the union (with multiplicity) of the orbits of the d^2 SIC states under $\mathcal{U}'$, we have an unbiased 3-design which contains not only a SIC but also s orthogonal states. □

7.12 Conclusion

And so we have come full circle. Let us retrace our steps. We began with the gambler trying to make better decisions, and supposed they have identified a reference measurement for a domain they are invested in. As we saw in Chapter 2, on very general grounds, one may derive a nonclassical coherence condition $P(E|\rho) = P(E|R)\Phi P(R|\rho)$ where Φ is any $\{1\}$-inverse of the conditional probability matrix $P(R|R)$ which characterizes the reference measurement, taken to be a preparatory measurement, itself. We then proceeded to make increasingly severe restrictions on the structure of this matrix $P \equiv P(R|R)$. We assumed that it is symmetric, and so both states $P(R|\rho)$ and response functions $P(E|R)$, as well as

valuations upon the reference outcomes, ought to be confined to $\mathrm{col}(P)$. We also assumed the P is constant along its diagonal. We presumed that the Born matrix Φ has the form of a Protourgleichung, $\Phi = \alpha I + \beta J$, and assumed it to be quasistochastic, fixing the value of β. This in fact implies that P depolarizes vectors in its column space, and that P has a simple eigenstructure, which fixes the value along the diagonal of P in terms of $r = \mathrm{rank}(P)$ and α.

We then discussed how the gambler can transfer valuations on arbitrary measurements to equivalent valuations on the reference measurement. Reference valuations have a natural product, the elementwise or Hadamard product $\circ$, which in fact is the Jordan product on $\mathbb{R}^n \oplus \cdots \oplus \mathbb{R}^n$. Just as Φ represents a minimal deformation of the classical law of total probability, we attempted to deform the classical product in as minimal a way as possible, in order to define a product on valuations which we hoped would be, in some sense, independent of the choice of reference measurement. We show that any product $\odot$ which is commutative, permutation equivariant, unital for $(1, \ldots, 1)$ and is a one parameter mixture with the classical product with as few terms as possible leads uniquely to a product of the form

$$x \odot y = \gamma(x \circ y) + (1 - \gamma)(\overline{y}x + \overline{x}y - (x \cdot y)u)/n. \tag{7.192}$$

But since valuations ought to be confined to $\mathrm{col}(P)$, we ultimately adopted $x \bar{\odot} y = \Pi(x \odot y)$, where Π is the projector onto $\mathrm{col}(P)$, to be the proper definition of our minimal extension of the classical product on valuations. Demanding compatibility with the Protourgleichung in fact fixes the value of γ in terms of α, and shows that the product is self-adjoint with respect to an inner product defined in terms of P, and the algebra it defines is formally real.

We then supposed that there exists a special class of measurements which satisfy a moment-matching condition. If $\mathfrak{x}$ is a valuation on $\{X_i\}$, then $\sum_i \mathfrak{x}_i^m P(X_i|\rho) = \sum_i x_i^{\bar{\odot} m} P(R_i|\rho)$, that is, that the mth moment of the reference valuation with respect to $\bar{\odot}$ in fact reproduces the mth moment of the valuation on the original measurement. For this moment-matching condition to make sense, we further demanded that $\bar{\odot}$ be power associative so that $\bar{\odot}$ powers have a univocal meaning. We thus confronted a finite dimensional real and formally real commutative power associative algebra of observables defined through the bilinear product $\bar{\odot}$: we then appealed to Jordan, von Neumann, and Wigner's classic result that the algebra we have defined is in fact a Euclidean Jordan algebra, and indeed, $\bar{\odot}$ satisfies the Jordan identity. Conversely, if the product satisfies the Jordan identity, it must be power associative, and so the Jordan identity (in its third order polarized form) gives us a new constraint on $P(R|R)$.

Next, we observed that $\bar{\odot}$ is defined on valuations, which may be interpreted as dual coordinates: we therefore lifted $\bar{\odot}$ to $\hat{\odot}$ a product on regular coordinates e.g., probability vectors and response functions, working out the form of the matrix $\mathcal{L}_x$ which satisfies $\mathcal{L}_x \Phi y = x \hat{\odot} y$, and for good measure, reexpressed the Jordan identity in terms of $\hat{\odot}$. We then appealed to the self-duality of the cone of squares of an EJA to derive the shape of state space via an uncertainty principle. We showed that in fact the positive semidefiniteness of the matrix $\mathcal{L}_{P(R|\rho)}$ equivalently diagnoses the validity of the state. Self-duality also implies that states $P(R|\rho)$ can be rescaled into effects $P(\rho|R)$, and we derived a self-duality constant κ

which does just this. The assumption that if we turn a reference state into an effect in this way, we should expect the corresponding outcome with certainty fixes κ in terms of r, n, and α, and shows that if $P(\rho|\rho) = 1$, the corresponding reference distribution $P(R|\rho)$ must live on the surface of a 2-sphere with a particular radius. Finally, we rescaled $\hat{\odot}$ into $\check{\odot}$ using the self-duality constant, readying us for the discussion of the idempotents of the Jordan algebra.

Indeed, we then noted that any element of the algebra has a spectral decomposition into a linear combination of primitive idempotents, and sketched a recipe for doing so. This in fact realizes our original ambition: the spectral decomposition into primitive idempotents implies that any reference valuation corresponds to a valuation on a measurement whose effects are precisely a set of mutually orthogonal idempotents—these are the types of measurements whose higher moments may be derived simply from the nonclassical product on the reference. We then derived explicit equations which characterize idempotents, and in particular a vector constraint which is a necessary and sufficient condition for a distribution $P(R|\rho)$ to be a primitive idempotent as well as a set of equivalent scalar constraints: the former 2-norm sphere constraint, as well as a 3-norm sphere constraint. Moreover, we observed that our assumptions have already implied that the reference states are primitive idempotents. Finally, we derived explicit conditions for two states to be orthogonal with respect to $\check{\odot}$, noting that the rank of a Euclidean Jordan algebra is the maximum number of mutually orthogonal primitive idempotents. Given the matrix $P(R|R)$, one may check computationally what is the rank of the algebra it gives birth to. (And by the same token, the same algorithm could be used to search for a $P(R|R)$ which has a particular rank, subject to the other constraints.)

The Jordan-von Neumann-Wigner theorem characterizes the simple Euclidean Jordan algebras, the atomic building blocks of all EJA's, and we give a recipe for checking computationally whether our EJA is simple in terms of $P(R|R)$—or equivalently, for searching for such a $P(R|R)$. Assuming simplicity, we may then appeal to the classification given in the aforementioned theorem. Supposing that the $r = \text{rank}(P) = d^2$ where d is the maximum number of mutually orthogonal idempotents, our product $\check{\odot}$ must in fact correspond to the Jordan product on $\text{Herm}_d(\mathbb{C})$, that is, $d \times d$ Hermitian matrices over the complex numbers. The state and effect spaces may be understood in terms of the positive semidefinite cone: the states must have representations as $d \times d$ positive semidefinite matrices over $\mathbb{C}$ satisfying $\text{tr}(\rho) = 1$ and the effects must be $d \times d$ positive semidefinite matrices satisfying $0 \leq E \leq I$. In other words, we have arrived at quantum mechanics. This fixes the value of the self-duality constant κ, which in turn fixes the value of α. Coming full circle, we show that with these values, the reference states must minimize the 3rd order frame potential, and so the reference measurement corresponds to an unweighted complex projective 3-design. We then showed that beyond its value as a reconstruction of quantum mechanics, our considerations amount to a theory agnostic scheme for certifying complex projective t-designs for $t \geq 3$. We closed by showing that the existence of a symmetric informationally complete set excludes composite EJA's and moreover, narrows down the possible simple EJA's dramatically.

In the previous chapter, we proceeded within quantum mechanics, assumed our reference measurement was built from an unbiased complex projective 3-design, and then

showed how the shape of quantum state space ccould be unpacked from this single matrix $P(R|R)$. In this chapter we reverted back to the origin: after wiping the slate clean, by placing increasingly poweful constraints on $P(R|R)$, we were ultimately able to narrow our focus *back* to quantum mechanics, showing that our assumptions take us full circle, back to a 3-design. The key step was the introduction of the nonclassical product $\odot$, guided by the idea that the classical product on valuations should be deformed as minimally as possible. Indeed, this is what is remarkable about 3-designs: just as 2-designs make the Born rule look as close as possible to the law of total probability, 3-designs make the classical Jordan product $\circ$ on valuations look as close as possible to the quantum Jordan product. Since the entire theory can be defined in terms of the Jordan product, this in fact buys us everything. Moreover, our entire construction may be viewed as a means of *self-testing* unbiased t-design measurements for $t \geq 3$. One may of course calculate a value which amounts to the frame potential from $P(R|R)$, but there is no guarantee that $P(R|R)$ has a Hilbert space representation. Checking the rank, the Jordan identity, and so forth (Theorem 7.10.1), ensures that $P(R|R)$ does indeed have a Hilbert space representation, and so is certifiably quantum t-design.

Until the final step fixing simplicity and the rank of the algebra, our construction is compatible with any Euclidean Jordan algebra, giving a new perspective on both their geometry and their relation to classical probability theory. That said, our reconstruction is by intention modest: its main goal, one might say, is to help one understand precisely in what sense quantum theory can be understood as a probability theory supplemented with nonclassical coherence conditions which are gentle modifications of classical rules. Indeed, this is the general strategy of QBist thinking—that and putting the limelight on a wisely chosen reference measurement so that consistency with the reference measurement properly understood implies consistency across all measurements. Of course, simply stipulating the rank of the algebra and the rank of P may appear ad hoc, or ulteriorly motivated: the real point, however, is to dramatize the fact that these are precisely stipulations about the reference measurement itself. Giving fuller motivation to those choices we leave to future work. Besides assuming a SIC, many plausible routes have been already taken to narrow down the simple EJA's to quantum mechanics over $\mathbb{C}$: for example, supposing observables are generators of time evolution [45] or assuming local tomography [169, 180]. We take here a pragmatic point of view, if you will. Confronted with some phenomenon, the gambler makes many experiments, identifying a reference measurement which they characterize by $P(R|R)$. If this matrix satisfies the constraints we have developed in this chapter, then the gambler can conclude: they may use the quantum mechanical formalism to facilitate their calculations.

A different kind of motivation, however, may become more apparent through further study: much is still not understood about the nature of our construction. For example, we showed how the pure states or primitive idempotents correspond to distributions $P(R|\rho)$ on n

outcomes satisfying

$$\forall i : P(R_i|\rho) \geq 0 \tag{7.193}$$

$$\sum_i P(R_i|\rho) = 1 \tag{7.194}$$

$$\sum_i P(R_i|\rho)^2 = \frac{d}{n}\frac{2}{d+1} \tag{7.195}$$

$$\sum_i P(R_i|\rho)^3 = \frac{d^2}{n^2}\frac{6}{(d+1)(d+2)} \tag{7.196}$$

$$P(R|\rho) \in \mathrm{col}(P), \tag{7.197}$$

where $\dim \mathrm{col}(P) = d^2$. Somehow these constraints pick out precisely the $2d-2$ dimensional manifold of pure quantum states[8]: Somehow the hyperplane defined by P intersects the spheres in a very peculiar nontrivial way. Better understanding this in purely geometrical terms might lead to a novel characterization of Euclidean Jordan algebras themselves.

Indeed, it may be in our interest to relax our assumptions in order to study this question. For example, let us assume merely that the state space is fully characterized by the uncertainty principle,

$$P(R|\rho) \text{ valid} \iff \forall x \in \mathrm{col}(P) : \langle x^2\rangle_\rho \geq \frac{\alpha-1}{\alpha+1}\Big(\langle x^2\rangle_\mu - 2\langle x\rangle_\mu\langle x\rangle_\rho\Big). \tag{7.198}$$

As we have seen, assuming that our nonclassical product is power associative gives us the Jordan identity, which justifies this characterization of the state space. But let us instead take the uncertainty principle as given. One can then explore the possible state spaces which emerge from different choices of $P(R|R)$. Compatibility with the uncertainty principle is still diagnosed by the positive semidefiniteness of $\mathcal{L}_{P(R|\rho)}$: nothing in that argument depended on the Jordan structure. Thus the state space in fact forms a so-called *spectrahedron*.

A *spectrahedron* is the intersection of an affine-linear space with the convex cone of positive semidefinite matrices [181]. An n-dimensional affine-linear subspace of real symmetric matrices may be parameterized by a vector $x \in \mathbb{R}^n$ and a collection of $n+1$ matrices $\{A_i\}_{i=0}^n$ as

$$A_x = A_0 + \sum_i x_i A_i. \tag{7.199}$$

[8] A state vector $|\psi\rangle$ is defined by $2d$ complex numbers, and after fixing phase and overall normalization, we end up with $2d-2$ parameters.

The spectrahedron itself $A \subseteq \mathbb{R}^n$ consists of those x such that $A_x \geq 0$. Now let

$$T_{ijk} = \frac{1}{2}\Bigg\{\alpha(\alpha+1)\sum_m P(R_m|R_i)P(R_m|R_j)\delta_{mk} \\ - (\alpha-1)\Big(P(R_i|R_j) + \delta_{ik} + \delta_{jk} + (\alpha-1)/n\Big)/n\Bigg\}, \tag{7.200}$$

so that $\mathcal{L}_x = \sum_i T_{ijk}x_k$. Letting $A_0 = 0$ and otherwise $A_k = T_{ijk}$, it is clear then that our state space is indeed a spectrahedron (in fact the intersection of a *linear space* with the positive semidefinite cone. We ought also to intersect the spectrahedron with the probability simplex as well as $\mathrm{col}(P)$ so that the state space is $\mathcal{S} = A \cap \Delta_n \cap \mathrm{col}(P)$.

One may then study this broader class of "variance-bounded" state spaces, which are fully defined by their compatibility with the uncertainty principle. In particular, when are such state spaces self-dual? When does there in fact exist a class of measurements such that if $\mathfrak{x}$ is the valuation on such a measurement $\{X_i\}$ and $x = \mathfrak{x}P(X|R)\Phi$ is the equivalent reference valuation, $\langle \mathfrak{x}^2 \rangle_\rho = \langle x^{\odot 2} \rangle_\rho$, giving some justification for the uncertainty principle itself? When does $x \odot y$ in fact have an equivalent valuation on a measurement which is reference independent? When can the extremal states of the theory be given a simple characterization? Given a choice of $P(R|R)$, what properties does the product $\mathcal{L}_x \Phi y = x \hat{\odot} y$ have, and how do they reflect the geometry of the state space? Conversely, if one chooses a $P(R|R)$ so that $\hat{\odot}$ has some interesting property, what geometry does this imply? Assuming power associativity delivered us an embarrassment of riches, the highly constrained geometry implied by the Jordan product. But there remain deep questions about how exactly the constraints we put on $P(R|R)$ allow a particular state space to emerge, through the interplay between the bare nonclassical product $\odot$ and its projection into $\mathrm{col}(P)$. By studying these variants, one may be better placed to understand the relationship between the algebra of observables and the geometry of the allowed reference distributions, in particular the restriction to $\mathrm{col}(P)$.

Indeed, one final, parting theorem can help dramatize the importance of the subspace restriction. Suppose we defined the product on $\mathbb{R}^n$ in its entirety. It turns out that the parameter γ interpolates between two Jordan algebras: when $\gamma = 1$, we have the Jordan algebra provided by the Hadamard product; when $\gamma = 0$, we have in fact a spin-factor Jordan algebra, although with a negative definite inner product. When $n = 2$, any choice of γ leads to a Jordan algebra. But for $n > 2$, only $\gamma = 0$ or $\gamma = 1$ give rise to Jordan algebras. But if we project the product into a subspace, as we have seen, then the Jordan identity may be satisfied for intermediate values of γ.

Theorem 7.12.1. *Let*

$$x \odot_\gamma y = \gamma(x \circ y) + (1-\gamma)\left(\overline{y}x + \overline{x}y - (x \cdot y)u\right)/n \tag{7.201}$$

on $\mathbb{R}^n$, where $u = (1, \ldots, 1)$ and $\overline{x} = \sum_i x_i$. Then $\odot_\gamma$ satisfies the Jordan identity for every γ when $n = 2$, and for $n > 2$ if and only if $\gamma = 0$ or $\gamma = 1$.

Proof. First, one may confirm that u is the identity element. If $\gamma = 1$, then $\odot_\gamma$ is the elementwise or Hadamard product and hence is associative, so satisfies the Jordan identity: indeed, it is the Jordan product on $\mathbb{R} \oplus \cdots \oplus \mathbb{R} \simeq \mathbb{R}^n$. If $\gamma = 0$, let $a = \overline{x}/n$ and $b = \overline{y}/n$ so that

$$x = au + x_0, \qquad y = bu + y_0, \qquad \overline{x_0} = \overline{y_0} = 0. \tag{7.202}$$

Then

$$x \odot_0 y = ay_0 + bx_0 + (ab - (x_0 \cdot y_0)/n)\, u, \tag{7.203}$$

or, separating out the traceful and traceless components,

$$(a, x_0) \odot_0 (b, y_0) = \Big(ab + B(x_0, y_0), ay_0 + bx_0\Big), \tag{7.204}$$

for $B(x_0, y_0) = -(x_0 \cdot y_0)/n$. In fact, this is the spin-factor Jordan product with a negative definite inner product. Hence $\odot_0$ satisfies the Jordan identity.

For $n = 2$, let $h = e_1 - e_2$ for computational basis vectors $\{e_i\}_{i=1}^n$. Then every element in the algebra is of the form $au + bh$, and since

$$u \odot_\gamma x = x, \qquad h \odot_\gamma h = (2\gamma - 1)u, \tag{7.205}$$

the algebra is generated entirely by h. Thus it suffices to check

$$(h^{\odot_\gamma 2}) \odot_\gamma h = \Big(2\gamma - 1)u\Big) \odot_\gamma h = (2\gamma - 1)h = h \odot_\gamma \Big((2\gamma - 1)u\Big) = h \odot_\gamma (h^{\odot_\gamma 2}). \tag{7.206}$$

Thus Jordan identity holds for all γ when $n = 2$. Now assume $n > 2$. Let $x = e_1 - e_2$ and $y = e_1$ so that

$$\overline{x} = 0, \qquad \overline{y} = 1, \qquad x \cdot x = 2, \qquad x \cdot y = 1, \qquad x \circ x = e_1 + e_2, \qquad x \circ y = e_1. \tag{7.207}$$

We then have

$$x^{\odot_\gamma 2} = \gamma(e_1 + e_2) - 2(1 - \gamma)u/n, \qquad x \odot_\gamma y = \gamma e_1 + (1 - \gamma)(x - u)/n. \tag{7.208}$$

Checking the Jordan identity gives

$$x^{\odot_\gamma 2} \odot_\gamma (x \odot_\gamma y) - x \odot_\gamma (x^{\odot_\gamma 2} \odot_\gamma y) = -2\gamma(1 - \gamma)^2 \left((n - 2)e_2 - \sum_{k=3}^{n} e_k\right)/n^2. \tag{7.209}$$

Since the parenthesized vector is nonzero, the Jordan identity forces

$$\gamma(1-\gamma)^2 = 0, \tag{7.210}$$

so $\gamma = 0$ or $\gamma = 1$. These two values were already shown to satisfy the Jordan identity, and this completes the proof. □

We leave it to future work to more fully explore the interplay between the "simplest modification of the classical product on valuations" and restriction to $\mathrm{col}(P)$. As we saw, the latter can be derived from the fundamental nonclassical coherence relation itself. At the same time, it is only a meaningful restriction because the representation of quantum mechanics afforded by a 3-design must be overcomplete. We hope that further study of the role played by this "gauge freedom" in representation will ultimately allow for an ever more transparent reconstruction of quantum mechanics, which centers the dual freedom of agent and world, mediated by a well-chosen reference measurement.

LIST OF REFERENCES

[1] Christopher A. Fuchs and Blake C. Stacey. QBism: Quantum Theory as a Hero's Handbook. *Proceedings of the International School of Physics "Enrico Fermi"*, 197(Foundations of Quantum Theory):133–202, 2019.

[2] Chris Fuchs. QBism, Where Next? In Philipp Berghofer and Harald A. Wiltsche, editors, *Phenomenology and Qbism: New Approaches to Quantum Mechanics*. Routledge, 2023.

[3] John B. DeBrota, Christopher A. Fuchs, and Rüdiger Schack. QBism's account of quantum dynamics and decoherence. *Physical Review A*, 110(5):052205, November 2024.

[4] Markus Müller. Probabilistic theories and reconstructions of quantum theory. *SciPost Physics Lecture Notes*, page 28, March 2021.

[5] Marcus Appleby, Christopher A. Fuchs, Blake C. Stacey, and Huangjun Zhu. Introducing the Qplex: A novel arena for quantum theory. *The European Physical Journal D*, 71(7), July 2017.

[6] Wojciech Słomczyński and Anna Szymusiak. Morphophoric POVMs, generalised qplexes, and 2-designs. *Quantum*, 4:338, September 2020.

[7] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020.

[8] Chengsi Mao, Changhao Yi, and Huangjun Zhu. The Magic in Qudit Shadow Estimation based on the Clifford Group, 2024.

[9] Martin Kliesch and Ingo Roth. Theory of Quantum System Certification. *PRX Quantum*, 2(1), January 2021.

[10] Datong Chen and Huangjun Zhu. Nonstabilizerness enhances thrifty shadow estimation, 2024.

[11] Philipp Berghofer. Defending the quantum reconstruction program. *European Journal for Philosophy of Science*, 14(3), September 2024.

[12] Gilles Brassard. Is information the key? *Nature Physics*, 1(1):2–4, October 2005.

[13] Christopher A. Fuchs. Quantum mechanics as quantum information (and only a little more), 2002.

[14] Alexei Grinbaum. On the notion of reconstruction of quantum theory, 2005.

[15] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Informational derivation of quantum theory. *Physical Review A*, 84(1), July 2011.

[16] Lucien Hardy. Quantum theory from five reasonable axioms, 2001.

[17] Lucien Hardy. Reconstructing quantum theory. In Giulio Chiribella and Robert W. Spekkens, editors, *Quantum Theory: Informational Foundations and Foils*, pages 223–248. Springer Netherlands, Dordrecht, 2016.

[18] Philipp Andres Höhn. Toolbox for reconstructing quantum theory from rules on information acquisition. *Quantum*, 1:38, December 2017.

[19] Lluís Masanes and Markus P Müller. A derivation of quantum theory from physical requirements. *New Journal of Physics*, 13(6):063001, June 2011.

[20] John H. Selby, Carlo Maria Scandolo, and Bob Coecke. Reconstructing quantum theory from diagrammatic postulates. *Quantum*, 5:445, April 2021.

[21] John van de Wetering. An effect-theoretic reconstruction of quantum theory. January 2018.

[22] Alexander Wilce. A Royal Road to Quantum Theory (or Thereabouts). *Entropy*, 20(4):227, March 2018.

[23] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.

[24] Blake C. Stacey. Quantum Theory as Symmetry Broken by Vitality, 2019.

[25] Detlef Dürr, Sheldon Goldstein, and Nino Zanghí. Quantum equilibrium and the origin of absolute uncertainty. *Journal of Statistical Physics*, 67(5–6):843–907, June 1992.

[26] Lev Vaidman. Why the many-worlds interpretation? *Quantum Reports*, 4(3):264–271, August 2022.

[27] Jacques Faraut and Adam Korányi. *Analysis on Symmetric Cones*. Oxford University PressOxford, December 1994.

[28] P Jordan, J v. Neumann, and E Wigner. On an algebraic generalization of the quantum mechanical formalism. *Annals of Mathematics*, 35(1):29, January 1934.

[29] Garrett Birkhoff and John Von Neumann. The logic of quantum mechanics. *The Annals of Mathematics*, 37(4):823, October 1936.

[30] A. Hartkämper, H. Neumann, J. Ehlers, K. Hepp, H. A. Weidenmüller, and W. Beiglböck, editors. *Foundations of Quantum Mechanics and Ordered Linear Spaces: Advanced Study Institute Marburg 1973*, volume 29 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1974.

[31] E. B. Davies and J. T. Lewis. An operational approach to quantum probability. *Communications in Mathematical Physics*, 17(3):239–260, September 1970.

[32] Günther Ludwig. *An Axiomatic Basis for Quantum Mechanics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985.

[33] G.W. Mackey. *Mathematical Foundations of Quantum Mechanics*. Dover Books on Mathematics. Dover Publications, 2004.

[34] C. H. Randall and D. J. Foulis. The operational approach to quantum mechanics. In *Physical Theory as Logico-Operational Structure*, pages 167–201. Springer Netherlands, 1978.

[35] Howard Barnum and Alexander Wilce. Information processing in convex operational theories. *Electronic Notes in Theoretical Computer Science*, 270(1):3–15, 2011.

[36] Jonathan Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3):032304, March 2007.

[37] Peter Janotta and Haye Hinrichsen. Generalized probability theories: What determines the structure of quantum theory? *Journal of Physics A: Mathematical and Theoretical*, 47(32):323001, July 2014.

[38] Ludovico Lami. Non-classical correlations in quantum mechanics and beyond, 2018.

[39] Martin Plávala. General probabilistic theories: An introduction. *Physics Reports*, 1033:1–64, 2023.

[40] Guillaume Aubrun, Ludovico Lami, Carlos Palazuelos, and Martin Plávala. Entanglement and superposition are equivalent concepts in any physical theory. *Physical Review Letters*, 128(16), April 2022.

[41] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Cloning and Broadcasting in Generic Probabilistic Theories, 2006.

[42] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Teleportation in general probabilistic theories, 2008.

[43] Andrew J. P. Garner, Markus P. Müller, and Oscar C. O. Dahlsten. The complex and quaternionic quantum bit from relativity of simultaneity on an interferometer. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2208):20170596, December 2017.

[44] Martin Plávala. All measurements in a probabilistic theory are compatible if and only if the state space is a simplex. *Physical Review A*, 94(4), October 2016.

[45] Howard Barnum, Markus P Müller, and Cozmin Ududec. Higher-order interference and single-system postulates characterizing quantum theory. *New Journal of Physics*, 16(12):123029, December 2014.

[46] Teiko Heinosaari, Leevi Leppäjärvi, and Martin Plávala. No-free-information principle in general probabilistic theories. *Quantum*, 3:157, July 2019.

[47] Marius Krumm and Markus P. Müller. Quantum computation is the unique reversible circuit model for which bits are balls. *npj Quantum Information*, 5(1), January 2019.

[48] Markus P. Müller and Cozmin Ududec. Structure of reversible computation determines the self-duality of quantum theory. *Physical Review Letters*, 108(13):130401, March 2012.

[49] Victoria J Wright and Stefan Weigert. General probabilistic theories with a Gleason-type theorem. *Quantum*, 5:588, November 2021.

[50] Mayalakshmi Kolangatt, Thigazholi Muruganandan, Sahil Gopalkrishna Naik, Tamal Guha, Manik Banik, and Sutapa Saha. Bipartite polygon models: Entanglement classes and their nonlocal behaviour. *Quantum*, 9:1599, January 2025.

[51] Thomas Scheidl, Rupert Ursin, Johannes Kofler, Sven Ramelow, Xiao-Song Ma, Thomas Herbst, Lothar Ratschbacher, Alessandro Fedrizzi, Nathan K. Langford, Thomas Jennewein, and Anton Zeilinger. Violation of local realism with freedom of choice. *Proceedings of the National Academy of Sciences*, 107(46):19708–19713, November 2010.

[52] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, October 1969.

[53] John D. Trimmer. The present situation in quantum mechanics: A translation of schrödinger's "cat paradox" paper. *Proceedings of the American Philosophical Society*, 124(5):323–338, 1980.

[54] Lucien Hardy and William K. Wootters. Limited holism and real-vector-space quantum theory. *Foundations of Physics*, 42(3):454–473, December 2011.

[55] B Dakić, T Paterek, and Č Brukner. Density cubes and higher-order interference theories. *New Journal of Physics*, 16(2):023028, February 2014.

[56] Ian Hacking. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge University Press, 2 edition, July 2006.

[57] John B. DeBrota, Christopher A. Fuchs, Jacques L. Pienaar, and Blake C. Stacey. Born's rule as a quantum extension of Bayesian coherence. *Physical Review A: Atomic, Molecular, and Optical Physics*, 104(2):022207, August 2021.

[58] Christopher A. Fuchs and Rüdiger Schack. Quantum-Bayesian coherence. *Reviews of Modern Physics*, 85(4):1693–1715, December 2013.

[59] P. Berghofer and H.A. Wiltsche. *Phenomenology and Qbism: New Approaches to Quantum Mechanics*. Routledge Studies in the Philosophy of Mathematics and Physics. Taylor & Francis, 2023.

[60] Amanda Gefter. Enaction for qbists. November 2024.

[61] Mark Rowlands. *The New Science of the Mind*. A Bradford Book. Bradford Books, Cambridge, MA, September 2010.

[62] Shayne F. D. Waldron. *An Introduction to Finite Tight Frames*. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY, 2018.

[63] Valentin Obst, Arne Heimendahl, Tanmay Singal, and David Gross. Wigner's Theorem for stabilizer states and quantum designs. *Journal of Mathematical Physics*, 65(11), November 2024.

[64] Aloys Krieg. *The Minnesota Notes on Jordan Algebras and Their Applications*. Number v.1710 in Lecture Notes in Mathematics Ser. Springer Berlin / Heidelberg, Berlin, Heidelberg, 1999.

[65] F. Alizadeh. An Introduction to Formally Real Jordan Algebras and Their Applications in Optimization. In Miguel F. Anjos and Jean B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166, pages 297–337. Springer US, New York, NY, 2012.

[66] Howard Barnum and Joachim Hilgert. Strongly symmetric spectral convex bodies are Jordan algebra state spaces, 2019.

[67] Carlton Caves. Probabilities as betting odds and the Dutch book.

[68] C. Van Fraassen. Belief and the Will. *The Journal of Philosophy*, 81(5):235, May 1984.

[69] Bruno De Finetti. *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons, Chichester, UK Hoboken, NJ, 2017.

[70] Bruno de Finetti. Foresight: Its Logical Laws, Its Subjective Sources. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 134–174. Springer, New York, NY, 1992.

[71] Adi Ben-Israel and Thomas N. E. Greville. *Generalized Inverses*. CMS Books in Mathematics. Springer-Verlag, New York, 2003.

[72] R. L. Hudson and G. R. Moody. Locally normal symmetric states and an analogue of de Finetti's theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 33(4):343–351, December 1976.

[73] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Unknown quantum states: The quantum de Finetti representation. *Journal of Mathematical Physics*, 43(9):4537–4559, September 2002.

[74] Ludovico Lami. Non-classical correlations in quantum mechanics and beyond, March 2018.

[75] Matthew B. Weiss. Depolarizing reference devices in generalized probabilistic theories, 2024.

[76] Anna Szymusiak and Wojciech Słomczyński. Can QBism exist without Q? Morphophoric measurements in generalised probabilistic theories. *Quantum*, 9:1598, January 2025.

[77] A J Scott. Tight informationally complete quantum measurements. *Journal of Physics A: Mathematical and General*, 39(43):13507–13530, October 2006.

[78] John B. DeBrota, Christopher A. Fuchs, and Blake C. Stacey. Symmetric informationally complete measurements identify the irreducible difference between classical and quantum systems. *Physical Review Research*, 2(1):013074, January 2020.

[79] David Schmid, John H. Selby, Elie Wolfe, Ravi Kunjwal, and Robert W. Spekkens. Characterization of Noncontextuality in the Framework of Generalized Probabilistic Theories. *PRX Quantum*, 2(1):010331, February 2021.

[80] David Schmid, John H. Selby, Matthew F. Pusey, and Robert W. Spekkens. A structure theorem for generalized-noncontextual ontological models. *Quantum*, 8:1283, March 2024.

[81] Farid Shahandeh, Theodoros Yianni, and Mina Doosti. A Unified Linear Algebraic Framework for Physical Models and Generalized Contextuality, December 2025.

[82] John H. Selby, Elie Wolfe, David Schmid, Ana Belén Sainz, and Vinicius P. Rossi. Linear Program for Testing Nonclassicality and an Open-Source Implementation. *Physical Review Letters*, 132(5):050202, January 2024.

[83] Bit-Shun Tam. A geometric treatment of generalized inverses and semigroups of nonnegative matrices. *Linear Algebra and its Applications*, 41:225–272, December 1981.

[84] Matthew B. Weiss. Characterizing quantum state-space with a single quantum measurement. *Physical Review A*, 111(5):052205, May 2025.

[85] Matthew Weiss. Heyredhat/redesigning: V1. Zenodo, July 2026.

[86] Jan Von Plato. *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*. Cambridge University Press, 1 edition, January 1994.

[87] Richard C. Jeffrey. *Subjective Probability: The Real Thing*. Cambridge University Press, Cambridge, U.K New York, 2004.

[88] Ian Hacking. Slightly More Realistic Personal Probability. *Philosophy of Science*, 34(4):311–325, December 1967.

[89] John B. DeBrota, Christopher A. Fuchs, and Rüdiger Schack. Quantum Dynamics Happens Only on Paper: QBism's Account of Decoherence. *Physical Review A: Atomic, Molecular, and Optical Physics*, 110(5):052205, November 2024.

[90] Michael Goldstein. The Prevision of a Prevision. *Journal of the American Statistical Association*, 78(384):817–819, December 1983.

[91] Glenn Shafer. A Subjective Interpretation of Conditional Probability. *Journal of Philosophical Logic*, 12(4):453–466, 1983.

[92] Christopher A. Fuchs and Rüdiger Schack. Bayesian Conditioning, the Reflection Principle, and Quantum Decoherence. In Yemima Ben-Menahem and Meir Hemmo, editors, *Probability in Physics*, pages 233–247. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[93] Andrej Nikolaevič Kolmogorov and A. T. Bharucha-Reid. *Foundations of the Theory of Probability*. Dover Books on Mathematics. Dover Publications, Mineola, New York, second english edition, dover edition, republication of the 1956 second edition of the work originally published in 1950 by chelsea publishing, new york edition, 2018.

[94] R. Tyrrell Rockafellar. *Convex Analysis*. Number 28 in Princeton Mathematical Series. Princeton University Press, Princeton, N.J, 1970.

[95] Matthew Macauley, Brian Rabern, and Landon Rabern. A Novel Proof of the Heine-Borel Theorem, August 2008.

[96] Walter Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, New York, 3d ed edition, 1976.

[97] Walter Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, New York, 2nd ed edition, 1991.

[98] P. Busch. Quantum States and Generalized Observables: A Simple Proof of Gleason's Theorem. *Physical Review Letters*, 91(12):120403, September 2003.

[99] Carlton M. Caves, Christopher A. Fuchs, Kiran K. Manne, and Joseph M. Renes. Gleason-type derivations of the quantum probability rule for generalized measurements. *Foundations of Physics*, 34:193–209, 2003.

[100] Stephen L. Campbell and Carl D. Meyer. *Generalized Inverses of Linear Transformations*. Society for Industrial and Applied Mathematics, January 2009.

[101] Shidong Li. On general frame decompositions. *Numerical Functional Analysis and Optimization*, 16(9-10):1181–1191, January 1995.

[102] Robert Piziak and P.L. Odell. *Matrix Theory*. Chapman and Hall/CRC, 0 edition, February 2007.

[103] R. Piziak and P. L. Odell. Full Rank Factorization of Matrices. *Mathematics Magazine*, 72(3):193–201, 1999.

[104] Lami, Ludovico. General probabilistic theories and meta-theoretical knowledge, December 2018.

[105] Sheldon Jay Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer Nature, Cham, 2024.

[106] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[107] Jonathan Barrett and Matthew Leifer. The de Finetti theorem for test spaces. *New Journal of Physics*, 11(3):033024, March 2009.

[108] Christopher A. Fuchs, Rüdiger Schack, and Petra F. Scudo. De Finetti representation theorem for quantum-process tomography. *Physical Review A*, 69(6):062305, June 2004.

[109] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Probabilistic theories with purification. *Physical Review A*, 81(6):062348, June 2010.

[110] Carlo Maria Scandolo. Information-theoretic foundations of thermodynamics in general probabilistic theories, 2019.

[111] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1 edition, December 1985.

[112] John B. DeBrota and Blake C. Stacey. L\"uders channels and the existence of symmetric-informationally-complete measurements. *Physical Review A*, 100(6):062327, December 2019.

[113] Pranaw Rungta, V. Bužek, Carlton M. Caves, M. Hillery, and G. J. Milburn. Universal state inversion and concurrence in arbitrary dimensions. *Physical Review A*, 64(4):042315, September 2001.

[114] Joseph M. Renes, Robin Blume-Kohout, A. J. Scott, and Carlton M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, June 2004.

[115] Marcus Appleby, Steven Flammia, Gary McConnell, and Jon Yard. SICs and Algebraic Number Theory. *Foundations of Physics*, 47(8):1042–1059, August 2017.

[116] Ingemar Bengtsson. SICs: Some Explanations. *Foundations of Physics*, 50(12):1794–1808, December 2020.

[117] Christopher Fuchs, Michael Hoang, and Blake Stacey. The SIC Question: History and State of Play. *Axioms*, 6(3):21, July 2017.

[118] Henry Cohn, Abhinav Kumar, and Gregory Minton. Optimal simplices and codes in projective spaces. *Geometry & Topology*, 20(3):1289–1357, July 2016.

[119] Aram W. Harrow. The Church of the Symmetric Subspace, 2013.

[120] Christopher A. Fuchs, Maxim Olshanii, and Matthew Benjamin Weiss. Quantum mechanics? It's all fun and games until someone loses an i. *Asian Journal of Physics*, 30(12):1707–1726, 2021.

[121] Daniel Hughes and Shayne Waldron. Spherical (t,t)-designs with a small number of vectors. *Linear Algebra and its Applications*, 608:84–106, January 2021.

[122] A. Schrijver. *Theory of Linear and Integer Programming*. A Wiley-Interscience Publication. Wiley, 1986.

[123] Markus P. Müller and Andrew J. P. Garner. Testing Quantum Theory by Generalizing Noncontextuality. *Physical Review X*, 13(4):041001, October 2023.

[124] David Schmid, John H. Selby, Vinicius P. Rossi, Roberto D. Baldijão, and Ana Belén Sainz. Shadows and subsystems of generalized probabilistic theories: When tomographic incompleteness is not a loophole for contextuality proofs. *Quantum*, 9:1880, October 2025.

[125] I. Markovsky. *Low Rank Approximation: Algorithms, Implementation, Applications*. Communications and Control Engineering. Springer London, 2011.

[126] Steven Diamond and Stephen Boyd. CVXPY: A Python-Embedded Modeling Language for Convex Optimization, 2016.

[127] G.M. Ziegler. *Lectures on Polytopes*. Graduate Texts in Mathematics. Springer New York, 2012.

[128] Pycddlib: A Python wrapper for cddlib.

[129] J. Ding and N. H. Rhee. Teaching Tip: When a Matrix and Its Inverse Are Stochastic. *The College Mathematics Journal*, 44(2):108–109, March 2013.

[130] Christopher Ferrie and Joseph Emerson. Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations. *Journal of Physics A: Mathematical and Theoretical*, 41(35):352001, September 2008.

[131] Christopher Ferrie and Joseph Emerson. Framed Hilbert space: Hanging the quasi-probability pictures of quantum theory. *New Journal of Physics*, 11(6):063040, June 2009.

[132] Christopher Ferrie, Ryan Morris, and Joseph Emerson. Necessity of negativity in quantum theory. *Physical Review A*, 82(4):044103, October 2010.

[133] Christopher Ferrie. Quasi-probability representations of quantum theory with applications to quantum information science. *Reports on Progress in Physics*, 74(11):116001, November 2011.

[134] R. F. Werner and M. M. Wolf. All-multipartite Bell-correlation inequalities for two dichotomic observables per site. *Physical Review A*, 64(3):032112, August 2001.

[135] Stefano Pironio. All Clauser–Horne–Shimony–Holt polytopes. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424020, October 2014.

[136] Christopher A. Fuchs and Blake C. Stacey. QBism, Polishing Some Points, 2025.

[137] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419–478, April 2014.

[138] Ravi Kunjwal and Sibasish Ghosh. Minimal state-dependent proof of measurement contextuality for a qubit. *Physical Review A*, 89(4):042118, April 2014.

[139] Niels Bohr. The quantum postulate and the recent development of atomic Theory. *Nature*, 121(3050):580–590, April 1928.

[140] J. Baggott. *The Quantum Story: A History in 40 Moments*. OUP Oxford, 2011.

[141] Jinchuan Hou and Kan He. Uncertainty relations for any multi observables. 2016.

[142] D. M. Appleby, Asa Ericsson, and Christopher A. Fuchs. Properties of QBist State Spaces. *Foundations of Physics*, 41(3):564–579, April 2010.

[143] Michael D. Mazurek, Matthew F. Pusey, Kevin J. Resch, and Robert W. Spekkens. Experimentally Bounding Deviations From Quantum Theory in the Landscape of Generalized Probabilistic Theories. *PRX Quantum*, 2(2):020302, April 2021.

[144] Michael J. Grabowecky, Christopher A. J. Pollack, Andrew R. Cameron, Robert W. Spekkens, and Kevin J. Resch. Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography. *Physical Review A*, 105(3):032204, March 2022.

[145] Howard Barnum and Alexander Wilce. Post-classical probability theory, 2013.

[146] Hans Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60(12):1103–1106, March 1988.

[147] Markus Heinrich and David Gross. Robustness of Magic and Symmetries of the Stabiliser Polytope. *Quantum*, 3:132, April 2019.

[148] D. Marcus Appleby, Hoan Bui Dang, and Christopher A. Fuchs. Symmetric informationally-complete quantum states as analogues to orthonormal bases and minimum-uncertainty states. *Entropy. An International and Interdisciplinary Journal of Entropy and Information Studies*, 16(3):1484–1492, 2014.

[149] Gianluca Cuffaro and Christopher A. Fuchs. Quantum states with maximal magic, 2024.

[150] John B. DeBrota, Christopher A. Fuchs, and Blake C. Stacey. The Varieties of Minimal Tomographically Complete Measurements. *International Journal of Quantum Information*, 19(07):2040005, 2021.

[151] Dave Bacon, Isaac L. Chuang, and Aram W. Harrow. The Quantum Schur Transform: I. Efficient Qudit Circuits, 2005.

[152] P.D Seymour and Thomas Zaslavsky. Averaging sets: A generalization of mean values and spherical designs. *Advances in Mathematics*, 52(3):213–240, 1984.

[153] Ingemar Bengtsson and Karol Zyczkowski. On discrete structures in finite Hilbert spaces, 2017.

[154] Huangjun Zhu and Berthold-Georg Englert. Quantum state tomography with fully symmetric measurements and product measurements. *Physical Review A*, 84(2), August 2011.

[155] David Gross, Sepehr Nezami, and Michael Walter. Schur–Weyl Duality for the Clifford Group with Applications: Property Testing, a Robust Hudson Theorem, and de Finetti Representations. *Communications in Mathematical Physics*, 385(3):1325–1393, June 2021.

[156] Huangjun Zhu, Chengsi Mao, and Changhao Yi. Third moments of qudit Clifford orbits and 3-designs based on magic orbits, 2024.

[157] Huangjun Zhu. Multiqubit Clifford groups are unitary 3-designs. *Physical Review A*, 96(6), December 2017.

[158] Nick S. Jones and Noah Linden. Parts of quantum states. *Physical Review A*, 71(1), January 2005.

[159] Christopher A. Fuchs, Maximilian Schlosshauer, and Blake C. Stacey. My Struggles with the Block Universe, 2015.

[160] Michał Oszmaniec, Daniel J. Brod, and Ernesto F. Galvão. Measuring relational information between quantum states, and applications. *New Journal of Physics*, (1):013053, January 2024.

[161] Paul K. Townsend. The Jordan formulation of Quantum Mechanics: A review, 2016.

[162] Dmitri Akhiezer. On the Commutator Map for Real Semisimple Lie Algebras. *Moscow Mathematical Journal*, 15(4):609–613, 2015.

[163] John Baez. Dynamics in Jordan Algebras.

[164] Kevin McCrimmon. *A Taste of Jordan Algebras*. Universitext. Springer, New York, NY, 2004 edition, 2003.

[165] N. Jacobson. *Structure and Representations of Jordan Algebras*. Colloquium Publications. American Mathematical Society, 1968.

[166] Ivan Šupić and Joseph Bowles. Self-testing of quantum systems: A review. *Quantum*, 4:337, September 2020.

[167] P. Jordan. *Über Verallgemeinerungsmöglichkeiten Des Formalismus Der Quantenmechanik*. Sonderdrucke Aus Den Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen : Mathematisch-physikalische Klasse. Weidmann, 1933.

[168] Richard D. Schafer. *An Introduction to Nonassociative Algebras*. Dover Publications, 1995.

[169] Howard Barnum and Alexander Wilce. Local tomography and the Jordan structure of quantum theory, 2013.

[170] Michael Orlitzky. Rank computation in Euclidean Jordan algebras. *Journal of Symbolic Computation*, 113:181–192, November 2022.

[171] M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. Beyond the Swap Test: Optimal Estimation of Quantum State Overlap. *Physical Review Letters*, 124(6):060503, February 2020.

[172] Xavier Gitiaux, Ian Morris, Maria Emelianenko, and Mingzhen Tian. SWAP test for an arbitrary number of quantum states. *Quantum Information Processing*, 21(10):344, October 2022.

[173] Yoshifumi Nakata, Yuki Takeuchi, Martin Kliesch, and Andrew Darmawan. Computational Complexity of Unitary and State Design Properties. *PRX Quantum*, 6(3):030345, September 2025.

[174] Max McGinley. Postselection-Free Learning of Measurement-Induced Quantum Dynamics. *PRX Quantum*, 5(2):020347, May 2024.

[175] Markus Grassl: Computing Numerical and Exact SIC-POVMs – nisq.pl.

[176] Marcus Appleby, Steven T Flammia, and Gene S Kopp. A Constructive Approach to Zauner's Conjecture via the Stark Conjectures, 2025.

[177] Aidan Roy and A. J. Scott. Unitary designs and codes. *Designs, Codes and Cryptography*, 53(1):13–31, October 2009.

[178] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: On the structure of unitary designs. *Journal of Mathematical Physics*, 48(5):052104, May 2007.

[179] Eiichi Bannai, Yoshifumi Nakata, Takayuki Okuda, and Da Zhao. Explicit construction of exact unitary designs. *Advances in Mathematics*, 405:108457, August 2022.

[180] Howard Barnum, Cozmin Ududec, and John van de Wetering. Self-duality and Jordan structure of quantum theory follow from homogeneity and pure transitivity, 2023.

[181] Cynthia Vinzant. What is... a spectrahedron. *Notices of the American Mathematical Society*, 61:492–494, 2014.