Fall 2024

Informational Principles and Structures in Quantum Theory

Yang Yu
University at Albany, State University of New York, yyu9@albany.edu

INFORMATIONAL PRINCIPLES AND STRUCTURES IN QUANTUM THEORY

by

Yang Yu

A Dissertation

Submitted to the University at Albany, State University of New York

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

College of Arts & Sciences

Physics

Fall 2024

ABSTRACT

This thesis examines the concept of information within the realm of quantum physics, investigating the nuanced relationship between information and physical laws as applied to quantum systems. With no consensus on a single definition of information in the physical sciences, our exploration is partitioned into two significant studies, each addressing distinct aspects of information in quantum contexts.

In the first study, we concentrate on the information that can be obtained from measurement results on identical quantum systems. The traditional approach of using Shannon entropy is limited due to its applicability primarily to discrete probability distributions. By incorporating a Bayesian update framework, we redefine the process of information gain, which allows for a more nuanced understanding of information dynamics within quantum measurements. Key findings from this approach include a novel expression for quantifying information gain and a principle for the selection of appropriate priors, specifically employing Jeffreys' prior for binomial distributions. The study also highlights the effectiveness of Jeffreys' binomial prior in optimizing quantum communication scenarios, such as maximizing the information deciphered by a receiver (Bob) from a sender's (Alice) message-encoded qubits.

The second study shifts focus to the foundational aspects of quantum theory itself, employing an informational approach to reconstruct the theory's underlying structure. Here, quantum measurements are conceptualized as finite outcome questions linked through classical logical operations within systems extending beyond binary dimensions. By introducing intuitive informational postulates, we achieve a partial reconstruction of quantum theory, particularly within systems characterized by prime number dimensions. This reconstruction yields rich connections between classical logical gates, generalized Pauli matrices, and mutually unbiased bases, enhancing our comprehension of how information flows during measurements on maximally entangled systems.

ACKNOWLEDGMENTS

I want to thank many people who helped me with my thesis.

First, I want to thank the Physics Department at University at Albany for accepting me into the program and giving me a scholarship. This support made it possible for me to focus on my research. The department has been a great place to learn and teach.

I am also very grateful to my parents. Their support and belief in me have been a constant source of strength. I could not have done this without them.

Most importantly, I want to thank my advisor, Prof. Philip Goyal. Your help, advice, and guidance have been invaluable. You have always been patient and dedicated, and I am very grateful for your mentorship and support.

Thank you all for making this possible.

CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Background and Significance

Applying information theory to quantum mechanics has opened new avenues for understanding the foundations of quantum theory. Understanding how information is quantified and processed in quantum systems has profound implications for both theoretical research and practical applications, such as quantum computing and quantum cryptography. This thesis presents two contributions to this evolving field, each addressing fundamental aspects of information in quantum systems.

1.2 Research Problem and Questions

The first piece of work, published in *Information*[1], delves into the operational perspective of information gained from quantum measurements. It scrutinizes the appropriate measures for quantifying information, especially when dealing with continuous probability distributions, and proposes new informational postulates to guide the selection of these measures. The second piece of work, available on arXiv[2], and presented[3] at "Quantum Reconstruction and Beyond" in August 2023, Graz, Austria, extends the discussion to higher-dimensional quantum systems. It explores the structure of quantum measurements through the lens of information theory, proposing a new construct termed "quantum question structure" to understand the complex relationships between measurements and the states they reveal.

[1] https://doi.org/10.3390/info15050287

[2] https://arxiv.org/abs/2402.19448

[3] An online recording of the presentation is also available: https://youtu.be/pBgIEX1j9vg?si=UsqLJa8kbgBLDrsg

1.3 Objectives and Scope

The objective of this thesis is to deepen our understanding of how information is quantified and utilized in quantum systems. Specifically, it aims to:

1. Propose and validate a physically intuitive postulate for determining the information gained from quantum measurements.
2. Investigate the applicability of two different measures of information gain, differential and relative information gain, in quantum tomography and in the foundations of quantum theory.
3. Apply the formalism of information theory to higher-dimensional quantum systems, introducing and exploring the concept of quantum question structures.
4. Establish connections between quantum question structures and conventional quantum mechanics, particularly in the context of mutually unbiased bases (MUBs) and generalized Pauli matrices.

1.4 Structure of the Thesis

This thesis is structured to guide the reader from the fundamental principles of information theory to advanced applications in quantum mechanics, organized as follows:

- **Chapter 2: Brief Introduction of Information Theory**
  This chapter provides an overview of information theory, covering essential concepts such as probability, Shannon entropy, mutual information, and Kullback-Leibler divergence. It also introduces the mathematical foundations and key theorems that underpin the subsequent discussions and analyses in the thesis.
- **Chapter 3: Importance of Information Theory in Physics**
  This chapter explores the historical development and interplay between information theory and physics, with a particular focus on the impact of quantum theory. It examines how information-theoretic concepts have been integrated into the study of physical systems and highlights key milestones and contributions in the field. The

chapter also addresses the role of information in foundational questions of quantum mechanics and the evolution of these ideas over time.

- **Chapter 4: Operational Perspective on Quantum Information Gain**
  Here, we present the first piece of work, focusing on the quantification of information gained from quantum measurements. The chapter discusses potential information measures based on Kullback-Leibler divergence and introduces new informational postulates to resolve ambiguities in different choices of information measure. It also examines the differential and relative information gain measures, analyzing their applicability and limitations in both tomographic applications and the reconstruction of quantum theory. Numerical and asymptotic analyses are provided to illustrate the behavior of these measures under different conditions, and the chapter concludes with a comparison of the two measures and the proposal of the Principle of Information Increase.

- **Chapter 5: Quantum Question Structures**
  This chapter presents the second piece of work, applying information theory to higher-dimensional quantum systems. It introduces the concept of quantum question structures and explores their implications for understanding the relationship between measurements and the states they reveal. The chapter begins with a discussion of spin-$\frac{1}{2}$ particles and the formalism of quantum tomography, then generalizes these ideas to higher-dimensional systems. A novel informational approach to the reconstruction of quantum theory is discussed without using traditional linear space language. The chapter also provides a connection between the new theoretical constructs and traditional quantum mechanics, using mutually unbiased bases (MUBs) and generalized Pauli matrices to translate abstract results into more familiar terms.

In Chapter 6, we provide a concise summary and suggest potential paths for future research. This conclusion chapter encapsulates the thesis's main contributions, detailing key findings and their significance in the integration of information theory with quantum mechanics.

PART I

Background

CHAPTER 2

A Brief Introduction to Information Theory

2.1 The Concept of Probability

The main focus of this thesis is the relation between information theory and physics. The key connection between information theory and physics is probability which plays an important role in both fields. In this introduction, we introduce the basis of probability as well as the conventions we are using.

2.1.1 Bayesian Probability

In the following context, when we refer to probability, we are specifically addressing the degree of belief in the truth of a proposition under certain conditions. To illustrate, consider the proposition 'It will rain tomorrow.' Initially, one might assign a 50% probability based on typical weather patterns. However, upon receiving a weather forecast predicting rain, one might update this probability to 80%, utilizing Bayes' theorem to adjust one's belief in light of new evidence. When receiving new related knowledge to the target proposition, we may update the degree of belief via Bayes' theorem:

$$\Pr(A|B) = \frac{\Pr(B|A)\ \Pr(A)}{\Pr(B)} \tag{2.1}$$

$\Pr(A)$ is the degree of belief about the truth of proposition A before knowing the condition B, which is also called the prior probability of A. Once we are given the condition B then the updated degree of belief on A, $\Pr(A|B)$, is the the posterior probability of A. $\Pr(B|A)$ is the likelihood, describing the degree of belief of B assuming A is true, and $\Pr(B)$ is the marginal likelihood or 'evidence'. Proposition B is a related proposition to A which usually plays the role as "data". For example proposition A could be "The value of a physical constant χ is equal to r" and B may be the observation results related to the constant χ.

This type of probability is Bayesian probability. It describes a dynamic process and enjoys a great benefit of generality. In comparison to the ensemble-based probability which is

defined over the frequencies over large number of identical trials, we can describe propositions that cannot be associated with a proper ensemble. Some examples are: "Tomorrow will be raining.", "The physics laws will be valid in the next 10 minutes." or "The nearby donut shop will be closed early in the rest of the week." We may gradually update our degree of belief on different propositions once receiving new evidence. The evidence could be the detailed data from measurements, or just another strong belief about a certain theory. In the situation that there is no confusion about the underlying conditions or assumptions, we may just use the usual phrase "probability of a proposition".

2.1.2 Cox's Axioms

Cox's axioms [15] lay the foundation for a mathematical framework of probability that aligns with the principles of Bayes' theorem, ensuring logical consistency and coherence in probabilistic reasoning. These axioms state that probabilities must satisfy certain logical properties. More specifically, Cox's axioms assert that probabilities are represented by real numbers, and that these numbers obey laws of combination that mirror the Boolean laws that apply to propositions.

Non-negativity and normalization The probability should be a non-negative real number. Moreover, the range of probability is between 0 and 1: $0 \leq \Pr(A|I) \leq 1$. By convention we use I to represent all our background knowledge about the proposition A. The two extreme cases, probability with values 0 and 1, are only happens when we have full certainty that the proposition is false or true. For example, if we include the Peano axioms as part of the background knowledge I, then we may have the following two probabilities: $\Pr(\text{“}1+1=2\text{”}|I) = 1$, $\Pr(\text{“}1+1!=2\text{”}|I) = 0$

Product rule Consider two propositions a and b and their logical combination a AND b or $a \cdot b$. Cox proposes an axiom that there exists a function F such that

$$\Pr(a \cdot b|I) = F(\Pr(a|b \cdot I), \Pr(b|I))$$

This function F is assumed to be consistent under boolean algebra of the propositions.

Sum rule Cox also proposes another function S such that a proposition A and its

negation $\neg A$ have the following relation

$$\Pr(A|I) = S(\Pr(\neg A|I))$$

This function S shall also be consistent under boolean algebra of the propositions.

These axioms are pivotal because they guarantee that probabilities derived from them are logically sound and applicable across various fields. For instance, violating the non-negativity axiom (assigning a probability less than 0) would lead to absurd outcomes, such as predicting events with negative chances of occurring, undermining the framework's logical structure. Followed by Cox's axioms, the Cox's theorems provide the basic mathematical justifications for Bayesian probability on reasoning the probability of propositions. Any probability measure that satisfies the above axioms will have the following results:

1. $0 \leq \Pr(A|I) \leq 1, \quad \Pr(\text{"}True\text{"}|I) = 1$
2. $\Pr(A, B|I) = \Pr(A|B, I) \times \Pr(B|I)$
3. $\Pr(A|I) + \Pr(\neg A|I) = 1$

2.1.3 Subjective Bayesianism vs Objective Bayesianism

Both Cox's theorems and Bayes' theorem do not specify how to assign the value for different probabilities, particularly how to choose prior probabilities.

Subjective Bayesian. For the same proposition, different agents might assign varying prior probabilities based on their unique perspectives and information. We must admit that there are many propositions that we cannot have common choice of prior, due to the various background knowledge of different agents. Even worse, sometimes we even cannot express and quantify the hidden assumptions or background knowledge. For example, we may choose an arbitrary coffee shop in the map and determine the probability that "this shop will be open tomorrow". The prior probability of this proposition reflects personal belief and prior experience. This situation is the subjective Bayesian approach, where the probability of a proposition is agent-dependent rather than objective fact.

Objective Bayesian. As long as we are dealing with physics which strives to minimize subjective variations, we tend to use objective Bayesian probability. That is, we consider agents and propositions which are such that, given the same background knowledge different agents will have the same degree of belief about a proposition. To achieve objectivity, we restrict the range of propositions, so that the propositions are only physics related and can be clearly interpreted. We also stipulate that, all the external knowledge we use for Bayesian updating consists of either the physical theory or measurement results. For example, we may set the non-relativistic quantum theory as background knowledge and give the knowledge that a qubit is prepared in the computational basis $\{|0\rangle, |1\rangle\}$, under this condition, different agents may determine the probability of the proposition "A Stern-Gerlach apparatus aligned at the same direction of computational basis and this qubit will be projected at up direction" with exactly the same result, which is 1.

Another key point is the choice of prior probabilities. In the situation that we cannot use physical laws to derive the probability of a proposition, we may need to use Bayes' theorem to update the probability from evidence, but the prior probabilities are needed to start this updating process. The prior probabilities can be determined via certain principles, depending upon the situation. Unfortunately there is no broadly accepted procedure for choosing priors. As long as we are mainly dealing with physical propositions, our least hope is to find some procedure that is physically meaningful. With a clear proposition, quantifiable external knowledge, and principle-based prior probabilities, we may expect different agents will arrive at the same result if given the same conditions, and the consequences of the derived posteriors will also be consistent with subsequent physical observations.

2.1.4 Probability in Classical Mechanics

The classical mechanics is essentially deterministic. When evolution of the parameters that characterize a system is not difficult to calculate, there is no room for probability.

When dealing with a system where the phase space dimensionality is vast, on the scale of Avogadro's number, tracking its evolution using Newtonian mechanics becomes impracticable. Instead, we have to use statistical mechanics to study the collective behavior of the system. Probability plays a fundamental role in statistical mechanics, The use of probability theory in statistical mechanics allows us to make predictions about the macroscopic behavior

of a system based on the behavior of its constituent particles. For example, the probability of a particle having a particular velocity distribution can be used to calculate the temperature of a gas, or the probability of a particular arrangement of particles can be used to calculate the entropy of a system.

One of the key concepts in statistical mechanics is the Boltzmann distribution, which gives the probability of finding a particle in a particular energy state. Using the Boltzmann distribution, we can calculate many important thermodynamic properties of a system, such as its temperature, pressure, and entropy. For example, the temperature of a gas can be related to the average kinetic energy of its particles, which can be calculated from the Boltzmann distribution. Similarly, the pressure of a gas can be related to the probability of particles colliding with the walls of a container, which can also be calculated from the distribution.

Another important application of probability in classical mechanics is the analysis of measurement uncertainty. Though ideally the value of every quantity in classical mechanics can be precisely measured, in practice there is always some uncertainty associated with the measurement. This uncertainty can arise from a variety of sources, including the limitations of the measurement instrument and environment fluctuations. Probability distributions are used to model these measurement uncertainties. In general the Gaussian distribution is widely chosen for modeling. For example, if we try to measure a certain quantity in many trials, the standard deviation of the results will be represented as the measurement uncertainty.

2.1.5 Probability in Quantum Mechanics

The Born Rule. In quantum mechanics, the state of a system is described by a wave function. The wave function itself is not directly observable, but encodes the information of the outcome probabilities of all observables. This connection is formalized by one of the fundamental postulates of quantum mechanics, the Born rule. Consider an observable $\hat{A} = \sum_i \lambda_i \hat{p}_i$ where λ_i is the eigenvalue of $\hat{A}$ and $\hat{p}_i$ is the corresponding projection operator for each λ_i. If we know the wave function $|\psi\rangle$ of a system very well, then we could calculate the outcome probability of $\hat{A}$ as:

$$\Pr\left(\text{“}\hat{A}, \lambda_i\text{”} \middle| \, |\psi\rangle, I\right) = \langle\psi| \, \hat{p}_i \, |\psi\rangle \tag{2.2}$$

The proposition "$\hat{A}, \lambda_i$" means "perform an measurement of $\hat{A}$, yielding outcome λ_i" and I denotes our background knowledge which include the postulates of quantum mechanics.

In the more general case, the system may not be in pure state. This happens when we consider a subsystem from an ensemble which contains different pure states of subsystems or the system is a subsystem of an entangled system. A density matrix $\hat{\rho}$ is then needed to describe the state of the system. In this case the Born rule becomes:

$$\Pr\left("\hat{A}, \lambda_i" | \hat{\rho}, I\right) = \operatorname{tr} \hat{\rho}\hat{p}_i \tag{2.3}$$

We note that that the Bayesian probability is still important in quantum mechanics, especially when applying Born rule to calculate outcome probabilities of observables. If different agents were given the same condition, say the wave function or density matrix of a system, they will arrive at the same results of outcome probabilities.

A key assumption that links quantum theorem to experiment is that the probability from Born rule is related to long-run frequencies, provided 1) we can prepare an ensemble of identical quantum systems; 2) each measurement is a projection and isolated from the environment; 3) the numbers of both prepared systems and performed measurements are sufficiently large. Indeed, if we really treating these conditions seriously, we might say that this assumption can be never valid. Of course we can always make reasonable approximations of these conditions, and the probabilities can then be checked via frequencies.

One common example would be single-photon double slit experiment. In this experiment, photons are sent one by one toward a barrier with two slits, and the locations where they strike a screen behind the slits are recorded. An interference pattern will be gradually formed on the screen over time due to the interference of photons, and this pattern can be predicted using the Born rule, which tells us that the probability of a photon arriving at a particular point on the screen is proportional to the square of the sum of the amplitudes of the two paths (through each of the slits). Another example will be the measurement of cross sections in particle collisions. If we collide particles many times, the relative frequency of a particular process happening should converge to the value given by the cross section.

However, preparing identical systems and isolating the effect from environment are very challenging for most quantum systems. Most of the direct verification of Born rule are

conducted over low-dimensional systems like qubits. Despite these difficulties, the Born rule has been indirectly confirmed in numerous experiments over the course of many decades. The success of quantum mechanics in predicting experimental results in a wide range of systems and scales, from microscopic particles to macroscopic superconductors, is seen as a strong confirmation of the Born rule.

Quantum Tomography. Quantum tomography, including state tomography and process tomography, are techniques for reconstructing the state or unitary process in a quantum system respectively. They can each be regarded as an inversion of the Born rule.

Quantum state tomography involves making many measurements on a quantum system that is prepared in the same state repeatedly [33, 23, 32, 58]. Each type of measurement gives some information about the state, and by making enough different types of measurements, one can reconstruct the full state. In other words, we use the Born rule in reverse: instead of using a known state to predict the probabilities of different measurement outcomes, we use known measurement outcomes to infer the state.

Similarly, quantum process tomography involves performing many different sequences of operations on a quantum system and making measurements to determine how the system evolves under these operations [44]. Again, this is effectively using the Born rule in reverse to infer the process from the measurement outcomes.

2.1.6 Relation to Kolmogorov Probability Theory

Besides Bayesian probability theory founded upon Cox's axioms, Kolmogorov probability theory is also widely used in many situations. Here we briefly summarize Kolmogorov's approach, and indicate that we do not adopt this approach to probability theory. Kolmogorov's framework is built on three axioms that define the probability space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra of events, and P is a probability measure.

One defect of Kolmogorov's probability theory is the limitation of conditional probability. The probability of event A under the condition B is defined as:

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)} \tag{2.4}$$

Noticing that both A and B are in the same sample space, which limits its applicability. In quantum mechanics, the probability of an event is always related to the context, and the context may not be encoded as an event in sample space. A typical example is that the given condition is "the state of a qubit is $|+\rangle$" and the event is "performing measurement of $\hat{\sigma}_z$ and obtaining outcome -1". This probability is calculated via Born's rule which is $|\langle 1|+\rangle|^2$ and cannot be expressed in terms of Kolmogorov's conditional probability.

Even if we assume A and B are two events of observable measurements, say $A =$"performing a measurement of $\hat{A}$ and obtaining outcome λ_A" $B =$"performing a measurement of $\hat{B}$ and obtaining outcome λ_B", we cannot use this conditional probability either, since $\hat{A}$ and $\hat{B}$ may not commute and the order of the measurements will yield different probabilities. This suggests that the sample space language may not be suitable for quantum measurements, and the flexibility of Bayesian probability is important.

2.1.7 Probability-Based Ideal Measurement

In practice there are no ideal measurements, the precision of measuring tools is bounded. Nevertheless it is still interesting to consider abstract models of such ideal devices.

Classical Viewpoint. Every measurement of a physical quantity can be abstracted into a process of comparison. A common example would be the measurement of length of a table, in which we compare the target with a standard meter ruler to obtain the length. This is one of the few examples in which we can take direct measurement of a quantity.

Most physical quantities cannot be measured directly, in which case we may use the relations between different quantities and measure the value of a quantity which can be directly measured to infer the value of the target quantity. A typical example would be the measurement of the table's mass. There are various tools to measure the table's mass, yet none of them can directly measure the mass. The gravitational mass is usually measured via the weight of the object ($F_g = mg$), and the inertial mass can be measured by angular frequency of simple harmonic motion ($\omega = \sqrt{\frac{k}{m}}$) by attaching the object on an ideal spring. In some sense, we can even convert the measurement of an object's mass into a measurement of length: the weight can be reflected by the change of an ideal spring's length, or the angular frequency of harmonic motion can be obtained via the change of length of a spring over a

fixed time period.

Noticing that the measurement of length is just a process of comparison, or even a process of counting how many integer multiple of the smallest scale of the measuring tool. This suggests that all measured values are in fact integers. More general, according to the latest definition of SI units, all units of physical quantities can be "derived" from time and some necessary constants. The unit of time is defined as an integer multiple of a constant value. This suggests that other quantities are also expressible in terms of an integer multiple of some constant value.

Yet usually we use real numbers to represent the value of a physical quantity. In an extreme case, if the value is equal to an irrational number, how could the ideal device obtain that number? One could imagine that such an ideal device could divide the smallest scale of a "length" indefinitely. However, such a number cannot be displayed since it has infinite many digits.

Quantum Viewpoint. From a quantum physical viewpoint, we could convert the measuring of length into a process of measuring probability. Consider a Stern-Gerlach measurement of a spin-$\frac{1}{2}$ particle. Assume we prepare a beam of particles in the state $|\psi\rangle = |0\rangle$ and the projection apparatus is configurated at an angle of θ relative to the z-axis of the particle. For each of the prepared particles, there is a probability $p = \cos^2 \frac{\theta}{2}$ for this particle to be deflected "up". This probability can be measured via counting and the value of this probability is related with the angle θ of the apparatus set up. This angle can be related to a length. Since every dimensional physical quantity can be indirectly measured via a measurement of some length, and this length is now corresponding to a probability.

This probability p is a real number between 0 and 1, and it is also "measurable". Now assume the ideal device could take infinite many trials of the projection, thus yielding an infinite sequence of frequencies. (This time we obtain a collection of rational numbers.) Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exists a sequence of rational numbers $\{f_n\}$ such that

$$\lim_{n\to\infty} f_n = p. \tag{2.5}$$

Each f_n will be one of the frequencies in these infinite many trials. This process is another

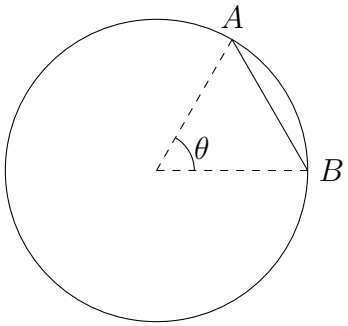

Figure 2.1: ***The Relation between an Angle and a Length*** The measurement of a length is equivalent to the measurement of an angle. In a circle, the distance between two points on the circle has a one-to-one relationship with the central angle, given that the radius is fixed. If the radius of this circle is R, the length $l_{AB} = 2R\sin\frac{\theta}{2}$.

way of measuring length. Instead of counting the multiples of the smallest scale, this method counts the frequencies of identical measurement results. This approach avoids the difficulty of infinite subdivision of a finite length.

2.2 Shannon Entropy

Shannon entropy [52] is a measure of uncertainty or information content in a probabilistic system. It has the same name with the concept in thermodynamics and statistical mechanics, and there is indeed a deep relation between the entropy in information theory and entropy in statistical thermodynamics. Here we want to focus on the information theory, which is defined on discrete random variables. Let X be a discrete random variable with possible outcomes $\{x_1, x_2, \ldots, x_n\}$ and associated probabilities $\{p_1, p_2, \ldots, p_n\}$, where $p_i = \Pr(\text{“}X = x_i\text{”}|B)$ represents the probability of the ith outcome of X under background knowledge B. The Shannon entropy of X is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i. \tag{2.6}$$

The Shannon entropy of X quantifies our degree of uncertainty about its possible outcomes. Noticing that Shannon entropy only concerns the probability distribution of a random variable, not the values themselves, it can also be defined in terms of discrete probability distribution $H(X) = H(\{p_i\})$. In classical information theory, Shannon entropy

has important applications in communication and data compression. We are more interested in the application of Shannon entropy in physics, especially for the usage as a measure of uncertainty.

When deriving the expression of the entropy function, Shannon introduced several conditions that any reasonable measure of uncertainty (information content) H should satisfy:

1. *Continuity.* H is a continuous function of the probabilities $\{p_i\}$.

2. *Monotonicity.* If all the probabilities are equal, say $p_i = \frac{1}{n}$ for some integer n, then $H(\frac{1}{n}, \cdots, \frac{1}{n})$ is a monotonic function of n.

3. *Additivity.* The entropy of a choice broken down into two successive choices should be the weighted sum of the individual entropies. Specifically, for choices with probabilities $p_1, p_2, \ldots, p_n$, if these are further broken down into sub-choices, the total entropy $H(p_1, p_2, \ldots, p_n)$ should equal the sum of the entropy of the initial choice and the weighted entropies of the subsequent choices, following the probability of each branch.

Shannon proved that, given these three conditions, the entropy of a discrete random variable X must take the form:

$$H(X) = -K \sum_{i=1}^{n} p_i \log p_i, \tag{2.7}$$

where K is a positive constant, and the logarithm can be taken in any base. In information theory, this constant K is taking to be 1 for the sake of convenience and the base of the logarithm is 2.

The significance of Shannon entropy extends beyond theoretical constructs into practical applications in coding theory, cryptography, and data compression algorithms. It serves as a fundamental limit on the best possible lossless compression of any communication, indicating the minimum number of bits required to encode a series of messages without loss of information. Furthermore, in cryptography, Shannon entropy measures the unpredictability of cryptographic keys, directly impacting their security against brute-force attacks.

2.3 Generalization of Shannon Entropy for Continuous Distributions

Shannon entropy has a wide importance. Yet it has a limitation, since Shannon entropy is defined on discrete probability distributions. It is natural to ask whether we apply such a measure of uncertainty to continuous probability distributions. In this section we want to discuss different ways of generalizing Shannon entropy to continuous probability distributions.

2.3.1 Differential Entropy

Differential entropy seems to be an intuitive approach to generalize Shannon entropy, by replacing the sum with integral:

$$H(X) = -\int_S p(x) \log p(x) dx \tag{2.8}$$

where $p(x)$ is the probability density of random variable X, and S is the support of $p(x)$.

Unlike the Shannon entropy, the differential entropy cannot be derived from axioms. In his 1948 founding paper [52], Shannon wrote it down without further derivation. There are several problems with differential entropy, it can be negative and it is not invariant under a change of variables.

Negativity. Consider the case where $p(x)$ is a uniform distribution over $[a, b]$. Then

$$H(X) = -\int_a^b p(x) \log p(x) dx = \log(b - a) \tag{2.9}$$

If $b - a < 1$, the differential entropy is negative. As long as we interpret or use entropy as a measure of uncertainty, it is not meaningful to have a negative degree of uncertainty about a variable.

Invariant under change of variables. If we change the variable x of the function $p(x)$ to some other variable, y, the differential entropy of $p(y)$ is:

$$-\int_{S_y} p(y) \log(p(y)) dy = -\int_S p(x) \log(p(x)) \frac{dx}{dy} dx \tag{2.10}$$

If $\frac{dx}{dy}$ is not equal to 1, then the differential entropy of $p(y)$ will be unequal to the differential entropy of $p(x)$. This coordinate-dependence suggests that the value of differential entropy is tied to the choice of coordinate.

2.3.2 Limiting Density of Discrete Points

Jaynes [36] has suggested an explicit procedure—the approach of limiting density of discrete points (LDDP)—to systematically generalize Shannon entropy to continuous distributions.

Assume the probability density $p(x)$ of random variable X is initially defined on a set of discrete points $x \in \{x_1, x_2, \cdots, x_n\}$. Jaynes proposes an invariant measure $m(x)$ such that when the collection of points $\{x_i\}$ becoming more and more numerous, in the limit $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n} \text{ (number of points in } a < x < b) = \int_a^b m(x)dx \tag{2.11}$$

With the help of $m(x)$, the entropy of X can then be represented as

$$H(X) = \lim_{n\to\infty} \log n - \int p(x) \log \frac{p(x)}{m(x)} dx \tag{2.12}$$

In this way, the weaknesses of differential entropy seems to be solved—LDDP is invariant under the change of variables and it is non-negative. However, we meet two new problems: (1) the entropy of continuous distributions is infinity due to the term $\log n$; (2) the measure function $m(x)$ is unknown.

For the infinity problem, there may be two solutions. (1) Assert that the entropy of continuous distribution is only meaningful when we consider the difference of two entropy; (2) Omit this infinite term.

1. **Entropy of continuous distribution as a difference**

   When variable X is updated to X' due to some actions, the change of entropy will be

equal to

$$X \to X', \quad \Delta H(X \to X') = H(X) - H(X') = \int p'(x) \log \frac{p'(x)}{m(x)} dx - \int p(x) \log \frac{p(x)}{m(x)} dx \tag{2.13}$$

where $p'(x)$ is the probability distribution of X'. The quantity ΔH quantifies the change of uncertainty about variable X in this action.

2. **Straightforward**

   Omitting the infinite term in equation (2.12), (this is Jaynes' approach), the entropy of continuous distribution is

$$H_{Jaynes}(X) = \int p(x) \log \frac{p(x)}{m(x)} dx \tag{2.14}$$

   To ensure the entropy is non-negative, the minus sign is also dropped. H_{Jaynes} is equal to the negative of the relative entropy of $p(x)$ to $m(x)$.

Both solutions are meaningful, and we currently have no preference to choose a unique generalization of Shannon entropy. There is a relation between the two approaches. In the special case that $p(x) = m(x)$,

$$\Delta H(X \to X') = H_{Jaynes}(X') \tag{2.15}$$

This leads to the second problem of limiting density of discrete points. What is the choice of measure $m(x)$? Intuitively one may think of using uniform measure function, i.e. $m(x)$ is constant function. In this way, the LDDP reduces to the differential entropy. However, we lack a criterion to ensure the form of this measure function. One way of resolving this issue is to apply Bayesian inference, interpreting $m(x)$ as the prior probability distribution of variable X and $p(x)$ as the posterior probability distribution of X. In this way we convert the problem of finding the measure function into a problem of finding a proper prior, a problem that can be investigated through fruitful information-theoretic approaches. We take up this problem below.

2.3.3 Relative Entropy

In the discussion of entropy of continuous distributions, we find one of that the solutions has the same form as the relative entropy up to sign or the Kullback-Leibler divergence, the latter of which is defined over two probability distributions p and q as:

$$D_{KL}(p(x)||q(x)) = \sum_x p(x) \log \frac{p(x)}{q(x)} \tag{2.16}$$

for discrete probability distributions, and as

$$D_{KL}(p(x)||q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx \tag{2.17}$$

for continuous probability distributions.

Noticing that in both cases of discrete and continuous distributions, the relative entropy is always non-negative, due to Jensen's inequality:

$$D_{KL}(p(x)||q(x)) \geq 0 \tag{2.18}$$

with equality if and only if $p = q$.

However, relative entropy is a function of two distributions and it may not be able to represent the degree of uncertainty for a single variable.

2.3.4 Information Gain from Measurement

In the discussion of entropy of continuous distribution above, it seems that we have to introduce an extra measure function $m(x)$ to ensure the invariance of entropy under change of variables. In Bayesian statistics, we can interpret $m(x)$ as the prior of the target variable, so that the necessity of $m(x)$ is very natural.

For a random variable X and we mainly consider one of its outcomes x, we may use $\Pr(p|D, I)$ to represent the posterior of X with outcome $X = x$ updated from the observed data D; and $\Pr(p|I)$ is the prior of $\Pr(X = x)$. The relative entropy of $\Pr(p|D, I)$ to $\Pr(p|I)$ can be used to represent the information gain from prior posterior, or the information gain of X from the data D. Both solutions of LDDP, yield the same information

gain.

From the viewpoint of the change entropy in terms of difference: initially we have no observed data about X, so the probability distribution of $\Pr(X = x)$ is just the prior $\Pr(p|I)$. Once we obtain the some data about the outcomes of X, the probability distribution of $\Pr(X = x)$ is changed to posterior $\Pr(p|D, I)$. The difference of entropy (2.13) is equal to

$$\begin{aligned} \Delta H &= \int \Pr(p|D, I) \log \frac{\Pr(p|D, I)}{\Pr(p|I)} dx - \int \Pr(p|I) \log \frac{\Pr(p|I)}{\Pr(p|I)} dp \\ &= \int \Pr(p|D, I) \log \frac{\Pr(p|D, I)}{\Pr(p|I)} dp \end{aligned} \tag{2.19}$$

This is the same with the Jaynes' entropy (2.14)

$$H_{Jaynes}(\Pr(X = x)) = \int \Pr(p|D, I) \log \frac{\Pr(p|D, I)}{\Pr(p|I)} dp \tag{2.20}$$

This suggests in Bayesian statistics, the relative entropy is the proper generalization of Shannon entropy. However, one issue remains: if we collect more data D' after the observation D, how can we represent the information gain from this extra observation D'?

Similar to the analysis of LDDP, we have two ways to quantify this information gain from data D'. The first is to take the idea of difference, that is, taking the difference of information gain from both D and D' and information gain from D; the second is more straightforward, via taking the relative entropy of posterior obtained from D, D' to the posterior obtained from D.

1. Difference of information gain

   The above discussion shows that we can just use the Jaynes' entropy to quantify the information gain from beginning to data D, D'.

   $$H_{Jaynes}(\Pr(X = x), \{D, D'\}) = \int \Pr(p|\{D, D'\}, I) \log \frac{\Pr(p|\{D, D'\}, I)}{\Pr(p|I)} dp$$

   Hence we take the difference of $H_{Jaynes}(\Pr(X = x), \{D, D'\})$ and $H_{Jaynes}(\Pr(X =$

$x), D)$

$$\begin{aligned}\Delta I(D') &= H_{Jaynes}(\Pr(X=x), \{D, D'\}) - H_{Jaynes}(\Pr(X=x), D) \\ &= \int \Pr(p|\{D, D'\}, I) \log \frac{\Pr(p|\{D, D'\}, I)}{\Pr(p|I)} dp - \int \Pr(p|D, I) \log \frac{\Pr(p|D, I)}{\Pr(p|I)} dp\end{aligned} \tag{2.21}$$

Noticing that this quantity could be negative.

2. Straightforward

The data D' is assumed to be collected after D, we can set a new beginning point to quantify the information gain from D'.

$$I(D'|D) = \int \Pr(p|\{D, D'\}, I) \log \frac{\Pr(p|\{D, D'\}, I)}{\Pr(p|D, I)} dp \tag{2.22}$$

We have two ways to represent the information gain in this extra observation. In general these two quantities are not the same: $I(D'|D)$ is always non-negative due to the property of relative entropy, but $\Delta I(D')$ could be negative. As each expression is well-motivated, it is worthwhile to systematically investigate the behavior and relations of these two different measures of information gain. We carry out this investigation in Chapter 4.

2.3.5 Rényi Entropy

Rényi entropy is a generalization of Shannon entropy, introduced by Alfréd Rényi [50] in 1961. It is a measure of the uncertainty or diversity of a probability distribution and is defined for a non-negative parameter α ($\alpha \geq 0$ and $\alpha \neq 1$) as follows:

$$H_\alpha(p_1, p_2, ..., p_n) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} p_i^\alpha \right), \tag{2.23}$$

where $\{p_1, p_2, ..., p_n\}$ is a discrete probability distribution, and the logarithm base that determines the unit of entropy (commonly 2 for bits or e for natural units). Rényi entropy represents a family of different entropies, parameterized by α. The Shannon entropy is included in this family in the limit $\alpha \to 1$:

$$\lim_{\alpha \to 1} H_\alpha(p_1, p_2, ..., p_n) = H(p_1, p_2, ..., p_n). \tag{2.24}$$

Another special case is the situation that α approaches to ∞. In the limit as $\alpha \to \infty$, H_α converges to min-entropy:

$$\lim_{\alpha \to \infty} H_\alpha(p_1, p_2, ..., p_n) = H_{min}(p_1, p_2, ..., p_n) = \log \frac{1}{p_{max}}, \tag{2.25}$$

where $p_{max} = \max_i p_i$.

In the case that $\alpha = 0$, Rényi entropy reduces to the Hartley entropy or max-entropy:

$$H_0(p_1, p_2, ..., p_n) = \log n, \tag{2.26}$$

where every p_i is non-zero.

As a generalization of Shannon entropy, Rényi entropy allows for the quantification of uncertainty in different ways dpending on values of α. Higher values of α ($\alpha > 1$) give more weight to higher values probabilities, the min-entropy being an extreme case where only the maximum value of the probabilities matters. Lower values of α tend to place equal weighting on every non-zero probability, the max-entropy being the extreme case that every non-zero probability contributes the same.

2.3.6 Logical Entropy

Ellerman's [22] conceptual foundation for logical entropy begins with the notion of a partition on a finite set. A partition divides a set into mutually exclusive subsets, and these subsets represent the classification of the set's elements based on a specific characteristic.

In this framework, the partitions represent how we distinguish elements in a set, and information is about distinctions between things. When we partition a set, we are essentially identifying and categorizing the differences among its elements.

Logical probability is the probability of making a distinction between elements of different blocks of the partition. If an element is chosen randomly from U, the logical probability $\Pr(S)$ is the probability that two randomly chosen elements belong to same block S, it is defined as a ratio of cardinalities:

$$\Pr(S) = \frac{|S|}{|U|} \tag{2.27}$$

Logical entropy is then defined in terms of these partitions. Let $(P_i)_{i\in I}$ be a partition of a finite set U. Then the logical entropy for this partition is:

$$H_L(\{P_i\}) = 1 - \sum_{i\in I}(\Pr(P_i))^2 \tag{2.28}$$

The definition and the idea behind logical entropy is very different with Shannon entropy. However, logical entropy has the same upper and lower bound as Shannon entropy, lower bound achieved by uniform distribution and the upper bound achieved by a peak distribution such as $(P_1, P_2, \cdots, P_n) = (1, 0, \cdots, 0)$. As mentioned by Brukner and Zeilinger [7], Shannon entropy may not be adequate to express the information content of a quantum system, while logical entropy seems to be a potential candidate for this measure of information especially for finite dimensional quantum systems 3.2.2.2.

CHAPTER 3

Importance of Information Theory in Physics

Before the foundation of modern information theory, the word "information" was rarely used in natural science. Nowadays the concept of information is widely used in many subjects. Yet there is a special relation between physics and information theory. This relation may date back to the origin of the foundations of both information theory and statistical mechanics, statistical entropy and Shannon entropy have very similar mathematical expressions.

In section 3.1 we discuss the special relation between the concept of entropy in information theory and physics. In section 3.2 we discuss several different information-theoretical approaches towards the foundation of quantum theory.

3.1 Classical Information and Physics

3.1.1 Information Entropy and Statistical Mechanics

The modern information theory introduces the concept of entropy from a mathematical rather than a physical perspective. That is, entropy defined not as pertaining to physical quantity, but as a general measure—a degree of uncertainty—that can be applied to any probabilistic source. Such a source may be physically instantiated. Take the example of a physical system in the canonical ensemble: we may be not sure about the energy of this system and all we know is that this system's energy is in one of the discrete energy spectrum $\{\epsilon_i\}$ and the average energy is $\bar{\epsilon}$. In the scheme of classical physics, in principle we may be able to know the exact state of this system, but in practical that is impossible. The entropy measures how much information we lack about the precise state of the system.

The pioneering work of Jaynes [34, 35] show that the Shannon entropy and Gibbs entropy are equivalent. Moreover, the problem of finding Boltzmann distribution can be regarded as an application of information theory, the principle of maximal entropy.

According to Jaynes, the probability distribution that best describe the existing knowledge is the one which maximize the Shannon entropy. This principle is particularly simple to

apply when the given conditions are expressed in terms of expected values. As an example, the Boltzmann distribution for a system in the canonical ensemble can be derived as follows. Assume we are given the conditions that the energy level of the system is discrete and the expected energy of the system is $\bar{\epsilon}$,

$$\epsilon \in \{\epsilon_i\}, \quad \sum_i p_i \epsilon_i = \bar{\epsilon} \tag{3.1}$$

The best probability distribution $\{p_i\}$ where p_i is the probability that the system in the energy level ϵ_i that is consistent with the given conditions should maximize the Gibbs-Shannon entropy:

$$S_{G-S} = -k \sum_i p_i \ln p_i \tag{3.2}$$

As we may know or anticipate, this k will be equal to the Boltzmann constant. However, at this stage, it is an undetermined quantity. To find the extreme value of S_{G-S}, we use the Lagrangian multiplier:

$$\delta[-k \sum_i p_i \ln p_i + \alpha(\sum_i p_i - 1) + \beta(\sum_i p_i \epsilon_i - \bar{\epsilon})] = 0 \tag{3.3}$$

Finally we arrive at the solution $p_i = \frac{e^{-\beta \epsilon_i}}{\sum_i e^{-\beta \epsilon_i}}$, where β is to be determined when connecting to thermodynamic equations and α is related to the normalization condition $e^{-\alpha} = \frac{1}{\sum_i e^{-\beta \epsilon_i}}$.

The great advantage of the maximal entropy principle approach is that this principle is the only assumption we need. In contrast, in traditional treatment of statistical mechanics, the derivation of Boltzmann distribution relies upon several assumptions, such as coarse graining of phase space, the ergodic hypothesis, equally likely states assumption etc [38]. The most important insgight here is that this physical problem can be regarded as a problem of pure statistical inference.

3.1.2 Classical Information and Thermodynamics

Classical information theory originated, and is heavily used in the field of communication and computing where the basic unit of information is a bit. A bit is an abstraction of a classical two-state system, and it can be physically instantiated in many different forms. However, the relation between classical information and physics is not only about the implementation—the operations on a bit directly relate to foundational problems in physics.

In 1961 [39], Landauer first proposed the principle that any logically irreversible operation is associated with a physical irreversible process. A famous example is his hypothesis that the deletion of one bit in a classical computer will lead to an unavoidable energy dissipation of the order of $k_B T \ln 2$, where T is the temperature. The Boltzmann constant is in the order of $10^{-23}\, J \cdot K$, so that at room temperature, the energy dissipation is around the order of $10^{-21}\, J$. Although this seems like a very small number, modern digital computers will run in excess of billions of bits per second, so this energy dissipation cannot be ignored. On the other hand, Landauer's principle indicates that a logically reversible operation may consume little or even no energy. Hence, this principle may on the one hand, impose a limit on current digital computers performance, but, on the other hand lead to the study of reversible computing design. In physics, this principle can be used to make sense of the Maxwell's demon paradox in statistical mechanics and thermodynamics.

Maxwell's Demon Paradox. The Maxwell's demon is an imaginary experiment that leads to a violation of the second law of thermodynamics. Assume there is an isolated container which is filled with gas molecules. There is a partition in the middle of the container such that the container is divided into two parts with same volume. On the partition there is a small door, and its opening and closing is controlled by an imaginary demon. When the door is closed the molecule on both sides cannot pass through the partition. The demon is able to measure the velocity of nearby molecules on both sides and calculate the average velocity of the whole collection of molecules. When a molecule approaches from the demon from the right with a velocity faster than the average, the demon can try to prevent this molecule from passing and keep it on the right part; while for nearby molecules have velocities slower than the average, the demon may keep them ending in the left part. Such a demon will eventually succeed in arranging the molecules such that the right side is hotter than the

left side. Hence, the entropy of the whole system has decreased with no external work input, which violates the Second Law of Thermodynamics.

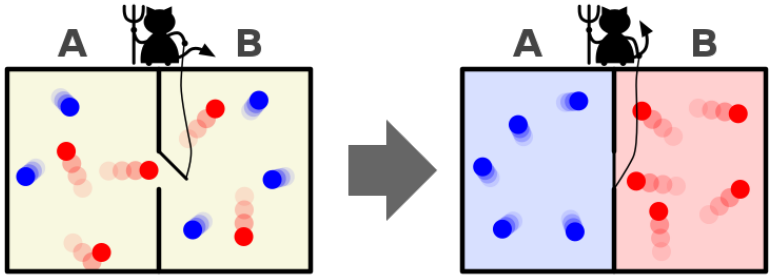

Figure 3.1: ***Maxwell's Demon*** Two chambers filled with gases are connected by a small door. An imaginary demon can measure the speed of molecules approaching the door. The demon allows only molecules with speeds exceeding a certain threshold to pass through the door into the right chamber, while slower molecules remain in the left chamber. After some time, this process results in the left chamber becoming cooler and the right chamber becoming warmer due to the difference in average molecular speed. Image source:`https://en.wikipedia.org/wiki/File:Maxwell%27s_demon.svg` Licensed under CC BY-SA 3.0.

Landauer's principle provides a solution to paradox. If we treat the gas molecules as classical particles, the demon can measure the exact state of particles as an ideal observer. Moreover, the demon may record each particle's velocity and calculate the average of them. This calculation will be a reversible operation and consumes no energy. Bennett [4] shows that the memory space that the demon uses to store the state of each molecule will be eventually run out, and the demon has to erase some recorded data. This erasure process is an irreversible operation which is associated with energy dissipation and an increase of entropy.

There are different ways defending the validity of the Second Law. Another famous argument is provided by Szilar, which we describe below.

Szilard Engine. In this scenario, the container is still divided into two parts, but the middle partition is very light and movable without any friction. Here we take the simplest example, the one-particle Szilard engine. (A) Initially, an imaginary demon can determine

whether a particle is in the left part or the right part of the container. (B) If the particle is in the left part, the demon can insert a light, frictionless partition in the middle of the container. (C) Since the partition is very light, the particle may push the partition into the right part when it collides with it. This moving partition can be used to do work, for example by connecting it to a small mass via a frictionless pulley. The particle will eventually push the partition to the rightmost part of the container. (D) The loss of energy of the particle is used to do work on the attached small mass, and the container can absorb a certain amount of energy from a heat bath to restore the particle to its initial state. In this whole process, the absorbed energy is equal to the net work, thereby violating the Second Law of Thermodynamics.

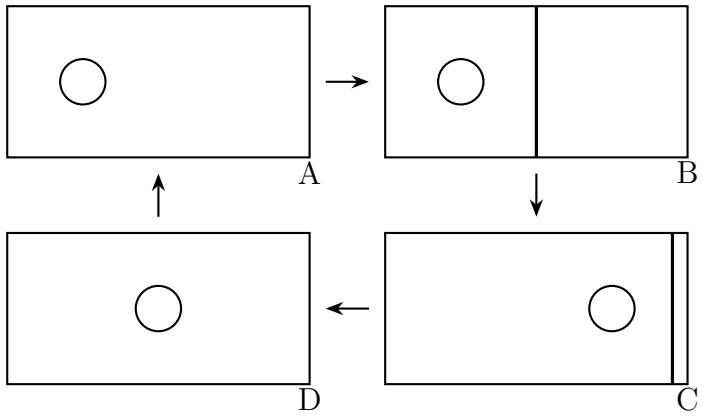

Figure 3.2: ***Four Phases of a Szilard Engine*** Assume there is only one particle in the container, and this particle can move freely inside. (A) An imaginary demon can determine the position of the particle inside the container, say, in the left part or the right part. (B) If the particle is in the left part, we can insert a light, frictionless partition in the middle of the container. (C) The particle may push the partition into the right part when it collides with it. This moving partition can be used to do work. (D) The energy lost by the particle is used to work on the moving partition, and the container can absorb a certain amount of energy from a heat bath to restore the particle to its initial state.

In the Szilard engine, the key step is the initial determination of the particle's position. The energetic and entropic cost of this position measurement cannot be ignored. Brillouin [6] shows that if using photon scattering to detect the particle, there is a minimum entropy increase due to this measurement, of the order of $k_B \ln 2$. In other words, if the demon's

behavior is restricted by quantum theory, obtaining 1 bit information will lead to an increase of entropy.

3.1.3 Information Theory and Quantum Theory

There is an intimate relation between information theory and quantum theory. First, the basis of information theory, the Shannon entropy is a function of probability. The intrinsic probabilistic nature of quantum systems suggests that the uncertainties associated with quantum systems could be described in terms of information theory. Second, another basic element of modern information theory, the bit, has a perfect analogy in quantum theory: a qubit. The uniqueness of quantum information theory is directly revealed in the difference between a qubit and a bit.

Intrinsic Probabilistic Nature. In some sense, a bit and a qubit are both two-outcome models. However, the superposition principle makes a qubit more complex than a bit. Typically we may choose two orthonormal basis $\{|0\rangle, |1\rangle\}$ as computational basis. For the sake of convenience we only consider the pure states of qubit, which can be represented via Bloch sphere,

$$|\psi\rangle = e^{i\delta} \cos\frac{\theta}{2} |0\rangle + e^{i(\phi+\delta)} \sin\frac{\theta}{2} |1\rangle \tag{3.4}$$

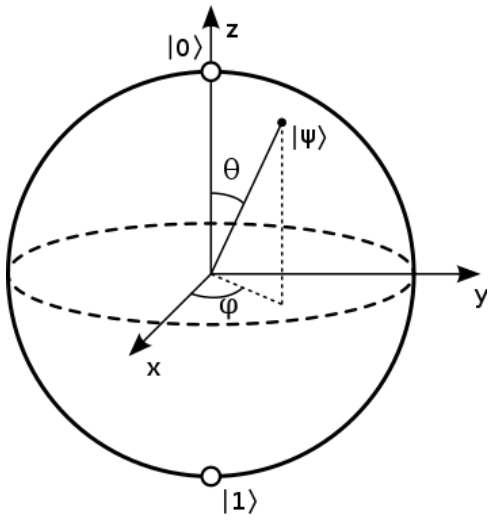

Figure 3.3: ***Bloch Sphere*** The north pole corresponds to the $|0\rangle$ state, while the south pole represents the $|1\rangle$ state. A point labeled as the state $|\psi\rangle$ is depicted, characterized by the polar angle θ and azimuthal angle ϕ, illustrating the geometric representation of a qubit's state. Image source: `https://en.wikipedia.org/wiki/File:Bloch_sphere.svg` Licensed under CC BY-SA 3.0.

Every point on the surface of the Bloch sphere represents a unique pure state, and we parameters θ, ϕ have straightforward geometrical meanings. The parameter δ denotes a global phase, which cannot be observed. In general, the state of qubit is represented via a density matrix. For a pure state qubit, the form of density matrix is simple, $\rho = |\psi\rangle \langle\psi|$. The density matrix representation is more convenient for qubits in a statistical ensemble. The distance between a non-pure state to a pure state can be measured by the von Neumann entropy:

$$S(\rho) = -Tr(\rho \ln \rho) \tag{3.5}$$

A pure state qubit has zero Von Neumann entropy and the maximal mixed state will have Von Neumann entropy $\ln 2$, for example $\rho = \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1|$ is a maximal mixed state.

As the state of a qubit is determined via three real parameters (or two parameters if we ignore the global phase), the state of a qubit cannot be represented by classical bits. The converse, however, is possible: a bit can be represented by the two poles on the sphere. In this sense we may see that a qubit can carry more information than a bit. The price is that retrieving the information from a qubit is much more difficult.

The physical meaning of the parameters is directly related to the probabilistic results of the projection measurement on the qubit. For example if we just take the projection measurement within the computational basis $\{|0\rangle, |1\rangle\}$, the associated probabilities are $\cos^2\frac{\theta}{2}, \sin^2\frac{\theta}{2}$. Moreover, after the projection, the post-measurement state of the qubit is one of the two eigenstates of the projection. Therefore, it is impossible to obtain all of these parameters in single measurement. We have to perform many projection measurement on identical copies to retrieve these parameters.

Due to the superposition principle, it is impossible to clone an unknown qubit [56, 20]. In order to retrieve information from a qubit, we need to create a collection of qubits prepared in identical states. If we find a qubit in the wild, and do not know its source, we can never determine its state.

Entanglement. Another important property of qubits that cannot be simulated by bits is entanglement. Take the example of a pair of maximal entangled qubits. The state of two qubits a, b are prepared to be maximally entangled and the state of this two-body system is:

$$|\psi\rangle_{ab} = \frac{1}{\sqrt{2}}(|0\rangle_a |0\rangle_b + |1\rangle_a |1\rangle_b) \tag{3.6}$$

This state is one of the famous Bell states. Outcomes of measurements performed on the two qubits now are strongly dependent upon each other. Assume Alice takes qubit a and Bob takes qubit b, and they carefully maintain the entanglement between the qubits. If Alice performs a projection measurement on qubit a in the computational basis, the post-measurement state of qubit a is either $|0\rangle_a$ or $|1\rangle_a$. Even though Alice did not operate on qubit b, the state of qubit b will also be affected by the projection on qubit a:

$$|\psi\rangle_{ab} \rightarrow \quad |0\rangle_a |0\rangle_b \quad or \quad |1\rangle_a |1\rangle_b \tag{3.7}$$

It seems that there is a "spooky action" on qubit b when Alice performs the projection on qubit a, and the speed of this interaction is theoretically infinite. Experiments in recent decades demonstrate that this interaction will be faster than speed of light [59]. However, this interaction cannot be used to transmit effective information, and cannot be regarded as superluminal motion. In this example, when Alice performs the projection on a, and Alice

immediately knows that the state of qubit b is changed, yet Bob does not know about this and Alice has to transfer a message to Bob about this change whose speed is constrained by the speed of light. This situation is generalized as the no-signaling theorem, which states that measurement on a subsystem of an entangled system (which may or may not be maximal entangled) cannot be used to communicate information to other observer.

The degree of entanglement of a two-body system can be quantified via the von Neumann entropy of the system:

$$S(\rho_a) = -Tr\rho_a \ln \rho_a \tag{3.8}$$

where $\rho_a = Tr_b\rho_{ab}$ is the reduced density matrix of subsystem a. When the system is maximally entangled, the von Neumann entropy will achieve its maximal value 1. The minimum value of 0 occurs when the system is non-entangled, i.e. the state can be written in terms of tensor product, $|\psi\rangle_{ab} = |\psi\rangle_a \otimes |\psi\rangle_b, \rho_{ab} = |\psi\rangle_{ab} \langle\psi|_{ab}$.

Classical information via Qubits. Consider a probabilistic source that generates messages drawn from a collection of letters $X = \{x\}$, with each letter x emitted from the source with a probability p_x. Shannon entropy $H(X)$ describes the minimum number of bits required to losslessly compress the message per letter. The quantum version of this information source is to replace each letter x with a quantum state ρ_x. For the sake of convenience, assume each state is a pure state, $\rho_x = |\psi_x\rangle \langle\psi_x|$. Alice will be using this quantum information source to generate a qubit to be sent to Bob, and Bob will represent the state of every qubit as $\rho = \sum_x p_x\rho_x$. The Von Neumann entropy $S(\rho)$ quantifies the minimum compressed information content of this source [47]. In the special case that states sent by Alice are all mutually orthogonal, $\langle\psi'_x|\psi_x\rangle = \delta_{x,x'}$, the Von Neumann entropy is equal to the Shannon entropy $H(X)$.

As classical messages are all expressed in terms of distinguishable units (alphabet letters, numbers, etc.), these units can be regarded as the mutually orthogonal quantum states. The special feature of quantum information emerges in this communication process. Assume Alice is preparing a collection of non-mutually orthogonal states $\{\rho_x\}$, each state generated with probability p_x. When Bob receives the state and tries to recover the information encoded about X, Bob can only perform measurements on the received state and obtain information about X via the measurement result Y. The mutual information $H(X : Y)$ describes the

information obtained by Bob over all possible measurement results. The maximal possible $H(X:Y)$ is the accessible information, which can also be regarded as the amount of classical information that can be obtained from quantum systems via optimal measurements. It was proved by Holevo that this accessible information has an upper bound,

$$H(X:Y) \leq S(\rho) - \sum_x p_x S(\rho_x) \tag{3.9}$$

where $\rho = \sum_x p_x \rho_x$. In the special case that the collection of quantum states ρ_x are mutually orthogonal, the right side of 3.9 reduces to $H(X)$, which is the classical bound.

3.2 Information in the Foundations of Physics

Landauer's principle treats information processing as a physical process. The special features of quantum theory show how one can extend classical information theory. In this section we wish to discuss whether it is possible to go in the reverse direction, viz. can we derive quantum theory itself from an information-theoretical viewpoint?

3.2.1 It From Bit

The success of Jaynes maximal entropy approach to recovering statistical mechanics suggests that physical theory can sometimes be derived from informational postulates. The idea of treating physics as informational is strongly inspired by John Wheeler's famous quote [55], "*It from Bit*":

> Every it—every particle, every field of force, even the spacetime continuum itself—derives its function, its meaning, its very existence entirely—even if in some contexts indirectly—from the apparatus-elicited answers to yes or no questions, binary choices, bits.

This suggests that a physical quantity is meaningful if and only we can use a device to measure it. Moreover, this device yields binary outcomes. It may not be intuitive that arbitrary physical quantities can be measured through a two-outcome device. Although there are observables in quantum mechanics that have a discrete spectrum, most physical quantities

are continuous. However, from an operational viewpoint, we may find these quantities are not as continuous as we expect.

We could consider a world without quantum theory, in which a physical quantity could be measured with arbitrarily high precision. In this scenario, a physical quantity is measured in finite digits would seem to be only a practical one. Yet if we accept the assumption that the universe is finite, and we do not have enough space to store as many digits as possible, then we have to erase some digits that we have already memorized. Yet according to Landauer's principle, erasing the digits consumes a certain amount of energy, and we do not possess an infinite source of energy. We may have to admit that there exist a lower bound to the precision, in which case we could use finite many yes-no device to measure any physical quantity.

In this sense, we could accept the postulate that all physical quantities are essentially discrete—physical quantities expressed as a sequence of binary outcomes would be one expression of "it as bit". Yet if information constitutes the foundation of physics, it should be possible to derive physical theory from an informational perspective.

Mixed Information. Clifton, Bub and Halvorson [14] proposed to reconstruct quantum theory from three informational constraints:

1. The impossibility of superluminal information transfer between two physical systems by performing measurements on one of them.

2. The impossibility of perfectly broadcasting the information contained in an unknown physical state.

3. The impossibility of unconditionally secure bit commitment.

These three no-go theorems are consequences of standard non-relativistic quantum theory, but here they are elevated to the level of foundational postulates. The concept of information differs amongst these three constraints.

The information in first constraint refers to classical information. This constraints suggests that if Alice and Bob are both performing local measurements, then Alice's measurements have no influence on the statistics of the outcomes of Bob's measurements and vice

versa. Indeed, Alice and Bob may be performing quantum measurements, yet the possible information transferred is via the statistics of measurement. In some sense this information may be just a real number which can be encoded in bits. The information in the third constraint also refers to classical information. As the name suggests, bit commitment describes a scheme of securely transmitting a classical bit that cannot be changed or viewed without permission.

The second constraint describes quantum information only. The information may refer to the parameters that uniquely characterize the state of a physical system, especially a quantum state. In some sense, we can also regard quantum information as a collection of numbers since the parameters that determine the quantum state must have a value. To differentiate it from classical information, quantum information is the collection of these parameters and also includes the postulates of quantum theory which provide the relationship between these parameters.

It from Qubit. Deutsch [18, 19] argued that classical information, often represented by bits in binary form, proves inadequate even within classical physics. This limitation arises from the continuous nature of most quantities in classical physics, allowing a physical process to be infinitely subdivided. The absence of Planck units during this era implies the potential need for an infinite amount of information in describing any physical process, posing a significant challenge.

In the realm of quantum physics, most systems can be represented using a finite number of qubits. Even some quantum field theories, such as lattice quantum field theory, are ideally conceptualized with finite degrees of freedom. Considering the qubit as the smallest unit of quantum information provides a natural starting point for information processing within this framework.

Furthermore, the inherent differences between classical and quantum physics imply that classical information can be encoded in terms of quantum information, but not the other way around. The proposition of a universal quantum gate suggests that if any physical process can be viewed as a computational process, a universal quantum computing device could theoretically represent any physical process. However, the practical realization of such a quantum computing device remains exceedingly challenging with current technology.

Quantum Probabilities as Bayesian Probabilities. In the approach of Quantum Bayesianism (QBism) [11, 24, 25], the state of a quantum system is treated as a collection of probabilities that are essentially subjective and updated via Bayesian rule. These probabilities correspond to different projective measurements and reflect an agent's degrees of belief about the possible outcomes of these projective measurements.

Information in QBism is mainly used to describe an agent's degree of belief, and it is applied in two different senses:

- *Degree of Belief about Measurement Outcomes* refers to an agent's subjective belief about the possible outcomes of a measurement. For instance, before performing a measurement, the agent holds a degree of belief about the probabilities associated with each possible outcome;
- *Degree of Belief regarding a Quantum State* signifies an agent's subjective belief regarding the essential properties of a quantum system. The quantum state is interpreted as a representation of an agent's degrees of beliefs about the system's possible states or outcomes.

However, these degrees of belief are not quantified by an information measure like Shannon entropy. The probability of a certain outcome of measurement already represents an agent's degree of belief. The quantum state itself is understood as an expression of an agent's belief about the system's state, without a formal quantification via a measure like Shannon entropy.

3.2.2 Different Informational-Theoretic Approaches towards Quantum Reconstruction

The concept of information may denote different meanings even in the same literature, depending on the context, especially on the usage of it. One motivation for introducing information is to apply informational principles to physics to interpret or derive current theory, especially quantum theory. The candidate we choose is from the variant of information in classical information theory. On one hand, those information-related quantities in classical information theory have a clear mathematical form and it is easier for us to propose quantifiable postulates on that; on the other hand, the success of Jaynes' approach show that

measures like Shannon entropy may play a big role in physics. In the following content, we want to discuss information from two different perspectives, and the context will be clearly specified. Both types can be regarded as variants of Shannon entropy, but with different interpretations and expressions.

3.2.2.1 Information associated with the Value of Physical Quantity

The first situation will focus on the value of a physical quantity, including measuring the value of a quantity (typically indirectly) from a collection of identical measurements or transmitting a real number which is encoded as a physical quantity through a quantum channel and then decoding it from measurement results on identical copies.

To make the understanding clearer, we do not use the word information alone but use suitable modifiers. In the first scenario, the measured value of a quantity may change depending on the results of measurements, especially on the number of repetitive measurements performed. We may hope that the results from more measurements lead to a more accurate value of the quantity, hence more 'information' obtained. We may describe this kind of information as information gained from data about the value of a quantity, where the data denotes the result of measurements on a collection of identical copies.

The following are two examples of applying this type of information in the foundations of quantum theory.

Information as Range of Measurement Uncertainty. We may start from an intuitive idea about physics and information from Summhanmmer[53, 54]: "more data from measurements lead to more knowledge about the system". The term "knowledge" can be defined via information theory, yet Summhammer adopts a primitive and straightforward approach. If we consider a system that has a single physical quantity, our knowledge of this quantity can be represented as an uncertainty range around the true value.

Summhammer proposes a similar scenario as Wheeler's: the physical quantity can be measured through a two-outcome device if we can associate this quantity with a probability. A partial reconstruction of quantum theory is derived from this idea.

Consider a two-outcome system which yields either event 1 or event 2. In total N

measurements, suppose event 1 occurs n_1 times, and event 2 n_2 times. Now the probability p_1 of event 1 can be estimated as:

$$p_1 = \frac{n_1}{N} \tag{3.10}$$

with the uncertainty

$$\Delta p_1 = \sqrt{\frac{p_1(1-p_1)}{N}} \tag{3.11}$$

Let event 1 correspond to a physical quantity χ_1 which is determined by the probability p_1. The uncertainty of χ_1 is:

$$\Delta\chi_1 = \left|\frac{\partial\chi_1}{\partial p_1}\right| \Delta p_1 = \left|\frac{\partial\chi_1}{\partial p_1}\right| \sqrt{\frac{p_1(1-p_1)}{N}} \tag{3.12}$$

According to the Summhammer's assumption, more data will lead to more knowledge about χ_1. Since the accuracy is inversely proportional to $\Delta\chi_1$, we expect the uncertainty interval will be a decreasing function depending on the total number of measurements N, that is:

$$\Delta\chi_1(N+1) < \Delta\chi_1(N) \tag{3.13}$$

But χ_1 is determined by p_1, and not all functions $\chi_1(p_1)$ satisfy this inequality. Under the above assumption, a physical quantity must satisfy this inequality. So it is natural to ask what kind of function is allowed. Summhammer now introduces the concept of "maximum predictive power". Although χ_1 is determined by p_1, $\Delta\chi_1$ depends on both N and p_1 (the latter is estimated by n_1). N is determined by the experimenter while p_1 is determined by nature. Summhammer asscerts that a quantity has the "maximum predictive power" if the uncertainty is maximally dominated by nature. According to (3.12),

$$\sqrt{N}\Delta\chi_1 = \left|\frac{\partial\chi_1}{\partial p_1}\right| \sqrt{p_1(1-p_1)} = constant \tag{3.14}$$

which yields Malus' law $p_1 = \cos^2(m\chi_1/2)$.

If the prior distribution of the quantity χ_1 is uniform, (3.14) implies that the prior distribution of p_1 is the Jeffreys prior.

Information as Classical Information in Communication. Wootters [57] investigated the possibility of a real-vector-space variant of quantum theory from a communication problem in which information is only communicated via probabilities and the receiver is only allowed to conduct a finite number of probabilistic measurements. This work does not deal with the informational origin of a physical quantity, but reveals the essential relation between quantum theory and information theory from a practical communication standpoint. The information to be communicated is just the classical information.

Suppose Alice wants to send a number to Bob indirectly, say θ, where $0 \leq \theta \leq \pi/2$. Alice encodes this number into a probability $p(\theta)$ and constructs a special coin with the probability of heads being $p(\theta)$. Now Alice sends the coin to Bob directly and Bob gains information about the coin by performing measurements on the coin. Assume Bob is only allowed to perform N tosses and gets n heads. We want to know what kind of function $p(\theta)$ would maximize the mutual information, $I(\theta : n)$, which is the average information Bob gains in many trials. Due to the symmetric property of mutual information, we can calculate the mutual information from another direction:

$$I(\theta : n) = I(n : \theta) = H(n) - H(n|\theta) \tag{3.15}$$

where $H(n)$ is the entropy of the number of heads and $H(n|\theta)$ is the entropy of the number of heads conditioned on θ. The full expression is given by:

$$I(\theta : n) = -\sum_{n=0}^{N} p(n) \ln P(n) - \left\langle -\sum_{n=0}^{N} p(n|p(\theta)) \ln p(n|p(\theta)) \right\rangle \tag{3.16}$$

To perform more detailed calculation, we may need to ensure the form of $p(\theta)$. For finite N, Wootters approximates $p(\theta)$ as a decreasing step function with $N + 1$ intervals. Since there would be $N + 1$ probabilities of the value of n, each interval corresponds to a different n. The length of each interval may vary, and a weighted function w is introduced to denote the length of each interval. In this way the mutual information becomes:

$$I(\theta : n) = -\sum_{n=0}^{N} p(n) \ln P(n) + \sum_{k=1}^{L} w_k \sum_{n=0}^{N} p(n|p_k) \ln p(n|p_k) \tag{3.17}$$

In finite cases, $p(\theta)$ and the weighted function change with different n. In the limit as N goes infinity, the weighted function has the form:

$$w(p) = \frac{1}{\pi\sqrt{p(1-p)}} \tag{3.18}$$

Here we find that the weighted function actually plays the role of the "prior of p". This result is then generalized into the N-dimensional case, and one arrives at a similar result to Summhammer's approach—the weighted function has the form of Jeffreys prior. This prior also acts like the distribution of pure states in a real-vector-space quantum theory.

We note that Summhammer and Wootters both arrive at Jeffreys prior from different motivations and different notions of information. In Chapter 4, we discuss how Jeffreys prior arises from a different intuitive idea.

3.2.2.2 Information of Observables and Systems

The second situation is that we want to focus on the relations between different quantities in the same system, especially the observables in a quantum system. The relation between observables will be reflected by the information of observables. The information of a system may be determined by the information of observables. We may put informational intuitive restrictions on the information of quantities and the information of the system to recover the relations between quantities and system. In this way, one hopes to reconstruct quantum theory from informational postulates.

Rovelli. Rovelli [51] suggested that there is no absolute independent observer so that all quantities observed in a quantum system are related to a specific observer. The connection between different observers is information.

The basic unit of information is the outcome of a "yes/no question". In a finite dimension quantum system, those "yes/no questions" can be regarded as projective measurements which have only two eigenvalues. If we ask a system some binary outcome question Q and obtain an outcome "yes", then we obtain one unit of information about this question Q, otherwise zero information is obtained. A system could be decomposed into a set of those binary outcomes questions, say system S can be described as a set of ques-

tions $(Q_1, Q_2, Q_3, \cdots)$ and a full description of S will be the collection of outcomes of those questions, $(e_1, e_2, e_3, \cdots), e_i = 0, 1$.

Rovelli made two postulates about this information:

1. There is a maximum amount of relevant information that can be extracted from a system.

2. It is always possible to acquire new information about a system.

The first postulate restricts the degree of freedom of a system to be finite. Though a full description of system S may contain infinitely many terms, finite outcomes of those questions could determine all the others. For example, the first n terms, $(e_1, e_2, \cdots, e_n)$ may determine the full infinitely long sequence $(e_1, e_2, e_3, \cdots)$.

It seems that the second postulate may violate the first, yet it does not since newly acquired information is not accumulated, some existing information can be erased. Consider a spin-$\frac{1}{2}$ system: if we first perform a Stern-Gerlach projection in the x-direction and then perform another projection in the y-direction, we will obtain new information according to the second projection and the information of projection in the x-direction will be lost.

Together with the relational hypothesis and other postulates, Rovelli tried to derive quantum mechanics in terms of information.

Brukner and Zeilinger. When characterizing information of a quantum system, Brukner and Zeilinger [8] imposed an assumption of finiteness of information:

The information content of a quantum system is finite.

The carrier of the basic unit of information can be defined as an elementary system, where each elementary system is simply a "yes or no" question. If we obtain an outcome "yes" then we obtain 1 bit information about this elementary system otherwise 0 bits information.

In real quantum systems, the outcome of such an elementary system is probabilistic, which means we may obtain information between 0 bits and 1 bit. For binary outcome measurement, if p_1, p_2 are the probabilities of the two possible outcomes, the information of

this measurement is:

$$I(p_1, p_2) = (p_1 - p_2)^2 \tag{3.19}$$

A real quantum system may also contain infinitely many such elementary systems. For a spin-$\frac{1}{2}$ system, every Stern-Gerlach projection is an elementary system. This means the information of a system may not be simply taking the sum of all information of its elementary systems. The mutually complementary measurements are proposed to quantify the information of the whole system. For a qubit, at most three projections of them could be mutually complementary, e.g. $\{\hat{S}_x, \hat{S}_y, \hat{S}_z\}$. The information of a single qubit is then represented as the sum of the information of these three projections:

$$I_{total} = I_x + I_y + I_z \tag{3.20}$$

Information of a composite system, especially an entangled system, may not be the same form as the information of a single system. The correlation between the individual systems may also contribute to the information of the whole system. In the case of two spin-$\frac{1}{2}$ systems, one of the correlation terms could be represented as a joint of two elementary systems, and the joint term is still an elementary system. For example, the question "Will the spin of particle 1 along x and the spin of particle 2 along y the same?" has two outcomes, and we can apply the above information measure of binary outcome measurement to quantify the information of this correlation term, labeled as I_{xy}. If one restricts the observation to the $x-y$ plane on both individual systems, the information contained in correlations is defined as:

$$I_{corr} = I_{xx} + I_{xy} + I_{yx} + I_{yy} \tag{3.21}$$

Consider a maximally entangled Bell state,

$$\left|\psi^-\right\rangle = \frac{1}{\sqrt{2}}(\left|+x\right\rangle_1 \left|-x\right\rangle_2 - \left|-x\right\rangle_1 \left|+x\right\rangle_2) \tag{3.22}$$

$$= \frac{1}{\sqrt{2}}(\left|+y\right\rangle_1 \left|-y\right\rangle_2 - \left|-y\right\rangle_1 \left|+y\right\rangle_2) \tag{3.23}$$

According to the state $|\psi^-\rangle$, the information of each correlation term are $I_{xx} = I_{yy} = 1, I_{xy} = I_{yx} = 0$. This means the information contained in correlations is $I_{corr} = 2$. For

such a maximally entangled state, projections on each individual system alone would not contribute any information to the whole system, hence the information of this two-body system is just equal to I_{corr} which is 2 bits. This result is anticipated and it does not violate the finiteness condition.

PART II

Information from Measurement Results

CHAPTER 4

Operational Perspective on Quantum Information Gain

4.1 Introduction

A measurement performed on a quantum system is an act of acquiring information about its state. This informational perspective on quantum measurement is widely embraced in practical applications such as quantum tomography [45, 41, 48, 29], Bayesian experimental design [43], and informational analysis of experimental data [46, 40]. It is also embraced in foundational research.

In particular, information assumes a central role in the quantum reconstruction program [28], which seeks to elucidate the fundamental physical origins of quantum theory by deriving its formalism from information-inspired postulates [9, 27, 10, 42, 17, 30, 1, 16, 13]. Nonetheless, in the foundational exploration of quantum theory, the concept of information is articulated and formalized in many different ways, which raises the question of whether there exists a more systematic basis for choosing how to formalize the concept of information within this domain.

In this chapter, we scrutinize the notion of information from an operational standpoint and propose a physically intuitive postulate to determine the appropriate information gained from measurements.

In both tomographic applications and reconstruction of quantum theory, the focus often lies on probability distributions of physical parameters or quantities, which are updated based on the measurement results. In these contexts, the outcomes of a measurement performed on a quantum system are modelled as the interrogation of an n-outcome probabilistic source characterised by a set of parameters. For example, a given measurement on a given system can be described by a probability distribution $\Pr(x|D)$ of a quantity x, which is updated from a prior probability distribution given the results D obtained from a series of measurements performed on identical copies of a system. It is natural to consider using Shannon entropy to quantify the information gained from this updated distribution. However, Shannon entropy is

limited to discrete distributions, whereas physical quantities and their associated probability distributions can be continuous.

The question thus arises: What is a suitable measure for quantifying the information obtained from real data, especially for quantities associated with continuous probability distributions?

One potential solution is to employ Kullback–Leibler (KL) divergence, also known as the relative entropy, $H(x|D) = \int \Pr(x|D) \ln \frac{\Pr(x|D)}{\Pr(x|I)} dx$, where $\Pr(x|I)$ represents the prior distribution of x, and $\Pr(x|D)$ represents the posterior distribution of x updated with the data D. This quantity is commonly referred to as the information gain from the prior distribution to the posterior distribution, and is widely used.

Since the KL divergence is non-negative and invariant under changes of coordinates, it appears to be a reasonable generalization of the Shannon entropy for continuous probability distributions. However, there are situations where *information gain* defined in terms of the KL divergence does not have a unique representation. Consider a scenario where one has acquired a series of data D, and one proceeds to take *additional* measurements, obtaining additional data D'. What is the additional information gain pertaining to D'? Using the KL divergence, there are two distinct ways to express the information related to this additional data. The first, to which we refer henceforth as the *differential information gain*, is simply the difference between the information gain from the combined dataset $\{D, D'\}$ and the information gain from D alone (see Figure 4.1). The second, which we refer to as the *relative information gain*, is given by the KL divergence of the posterior distribution after obtaining the complete dataset $\{D, D'\}$ compared to the posterior distribution after receiving data D alone (see Figure 4.2). These two measures of information gain exhibit notably different characteristics. For instance, whether the differential information gain increases or decreases when data D' is acquired depends on the choice of the prior distribution over the parameter, while the relative information gain consistently increases regardless of the choice of prior.

As we shall discuss in Section 4.2, both of these measures can be viewed as arising as a consequence of seeking to generalize the Shannon entropy to continuous probability distributions. In order to determine which of these options is most appropriate for our purposes, we seek a physically intuitive informational postulate to guide our selection. The first criterion comes from the intuitive notion proposed by Summhammer [53, 54] that *more*

data from measurements leads to more knowledge about the system. This idea has its origin in the observation that, as we conduct more measurements to determine the value of a physical quantity, the measurement uncertainty tends to decrease. In the following, we employ information theory to formalize and explore the plausibility of this idea. We find that relative information gain is consistently non-negative, whereas the positivity of differential information gain hinges on the choice of the prior distribution.

Contrary to Summhammer's criterion, we argue that under certain circumstances, negative information gain due to acquisition of additional data D' is also meaningful. Take, for instance, the occurrence of a *black swan event*: an event so rare and unexpected that it significantly increases one's uncertainty about the colour of swans. If the gain of information is considered to result from a reduction in the degree of uncertainty, the information gain associated with the observation of a black swan should indeed be negative. By combining this observation with Summhammer's criterion, we are led to the *Principle of Information Increase:* the information gain from additional data should be positive asymptotically and negative in extreme cases. On the basis of the Principle of Information Increase, in the case of a two-outcome probabilistic source, we show that differential information gain is the more appropriate measure.

In addition, we formulate a new criterion, *the robustness of information gain*, for selecting priors to use with the differential information gain. The essential idea behind this criterion is as follows. If the result of the additional data D' is fixed, then the information gain due to D' will vary for different D. Robustness quantifies this difference in information gain across all possible data D. We show that for a two-outcome probabilistic source amongst the symmetric beta distributions, the Jeffreys binomial prior exhibits the highest level of robustness.

The quantification of knowledge gained from additional data is a topic that has received limited attention in the literature. In the realm of foundational research on quantum theory, this issue has been acknowledged but not extensively explored. Summhammer initially proposed the notion that "more data from measurements lead to more knowledge about the system" but did not employ information theory to address this problem, instead using changes in measurement uncertainty to quantify knowledge obtained in the asymptotic limit. This approach limits the applicability of the idea, as it excludes considerations pertaining to

prior probability distributions and does not readily apply to finite data.

Wootters demonstrated the significance of the Jeffreys prior in the context of quantum systems from a different information-theoretical perspective [57]. In the domain of communication through quantum systems, the Jeffreys prior can maximize the information gained from measurements. Wootters approaches the issue from a more systematic perspective, utilizing mutual information to measure the information obtained from measurements. However, mutual information quantifies the *average information gain over all possible data sequences,* which is not suitable for addressing the specific scenario we discussed earlier, for which the focus is on the information gain from a fixed data sequence.

More broadly, the question of how much information is gained with the acquisition of additional data has been a relatively under-explored topic in both practical applications and foundational research on quantum theory. Commonly, mutual information is employed as a utility function. However, as noted above, mutual information essentially represents the expected information gain averaged over all possible data sequences. Consequently, it does not address the specific question of how much information is gained when a particular additional data point is obtained. From our perspective, this averaging process obscures essential edge effects, including black swan events, which, as we will discuss, serve as valuable guides for selecting appropriate information measures.

While our investigation primarily focuses on information gain in quantum systems, we conjecture that the principles and conclusions we draw can be extended to general probabilistic systems. Based on our analysis, we recommend quantification using differential information gain and the utilization of the Jeffreys multinomial prior. If one seeks to calculate the *expected* information gain in the next step, both the expected differential information gain and the expected relative information gain can be employed since, as we demonstrate for the two-outcome probabilistic case, they yield the same result.

The chapter is organized as follows. In Section 4.2, we detail the two information gain measures, both of which have their origins in the generalization of Shannon entropy to continuous probability distributions. Sections 4.3 and 4.4 focus on the numerical and asymptotic analysis of differential information gain and relative information gain for two-outcome probabilistic sources. Our primary emphasis is on how these measures behave under different prior distributions. We will explore black swan events, where the additional

data D' are highly improbable given D. In this unique context, we will assess the physical meaningfulness of the two information gain measures. In Section 4.5, we will discuss expected information gain under the assumption that data D' from additional measurements have not yet been received. Despite the general differences between the two measures, it is intriguing to note that the two expected information gain measures are equal. Section 4.6 presents a comparison of the two information gain measures and the expected information gain. It is within this section that we propose the *Principle of Information Increase*, which crystallises the results of our analysis of the two measures of information gain. Finally, Section 4.7 explores the relationships between our work and other research in the field.

4.2 Continuous Entropy and Bayesian Information Gain

In a coin-tossing model, let p denote the probability of getting a head in a single toss, and let N be the total number of tosses. After N tosses, the outcomes of these N tosses can be represented by an N-tuple, denoted as $T_N = (t_1, t_2, \cdots, t_N)$, where each t_i represents the result of the ith toss, with t_i taking values in the set $\{\text{Head}, \text{Tail}\}$. Applying the Bayes rule, the posterior probability for the probability of getting a head is given by:

$$\Pr(p|N, T_N, I) = \frac{\Pr(T_N|N, p, I)\ \Pr(p|I)}{\int \Pr(T_N|N, p, I)\ \Pr(p|I)dp} \tag{4.1}$$

where $\Pr(p|I)$ represents the prior. The information gain after N tosses would be the KL divergence from the prior distribution to the posterior distribution:

$$I(N) = D_{\text{KL}}(\Pr(p|N, T_N, I)||\ \Pr(p|I)) = \int_0^1 \Pr(p|N, T_N, I) \ln \frac{\Pr(p|N, T_N, I)}{\Pr(p|I)} dp \tag{4.2}$$

Based on the earlier discussion on continuous entropy, this quantity can be interpreted in two ways, either as the difference between the information gain after N tosses and the information gain without any tosses or as the KL divergence from the posterior distribution to the prior distribution.

When considering the information gain of additional tosses based on the results of the previous N tosses, we may observe two different approaches to represent this quantity.

Let t_{N+1} represent the outcome of the $(N+1)$th toss, and let $T_{N+1} = (t_1, t_2, \ldots, t_N, t_{N+1})$

denote the combined outcomes of the first N tosses and the $(N+1)$th toss. The posterior distribution after these $N+1$ tosses is given by:

$$\Pr\left(p|N+1, T_{N+1}, I\right) = \frac{\Pr\left(T_{N+1}|N+1, p, I\right) \Pr\left(p|I\right)}{\int \Pr\left(T_{N+1}|N+1, p, I\right) \Pr\left(p|I\right) dp} \tag{4.3}$$

When considering information gain as a difference between two quantities, the first form of information gain for this single toss t_{N+1} can be expressed as:

$$I_{\text{diff}} = D_{\text{KL}}(\Pr\left(p|N+1, T_{N+1}, I\right)||\Pr\left(p|I\right)) - D_{\text{KL}}(\Pr\left(p|N, T_N, I\right)||\Pr\left(p|I\right)) \tag{4.4}$$

In this expression, the first term $H(\Pr\left(p|N+1, t_{N+1}, I\right)||\Pr\left(p|I\right))$ represents the information gain from 0 tosses to $N+1$ tosses, while the second term $H(\Pr\left(p|N, T_N, I\right)||\Pr\left(p|I\right))$ represents the information gain from 0 tosses to N tosses. The difference between these terms quantifies the information gain in the single $(N+1)$th toss (see Figure 4.1). In this context, we can refer to I_{diff} as the *differential information gain in a single toss.*

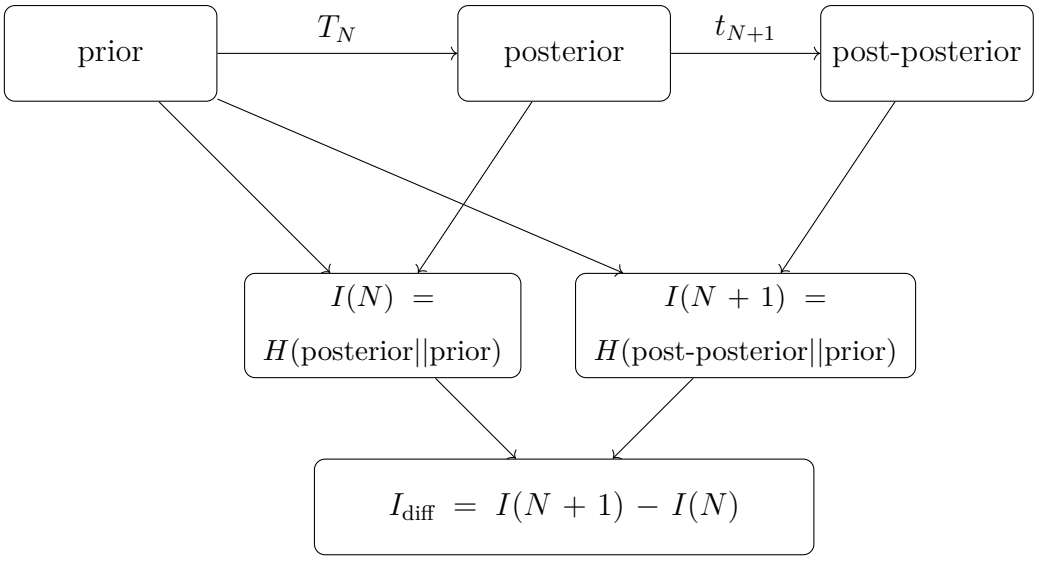

Figure 4.1: ***Differential Information Gain in a Single Toss*** Assuming we have data from the first N tosses, denoted as T_N. Using a specific prior distribution, we can calculate the information gain for these first N tosses, denoted as $I(N)$. If we now consider the $(N+1)$th toss and obtain the result t_{N+1}, we can repeat the same procedure to calculate the information gain for a total of $N+1$ tosses, denoted as $I(N+1)$. The information gain specific to the $(N+1)$th toss can be obtained as the difference between $I(N+1)$ and $I(N)$.

Alternatively, we directly calculate the information gain from the Nth toss to the $(N+1)$th toss. Hence, the second form of information gain is defined as follows:

$$I_{\text{rel}} = D_{\text{KL}}(\Pr(p|N+1, T_{N+1}, I)||\Pr(p|N, T_N, I)), \tag{4.5}$$

which is simply the KL divergence from the posterior distribution after N tosses to the posterior distribution after $N+1$ tosses (see Figure 4.2). We refer to I_{rel} as the *relative information gain in a single toss.*

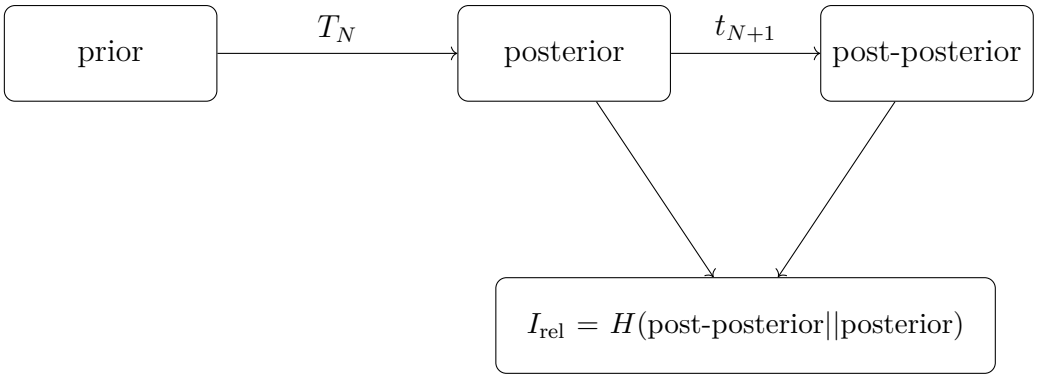

Figure 4.2: ***Relative Information Gain in a Single Toss*** The posterior distribution calculated from the results of the first N tosses serves as the prior for the $(N+1)$th toss. The KL divergence between this posterior and the subsequent posterior represents the information gain in the $(N+1)$th toss.

In general, these two quantities, I_{diff} and I_{rel}, are not the same unless $N = 0$, which implies that no measurements have been performed. I_{diff} could take on negative values, while I_{rel} is always non-negative due to the properties of the KL divergence. (This non-negativity is a consequence of Jensen's inequality applied to the convex logarithmic function, ensuring that the expected logarithmic difference between two probability distributions, which constitutes the KL divergence, cannot be negative.) Although KL divergence is not a proper distance metric between probability distributions (as it does not satisfy the triangle inequality), it is a valuable tool for illustrating the analogy of displacement and distance in a random walk model. (In a random walk, the change in total distance after $N+1$ steps compared to after N steps could be either positive or negative, analogous to how I_{diff} can have positive or negative values. On the other hand, the net displacement between the positions at step N and step $N+1$ represents the absolute change in position, which is analogous to I_{rel} always having a non-negative value.) This analogy helps elucidate the subtle difference between the two types of information gain.

Our goal is to determine which information gain measure is a more suitable choice. To do so, we use Summhammer's aforementioned postulate—"more measurements lead to more knowledge about the physical system" [53, 54]—as our point of departure. If we quantify "knowledge" in terms of information gain from data, this notion suggests that the information gain from additional data should be positive if it indeed contributes to our

understanding. This consideration makes relative information gain seem an appealing choice, as it is always non-negative. However, the derivation of differential information gain also carries significance. This leads to the question of whether Summhammer's intuitive idea is sufficient, and if not, what can replace it. In the following sections, we first will investigate differential information gain in both the finite N and asymptotic cases. We will explore the implications of negative values of differential information gain, particularly in extreme situations. We will then conduct numerical and asymptotic analyses of relative information gain. After analysing both measures of information gain, we will be better equipped to compare and establish connections between them and to assess the physical meaningfulness of Summhammer's proposal.

4.3 Differential Information Gain

4.3.1 Finite Number of Tosses

For the prior distribution, we employ the symmetric beta distribution, which serves as the conjugate prior for the binomial distribution:

$$\Pr(p|I) = \frac{p^\alpha (1-p)^\alpha}{B(\alpha+1, \alpha+1)} \tag{4.6}$$

where $\alpha > -1$, and $B(\cdot, \cdot)$ is the beta function.

In general, the beta distribution is characterized by two parameters. However, as the prior over p is invariably taken to be symmetric about $p = 1/2$ (which follows from the desideratum that the prior be invariant under outcome relabelling), we use a symmetric, single-parameter beta distribution. This distribution encompasses a wide spectrum of priors, including the uniform distribution (when $\alpha = 0$) and the Jeffreys binomial prior (when $\alpha = -0.5$).

The differential information gain of the $(N+1)$th toss is (see Appendix A.1)

$$\begin{aligned} I_{\text{diff}} = {} & \psi(h_N + \alpha + 2) - \psi(N + 2\alpha + 3) \\ & + \frac{h_N}{h_N + \alpha + 1} - \frac{N}{N + 2\alpha + 2} + \ln \frac{N + 2\alpha + 2}{h_N + \alpha + 1} \end{aligned} \tag{4.7}$$

where ψ is the digamma function (the digamma function can be defined in terms of the

gamma function: $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$), and h_N is the number of heads in the first N tosses.

In this context, we assume that t_{N+1} = 'Head'. There is also a corresponding $I_{\text{diff}}(t_{N+1} =$ 'Tail'), but there is no loss of generality since we consider all possible values of T_N and since the expressions for both cases (Head and Tail) are symmetric.

I_{diff} is a function of h_N and α, and h_N ranges from 0 to N. In the following, we select a specific value for α and calculate all the $N + 1$ values of I_{diff} for each N (see Figure 4.3).

4.3.1.1 Positivity of I_{diff}

Returning to our initial question—"Will more data lead to more knowledge?"—if we use the term "knowledge" to represent the differential information gain and use I_{diff} to quantify the information gained in each measurement, the question becomes rather straightforward: "Is I_{diff} always positive?"

In Figure 4.3, we present the results of numerical calculations for various values of N. Upon close examination of the graph, it becomes evident that I_{diff} is not always positive, except under specific conditions. In the following sections, we will investigate the conditions that lead to exceptions.

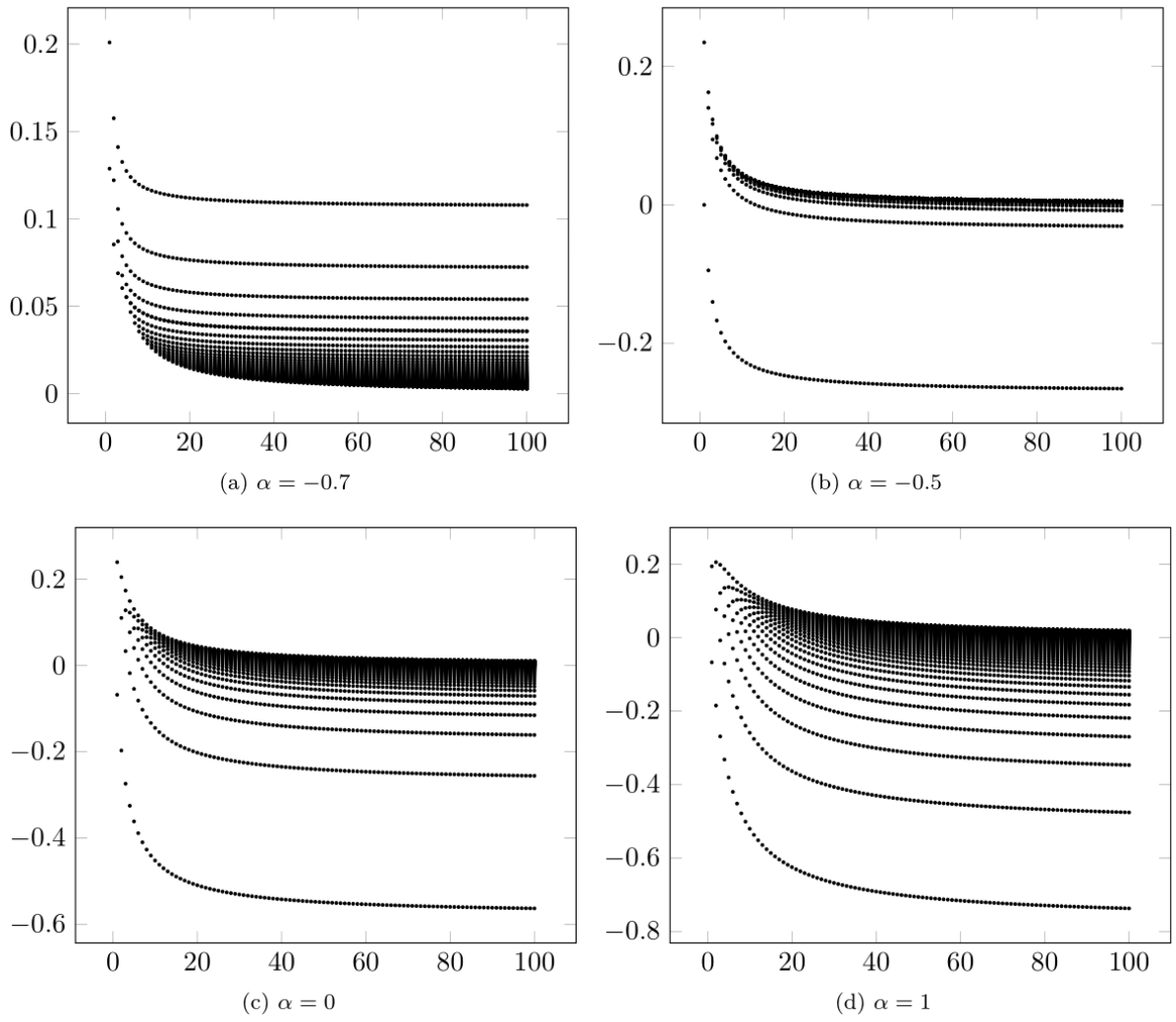

Figure 4.3: ***Differential Information Gain (I_{diff}) vs. N for Different Priors*** Here, the y-axis represents the value of I_{diff}, and the x-axis corresponds to the value of N. In each graph, we fix the value of α to allow for a comparison of the behaviour of I_{diff} under different priors. Given N, there are $N+1$ points in the vertical direction as h_N ranges from 0 to N. Notably, for $\alpha = -0.7$, all points lie above the x-axis, while for other priors, negative points are present, and the fraction of negative points becomes constant as N increases. The asymptotic behaviour of this fraction is shown in Figure 4.4. Moreover, it appears that the graph is most concentrated when $\alpha = -0.5$, whereas for $\alpha < -0.5$ and $\alpha > -0.5$, the graph becomes more dispersed.

For certain priors, the differential information gain is consistently positive (Figure 4.3a),

while for other priors, both positive and negative regions exist (Figure 4.3b–d). We note that for priors leading to negative regions, the lowest line exhibits greater dispersion compared to the other data lines. This lower line represents the scenario where the first N tosses all result in tails, but the $(N+1)$th toss yields a head. This situation is akin to a black swan event, and negative information gain in this extreme case holds significant meaning—if we have tossed a coin N times and obtaining all tails, we anticipate another tail in the next toss; hence, receipt of heads on the next toss raises the degree of uncertainty about the outcome of the next toss, leading to a reduction in information about the coin's bias.

4.3.1.2 Fraction of Negatives

In order to illustrate the variations in the positivity of information gain under different priors, we introduce a new quantity that we refer as to as the *Fraction of Negatives* (FoN), which represents the ratio of the number of h_N values that lead to negative I_{diff} and $N+1$. For instance, if, for a given α, $N = 10$ and $I_{\text{diff}} < 0$ when $h_N = 0, 1, 2, 3$, the FoN under this α and N is $\frac{4}{11}$.

From Figure 4.4, we identify a critical point, denoted as α_p, which is approximately -0.7. For any $\alpha \leq \alpha_p$, I_{diff} is guaranteed to be positive for all N and h_N values.

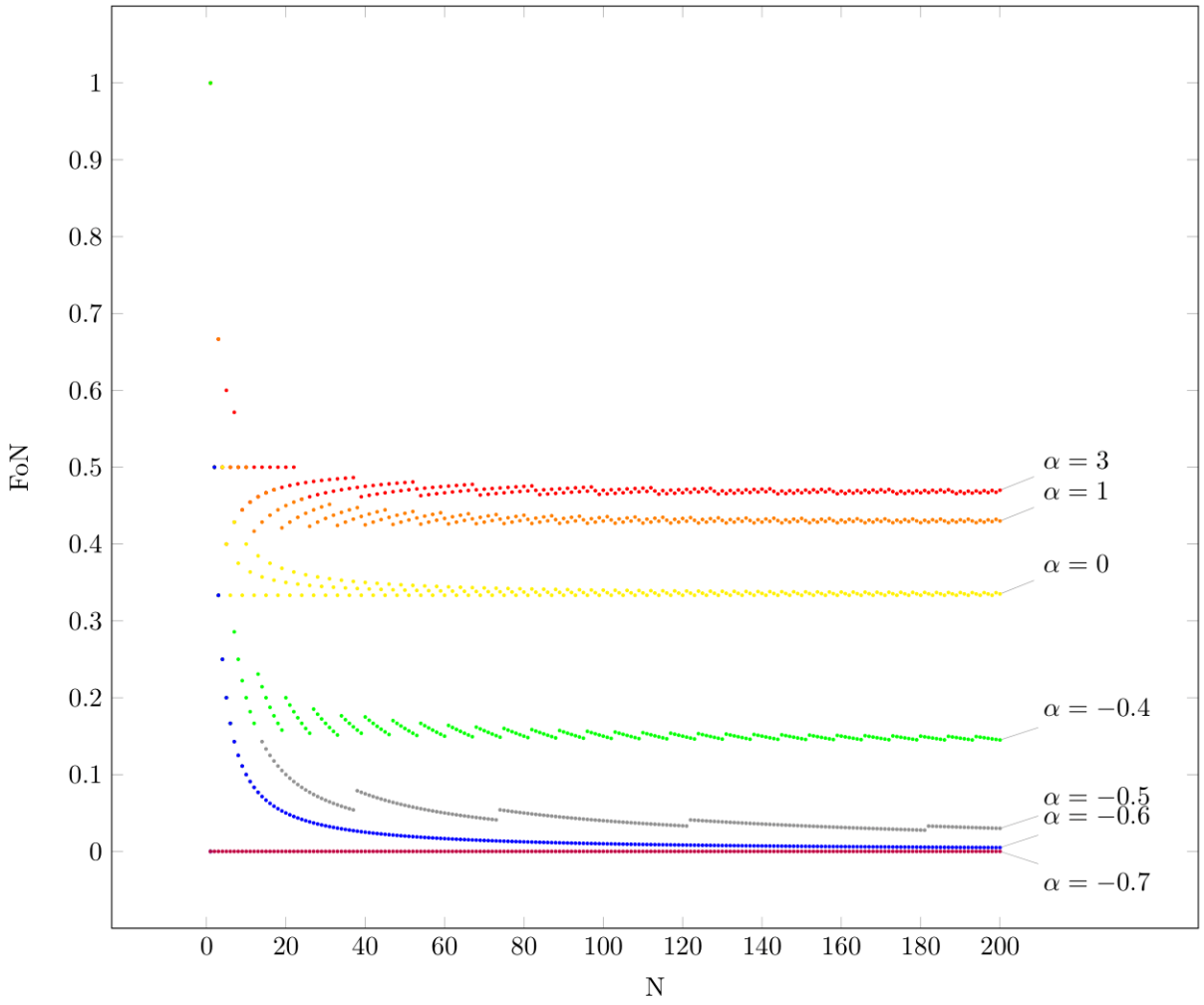

Figure 4.4: ***Fraction of Negatives (FoN) vs.*** N ***under Different Values of*** α In Figure 4.3, we can observe that larger α values lead to more dispersed lines and an increased number of negative values for each N. We use FoN to quantify this fraction of negative points. It appears that for $\alpha \leq -0.7$, FoN is consistently zero, indicating that I_{diff} is always positive. For $\alpha \leq -0.5$ FoN decreases and tends to be zero as N becomes large, while for $\alpha > -0.5$, FoN tends to a constant as N increases, and this constant grows with increasing values of α.

If $\alpha > \alpha_p$, negative terms exist for some h_N; however, the patterns of these negative terms differ across various α values.

Additionally, we notice the presence of a turning point, $\alpha_0 = -0.5$. For $\alpha \leq \alpha_0$, FoN tends to zero as N increases, whereas for $\alpha > \alpha_0$, FoN approaches a constant as N grows.

A clearer representation of the critical point α_p and the turning point α_0 can be found in Figure 4.5, where the critical point α_p is approximately -0.68.

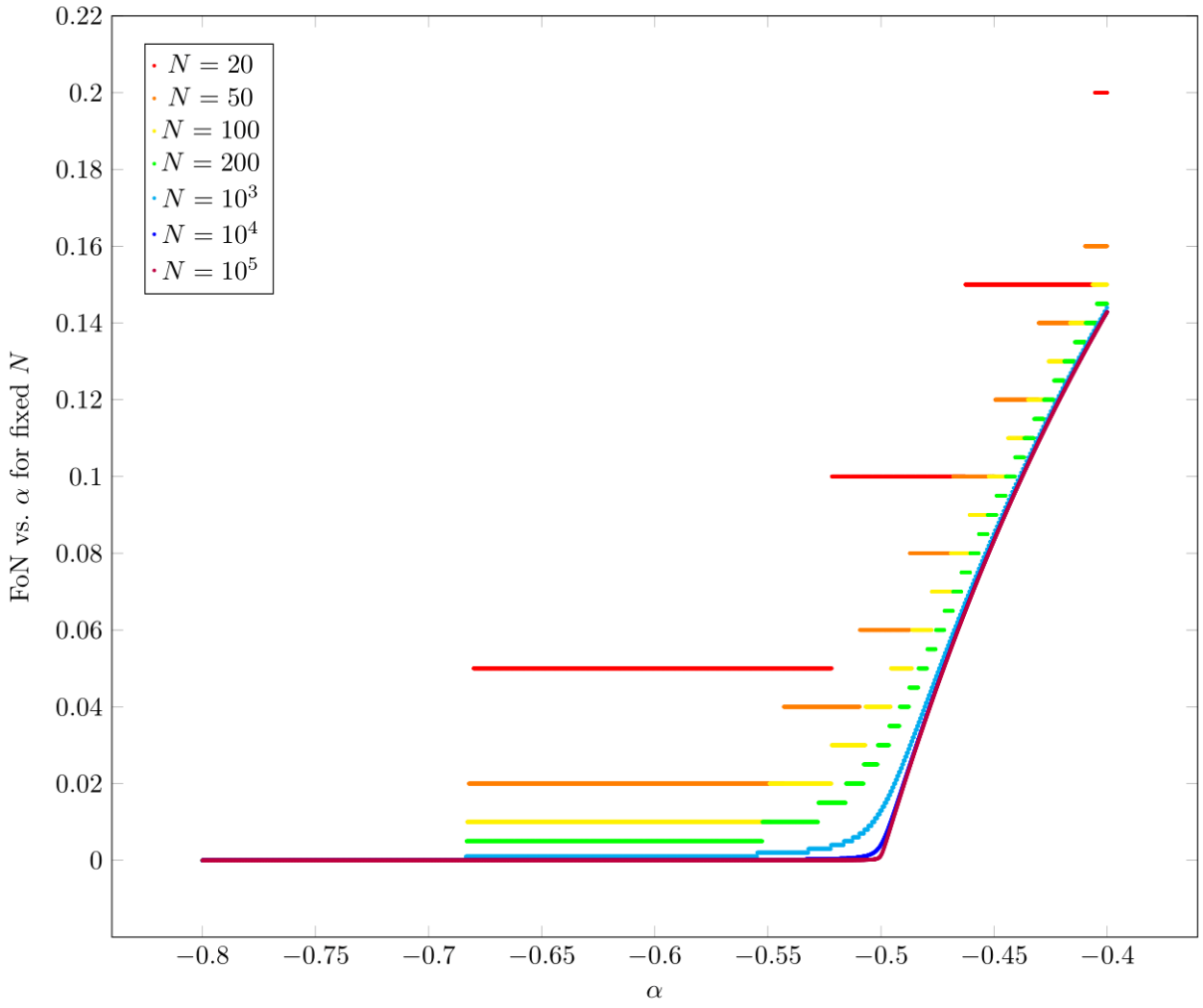

Figure 4.5: ***Fraction of Negatives (FoN) vs.*** α ***for Different Values of*** N We identify a critical point, denoted as α_p, where the FoN equals zero when $\alpha \leq \alpha_p$. The critical point exhibits a gradual variation with respect to N following these patterns: (i) for small N, α_p is in close proximity to -0.68; (ii) for large N, α_p tends to -0.5.

4.3.1.3 Robustness of I_{diff}

In Figure 4.3, different priors not only exhibit varying degrees of positivity but also display varying degrees of variation in I_{diff} for different values of h_N; we refer to this as *divergence*. The divergence depends upon the choice of prior. To better understand this

dependence, we quantify the dependence of I_{diff} on h_N by the standard deviation of I_{diff} across different values of h_N. Figure 4.6 illustrates how the standard deviation changes with respect to α while keeping N constant.

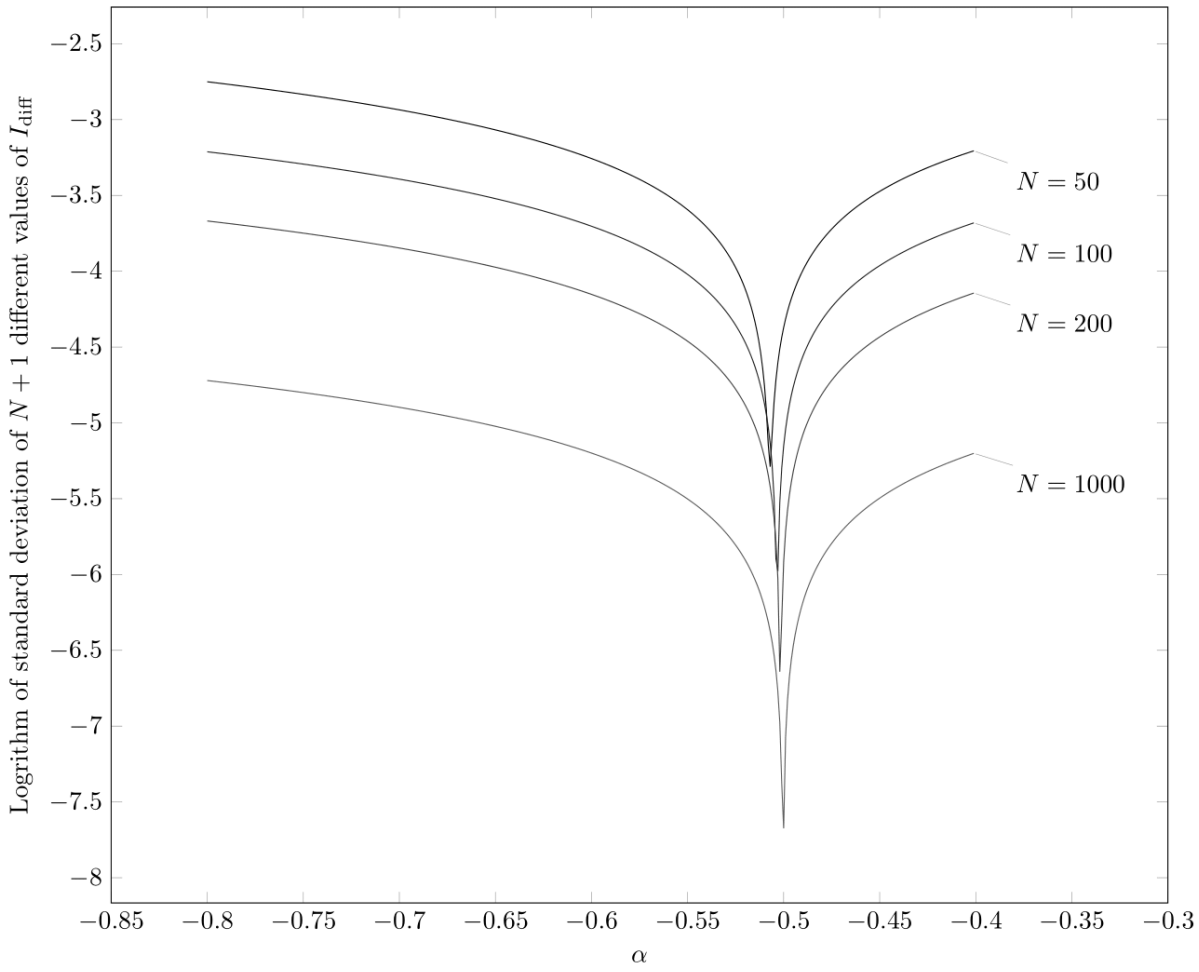

Figure 4.6: ***Robustness of Differential Information Gain (I_{diff})*** The y-axis represents the logarithm of the standard deviation of I_{diff} over all possible h_N values, while the x-axis depicts various selections of α. A smaller standard deviation indicates that different h_N values lead to the same result, implying greater independence of I_{diff} from h_N. This independence signifies the robustness of I_{diff} with respect to the natural variability in h_N, as we consider h_N to be solely determined by nature. The standard deviation, given a fixed N, is notably influenced by α, and there exists an α value at which the dependence on h_N is minimized. This particular α value approaches -0.5 as N increases.

It is evident that when α is close to -0.5, the standard deviation is at its minimum.

Reduced dependence of I_{diff} on h_N enhances its robustness against the effects of nature, as we attribute h_N to natural factors, while N is determined by human measurement choices. As N increases, the minimum point approaches -0.5. In the limit of large N, this minimum point will eventually converge to $\alpha = -\frac{1}{2}$, which means that under this specific choice of prior, I_{diff} depends minimally on h_N and primarily on N.

4.3.2 Large N Approximation

Utilizing a recurrence relation and a large x approximation, the digamma function can be approximated as:

$$\psi(x) = \frac{1}{x-1} + \psi(x-1) \approx \frac{1}{x-1} + \ln(x-1) - \frac{1}{2(x-1)} = \frac{1}{2(x-1)} + \ln(x-1) \quad (4.8)$$

As a result, the large N approximation for the differential information gain in Equation (4.7) becomes:

$$I_{\text{diff}} = \frac{2h_N + 1}{2(h_N + \alpha + 1)} - \frac{2N + 1}{2(N + 2\alpha + 2)} \quad (4.9)$$

Using this approximation, when $\alpha = -\frac{1}{2}$, $I_{\text{diff}} = \frac{1}{2(N+1)}$, which shows that I_{diff} solely depends on N. This finding aligns with Figure 4.3, which demonstrates that I_{diff} is most concentrated when $\alpha = -0.5$ and is also consistent with the results of [26].

In Figure 4.4, we observe that the FoN tends to become constant for very large values of N. These constants can be estimated using the large N approximation of I_{diff} in Equation (4.9) (see Table 4.1). If $I_{\text{diff}} \leq 0$, then

$$h_N \leq \frac{2N\alpha + N + \alpha + 1}{4\alpha + 3}, \quad (4.10)$$

and we obtain:

$$\text{FoN} = \frac{1}{N+1} \frac{2N\alpha + N + \alpha + 1}{4\alpha + 3} \approx \frac{2\alpha + 1}{4\alpha + 3} \quad (4.11)$$

This equation aligns with the asymptotic lines in Figure 4.4, providing support for the observation mentioned in Figure 4.3: namely, that for $\alpha = -0.7$, all points lie above the x-axis, while for other priors, negative points are present, and the fraction of negative points becomes constant.

α	FoN (Numerical Result, $N = 1000$)	FoN (Asymptotic Result)	Discrepancy between the Two Results
-0.7	0	0	0
-0.6	0.001	0	0.1%
-0.5	0.013	0	1.3%
-0.4	0.144	0.143	0.1%
0	0.334	0.333	0.1%
1	0.429	0.429	0
3	0.467	0.467	0

Table 4.1: ***Fraction of Negatives (FoN) under Selected Priors*** Comparison between numerical results and asymptotic results show that they agree with each other.

4.4 Relative Information Gain

The second form of information gain in a single toss is relative information gain, which represents the KL divergence from the posterior after N tosses to the posterior after $N + 1$ tosses. We continue to use the one-parameter beta distribution prior in the form of Equation (4.6). The relative information gain is (see Appendix A.2):

$$I_{\text{rel}}(t_{N+1} = \text{'Head'}) = \psi(h_N + \alpha + 2) - \psi(N + 2\alpha + 3) + \ln \frac{N + 2\alpha + 2}{h_N + \alpha + 1} \tag{4.12}$$

Relative information gain exhibits entirely different behaviour compared to differential information gain. Due to the properties of KL divergence, relative information gain is always non-negative, eliminating the need to consider negative values. We explore the dependence of relative information gain on priors and the interpretation of information gain in extreme cases.

In Figure 4.7, it becomes evident that, under different priors, the data lines exhibit similar shapes. This suggests that relative information gain is relatively insensitive to the choice of priors. On each graph, the top line represents the extreme case where the first N tosses result in tails and the $(N + 1)$th toss results in a head. This line is notably separated from the other data lines, indicating that relative information gain behaves more like a

measure of the degree of surprise associated with this additional data. In this black swan event, the posterior after $N+1$ tosses differs significantly from the posterior after N tosses.

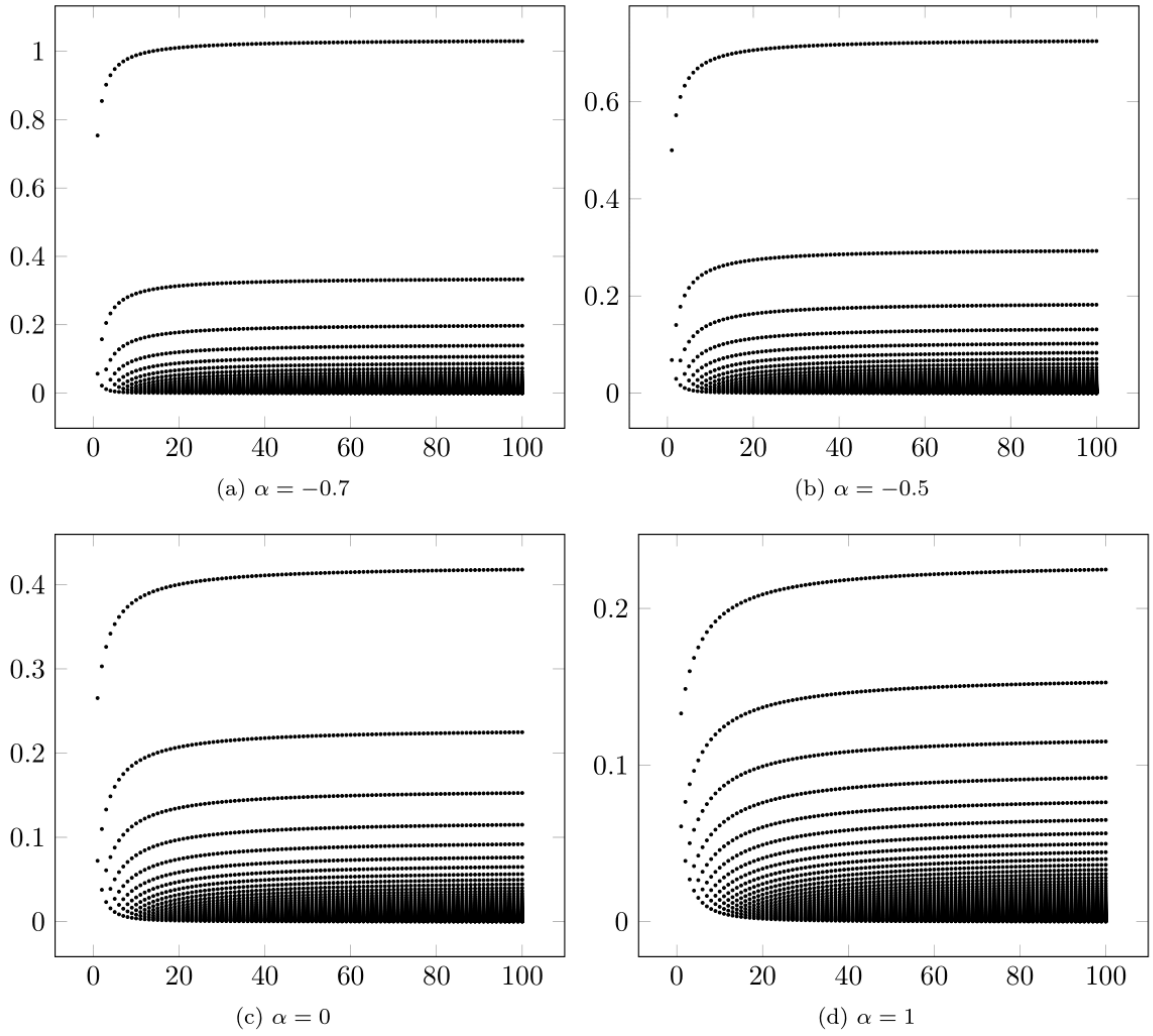

Figure 4.7: ***Relative Information Gain (I_{rel}) over Different Priors*** The y-axis represents the value of I_{rel}, while the x-axis represents N. For each N, there are $N+1$ different values of I_{rel}. It is important to note that I_{rel} is consistently positive across these selected priors. Similar to the differential information gain, each graph displays numerous divergent lines. However, the shape of these divergent lines remains remarkably consistent across varying values of α. The majority of these lines fall within the range of I_{rel} between 0 and 0.2.

For small values of N, both the average value and the standard deviation of I_{rel} exhibit

a clear monotonic relationship with α, meaning that larger values of α result in smaller average values and standard deviations. However, as N becomes large, all priors converge and become indistinguishable. Nonetheless, it is important to note that relative information gain remains heavily independent on the specific data sequences (h_N). Figure 4.8 illustrates how the standard deviation of I_{rel} under different priors converges to the same value as N increases.

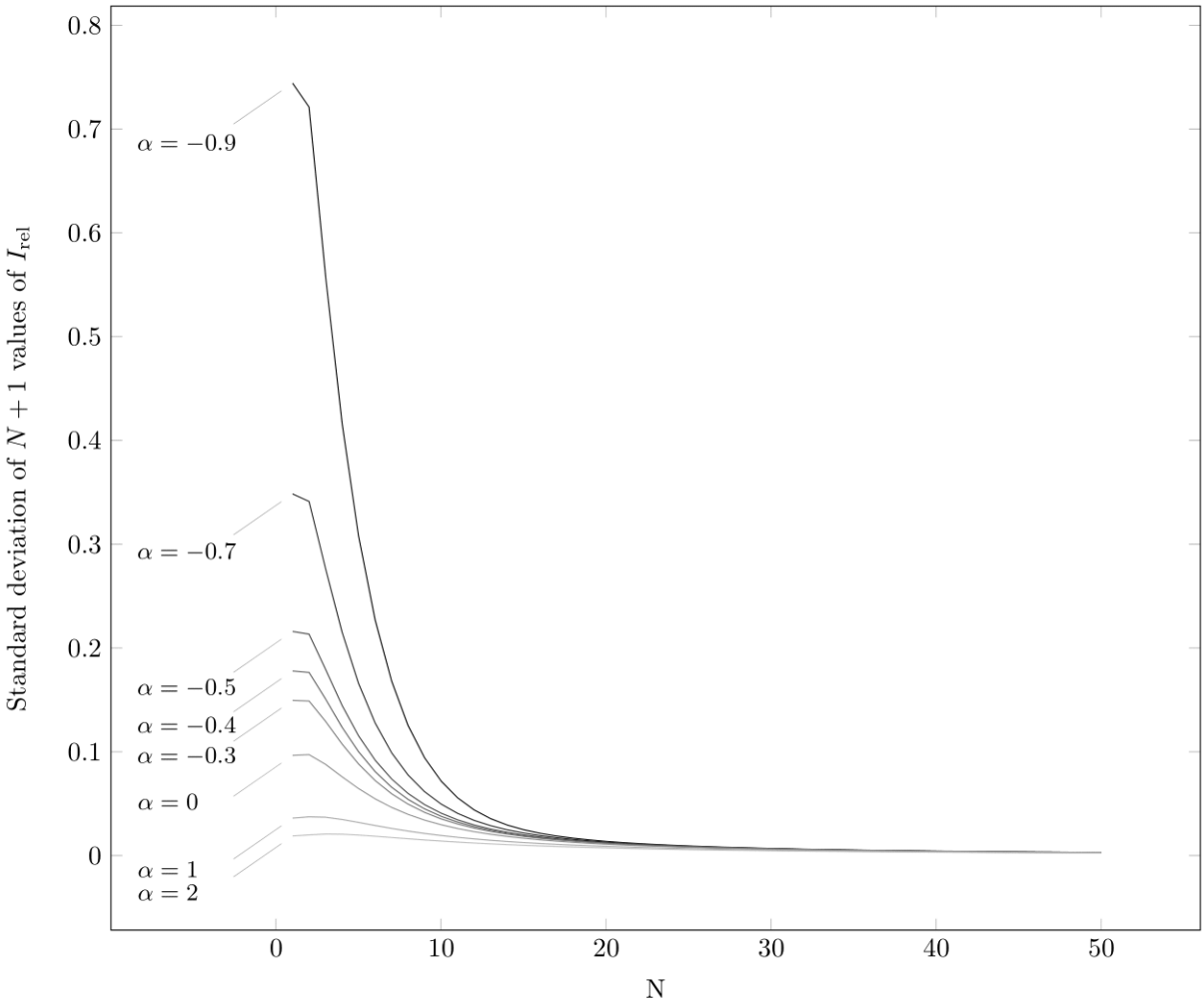

Figure 4.8: ***Robustness of Relative Information Gain (I_{rel})*** The y-axis represents the standard deviation of I_{rel} across all possible values of h_N. This demonstrates the substantial independence of I_{rel} from h_N. Additionally, as N increases, the standard deviations tend to approach zero for all priors.

By utilizing the aforementioned approximation of the digamma function, we obtain:

$$\begin{aligned} I_{\text{rel}}(t_{N+1} = \text{`Head'}) &\approx \frac{1}{2(h_N + \alpha + 1)} - \frac{1}{2(N + 2\alpha + 2)} \\ &= \frac{N - h_N + \alpha + 1}{2(h_N + \alpha + 1)(N + 2\alpha + 2)} \end{aligned} \tag{4.13}$$

In the large N limit, I_{rel} becomes:

$$I_{\text{rel}}(t_{N+1} = \text{`Head'}) \approx \frac{1}{2N}\left[\left(\frac{h_N}{N}\right)^{-1} - 1\right], \tag{4.14}$$

which is independent of α. Thus, it appears that the properties of relative information gain and differential information gain are complementary to each other. The differences between them are summarized in Table 4.2.

Information Gain Measure	Asymptotic forms ($t_{N+1} = $ 'Head')	Asymptotic sensitivity to prior
Differential Information Gain	$I_{\text{diff}} \approx \frac{2h_N+1}{2(h_N+\alpha+1)} - \frac{2N+1}{2(N+2\alpha+2)}$	Heavily dependent upon prior. Independent of h_N for certain priors ($\alpha = -1/2$).
Relative Information Gain	$I_{\text{rel}} \approx \frac{1}{2(h_N+\alpha+1)} - \frac{1}{2(N+2\alpha+2)}$	Insensitive to prior. For large N, only affected by h_N.

Table 4.2: ***Comparison of Characteristics of Two Measures of Information Gain***

4.5 Expected Information Gain

In this section, we discuss a new scenario: after N tosses but before the $(N+1)$th toss has been taken, can we predict how much information gain will occur in the next toss? The answer is affirmative, as discussed earlier.

After N tosses, we obtain a data sequence T_N with h_N heads. However, we can only estimate the probability p based on the posterior $\Pr(p|N, T_N, I)$. The expected value of p

can be expressed as:

$$\langle p \rangle = \int_0^1 p \;\; \Pr\left(p|N, T_N, I\right) \, dp = \frac{h_N + \alpha + 1}{N + 2\alpha + 2} \tag{4.15}$$

Based on this expected value of p, we can calculate the average of the information gain in the $(N+1)$th toss. We define the expected differential information gain in the $(N+1)$th toss as:

$$\begin{aligned} \overline{I_{\text{diff}}} &= \langle p \rangle \times I_{\text{diff}}(t_{N+1} = \text{`Head'}) + \langle 1 - p \rangle \times I_{\text{diff}}(t_{N+1} = \text{`Tail'}) \\ &= \frac{h_N + \alpha + 1}{N + 2\alpha + 2} \psi(h_N + \alpha + 2) + \frac{N - h_N + \alpha + 1}{N + 2\alpha + 2} \psi(N - h_N + \alpha + 2) - \psi(N + 2\alpha + 3) \\ &\quad + \frac{h_N + \alpha + 1}{N + 2\alpha + 2} \ln \frac{N + 2\alpha + 2}{h_N + \alpha + 1} + \frac{N - h_N + \alpha + 1}{N + 2\alpha + 2} \ln \frac{N + 2\alpha + 2}{N - h_N + \alpha + 1} \end{aligned} \tag{4.16}$$

$\overline{I_{\text{diff}}}$ represents the expected value of differential information gain in the $(N+1)$th toss. Similarly, we can define the expected relative information gain as:

$$\begin{aligned} \overline{I_{\text{rel}}} &= \langle p \rangle \times I_{\text{rel}}(t_{N+1} = \text{`Head'}) + \langle 1 - p \rangle \times I_{\text{rel}}(t_{N+1} = \text{`Tail'}) \\ &= \frac{h_N + \alpha + 1}{N + 2\alpha + 2} \psi(h_N + \alpha + 2) + \frac{N - h_N + \alpha + 1}{N + 2\alpha + 2} \psi(N - h_N + \alpha + 2) - \psi(N + 2\alpha + 3) \\ &\quad + \frac{h_N + \alpha + 1}{N + 2\alpha + 2} \ln \frac{N + 2\alpha + 2}{h_N + \alpha + 1} + \frac{N - h_N + \alpha + 1}{N + 2\alpha + 2} \ln \frac{N + 2\alpha + 2}{N - h_N + \alpha + 1} \end{aligned} \tag{4.17}$$

Surprisingly, $\overline{I_{\text{diff}}} = \overline{I_{\text{rel}}}$. This relationship holds true for any prior, not being limited to the beta distribution type prior, and furthermore holds for an arbitrary n-outcome probabilistic source. Please refer to Appendix A.3 for a detailed proof. This suggests that there is only one choice for the expected information gain.

We first show the numerical results of expected information gain under different priors. It is evident that all data points are above the x-axis, indicating that the expected information gain is positive-definite, as anticipated. Since both I_{rel} and $\langle p \rangle$ are positive, it follows that $\overline{I_{\text{rel}}}$ must also be positive.

As with the discussions of differential information gain and relative information gain, we are also interested in examining the dependence of expected information gain on α or h_N. However, such dependence appears to be weak, as illustrated in Figures 4.9 and 4.10. Expected information gain demonstrates strong robustness concerning variations in α and h_N.

The asymptotic expression of expected information gain is

$$\overline{I_{\text{diff}}} = \overline{I_{\text{rel}}} = \frac{1}{2N} \tag{4.18}$$

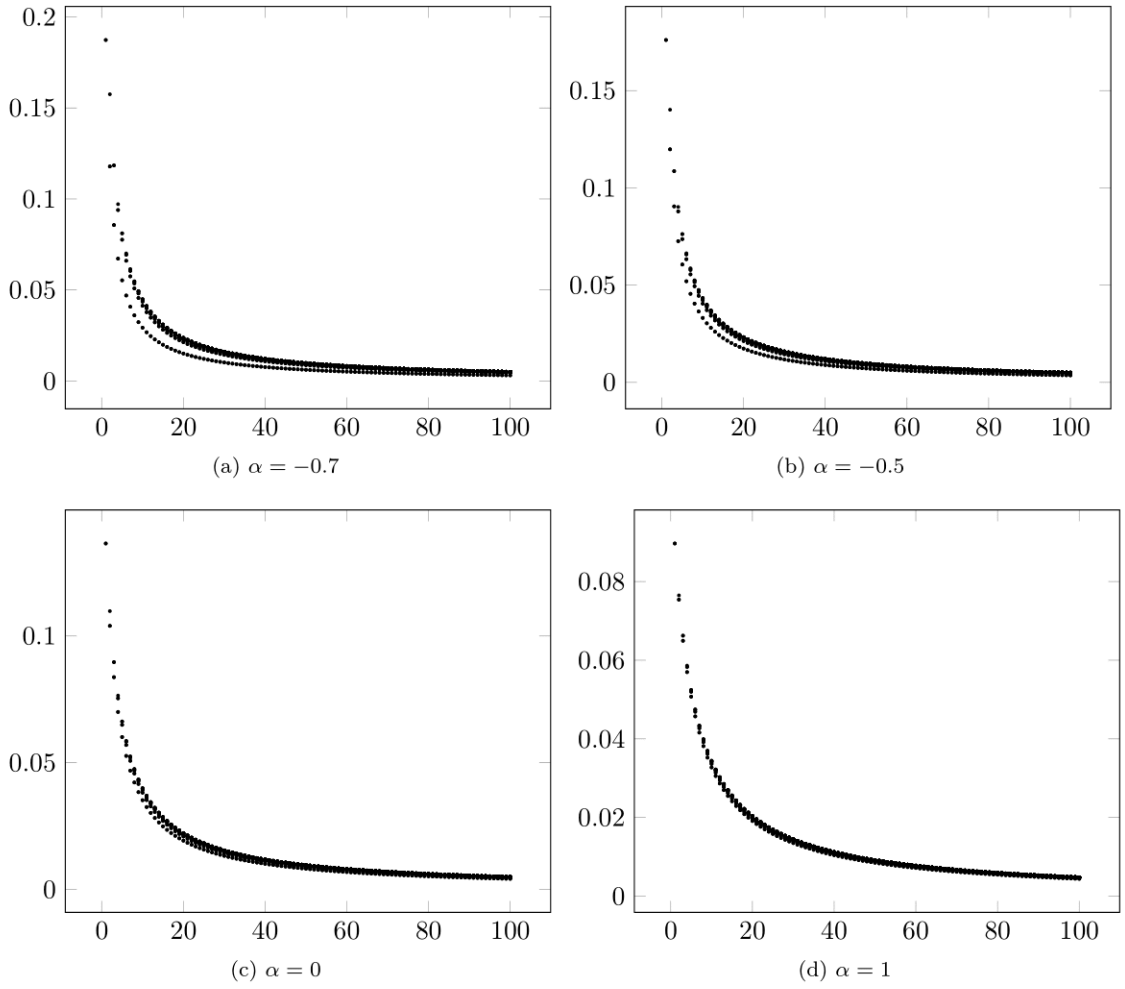

Figure 4.9: ***Expected Information Gain vs.*** N ***for Fixed*** α The y-axis represents the value of expected information, while the x-axis represents the value of N. Notably, all expected information gain values are positive. The shapes of each graph exhibit remarkable similarity, with a limited number of divergent lines. As α increases, the number of divergent lines decreases.

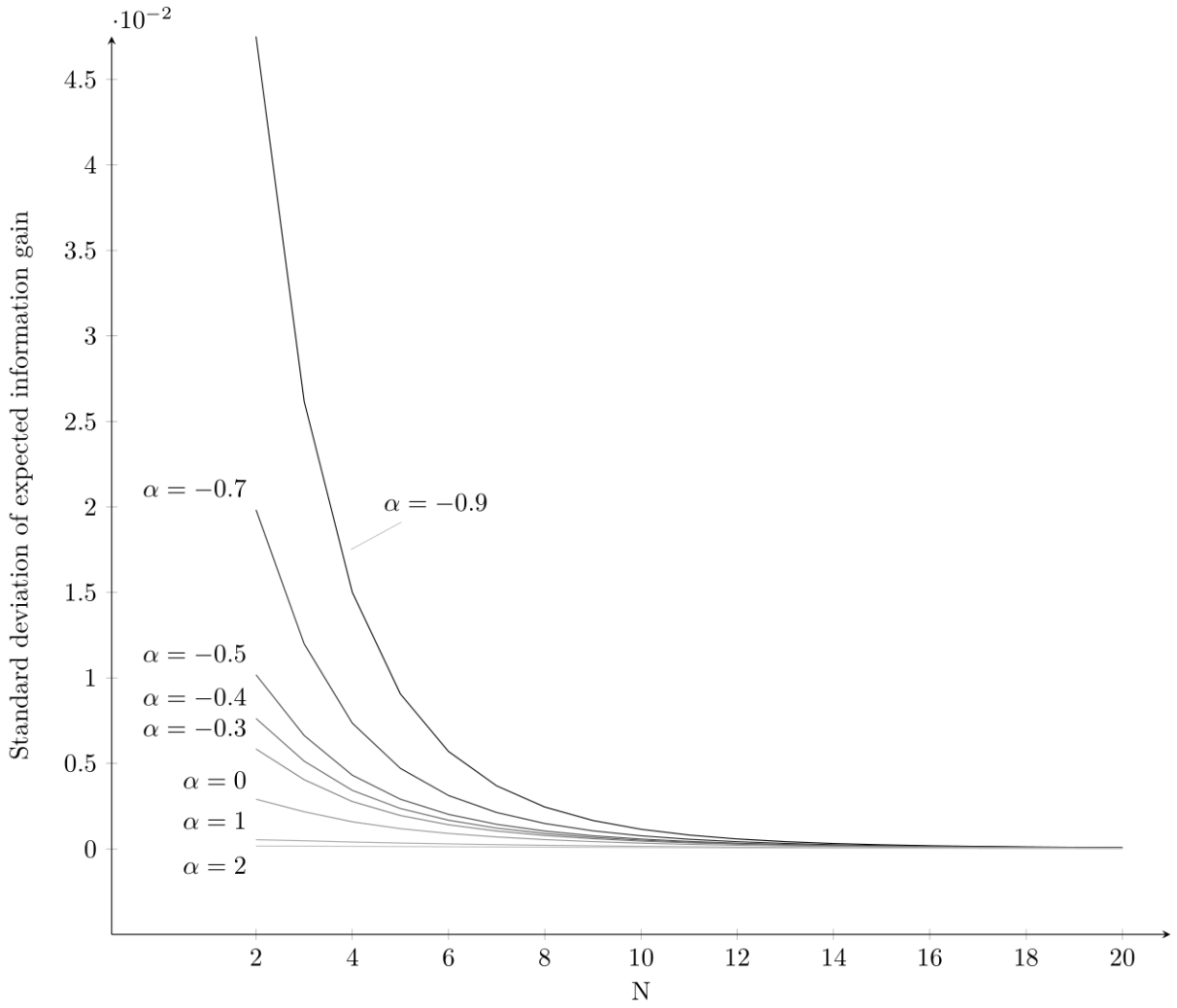

Figure 4.10: ***Robustness of Expected Information Gain*** The y-axis represents the standard deviation of the expected information gain over all possible values of h_N, while the x-axis represents the value of N. As N increases, and even for relatively small values of N, the standard deviation tends toward zero for all priors.

4.6 Comparison of Three Measures, and the Information Increase Principle

From an operational perspective, the information measures we have considered can be categorized into two types: differential information gain and relative information gain pertain to a measurement that has *already been made*, while expected information gain pertains to a measurement that has *yet to be conducted.*

Regarding positivity, which is tied to the fundamental question of "Will acquiring

more data from measurements lead to a deeper understanding of the system?": for relative information gain and expected information gain, the answer is affirmative, but differential information gain is positive only under certain specific prior conditions.

All three measures are functions of variables denoted as N, α, and h_N, which characterize the size of the data sequences, the prior information, and the existing data sequence, respectively. How sensitive are these measures to these parameters, particularly for large values of N? As we have shown, differential information gain is heavily influenced by all three parameters. It becomes nearly independent of h_N only when $\alpha = -0.5$. Relative information gain is not highly sensitive to the choice of priors. In the case of large values of N, relative information gain is affected by both h_N and N, whereas expected information gain depends solely on N. The comparison between them is summarized in Table 4.3.

Information Gain	Positivity	Robustness about T_N
Differential	Strictly positive when $\alpha < \alpha_p$ where $\alpha_p \approx -0.68$. Asymptotically positive when $\alpha \leq -0.5$.	Robustness exists only when $\alpha = -0.5$ of beta distribution prior.
Relative	Strictly positive for all priors.	No significant differences of robustness among beta distribution priors.
Expected	Strictly positive for all priors.	No significant differences of robustness among beta distribution priors.

Table 4.3: ***Comparison of Three Information Gain Measures***

At first, one might have expected that the idea that *more data from measurements lead to more knowledge about the system* would hold strictly: namely, that the information gain from additional data would always be strictly positive. However, our perspective has been challenged by the observation of black swan events. In the extreme scenario where the first N tosses all result in tails and the $(N+1)$th toss yields a head, a negative information gain in this $(N+1)$th toss may be a more reasonable interpretation. To address this, we propose the

Principle of Information Increase: *In a series of interrogations of an n-outcome probabilistic source, the information gain from additional data should tend towards positivity in the asymptotic limit. However, in the extreme case where the first N data points are identical and the data of the $(N+1)th$ trial is contrary to the previous data, the information gain in this exceptional case should be negative.*

Applying this criterion, the choice of using the differential information gain becomes more appropriate for measuring the extent of knowledge contributed by additional data. For the beta distribution prior, it should be constrained within the range of approximately $-0.68 \lesssim \alpha \leq -0.5$. If we also consider the robustness of information gain under various given data scenarios, then the Jeffreys binomial prior ($\alpha = -0.5$) emerges as the most favourable choice.

All three measures are based on KL divergence, however, strictly speaking none of them can be a "measure" since the triangle inequality cannot be satisfied. We can still use measure to denote their role when quantifying the information gain in measurements.

From the view of operational perspective, they can be divided into two types: differential information gain and relative information gain are evaluations of a measurement that has already been taken; expected information gain is a prediction of a measurement that hasn't been taken.

From the view of positivity, that is, will this quantity always be positive? This connects with our beginning question, "will more data from measurements lead to more knowledge of a system?" For relative information gain and expected information gain the answer is yes, while differential information gain is positive only under some certain priors.

All three measures are functions of N, α, h_N which characterize size of the data sequences, prior and existing data sequence respectively. How are they sensitive to the three parameters, especially for large N? As we know, differential information gain is heavily influenced by all three parameters, only when $\alpha = -0.5$ differential information gain will be nearly independent of h_N. Relative information gain is not sensitive to priors, in large N, relative information gain is affected by h_N and N, while expected information gain only depends on N.

Initially we may hope that the idea "more data from measurements lead to more knowledge about the system" is strictly hold, that is, the information gain of the additional data should be strictly positive under all cases. However, based on the observation of the "Black Swan Event", we find a strictly positive information gain may not be meaningful. In the extreme case that first N tosses are all tails and the $N+1$th toss is head, a negative information gain in this $N+1$th toss may be more physically reasonable. We propose the Principle of Information Gain as follows:

In a series of binomial distribution data, the information gain of the additional data should be positive asymptotically; in the extreme case that first N trials of data are all the same and the data of $N+1$th trial is opposite to previous data, then the information gain in this extreme case should be negative.

Under this criterion, the differential information gain should be a better choice to measure the degree of knowledge of the additional data. The beta distribution prior may be ranged between $-0.68 \lessapprox \alpha \leq -0.5$. If we also consider about the robustness of information gain over different given data, then the Jeffreys' binomial prior ($\alpha = -0.5$) would be the best choice.

4.7 Related Work

4.7.1 Information Increase Principle and the Jeffreys Binomial Prior

In [53, 54], Summhammer introduces the idea that *more measurements lead to more knowledge about a physical quantity* and quantifies the level of knowledge regarding a quantity by assessing its uncertainty range after a series of repeated measurements. Quantified in this manner, the notion can be summarized as: "The uncertainty range of a physical quantity should decrease as the number of measurements increases." For a quantity θ, the uncertainty range $\Delta\theta$ is a function of the number of measurements:

$$\Delta\theta(N+1) < \Delta\theta(N) \tag{4.19}$$

If this quantity is determined by the probability of a two-outcome measurement, such as the probability of obtaining heads (p) in a coin toss, then there exists a relationship

between the uncertainty range of θ and that of p,

$$\Delta\theta = \left|\frac{\partial\theta}{\partial p}\right| \Delta p \tag{4.20}$$

In large N approximation, $\Delta p = \sqrt{p(1-p)/N}$, so that

$$\Delta\theta = \left|\frac{\partial\theta}{\partial p}\right| \sqrt{p(1-p)/N}. \tag{4.21}$$

One intuitive way to ensure Equation (4.19) holds is by forcing $\Delta\theta$ to be purely a function of N. Observing the relationship between $\Delta\theta$ and Δp, the simplest solution would be to set $\Delta\theta = \frac{\text{const.}}{\sqrt{N}}$. Under this solution, the relationship between p and θ takes the following form:

$$\left|\frac{\partial\theta}{\partial p}\right| \sqrt{p(1-p)} = \text{const.}, \tag{4.22}$$

which yields Malus' law $p(\theta) = \cos^2(m(\theta - \theta_0)/2)$, with $m \in \mathbb{Z}$.

Summhammer does not employ information theory to quantify "knowledge about a physical quantity" but instead utilizes the statistical uncertainty associated with the quantity. However, viewed from the Bayesian perspective, if we assume that the prior distribution of the physical quantity, θ, is uniform, the difference between θ and p in Equation (4.21) implies that the prior distribution of the probability follows the Jeffreys binomial prior:

$$\Pr(p|I) = \left|\frac{\partial\theta}{\partial p}\right| \Pr(\theta|I) = \frac{1}{\pi}\frac{1}{\sqrt{p(1-p)}} \tag{4.23}$$

Thus, in the large N approximation, Summhammer's result can be interpreted to mean that the prior associated with the probability of a uniformly distributed physical quantity must adhere to the Jeffreys binomial prior.

Goyal [26] introduces an *asymptotic* Principle of Information Gain, which states that "In n interrogations of a N-outcome probabilistic source with an unknown probabilistic vector $\vec{P}$, the amount of Shannon–Jaynes information provided by the data about $\vec{P}$ remains independent of $\vec{P}$ for all $\vec{P}$ in the limit as $n \to \infty$." Goyal establishes the equivalence between this principle and the Jeffreys rule. Under his Principle of Information Gain, the Jeffreys multinomial prior is then derived. In the case of a two-outcome probabilistic model,

the Jeffreys multinomial prior reduces to the Jeffreys binomial prior. Asymptotic analysis reveals that Shannon–Jaynes information is not only independent of the probability vector $\vec{P}$ but also monotonically increases with the number of interrogations. It is worth noting that Shannon–Jaynes information can be viewed as the accumulation of differential information gain. This asymptotic result aligns with our findings: under the Jeffreys binomial prior, the differential information gain is solely dependent on the number of measurements.

4.7.2 Other Information-Theoretical Motivations of the Jeffreys Binomial Prior

Wootters [57] introduces a novel perspective on the Jeffreys binomial prior, where quantum measurement is employed as a communication channel. In this framework, Alice aims to transmit a continuous variable, denoted as θ, to Bob. Instead of directly sending θ to Bob, Alice transmits a set of identical coins to Bob, where the probability of getting heads, $p(\theta)$, in each toss is a function of θ. Bob's objective is to maximize the information about θ that he can extract from a finite number of tosses. The measure of information used in this context is the mutual information between θ and the total number of heads, n, in N tosses.

$$I(n:\theta) = H(n) - H(n|\theta) = -\sum_{n=0}^{N} p(n)\ln P(n) - \left\langle -\sum_{n=0}^{N} p(n|p(\theta))\ln p(n|p(\theta)) \right\rangle \tag{4.24}$$

However, the function $p(\theta)$ is unknown, and the optimization process begins with a set of discrete values, $p_1, p_2, \ldots, p_L$ rather than utilizing the continuous function $p(\theta)$. For each discrete value, p_k, there is an associated weight, w_k. The mutual information can be expressed as follows:

$$I(n:\theta) = -\sum_{n=0}^{N} p(n)\ln P(n) + \sum_{k=1}^{L} w_k \sum_{n=0}^{N} p(n|p_k)\ln p(n|p_k) \tag{4.25}$$

In the large N approximation, it is found that the weight w takes on a specific form:

$$w(p) = \frac{1}{\pi\sqrt{p(1-p)}} \tag{4.26}$$

which serves a role akin to the prior probability of p. Remarkably, this prior probability

aligns with the Jeffreys binomial prior. A similar procedure can be extended to the Jeffreys multinomial prior distribution. Wootters' approach shares similarities with the concept of a reference prior, where the selected prior aims to maximize mutual information, which can be viewed as the expected information gain across all data. The outcome is consistent with the reference prior for multinomial data [5], thus revealing another informational interpretation of the Jeffreys prior.

4.8 Conclusion

Motivated by recent work in quantum reconstruction and quantum state tomography, we have investigated the concept of information gain for a two-outcome probabilistic source from an operational perspective. We have introduced an informational postulate, the Principle of Information Increase, which serves as a criterion for selecting the appropriate measure to quantify the extent of information gained from measurements and to guide the choice of prior. We have shown that differential information gain is the most physically meaningful measure when compared to the other contender: the relative information gain. We have also uncovered the unanticipated and rather remarkable result that the *expected* value of these two measures of information gain are *equal* for any prior and for any n-outcome probabilistic source.

Within the set of symmetric beta distributions, we have shown that the Jeffreys binomial prior exhibits notable characteristics. Both Summhammer's work and ours demonstrate that, under this prior, the intuitive notion that *more data from measurements leads to more knowledge about the system* holds true, as confirmed by two distinct methods of quantifying knowledge. Additionally, Wootters shows that this prior enables the communication of maximal information, further highlighting its significance. Here, we have formulated the novel notion of *robustness* and have shown that the Jeffreys binomial prior displays maximal robustness within the set of symmetric beta distributions. Our work raises the intriguing question of whether this feature could be extended to the multinomial Jeffreys prior and whether it would be possible to lift the initial restriction to the set of beta distributions. We also speculate that a deeper understanding of the robustness of the Jeffreys prior remains to be uncovered.

PART III

Information about Observables

CHAPTER 5

Quantum Question Structures

5.1 Introduction

A fundamental difference between classical and quantum systems is the following: whereas a *single* measurement can be performed on a classical system which reveals the state of the system, many different and inequivalent measurements can be performed on a single qubit, each of which generally provides limited information about the state of the system. The set of possible measurements that can be performed on a quantum system has a rich internal structure.

The motivation for the present work comes from some facts observed on spin-$\frac{1}{2}$ particles. In quantum tomography, the state of a single spin-$\frac{1}{2}$ particle can be determined by probability of the Stern-Gerlach measurements. The density matrix of a single spin-$\frac{1}{2}$ particle can then be represented as:

$$\hat{\rho} = \frac{1}{2}(\hat{I} + \vec{r} \cdot \hat{\vec{\sigma}}) \tag{5.1}$$

Moreover, consider a two-body system composed of spin-$\frac{1}{2}$ system A and B, the state of this composite system can be determined by a set of local measurements and global measurements [12]:

$$\hat{\rho}_{AB} = \frac{1}{4}(\hat{I} \otimes \hat{I} + \vec{r}_A \cdot \hat{\vec{\sigma}}^A \otimes \hat{I} + \hat{I} \otimes \vec{r}_B \cdot \hat{\vec{\sigma}}^B + \sum_{i,j} \beta_{ij} \hat{\sigma}_i^A \otimes \hat{\sigma}_j^B), \tag{5.2}$$

where $\hat{\vec{\sigma}}^A, \hat{\vec{\sigma}}^B$ are Bloch vectors on the single spin-$\frac{1}{2}$ system A and B respectively, and β_{ij} are real numbers. Two interesting properties are observed in qubits systems: 1. The outcome probability of a joint measurement of $\hat{\sigma}_i^A \otimes \hat{\sigma}_j^B$ can be obtained by the combination of statistics of local measurement outcomes of $\hat{\vec{\sigma}}^A \otimes \hat{I}$ and $\hat{I} \otimes \hat{\vec{\sigma}}^B$, yet the single time measurement behavior of the joint measurement and local measurement are totally different. 2. For a single qubit, we could use a set of mutually complementary measurements, $\{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$, to do the state tomography. For two qubits, we could also do the state tomography via those locally mutually complementary measurements and the joint measurements. However, those joint measurements may not be mutually complementary; some joint measurements are even

mutually commuting to each other. The coincidence of commutativity and complementarity suggests there may be a deeper reason. We tend to use information theory to explore an explanation by introducing a new structure, *quantum question structure*. In one hand it is possible to provide a new viewpoint of understanding quantum mechanics, on the other hand we also find some new results of quantum mechanics.

Employing information theory into quantum mechanics is not a new thing. Rovelli[51] first proposed this idea within a new interpretation of finite dimensional quantum mechanics. Since every quantum measurement can be decomposed into many projections, the whole physical system can be regarded as a collection of binary outcome measurements. Each projection is a binary question. The state of the system is the collection of all the results of these binary questions. It is assumed that we may obtain 1 unit of information when obtaining the result of one question. The total information we can obtain about the system is assumed to be finite. Yet the detailed structure of questions is not mentioned.

Brukner and Zellinger [8, 9] proposed very similar assumptions of information as Rovelli's, and information is used to build the structure of measurements of quantum mechanics. They suggested that the information of a system is the sum of information of all complementary questions. In the case of a spin-$\frac{1}{2}$ particle, this assumption is very plausible. Since all complementary questions for a spin-$\frac{1}{2}$ particle are just the spin operators. Yet in the case of a two-body spin-$\frac{1}{2}$ particle, the definition of complementary questions is not clear. The complementary questions for single body system seems to be related to mutually unbiased operators while using mutually unbiased operators only may not be complete to describe a composite system.

Höhn[30, 31] proposed a new reconstruction of quantum mechanics. Rovelli's assumptions are included in this reconstruction. The relation between the joint measurements and local measurements is described by a logical gate. The information of the system is defined in a new way, it is the sum of a collection of finite questions. Together with some other fundamental assumptions, it is derived that the single system's information is determined by all complementary questions. This reconstruction claimed to recover qubit quantum mechanics and some important results are derived. However, generalizing to higher dimensions is not straightforward. It appears that the allowed logical gate, derived from specific rules, has only one form and is associative in a two-dimensional case. While in higher dimensions, there

exist multiple possible formations of the logical gate, among which only a few are associative.

The above discussions are all based on qubit quantum mechanics. Some results are really impressive, yet there is a small concern: qubit quantum mechanics may not be easily extended to higher-order dimensional cases. We may say that an arbitrary even number dimensional quantum system can be decomposed as a many-qubit system, but how about odd number dimensional cases, say three dimensions? Indeed, one may still possibly use qubits to represent odd number dimensional systems, yet this will definitely lead to some redundancies. Moreover, some results may be just a coincidence of the two dimensions. For example, one core task of the reconstruction of qubit quantum mechanics is to recover/derive the structure of Pauli matrices, since the Pauli matrices are good enough to describe a qubit. However, Pauli matrices themselves are really unique; they are pairwise complementary and anti-commuting[4]. It is natural to ask what's the analogy for generalization of Pauli matrices in higher dimension. Moreover, we want to deal with the odd number dimensional system without any redundancy. We are curious about what happens if applying the similar formalism of information theory to higher-dimensional cases.

Here are the main features of this chapter:

1. We generalize Höhn's formalism into higher dimensional cases and find several non-trivial results. The main difficulty of generalization emerges when dealing with composite systems. In quantum mechanics, we can use tensor product to compose measurement on more than one single system. By abstracting the act of performing a measurement on a physical system as asking a question to the system, we inherit Höhn's notion of a logical gate to connect questions about a single system, analogous to tensor products. In two dimensional case, there is only one choice of logical gate, the exclusive or gate. In higher dimensional case, the choice of logical gate may not be limited. We find a mathematical structure, *orthogonal array*, to describe the classification of different logical gates. The well discussed results on orthogonal array in prime number dimensional case lead us to focus on prime number dimensional quantum mechanics only.

[4]Anti-commutativity is not widely used. For example, $\hat{\sigma}_x$ and $\hat{\sigma}_z$ are mutually complementary, but $\hat{\sigma}_x \times \hat{\sigma}_x$ and $\hat{\sigma}_z \times \hat{\sigma}_z$ commute with each other. This local complementarity and global commutativity of Pauli matrix is directly due to anti-commutativity.

2. We clarify the notion and definition of information, especially information of measurement and information of system. The definition discussed by Rovelli and Brukner and Zeilinger is not very clear. We restrict the information of measurement to be a function of the probability distributions of the outcomes. In some sense this is a measure of uncertainty of the outcomes. Moreover, all the probabilities we deal with are written in the Bayesian style. In this way, we may show that there are two different understanding of information of measurement and in this paper we only use one of them. The information of system is proposed from the viewpoint of tomography, where a finite set of selected measurements can be used to determine the state of system. By holding a similar assumption, the combination of the information from those selected measurements characterizes all our knowledge about the system, and we name it as information of system.

3. We provide a connection between the quantum question structure we begin with and the ordinary quantum mechanics. In two dimensional case, all questions have binary outcomes and the system is corresponding to a qubit. Every question resembles a Pauli matrix, i.e. a Stern-Gerlach measurement on the qubit. For composite systems, every composite question resembles a joint measurement on the multi-qubit system where each joint measurement can be represented as a tensor product of Pauli matrices.

   For higher dimensional cases, this analogy is not clear. Indeed we may want to find an analogy of Pauli matrix in higher dimensional space such that every question is corresponding to a specific measurement on a qudit. We choose the generalized Pauli matrix that is build based on *mutually unbiased bases* (MUBs). The reason of this choice is that we think the complementarity is the most important property of Pauli matrix, which is perfectly revealed in MUBs from qubits to higher dimensional spaces. In this way, we could translate our main results and their derivations in terms of linear space language, making the abstract formalism not so abstract.

The goal of the current project is to elucidate the structure of these questions. This chapter is organized as follows. Section 5.2 introduces the new question set structure. We begin with several interesting properties among qubits and the abstract structure of quantum questions, then we show the whole construction of quantum questions. Section 5.3 introduces MUBs in ordinary quantum mechanics. It is the similarity between the properties in MUBs

and the consequences in quantum question structure lead to a possible connection. Such connection is discussed in Section 5.4, where examples are provided to illustrate the basic idea of quantum questions in terms of ordinary quantum mechanics language. The degree of freedom of the system agrees with the result in ordinary quantum mechanics. The proof of two new results are stated in Appendix B.

5.2 A physical system as set of questions

5.2.1 Motivation of the question structure

We begin our discussion by considering a qubit. There are an infinite number of projective measurements on the qubit, each of them can be represented as a unitary operator:

$$\hat{\sigma}_{\theta,\phi} = |+_{\theta,\phi}\rangle \langle +_{\theta,\phi}| - |-_{\theta,\phi}\rangle \langle -_{\theta,\phi}| \quad \theta \in [0,\pi], \phi \in [0,2\pi), \tag{5.3}$$

where

$$|+_{\theta,\phi}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ e^{i\phi} \sin\left(\frac{\theta}{2}\right) \end{pmatrix} \qquad |-_{\theta,\phi}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ -e^{-i\phi} \sin\left(\frac{\theta}{2}\right) \end{pmatrix}. \tag{5.4}$$

Once a projective measurement has been performed, the post-measurement state of the system will be the eigenstate of this operator, and we can infer the outcome probabilities of any other projective measurements performed on the system immediately afterwards. All these projections have the same eigenvalues, ± 1. Moreover, the state of the qubit can be reconstructed via these projections.

Property 1. The state of a single qubit can be reconstructed by state tomography over any three different projective measurements.

This property can be viewed as a natural consequence of the density matrix of qubit (5.1). Though theoretically we could choose any three different axes for the projections, three perpendicular axis are commonly chosen for the sake of calculational convenience. Moreover, the projections with perpendicular axes has another property.

Property 2. Two projective measurements, $\hat{\sigma}_{\theta,\phi}$ and $\hat{\sigma}_{\theta',\phi'}$, with perpendicular axes will

be mutually unbiased[5] to each other.

We can choose any two direction $\{\theta, \phi\}, \{\theta', \phi'\}$ which are mutually perpendicular and the corresponding measurements are mutually unbiased. A typical example is $\{\theta = \frac{\pi}{2}, \phi = 0\}, \{\theta' = 0, \phi' \in [0, 2\pi)\}$, which denotes to the mutually unbiased measurements of $\hat{\sigma}_x, \hat{\sigma}_z$,

$$|\langle +|0\rangle|^2 = |\langle -|0\rangle|^2 = |\langle +|1\rangle|^2 = |\langle -|1\rangle|^2 = \frac{1}{2}. \tag{5.5}$$

Property 1&2 are two common facts of single qubit. Noticing that those two properties are describing measurements on an individual system, and we may call them local measurements. The mutually unbiasedness between local measurements are very obvious and widely used. What attracts us more is the mutually unbiasedness between global measurements on a composite system. Here, composite system denotes a collection of identical systems, and global measurements are the tensor products of local measurements. For a composite system containing several qubits, the global measurement may be of the form of $\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \cdots \otimes \hat{\sigma}_{i_n}$ where each $\hat{\sigma}_{i_r}$ is a local measurement on a subsystem. These global measurements have the same number of outcomes as the local measurements, but the relations between global measurements will be more different, they could commute or be mutually unbiased to each other. We want to show two interesting properties about these global measurements by taking examples on a two-body qubit system.

Property 3. In a two-body qubit system, two global measurements $\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}$ and $\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}$ either commute or be mutually unbiased, and then they are non-informative [6] to each other , where $\hat{\sigma}_{i_1}, \hat{\sigma}_{i_2}, \hat{\sigma}_{j_1}, \hat{\sigma}_{j_2} \in \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$.

Two measurements are non-informative, which means the result of one measurement alone cannot determine the result of another measurement if performed subsequently. We will show this in detail for global measurements that commute or mutually unbiased.

[5] Two operators are mutually unbiased if their eigenstates or eigensubspaces are mutually unbiased, we will talk more about that in Section 5.3.

[6] Addressing the unclear definition of complementary questions in Brukner and Zellinger's work, we propose using the set of non-informative questions as a complete description of a composite system. This set contains both mutually unbiased operators as well as commuting operators, akin to the joint terms in the density matrix of two-body spin-$\frac{1}{2}$ system.

- $\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}$ and $\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}$ commute

  We first take the example on $\hat{\sigma}_{i_1} = \hat{\sigma}_{j_1} = \hat{\sigma}_x$ and $\hat{\sigma}_{i_2} = \hat{\sigma}_{j_2} = \hat{\sigma}_z$. Noticing that $\hat{\sigma}_x \otimes \hat{\sigma}_x$ and $\hat{\sigma}_z \otimes \hat{\sigma}_z$ commute, their common eigenstates are just the famous Bell states:

$$\begin{aligned}
\left|\Psi^+\right\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) \\
\left|\Psi^-\right\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle) \\
\left|\Phi^+\right\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) \\
\left|\Phi^-\right\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle).
\end{aligned} \tag{5.6}$$

  For an unknown system, if we first take measurement of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ at time t_1 and get some eigenvalue ± 1, after this measurement the state of the system would be projected into one of the two eigensubspaces of $\hat{\sigma}_x \otimes \hat{\sigma}_x$:

$$|\psi\rangle_{t>t_1} \in \begin{cases} Span(\{|\Psi^+\rangle, |\Phi^+\rangle\}) & \text{if outcome is 1} \\ Span(\{|\Psi^-\rangle, |\Phi^-\rangle\}) & \text{if outcome is -1.} \end{cases} \tag{5.7}$$

  If we immediately then take a measurement of $\hat{\sigma}_z \otimes \hat{\sigma}_z$ at time $t_2 > t_1$[7], what can we say about the outcome probabilities **before** t_2? The answer is that it depends on the initial state before t_1. The unknown initial state cannot infer anything about these outcome probabilities, nor can the measurement result of $\hat{\sigma}_x \otimes \hat{\sigma}_x$. The best we can say is that[8],

$$\Pr\left(\text{“}\hat{\sigma}_z \otimes \hat{\sigma}_z, \lambda, t_2\text{”} \middle| \text{“}\hat{\sigma}_x \otimes \hat{\sigma}_x, \lambda', t_1\text{”}, I\right) = \text{unknown}, \tag{5.8}$$

  where I denotes the fundamental postulates of quantum mechanics.

  For the sake of convenience, we could take a special choice of the initial state $\hat{\rho}_{t=t_0}$

[7] In the following discussions, the lower index moment always denotes earlier moments, i.e. $t_0 < t_1 < t_2 < t_3 < \cdots$

[8] Assume the operator $\hat{A}$ has distinct eigenvalues $\{a_1, a_2, \cdots, a_N\}$ then the action “take measurement $\hat{A}$ at time t and obtain an outcome of a_i” is abbreviated as “$\hat{A}, a_i, t$”

such that the "unknown" is replaced with an intuitive choice,

$$\begin{aligned} \Pr\left(\text{“}\hat{\sigma}_z \otimes \hat{\sigma}_z, \lambda, t_2\text{”} \middle| \text{“}\hat{\sigma}_x \otimes \hat{\sigma}_x, \lambda', t_1\text{”}, \hat{\rho}_{t=t_0}, I\right) &= \frac{1}{2}, \\ \Pr\left(\text{“}\hat{\sigma}_z \otimes \hat{\sigma}_z, \lambda, t_2\text{”} \middle| \hat{\rho}_{t=t_0}, I\right) &= \frac{1}{2}. \end{aligned} \tag{5.9}$$

This suggests that the measurement result of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ does nothing about the outcome probabilities of $\hat{\sigma}_z \otimes \hat{\sigma}_z$, and if we choose an uninformative initial state, the outcome probabilities of $\hat{\sigma}_z \otimes \hat{\sigma}_z$ before and after the measurement of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ are the same. In this situation, we may say the two measurements are independent. This state $\hat{\rho}_{t=t_0}$ turns out to be a maximally entangled state for a two-body qubits system, where $\hat{\rho}_{t=t_0} = \frac{1}{4}\hat{I} \otimes \hat{I}$. Moreover, this choice is also uninformative. Under this initial state, all these global measurements will have the same outcome probabilities:

$$\Pr\left(\text{“}\hat{\sigma}_{\theta,\phi} \otimes \hat{\sigma}_{\theta',\phi'}, \lambda, t_1\text{”} \middle| \hat{\rho}_{t=t_0}, I\right) = \frac{1}{2}. \tag{5.10}$$

We can generalize this idea to any two pairs of commuting global measurements $\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}$ and $\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}$. If we know nothing about the initial state, we could choose $\hat{\rho}_{t=t_0} = \frac{1}{4}\hat{I} \otimes \hat{I}$ to act as the prior of the initial state. This choice leads to the following relation:

$$\Pr\left(\text{“}\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}, \lambda, t_2\text{”} \middle| \text{“}\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}, \lambda', t_1\text{”}, \hat{\rho}_{t=t_0}, I\right) = \frac{1}{2}. \tag{5.11}$$

- $\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}$ and $\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}$ are mutually unbiased

  Now, let's consider the example of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ and $\hat{\sigma}_x \otimes \hat{\sigma}_z$, which do not commute, to illustrate this property. If we first take a measurement of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ at time t_1 and obtain an outcome of $+1$, the state $|\psi\rangle_{t>t_1}$ of the system will be projected into the eigensubspace of $\hat{\sigma}_x \otimes \hat{\sigma}_x$, denoted as $E(+1, \hat{\sigma}_x \otimes \hat{\sigma}_x)$. This eigensubspace can be expressed as a combination of Bell states:

$$|\psi\rangle_{t>t_1} = \alpha \left|\Psi^+\right\rangle + \beta \left|\Phi^+\right\rangle \quad \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1. \tag{5.12}$$

We may also decompose $|\psi\rangle_{t>t_1}$ as a combination of eigenstates of $\hat{\sigma}_x \otimes \hat{\sigma}_z$:

$$\begin{aligned}|\psi\rangle_{t>t_1} &= \alpha \left|\Psi^+\right\rangle + \beta \left|\Phi^+\right\rangle \\ &= \frac{\alpha}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) + \frac{\beta}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) \\ &= \frac{\alpha}{2}(|+\rangle|0\rangle - |-\rangle|1\rangle) + \frac{\beta}{2}(|-\rangle|1\rangle - |+\rangle|0\rangle) + \frac{\alpha}{2}(|-\rangle|0\rangle + |+\rangle|1\rangle) + \frac{\beta}{2}(|+\rangle|1\rangle - |-\rangle|0\rangle).\end{aligned} \tag{5.13}$$

This suggests after the measurement of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ with outcome $+1$, if we take measurement $\hat{\sigma}_x \otimes \hat{\sigma}_z$ at time t_2 then the probabilities of the two outcomes of $\hat{\sigma}_x \otimes \hat{\sigma}_z$ will be the same,

$$\Pr\left(\text{“}\hat{\sigma}_x \otimes \hat{\sigma}_z, +1, t_2\text{”} \middle| \, |\psi\rangle_{t>t_1}, I\right) = \Pr\left(\text{“}\hat{\sigma}_x \otimes \hat{\sigma}_z, -1, t_2\text{”} \middle| \, |\psi\rangle_{t>t_1}, I\right) = \frac{|\alpha|^2 + |\beta|^2}{2} = \frac{1}{2}. \tag{5.14}$$

The same situation happens if we obtain outcome -1 by measuring $\hat{\sigma}_x \otimes \hat{\sigma}_x$ first. More generally, if $\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}$ and $\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}$ do not commute, we will have the following relation:

$$\Pr\left(\text{“}\hat{\sigma}_{i_1} \otimes \hat{\sigma}_{j_1}, \lambda, t_2\text{”} \middle| \text{“}\hat{\sigma}_{i_2} \otimes \hat{\sigma}_{j_2}, \lambda', t_1\text{”}, h_{<t_1}, I\right) = \frac{1}{2} \quad \forall \lambda, \lambda' \in \{-1, +1\}, \tag{5.15}$$

where $h_{<t_1}$ denotes all the historical measurements performed before time t_1.

Both commuting and mutually unbiased global measurements yield very similar results, as revealed in the equiprobable relations of equations (5.11) and (5.15). The difference is that for mutually unbiased global measurements, the equiprobable relation (5.15) is valid for any prior measurement knowledge, while the two commuting operators may require a specific choice of an initial state and assume that no other measurements were performed before. In the following discussions, when the state of the system is not given, we tend to use the maximally mixed state as the initial state of the system. Under this initial state, both commuting and mutually unbiased measurements are non-informative; the result of one measurement cannot provide any information about the possible outcome of any subsequent measurement.

Property 4. In state tomography of a two-body qubit system, the outcome probability of a global measurement $\hat{\sigma}_i \otimes \hat{\sigma}_j$ can be determined by local statistics.

We can always decompose the local measurements as summations of projection operators:

$$\hat{\sigma}_i = \hat{P}_{i,1} - \hat{P}_{i,-1}, \hat{\sigma}_j = \hat{P}_{j,1} - \hat{P}_{j,-1}. \tag{5.16}$$

The global measurement $\hat{\sigma}_i \otimes \hat{\sigma}_j$ can also be decomposed into summation of projectors:

$$\hat{\sigma}_i \otimes \hat{\sigma}_j = (\hat{P}_{i,1} \otimes \hat{P}_{j,1} + \hat{P}_{i,-1} \otimes \hat{P}_{j,-1}) - (\hat{P}_{i,1} \otimes \hat{P}_{j,-1} + \hat{P}_{i,-1} \otimes \hat{P}_{j,1}). \tag{5.17}$$

According to Lüders' rule we may have the following relation:

$$\begin{aligned}\Pr\left(\text{“}\hat{\sigma}_i \otimes \hat{\sigma}_j, +1, t\text{”}|\hat{\rho}, I\right) = \Pr\left(\text{“}\hat{I} \otimes \hat{\sigma}_j, +1, t_2\text{”}|\text{“}\hat{\sigma}_i \otimes \hat{I}, +1, t_1\text{”}, \hat{\rho}, I\right) \Pr\left(\text{“}\hat{\sigma}_i \otimes \hat{I}, +1, t_1\text{”}|\hat{\rho}, I\right) + \\ \Pr\left(\text{“}\hat{I} \otimes \hat{\sigma}_j, -1, t_2\text{”}|\text{“}\hat{\sigma}_i \otimes \hat{I}, -1, t_1\text{”}, \hat{\rho}, I\right) \Pr\left(\text{“}\hat{\sigma}_i \otimes \hat{I}, -1, t_1\text{”}|\hat{\rho}, I\right).\end{aligned} \tag{5.18}$$

For any state $\hat{\rho}$ of a two-body qubit system, the probability of obtaining $+1$ for $\hat{\sigma}_i \otimes \hat{\sigma}_j$ is equal to the probability of obtaining the same outcome when two local measurements are performed separately on the same system. The time order of the two local measurements is not relevant. Similarly, the probability of obtaining -1 for $\hat{\sigma}_i \otimes \hat{\sigma}_j$ is equal to the probability that the two local measurements yield different outcomes.

5.2.2 Basic concepts and assumptions of question set structure

5.2.2.1 Single system

We will now abstract away from this quantum description and instead regard the physical system as a black box to which we can pose one of an infinite number of different binary questions. This black box can be represented as a set $\mathcal{Q}$ of questions. Each binary outcome question $Q_{\theta,\phi} \in \mathcal{Q}$ takes the form:

$$Q_{\theta,\phi} : \text{What's the result of projective measurement } \hat{\sigma}_{\theta,\phi}? \tag{5.19}$$

where

$$Q_{\theta,\phi} = 0(1) \text{ if result is up (down).} \tag{5.20}$$

The interrogations of these questions are formalized as propositions. We use the following convention to express an interrogation.

"$Q_{\theta,\phi}, q, t$" : Conduct an interrogation of $Q_{\theta,\phi}$ to the system at time t and obtain an outcome of q

The state of the system at some time t can then be regarded as the set of the outcome probabilities of all possible propositions in $\mathcal{Q}$ at time t.

In the case of binary outcomes, the two propositions, $Q_{\theta,\phi}, q, t$" and $\hat{\sigma}_{\theta,\phi}, \lambda, t$", are equivalent. We may generalize the notion of a question to correspond not just to a binary qubit but to an n-ary qunit. The qunit system may also be abstracted as a black box that contains many questions, and each question Q has an outcome q in the range of $0, 1, 2, \ldots, n-1$, which means the outcomes belong to the finite field $\mathbb{F}_n$.

Moreover, we may assume that after the interrogation of a question Q, and if we keep conducting the same interrogations, the results will be the same:

$$\Pr\left("Q, q', t_2" \,|\, "Q, q, t_1", h_{<t_1}, I\right) = \delta_{q,q'}, \tag{5.21}$$

where $h_{<t_1}$ denotes all the historical interrogations we conducted before time t_1 and I represents the basic structure of this quantum question system.

We abstract the system as a set $\mathbb{Q}$, which usually contains an infinite number of questions. On one hand, we don't want to deal with an infinite degree of freedom; on the other hand, our system is taking analogies from qunits where the state has a finite number of parameters. Similarly, we may assume the question structure also has a finite number of parameters.

Assumption 1. For a system represented as a question set $\mathcal{Q}$, there exists a maximal subset $\mathcal{Q}_M$ containing pairwise non-informative questions, such that the outcome probability distributions of all questions in $\mathcal{Q} \setminus \mathcal{Q}_M$ are determined by the outcome probability distributions of questions in $\mathcal{Q}_M$.

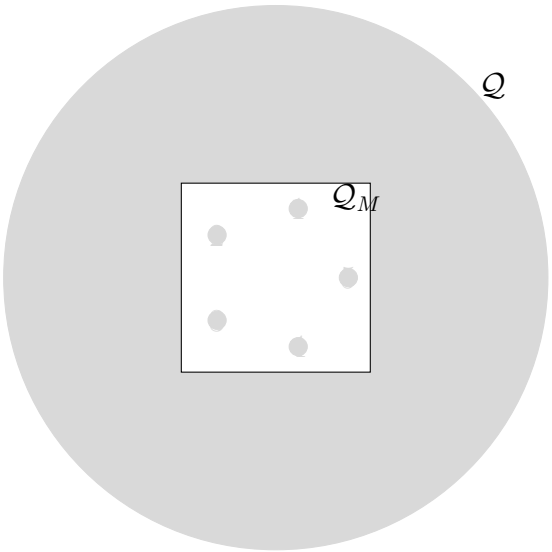

Figure 5.1: ***Question Set Representation of Quantum System*** A quantum system is abstracted as a set of questions, $\mathcal{Q}$, and it may contain an infinite number of questions. We assume that there is a finite collection of questions, $\mathcal{Q}_M$, such that the outcome probability distributions of the questions in $\mathcal{Q}_M$ determine all other outcome probabilities.

Two questions, Q_a and Q_b, are non-informative. This means that from the interrogation result of one question, we cannot obtain any information about the other question. Later, we will introduce a formal definition of information. In brief, non-informative means that if we know the interrogation result of one question and this is the only thing we know, then we cannot predict the possible result of the interrogation of the other question.

In the qubit case, the subset $\mathcal{Q}_M$ is analogous to Pauli matrices. For example, $\hat{\sigma}_x$ and $\hat{\sigma}_z$ are pairwise non-informative. Once we take a measurement of $\hat{\sigma}_x$, we cannot predict the exact outcome of $\hat{\sigma}_z$. The state of the qubit could be determined via the outcome probabilities of all three operators. Although, in principle, the state of a qubit could be determined via the outcome probabilities of Stern-Gerlach projective measurements along three different axes, the three mutually perpendicular axes are more special. The Pauli matrices are mutually unbiased, making certain calculations easier. We mimic this feature in the question structure as complementary questions.

Definition 5.2.1. Two n-outcome questions $Q_a, Q_b \in \mathcal{Q}$ are said to be *complementary* if

$$\begin{aligned} \Pr\left(\text{“}Q_a, q_a, t_2\text{”} \middle| \text{“}Q_b, q_b, t_1\text{”}, h_{<t_1}, I\right) &= \frac{1}{n} \quad \forall q_a, q_b \in \mathbb{F}_n, \\ \Pr\left(\text{“}Q_b, q_b, t_2\text{”} \middle| \text{“}Q_a, q_a, t_1\text{”}, h_{<t_1}, I\right) &= \frac{1}{n} \quad \forall q_a, q_b \in \mathbb{F}_n. \end{aligned} \tag{5.22}$$

In other words, no matter what interrogations have been conducted before t_1, the interrogation of one question will yield a uniform outcome probability distribution for the other question.

Another important relation between two projective measurements is commutativity. We can think of this feature as compatible questions expressed in terms of outcome probabilities.

Definition 5.2.2. Two questions $Q_a, Q_b \in \mathcal{Q}$ are said to be *compatible* if

$$\begin{aligned} \Pr\left(\text{“}Q_a, q'_a, t_3\text{”} \middle| \text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, h_{<t_1}, I\right) &= \delta_{q'_a, q_a}, \\ \Pr\left(\text{“}Q_b, q'_b, t_3\text{”} \middle| \text{“}Q_a, q_a, t_2\text{”}, \text{“}Q_b, q_b, t_1\text{”}, h_{<t_1}, I\right) &= \delta_{q'_b, q_b}. \end{aligned} \tag{5.23}$$

In other words, they don't affect each other's outcomes. For any system, we first ask question Q_a and obtain some outcome q_a. Subsequently, when we ask question Q_b and obtain some outcome q_b, if we continue to ask question Q_a, we will still get outcome q_a, and vice versa.

Compatibility is defined from an operational perspective, and while it's not exactly the same as the commutativity of operators in linear space, it is very similar. We will use compatibility as the analogy of commutativity in the following discussion.

5.2.2.2 Composite system

For composite systems, we can always regard them as a combination of individual subsystems. Moreover, we still treat the whole system as a set of questions, where each of them is a d-outcome question. It is natural to assume that the questions for a composite system contain questions from subsystems, as well as correlations between subsystems, which have a special form.

Assumption 2. Let $\mathcal{Q}_A$ and $\mathcal{Q}_B$ are the question sets of system A, B respectively. They

form a composite system with question set $\mathcal{Q}_{AB}$ such that

$$\mathcal{Q}_{AB} = \mathcal{Q}_A \cup \mathcal{Q}_B \cup \tilde{\mathcal{Q}}_{AB} \tag{5.24}$$

$\tilde{\mathcal{Q}}_{AB}$ contains composite questions in the form $\{Q_a *_1 Q_b, Q_a *_2 Q_b, \ldots, Q_{a'} *_1 Q_{b'}, Q_{a'} *_2 Q_{b'}, \ldots\}$, where $Q_a, Q_{a'}, \ldots \in \mathcal{Q}_A, Q_b, Q_{b'}, \ldots \in \mathcal{Q}_B$, and $*_1, *_2, \ldots$ are classical logical gates. Moreover, composite questions in the set $\tilde{\mathcal{Q}}_{AB}$ also have the same number of outcomes as questions in $\mathcal{Q}_A$ and $\mathcal{Q}_B$.

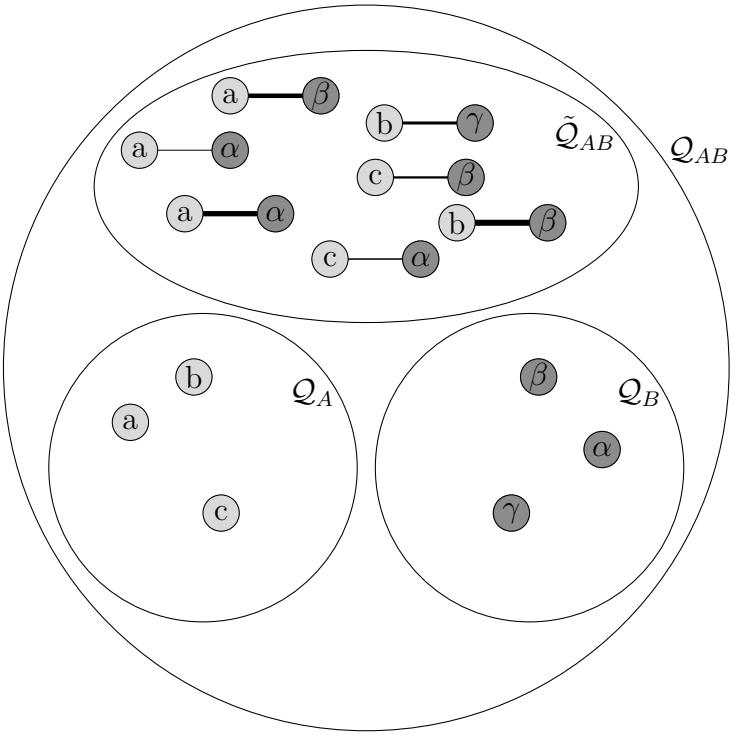

Figure 5.2: ***Question Set Structure in Composite System*** In a composite system composed of individual systems A and B, the question set $\mathcal{Q}_{AB}$ contains both individual questions from $\mathcal{Q}_A$ and $\mathcal{Q}_B$, as well as the composite questions between these two individual systems. Each composite question is in the form of $Q_a *_i Q_b$ where $*_i$ is a logical gate, and there could be different forms of logical gates. In the figure, the different thickness of lines that connect questions represent various logical gates.

The correlations are also questions with outcomes in the range of $\mathbb{F}_n$. A composite question $Q_a *_i Q_b$ represents a correlation between the questions Q_a and Q_b in subsystems. In the case where we know the exact outcomes of Q_a and Q_b at the same time, denoted as

q_a and q_b respectively, the outcome of $Q_a *_i Q_b$ is then uniquely determined by q_a and q_b.

In the qubit case, a composite question is analogous to the tensor product between two local measurements. For example, the global measurement $\hat{\sigma}_x \otimes \hat{\sigma}_z$ represents a correlation question between two local measurements $\hat{\sigma}_x \otimes \hat{I}$ and $\hat{I} \otimes \hat{\sigma}_z$. Once we know the exact outcomes of each local measurement at the same time, it means the system is in one of the eigenstates of those two local measurements. Notably, the eigenstate of the two local measurements is also the eigenstate of $\hat{\sigma}_x \otimes \hat{\sigma}_z$, making the outcome of this global measurement uniquely determined.

In the qubit case, the correspondence between composite questions and global measurements may seem trivial. The non-trivial aspect lies in the existence of different forms of logical gates, meaning there could be various binary functions in the form $f : \mathbb{F}_n \times \mathbb{F}_n \to \mathbb{F}_n$. The connection between composite questions with different logical gates and operators in quantum mechanics is not obvious. We will first discuss the possible formation of these logical gates and then explore their relationship with quantum mechanics.

Indeed, the degree of freedom of this composite system should be finite. The subset $\mathcal{Q}MAB$ that determines the composite system may have a structure similar to $\mathcal{Q}_{AB}$; it contains questions from subsystems as well as correlations.

Assumption 3. For a two-body composite system $\mathcal{Q}_{AB}$, the maximal subset $\mathcal{Q}_{M_{AB}}$ determines (the probability distribution of) outcomes of all questions in $\mathcal{Q}_{AB} \setminus \mathcal{Q}_{M_{AB}}$. $\mathcal{Q}_{M_{AB}}$ has the structure:

$$\mathcal{Q}_{M_{AB}} = \mathcal{Q}_{M_A} \cup \mathcal{Q}_{M_B} \cup \tilde{\mathcal{Q}}_{M_{AB}}, \tag{5.25}$$

where

$$\tilde{\mathcal{Q}}_{M_{AB}} = \{Q_a *_1 Q_b, Q_a *_2 Q_b, \cdots, Q_{a'} *_1 Q_{b'}, Q_{a'} *_2 Q_{b'}, \cdots\}, \tag{5.26}$$

with $Q_a, Q_{a'}, \cdots \in \mathcal{Q}_{M_A}, Q_b, Q_{b'}, \cdots \in \mathcal{Q}_{M_B}$. The questions in $\tilde{\mathcal{Q}}_{M_{AB}}$ are pairwise non-informative.

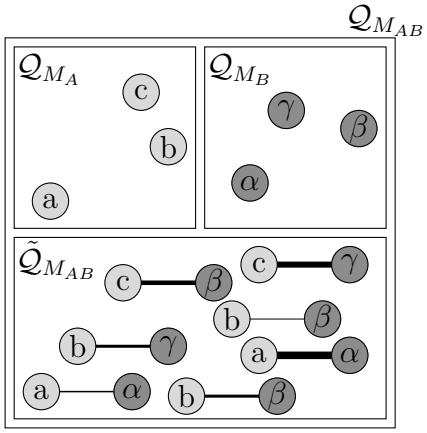

Figure 5.3: $\mathcal{Q}_M$ ***of Composite System*** We assume that the subset $\mathcal{Q}M$ of a composite system, which is formed by combining individual systems A and B, exhibits a structure akin to the question set $\mathcal{Q}AB$. $\mathcal{Q}M$ includes questions from both $\mathcal{Q}M_A$ and $\mathcal{Q}_{M_B}$, along with composite questions formed using various types of logical gates.

So far, we know little about the properties of these correlation questions and logical gates. We will first investigate the construction of these logical gates using the restriction that questions in $\mathcal{Q}_{M_{AB}}$ are pairwise non-informative. After that, we may propose information about questions and use assumptions on information to derive the detailed structure of the subset $\mathcal{Q}_M$ for both single systems and composite systems.

5.2.3 Restrictions of logical gates and generalizations in higher dimension

The key aspect to investigate in the structure of logical gates is the assumption that the questions in $\mathcal{Q}_{M_{AB}}$ are pairwise non-informative. Now, consider four questions in a composite system: $Q_a, Q_b, Q_a * iQ_b, Q_a * jQ_b \in \mathcal{Q}_{M_{AB}}$. They are pairwise non-informative. This implies there are two restrictions on the choices of logic gates:

Restriction 1. $Q_a *_i Q_b$ is non-informative to both Q_a, Q_b respectively;

Restriction 2. $Q_a *_i Q_b$ is non-informative to $Q_a *_j Q_b$ if $*_i$ and $*_j$ are different logical gates.

In the case of binary outcome systems, the choice of $*_i$ is unique (see Table 5.1). There are a total of 16 binary logical gates, while only XNOR or XOR satisfy the first restriction.

Although XNOR and XOR are different, they don't satisfy the second restriction. Notice that if we know the value of Q_a XNOR Q_b, we can immediately infer the value of Q_a XOR Q_b, and vice versa. Therefore, they are equivalent up to a negation operation.

Q_a	Q_b	Q_a AND Q_b
0	0	0
0	1	0
1	0	0
1	1	1

Q_a	Q_b	Q_a OR Q_b
0	0	0
0	1	1
1	0	1
1	1	1

Q_a	Q_b	Q_a XOR Q_b
0	0	0
0	1	1
1	0	1
1	1	0

Q_a	Q_b	Q_a XNOR Q_b
0	0	1
0	1	0
1	0	0
1	1	1

Table 5.1: ***Binary Logical Gates Exmaples*** The AND gate does not satisfy the first restriction, since if we know the result of Q_a AND Q_b is 1, then we immediately know both Q_a and Q_b must have outcome 1 which is not non-informative. The OR gate meets from a similar problem. Of the 16 possible two-input one-output logic gates, only XOR and XNOR satisfy the first restriction.

The logical gates in the n-ary case are more complex than binary logic gates. In binary case there is only one allowable logical gate which is XOR (or XNOR up to a negation). But, in the n-ary case, there could be more than one allowable logical gate. Table 5.2 shows an example of logical gates in the ternary case.

Q_a	Q_b	$Q_a \times Q_b$
0	0	0
2	1	2
1	2	2
2	0	0
1	1	1
0	2	0
1	0	0
0	1	0
2	2	1

Q_a	Q_b	$Q_a + Q_b$
0	0	0
2	1	0
1	2	0
1	0	1
0	1	1
2	2	1
2	0	2
1	1	2
0	2	2

Q_a	Q_b	$Q_a - Q_b$
0	0	0
1	1	0
2	2	0
1	0	1
2	1	1
0	2	1
2	0	2
0	1	2
1	2	2

Table 5.2: ***Ternary Logical Gates Exmaples*** Ternary multiplication does not satisfy restriction 1 since if we know the value of $Q_a \times Q_b$ is non-zero we immediately know the values of both Q_a and Q_b cannot be zero. While ternary addition and ternary subtraction satisfy both two restrictions.

In ternary case there are at least two different logical gates satisfy restriction 1. Before discussing different logical gates, we first exclude equivalent logical gates.

Q_a	Q_b	$Q_a + Q_b$
0	0	0
2	1	0
1	2	0
1	0	1
0	1	1
2	2	1
2	0	2
1	1	2
0	2	2

$\xrightarrow{0\to 2,1\to 0,2\to 1}$

Q_a	Q_b	$Q_a *_+ Q_b$
0	0	2
2	1	2
1	2	2
2	0	0
1	1	0
0	2	0
1	0	1
0	1	1
2	2	1

Table 5.3: ***Variant of Ternary Addition*** By knowing the value of $Q_a + Q_b$ we can immediately know the value of $Q_a *_+ Q_b$, and vice versa. These two logical gates are equivalent up to a permutation (021).

In the example shown in Table 5.3, if we know the value of $Q_a + Q_b$, then we know the value of $Q_a *_+ Q_b$, and vice versa. Just as in the binary case, where XNOR and XOR are equivalent up to a negation operation, the operations $Q_a + Q_b$ and $Q_a *_+ Q_b$ are also equivalent up to a permutation in the symmetric group S_3. This can be generalized to n-ary logical gates: if an n-ary logical gate satisfies restriction 1, then it is equivalent to $n! - 1$ variations, with each variation corresponding to a non-identity element in S_n.

Now, we can investigate different logical gates, and two problems need to be solved. First, what is the maximal number of different logical gates in the n-ary case? And how can we construct these different logical gates?

By excluding equivalent variations, the truth table of each logical gate can be written in the same pattern. As Table 5.4 shows, the truth table of a logical gate is divided into blocks, and in each block, the second column ranges orderly in $0, 1, \ldots, n-1$, and the first column remains constant.

Q_a	Q_b	$Q_a + Q_b$
0	0	0
0	1	1
0	2	2
1	0	1
1	1	2
1	2	0
2	0	2
2	1	0
2	2	1

Q_a	Q_b	$Q_a - Q_b$
0	0	0
0	1	2
0	2	1
1	0	1
1	1	0
1	2	2
2	0	2
2	1	1
2	2	0

Q_a	Q_b	$Q_a *_1 Q_b$
0	0	0
0	1	1
0	2	2
0	3	3
0	4	4
1	0	1
1	1	2
1	2	3
1	3	4
1	4	0
2	0	2
2	1	3
2	2	4
2	3	0
2	4	1
3	0	3
3	1	4
3	2	0
3	3	1
3	4	2
4	0	4
4	1	0
4	2	1
4	3	2
4	4	3

Q_a	Q_b	$Q_a *_2 Q_b$
0	0	0
0	1	2
0	2	4
0	3	1
0	4	3
1	0	1
1	1	3
1	2	0
1	3	2
1	4	4
2	0	2
2	1	4
2	2	1
2	3	3
2	4	0
3	0	3
3	1	0
3	2	2
3	3	4
3	4	1
4	0	4
4	1	1
4	2	3
4	3	0
4	4	2

Table 5.4: ***Example of Logical gates by Fixing Two Columns*** We rearrange the first two columns of a logical gate in the given order. For a n-ary logical gates, we divide the truth table into n different blocks. In the first column, each block contains only one number while in the second column each block contains numbers from 0 to $n-1$.

The benefit of this pattern is that we can combine different tables into one larger table. By doing so, the combined table becomes an orthogonal array, a concept well-investigated in mathematics. Table 5.5 provides an example of a combined table in the ternary case.

Q_a	Q_b	$Q_a + Q_b$
0	0	0
0	1	1
0	2	2
1	0	1
1	1	2
1	2	0
2	0	2
2	1	0
2	2	1

$\cup$

Q_a	Q_b	$Q_a - Q_b$
0	0	0
0	1	2
0	2	1
1	0	1
1	1	0
1	2	2
2	0	2
2	1	1
2	2	0

$\rightarrow$

Q_a	Q_b	$Q_a + Q_b$	$Q_a - Q_b$
0	0	0	0
0	1	1	2
0	2	2	1
1	0	1	1
1	1	2	0
1	2	0	2
2	0	2	2
2	1	0	1
2	2	1	0

Table 5.5: ***Logical Gate and Orthogonal Array*** The truth tables of ternary addition and ternary subtraction are both orthogonal arrays of 9 rows , 3 columns, level 3 and strength 2. Level 3 means there are 3 different elements. Strength 2 means it is a table of 9 rows and 3 columns and for every selection of 2 columns, all ordered 2-tuples of the elements appear exactly $\frac{row}{level^{strength}}$ times. These two tables can be combined into a larger orthogonal array with 4 columns.

The combined table of ternary addition and subtraction is an *orthogonal array* [49, 2]. An orthogonal array, denoted as $OA(N, k, s, t)$, is an array with N rows and k columns, where there are s different elements, and its strength is t. This means that every $N \times t$ subarray contains each t-tuple exactly λ times as a row, with $\lambda = N/s^t$. In this case, the orthogonal array is of size $3^2 \times 4$, with level 3 and strength 2.

As all possible logical gates can be combined into a single orthogonal array, the question of the maximal number of different logical gates can be rephrased as follows: What is the maximal number of columns in an orthogonal array with size n^2, level n, and strength 2? This problem is well-investigated when n is a power of a prime number but extremely difficult in other cases.

Fact 1. The maximal number of columns for an orthogonal array with n^2 rows, level n, and strength 2 is $n + 1$ if n is a prime power [2, p. 38].

This fact implies that when n is a prime power, there will be $n - 1$ different logical gates. The next problem is the construction of these $n - 1$ different logical gates. We have

found that, if n is a prime number, the $n-1$ logical gates, denoted as ${}_1, 2, \cdots, *n-1$, can be represented in this way:

$$Q_a *_i Q_b := Q_a + i \times Q_b \pmod{n} \quad \forall i \in \{1, 2, \cdots, n-1\}. \tag{5.27}$$

The example of the quinary case is shown in Table 5.6.

Q_a	Q_b	$Q_a + Q_b$
0	0	0
0	1	1
0	2	2
0	3	3
0	4	4
1	0	1
1	1	2
1	2	3
1	3	4
1	4	0
2	0	2
2	1	3
2	2	4
2	3	0
2	4	1
3	0	3
3	1	4
3	2	0
3	3	1
3	4	2
4	0	4
4	1	0
4	2	1
4	3	2
4	4	3

Q_a	Q_b	$Q_a + 2 \times Q_b$
0	0	0
0	1	2
0	2	4
0	3	1
0	4	3
1	0	1
1	1	3
1	2	0
1	3	2
1	4	4
2	0	2
2	1	4
2	2	1
2	3	3
2	4	0
3	0	3
3	1	0
3	2	2
3	3	4
3	4	1
4	0	4
4	1	1
4	2	3
4	3	0
4	4	2

Q_a	Q_b	$Q_a + 3 \times Q_b$
0	0	0
0	1	3
0	2	1
0	3	4
0	4	2
1	0	1
1	1	4
1	2	2
1	3	0
1	4	3
2	0	2
2	1	0
2	2	3
2	3	1
2	4	4
3	0	3
3	1	1
3	2	4
3	3	2
3	4	0
4	0	4
4	1	2
4	2	0
4	3	3
4	4	1

Q_a	Q_b	$Q_a + 4 \times Q_b$
0	0	0
0	1	4
0	2	3
0	3	2
0	4	1
1	0	1
1	1	0
1	2	4
1	3	3
1	4	2
2	0	2
2	1	1
2	2	0
2	3	4
2	4	3
3	0	3
3	1	2
3	2	1
3	3	0
3	4	4
4	0	4
4	1	3
4	2	2
4	3	1
4	4	0

Table 5.6: ***Four Different Quinary Logical Gates*** The colored numbers indicates how the four logical gates are non-informative to each other. If the value of $Q_a + Q_b$ is 0, there will be five different combinations of values of Q_a and Q_b, and each combination yields a different value in other gate.

Unfortunately, this construction fails when n is a prime power. Up to now, we have not found any elegant representations of logical gates when n is a prime power. Therefore, we shall henceforth focus on the case where all questions have prime number outcomes.

Once we obtain a relatively clear form of logical gates in a p-ary system, we immediately have the following consequence:

Corollary 1. In a two-body composite system $\mathcal{Q}_{AB}$, $Q_a, Q_b, Q_a * Q_b$ are mutually compatible, where $Q_a \in \mathcal{Q}_A, Q_b \in \mathcal{Q}_B$, and $*$ is any allowable logical gate. If the exact values of two of them are known, then the value of the remaining question will be ensured due to the specific form of logical gate $*$.

This result is an analogy of the commutativity between $\hat{\sigma}_i \otimes \hat{I}, \hat{I} \otimes \hat{\sigma}_j, \hat{\sigma}_i \otimes \hat{\sigma}_j$. If we take measurements of any two of the three operators, the state of the system will be ensured, and the possible outcome of the unmeasured operator can also be ensured.

In the binary case, since there is only one allowable logical gate, the relation between composite questions and composite operators seems natural when replacing $*$ with $\otimes$. In higher-order cases when we have more allowable logical gates, this correspondence may not be very clear. Later, we will introduce a general correspondence between composite questions and composite operators.

5.2.4 Information of questions and consequences

So far, we have obtained a nice property and expression of logical gates, specifically in prime number dimensional systems. However, this couldn't yield more information about the internal structure of $\mathcal{Q}_M$. Suppose we assume that $\mathcal{Q}_M$ contains a finite number of questions, but what is that number? In the following discussions, we will introduce a new concept: information of questions, to help construct the structure of quantum questions under informational postulates.

Like many existing information measures that are built upon probability distributions, we tend to define the information of a question based on its outcome probability. Given the background knowledge of the system, the historical interrogations $h_{<t}$ we have conducted on the system before time t, the outcome probability of a question Q at this moment t is denoted as $\Pr(\text{“}Q, q, t\text{”}|h_{<t}, I)$. For an n-outcome question, there will be n outcome probabilities.

The information of a question Q given the background knowledge $h_{<t}$ is a function of these outcome probabilities:

$$I(Q|h_{<t}) = H(\Pr(\text{“}Q, 0, t\text{”}|h_{<t}, I), \Pr(\text{“}Q, 1, t\text{”}|h_{<t}, I), \cdots, \Pr(\text{“}Q, n-1, t\text{”}|h_{<t}, I)). \tag{5.28}$$

H is a function of probability distributions and as a convention we set the following restrictions of information I and function H:

1. $0 \leq I(Q|h_{<t}) \leq 1$;
2. $H(\vec{p}) = 0$ if and only if $\vec{p} = (\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n})$;
3. $H(\vec{p}) = 1$ if and only if $\vec{p}$ is a n-tuple contains $n-1$ zeros and 1 one.

Instead of offering a formal measure of information, our approach focuses on considering two extreme cases concerning the information of questions. To find a detailed expression of this information, we would need more informational assumptions regarding the constraints of this measure. However, at the current stage, we cannot find more intuitive informational postulates. In fact, in the following calculations, the information of each question is either 0 or 1. The two extreme cases are already sufficient for us to demonstrate the subtle structure of quantum questions.

Now, we can define non-informative questions in terms of information. Two questions, Q_a and Q_b, are said to be pairwise non-informative if from the interrogation of one of them alone, we cannot obtain any information about the other question,

$$I(Q_a|\text{“}Q_b, q, t\text{”}) = 0, \quad I(Q_b|\text{“}Q_a, q, t\text{”}) = 0. \tag{5.29}$$

This definition lead to two subsequent consequences on complementary and compatible questions.

Corollary 2. If we perform interrogations on two complementary questions, the latter interrogation will erase the information about the question interrogated earlier.

Proof. Assume two questions Q_a and Q_b are pairwise complementary. If we take interrogation of question Q_a at time t_1 with outcome q_a, then after the interrogation we obtain 1 unit

information of Q_a and 0 unit information of Q_b:

$$\begin{aligned} &\Pr\left(\text{“}Q_a, q_a', t_2\text{”}\middle|\text{“}Q_a, q_a, t_1\text{”}, I\right) = \delta_{q_a', q_a}, \qquad && I(Q_a|\text{“}Q_a, q_a, t_1\text{”}, I) = 1; \\ &\Pr\left(\text{“}Q_b, q_b, t_2\text{”}\middle|\text{“}Q_a, q_a, t_1\text{”}, I\right) = \frac{1}{n}, && I(Q_b|\text{“}Q_a, q_a, t_1\text{”}, I) = 0. \end{aligned} \tag{5.30}$$

Then we take interrogation of question Q_b at time t_2 with outcome q_b. After t_2 we gain 1 unit information about question Q_b and 0 unit information about Q_a:

$$\begin{aligned} &\Pr\left(\text{“}Q_a, q_a', t_2\text{”}\middle|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I\right) = \frac{1}{n}, && I(Q_a|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I) = 0; \\ &\Pr\left(\text{“}Q_b, q_b', t_2\text{”}\middle|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I\right) = \delta_{q_b', q_b}, && I(Q_b|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I) = 1. \end{aligned} \tag{5.31}$$

This shows the interrogation “Q_b, q_b, t_2” erases the information we gain about Q_a at time t_1. Similar results will be yielded if we change the order of interrogations on Q_a and Q_b. □

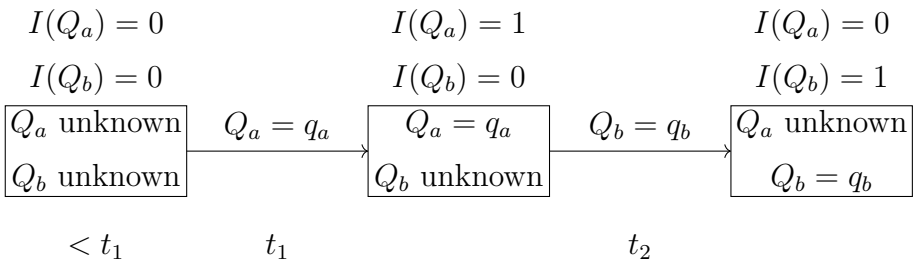

Figure 5.4: ***Mutually Complementary Questions in Interrogation*** Q_a and Q_b are mutually complementary. Initially, we have no knowledge about the state and two questions. At time t_1, we conduct an interrogation of Q_a with an outcome q_a. After this interrogation, we gain 1 unit of information about Q_a and cannot obtain any information about Q_b. At time t_2, we conduct another interrogation of Q_b with outcome q_b, and this interrogation may erase the information of Q_a.

Corollary 3. If we perform interrogations on two compatible questions, the later interrogation will retain the information about the question interrogated earlier.

Proof. For two compatible questions Q_a and Q_b, from the definition we may have the fol-

lowing relations:

$$\begin{aligned}\Pr\left(\text{“}Q_a, q'_a, t_3\text{”}\middle|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, h_{<t_1}, I\right) &= \delta_{q'_a,q_a},\\ \Pr\left(\text{“}Q_b, q'_b, t_3\text{”}\middle|\text{“}Q_a, q_a, t_2\text{”}, \text{“}Q_b, q_b, t_1\text{”}, h_{<t_1}, I\right) &= \delta_{q'_b,q_b}.\end{aligned} \tag{5.32}$$

This suggests that $I(Q_a|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, h_{<t_1}) = I(Q_b|\text{“}Q_a, q_a, t_2\text{”}, \text{“}Q_b, q_b, t_1\text{”}, h_{<t_1}) = 1$. If we have conducted an interrogation of a question, the latter interrogation of another compatible question will retain the information obtained in both interrogations. □

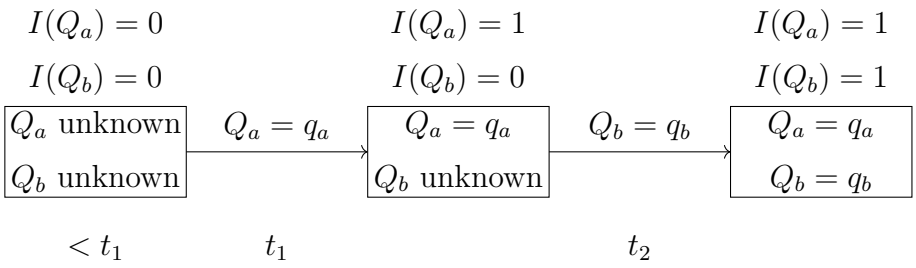

Figure 5.5: ***Mutually Compatible Questions in Interrogation*** Q_a and Q_b are mutually compatible. Initially, we have no knowledge about the state and two questions. At time t_1, we conduct an interrogation of Q_a with an outcome q_a. After this interrogation, we gain 1 unit information of Q_a and cannot obtain any information about Q_b. At time t_2, we conduct another interrogation of Q_b with outcome q_b, and information of Q_a is reserved.

Based on the information of questions, we can then try to define the information of the system. Since we assume that the outcome probabilities of questions in $\mathcal{Q}_M$ determine all other outcome probabilities of questions in the set $\mathcal{Q}$, it is intuitive to define the information of the system as a sum of the information of questions in $\mathcal{Q}_M$, say $\sum_{Q\in\mathcal{Q}_M} I(Q|h_{<t})$.

However, simply taking the summation over $\mathcal{Q}_M$ may not be very useful, especially when dealing with a composite system, and $\mathcal{Q}_M$ contains compatible questions. Interrogations on compatible questions will retain each other's information, and they may also 'derive' information about non-interrogated questions.

Consider a two-body qubit system where $\{\hat{\sigma}_x \otimes \hat{\sigma}_x, \hat{\sigma}_y \otimes \hat{\sigma}_y, \hat{\sigma}_z \otimes \hat{\sigma}_z\}$ is a set of pairwise commuting operators. Under certain situations, say $h_{\leq t_2} = \{\text{“}\hat{\sigma}_x \otimes \hat{\sigma}_x, +1, t_1\text{”}, \text{“}\hat{\sigma}_z \otimes \hat{\sigma}_z, +1, t_2\text{”}\}$, from these two interrogations we may have 1 unit information for both $\hat{\sigma}_x \otimes \hat{\sigma}_x$

and $\hat{\sigma}_z \otimes \hat{\sigma}_z$. Moreover, now the system is in the common eigenstate of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ and $\hat{\sigma}_z \otimes \hat{\sigma}_z$, which is just the Bell state $|\Psi^+\rangle$, and the outcome probabilities of $\hat{\sigma}_y \otimes \hat{\sigma}_y$ will be ensured, even if we haven't conducted an interrogation on it,

$$\Pr\left(\text{“}\hat{\sigma}_y \otimes \hat{\sigma}_y, 1, t_3\text{”}|h{\leq}\, t_2, I\right) = 1, \quad \Pr\left(\text{“}\hat{\sigma}_y \otimes \hat{\sigma}_y, 0, t_3\text{”}|h{\leq}\, t_2, I\right) = 0. \tag{5.33}$$

This suggests that the information of $\hat{\sigma}_y \otimes \hat{\sigma}_y$ is not independent, but can be derived from the results of the other two interrogations.

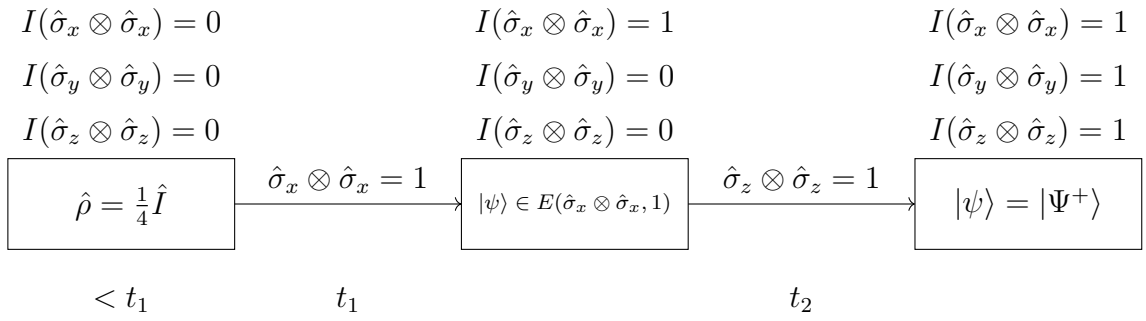

Figure 5.6: ***Information Acquisition in Joint Measurements*** Initially, we have no information about these joint measurements, and the initial state is taken as the maximal mixed state with no preference for any projection. At time t_1, a measurement of $\hat{\sigma}_x \otimes \hat{\sigma}_x$ is taken, projecting the system into the eigensubspace of $\hat{\sigma}_x \otimes \hat{\sigma}_x$. At time t_2, another measurement of $\hat{\sigma}_z \otimes \hat{\sigma}_z$ is conducted, ensuring that the system is in one of the four Bell states. All the information about the three joint measurements is obtained.

We want to focus on those non-derived information only, the information of the system will be sum of these independent information in the subset $\mathcal{Q}_M$:

$$I_{system}(h_{<t}) = \sum_{\substack{Q_i \in \mathcal{Q}_M \\ Q_i \ rel.\ ind.}} I(Q_i|h_{<t}). \tag{5.34}$$

Here Q_i *rel. ind.* denotes to all the questions such that the outcome probabilities cannot be derived from other questions in $\mathcal{Q}_M$. In the above example when $h_{<=t_2} = \{\text{“}\hat{\sigma}_x \otimes \hat{\sigma}_x, +1, t_1\text{”}, \text{“}\hat{\sigma}_z \otimes \hat{\sigma}_z, +1, t_2\text{”}\}$, then $\hat{\sigma}_y \otimes \hat{\sigma}_y$ will be excluded in the sum.

Assumption 4. For n-ary single system, the upper bound of information of system is 1

unit. For composite system composed of N subsystems, this upper bound is N units.

This assumption together with logical gates directly leads to the following corollaries:

Corollary 4. In single n-ary system, all questions in $\mathcal{Q}_M$ are mutually complementary.

Proof. Let $Q_a, Q_b \in \mathcal{Q}_M$, and we take interrogation on Q_a at time t_1 with outcome q_a and interrogation on Q_b at time t_2 with outcome q_b.

After t_2 we may obtain 1 unit information about question Q_b:

$$\Pr\left(\text{“}Q_b, q_b', t_3\text{”}|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I\right) = \delta_{q_b', q_b}, \quad I(Q_b|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I) = 1. \tag{5.35}$$

The information of the system is no more than 1 unit:

$$I_{system}(\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}) = I(Q_a|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}) + I(Q_b|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}) \leq 1. \tag{5.36}$$

This yields $I(Q_a|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}) = 0$ and
$\Pr\left(\text{“}Q_a, q_a', t_3\text{”}|\text{“}Q_b, q_b, t_2\text{”}, \text{“}Q_a, q_a, t_1\text{”}, I\right) = \frac{1}{n}, \forall q_a'$.

By applying the same argument, if we first take interrogation of Q_b and then take interrogation of Q_a we may have $\Pr\left(\text{“}Q_b, q_b', t_3\text{”}|\text{“}Q_a, q_a, t_2\text{”}, \text{“}Q_b, q_b, t_1\text{”}, I\right) = \frac{1}{n} \ \ \forall q_b'$.

The above procedures show that Q_a and Q_b are pairwise complementary, and choice of Q_a and Q_b are arbitrary. Therefore all questions in $\mathcal{Q}_M$ are mutually complementary. □

Corollary 5. Two composite questions $Q_a *_i Q_b, Q_{a'} *_j Q_{b'}$ are not compatible if $Q_a = Q_{a'}$ or $Q_b = Q_{b'}$, $*_i, *_j$ are any two allowable logical gates.

Proof. By contradiction, assume if $Q_a = Q_{a'}$ then $Q_a *_i Q_b, Q_a *_j Q_{b'}$ are compatible.

According to Assumption 4, the three questions $Q_a, Q_a *_i Q_b, Q_a *_j Q_{b'}$ are mutually compatible. This means we can make three interrogations on each of them and information on each question won't be lost.

The outcomes of $Q_a, Q_a *_i Q_b$ will yield the outcome of Q_b and outcomes of $Q_a, Q_a *_j Q_{b'}$ will yield outcome of $Q_{b'}$. Yet if $Q_b, Q_{b'} \in \mathcal{Q}_{M_B}$ then this violate Corollary 3 since we cannot know the outcomes of two complementary questions.

The same argument can be applied on the case that $Q_b = Q_{b'}$. □

Corollary 6. In two body p-ary system, there are at most $p+1$ mutually compatible composite questions, where every composite question is in the form $Q_a *_i Q_b$, $Q_a \in \mathcal{Q}_{M_A}, Q_b \in \mathcal{Q}_{M_B}, *_i$ is any allowable logical gate.

Proof. Assume there are k different mutually compatible composite questions, which are labeled as $Q_1 *_{i_1} Q_{j_1}, Q_2 *_{i_2} Q_{j_2}, \cdots, Q_k *_{i_k} Q_{j_k}$.

The compatibility of those questions means we may take k those different interrogations, say $h_{<=t_m} = \{\text{“}Q_m *_{i_m} Q_{j_m}, q_m, t_m\text{”}\}_{m=1}^{k}$, and after that information of each of the k different questions is retained.

Yet for two body system we can only obtain at most 2 units information of system. Of course it's possible to require $k \leq 2$ but it could be very trivial. We want to attain the maximal value of k. If $k > 2$, even if information of every question is retained, the information of system is still 2 units.

Assume we first take 2 interrogations, $\{\text{“}Q_2 *_{i_2} Q_{j_2}, q_2, t_2\text{”}, \text{“}Q_1 *_{i_1} Q_{j_1}, q_1, t_1\text{”}\}$, we will have 1 unit information for each of the question. Those two questions are independent to each other, this suggests that we must have 2 units information of the system. Yet if we take another $k-2$ consecutive interrogations, $\{\text{“}Q_n *_{i_m} Q_{j_m}, q_m, t_m\text{”}\}_{m=1}^{k}$, the information of the system is still 2 units, it won't violate the upper bound,

$$I_{system}(h_{<=t_m}) = I_{system}(\text{“}Q_2 *_{i_2} Q_{j_2}, q_2, t_2\text{”}, \text{“}Q_1 *_{i_1} Q_{j_1}, q_1, t_1\text{”}) = 2. \tag{5.37}$$

All the k questions are mutually compatible, and we could ensure their outcomes simultaneously. This means the remained $k-2$ interrogations are pre-determined, the outcomes of those $k-2$ interrogations must be determined from the outcomes of first 2 interrogations:

$$\{q_3, q_4, \cdots, q_k\} \text{ are determined by } \{q_1, q_2\}.$$

In other words, $\forall m \in \{3, 4, \cdots, k\}\ \exists f_m : \mathbb{F}_p \times \mathbb{F}_p \to \mathbb{F}_p$ s.t. $q_m = f_m(q_1, q_2)$. And there are two restrictions of function f_m:

1. q_m is independent with q_1 and q_m is independent with q_2;

2. q_m is independent with $q_{m'}$ if $m \neq m'$.

Those two restrictions come from the fact that $\{Q_m *_{i_m} Q_{j_m}\}_{m=1}^{k} \subset \mathcal{Q}_{M_{AB}}$ and they are mutually independent, so are their outcomes. The domain and image of f_m are both collection of discrete numbers and it is possible to write down a truth table of f_m. According to the discussion of logical gates above, in such a p-ary system, there are only at most $p-1$ such functions, which means $k-2 \leq p-1$. Therefore the maximal possible number of k is $p+1$. □

Assumption 5. In a two-body composite system, given two composite questions in the form $Q_i *_n Q_j$ and $Q_k *_m Q_l$, if $i \neq k$ and $j \neq l$, then for every logical gate $*_n$, there exists a unique logical gate $*_m$ such that $Q_i *_n Q_j$ and $Q_k *_m Q_l$ are compatible.

This assumption is not very intuitive. From the viewpoint of tomography, we may regard each composite question as a combination of two questions asked on individual systems, and different logical gates shouldn't be affected much. However, this assumption actually arises from a fact of correspondence in linear space. Later, when discussing the quantum mechanical correspondence of the quantum question structure, we will provide a detailed proof of that fact. With the help of this assumption and Corollary 6, we will derive an important result.

Theorem 5.1. *For single p-ary system, the size of $\mathcal{Q}_M$ is no more than $p+1$.*

Proof. By contradiction, assume size of $\mathcal{Q}_M$ is large than $p+1$, say equal to $p+2$.

Consider a two body p-ary system. According to Corollary 5, there are at most $p+1$ mutually compatible composite questions, and we may choose a set of them labeled as $Q_1 *_{i_1} Q_{j_1}, Q_2 *_{i_2} Q_{j_2}, \cdots, Q_{p+1} *_{i_{p+1}} Q_{j_{p+1}}$.

Let $Q_{j_{p+2}} \in \mathcal{Q}_M \setminus \{Q_{j_1}, Q_{j_2}, \cdots, Q_{j_{p+1}}\}$, then $Q_{p+2} *' Q_{j_{p+1}}$ is not in the collection of those $(p+1)$ mutually commuting operators for any logical gate $*'$.

From Assumption 6, for every $k \in \{1, 2, \cdots, p+1\}$, there exists a unique logical gate $*_{m_k}$ such that $Q_k *_{i_k} Q_{j_k}$ and $Q_{p+2} *_{m_k} Q_{j_{p+2}}$ are compatible.

Since there are at most $p-1$ different logical gates, this means there must be repetition among the collection of logical gates $\{*_{m_1}, *_{m_2}, \cdots, *_{m_{p+1}}\}$. Let the repetition logical gate be $*_{m_\alpha} = *_{m_\beta}$. Therefore $Q_\alpha *_{i_\alpha} Q_{j_\alpha}$ and $Q_\beta *_{i_\beta} Q_{j_\beta}$ are both compatible to $Q_{p+2} *_{m_\alpha} Q_{j_{p+2}}$.

Follow from the argument of Corollary 5, $Q_{p+2}*_{m_\alpha}Q_{j_{p+2}}$ must be a function of $Q_\alpha *_{i_\alpha} Q_{j_\alpha}$ and $Q_\beta *_{i_\beta} Q_{j_\beta}$. Yet in the collection $\{Q_1 *_{i_1} Q_{j_1}, Q_2 *_{i_2} Q_{j_2}, \cdots, Q_{p+1} *_{i_{p+1}} Q_{j_{p+1}}\}$ all operators other than $Q_\alpha *_{i_\alpha} Q_{j_\alpha}$ and $Q_\beta *_{i_\beta} Q_{j_\beta}$ are also different functions of $Q_\alpha *_{i_\alpha} Q_{j_\alpha}$ and $Q_\beta *_{i_\beta} Q_{j_\beta}$. This suggests that $Q_{p+2} *_{m_\alpha} Q_{j_{p+2}}$ will also be compatible with the collection of composite questions.

However, $Q_{p+2} \notin \{Q_1, Q_2, \cdots, Q_{p+1}\}$ and $Q_{j_{p+2}} \notin \{Q_{j_1}, Q_{j_2}, \cdots, Q_{j_{p+1}}\}$, $Q_{p+2}*_{m_\alpha}Q_{j_{p+2}}$ is different with any member of the collection $\{Q_l *_{i_l} Q_{j_l}\}_{l=1}^{p+1}$. Now there are $(p+2)$ different compatible composite questions, which contradicts with Corollary 5.

□

5.3 Correspondences in quantum mechanics

In the above discussions, the key concepts are compatible and complementary questions. The former corresponds to commuting operators while the latter are closely related to the concept of mutually unbiasedness [37, 21] in quantum mechanics.

Definition 5.3.1. Two non-degenerate operators $\hat{A}$ and $\hat{B}$ with d distinct eigenvalues are said to be *mutually unbiased* if there is a set of orthonormal eigenstates $\{|a_n\rangle\}$ of $\hat{A}$ and a set of orthonormal eigenstates $\{|b_n\rangle\}$ of $\hat{B}$ such that

$$|\langle a_i|b_j\rangle|^2 = \frac{1}{d} \qquad \forall i, j \in \mathbb{F}_d. \tag{5.38}$$

The mutually unbiasedness between two non-degenerate operators is in fact determined by their eigenstates. The concept of mutually unbiased operators can be extended to degenerate operators where the two set of eigensubspaces are mutually unbiased.

Definition 5.3.2. Two degenerate operators $\hat{A}$ and $\hat{B}$ with d distinct eigenvalues $\{\lambda_{Ai}\}_{i=1}^d$, $\{\lambda_{Bi}\}_{i=1}^d$ are said to be *mutually unbiased* if

$$|\langle a_i|b_j\rangle|^2 = \frac{1}{d} \qquad \forall i, j \in \mathbb{F}_d \quad |a_i\rangle \in E(\lambda_{Ai}, \hat{A}) \quad |b_j\rangle \in E(\lambda_{Bj}, \hat{B}). \tag{5.39}$$

Definition 5.3.3. In $\mathbb{C}^d$, two orthonormal bases $\{|a_n\rangle\}, \{|b_n\rangle\} (n \in \{0, 1, 2, \cdots, d-1\})$ are

said to be *mutually unbiased bases* if

$$|\langle a_i|b_j\rangle|^2 = \frac{1}{d} \quad \forall i,j \in \mathbb{F}_d. \tag{5.40}$$

In $\mathbb{C}^d$, we can always find a set of orthonormal basis $\{|i\rangle\}(i \in \{0,1,2,\cdots,d-1\})$ as *computational basis*. Another set of orthonormal basis $\{\left|\tilde{j}\right\rangle\}(j \in \{0,1,2,\cdots,d-1\})$ can be defined by quantum discrete Fourier transformation:

$$\left|\tilde{j}\right\rangle = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \omega_d^{-kj} |k\rangle \qquad \omega_d = e^{2i\pi/d}, \tag{5.41}$$

where $\{|i\rangle\}$ and $\{\left|\tilde{j}\right\rangle\}$ are unbiased since $\left\langle i\middle|\tilde{j}\right\rangle = \frac{1}{\sqrt{d}}\omega_d^{-ij}$.

Based on those two set of mutually unbiased bases, we can introduce *generalized Pauli matrix* $\hat{X}$ and $\hat{Z}$:

$$\hat{X}\left|\tilde{j}\right\rangle = \omega_d^j \left|\tilde{j}\right\rangle, \qquad \hat{Z}|i\rangle = \omega_d^i |i\rangle. \tag{5.42}$$

From the definition, it follows that $\hat{X}, \hat{Z}$ have the following important properties:

1. $\hat{X}^d = \hat{Z}^d = \hat{I}$;
2. $\hat{X}|i\rangle = |i+1\rangle$, $\left\langle\tilde{j}\right|\hat{Z} = \left\langle\widetilde{j+1}\right|$;
3. $\hat{Z}\hat{X} = \omega_d \hat{X}\hat{Z}$ (Weyl commutation relation).

$$\hat{X} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \hat{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \omega_5 & 0 & 0 & 0 \\ 0 & 0 & \omega_5^2 & 0 & 0 \\ 0 & 0 & 0 & \omega_5^3 & 0 \\ 0 & 0 & 0 & 0 & \omega_5^4 \end{bmatrix}$$

Table 5.7: ***Matrix Representation of $\hat{X}, \hat{Z}$ in the Computational Basis of Dimension 5***

Fact 2. In $\mathbb{C}^d$, there are at least three mutually unbiased bases which are the eigenstates of $\{\hat{X}, \hat{Z}, \hat{X}\hat{Z}\}$. If d is a prime number, $\mathbb{C}^d$ has the maximal number of of MUBs, which are

the eigenstates of $\{\hat{X}, \hat{Z}, \hat{X}\hat{Z}, \hat{X}\hat{Z}^2, \cdots, \hat{X}\hat{Z}^{d-1}\}$[21]. The eigenstate of $\hat{X}\hat{Z}^k$ is expressed as:

$$\left|e_k^j\right\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} \omega^{-ij} \omega^{ki(i-1)/2} \left|i\right\rangle, \tag{5.43}$$

where $|i\rangle$ is the eigenstate of $\hat{Z}$ and $\hat{X}\hat{Z}^k \left|e_k^j\right\rangle = \omega^j \left|e_k^j\right\rangle$.

5.4 Connections between question set structure and quantum mechanics

5.4.1 Single system

5.4.1.1 Relations between p-ary question set structure and quantum mechanics

In the following, we list some connections that we have established between quantum mechanics in $\mathbb{C}^p$ space and a system represented as set of p-ary questions.

p-ary question question structure	Quantum mechanics in $\mathbb{C}^p$
Question Q_a with p different outcomes	Unitary operator $\hat{U}_a$ with p different eigenvalues
Questions in $\mathcal{Q}_M$ are mutually complementary	Bases of corresponding operators in $\mathcal{Q}_M$ are MUBs
The probabilities of set $\mathcal{Q}_M$ determines state of system	Probabilities of projections of MUBs determines density matrix

Table 5.8: ***Comparison between Question Set Structure and Quantum Mechanics on Single System***

1. Each question A in the question set $\mathcal{Q}$ has p different outcomes, which are $0, 1, 2, \ldots, p-1$. Question A corresponds to a unitary operator $\hat{A}$ in $\mathbb{C}^p$ with p distinct eigenvalues $\omega_p^0, \omega_p^1, \omega_p^2, \ldots, \omega_p^{p-1}$. While $\hat{A}$ is not Hermitian, we can always decompose it in terms of its eigenstate projectors, such as $\hat{A} = \sum_i \omega^{i-1} \left|e_i\right\rangle \left\langle e_i\right|$. Every interrogation of question Q_a corresponds to a collection of p projections, with each projection onto an eigenstate of $\hat{A}$. Of course, we can find a set of p real values $\lambda_0, \lambda_1, \ldots, \lambda_{p-1}$ to create a Hermitian operator, denoted as $\hat{A}_{\text{Her.}} = \sum_i \lambda_i \left|e_i\right\rangle \left\langle e_i\right|$, where $\lambda_i \in \mathbb{R}$. This makes it

more natural to connect the interrogation of Q_a to the measurement of $\hat{A}_{\text{Her.}}$. However, such a $\hat{A}_{\text{Her.}}$ is not convenient for the following calculations, especially for a composite system. Therefore, we will continue to use the interrogation-projection connection.

2. The set $\mathcal{Q}_M$ corresponds to a maximal set of Mutually Unbiased Bases (MUBs) of $\mathbb{C}^p$. For each set of bases e_n in a maximal set of MUBs, we can define a unitary or Hermitian operator based on e_n. In fact, the generalized Pauli matrices introduced above are used to label different bases in a set of MUBs.

5.4.1.2 Example of information changed on different interrogations

In a single p-ary system, there are $p+1$ questions in $\mathcal{Q}_M$, say $\mathcal{Q}_M = \{Q_1, Q_2, \cdots, Q_{p+1}\}$. Their corresponding operators are just the generalized Pauli matrices in $\mathbb{C}^p$, $\{\hat{X}, \hat{Z}, \hat{X}\hat{Z}, \hat{X}\hat{Z}^2, \cdots, \hat{X}\hat{Z}^{p-1}\}$.

Quantum Question Scenario

Assume we are facing an unknown system, all the knowledge we have is that this is a p-ary outcome system. In this situation we may initialize the system being the state such that all the outcome probabilities of those questions are the same,

$$\Pr\left(\text{“}Q_i, q_i, t_1\text{”}|h_{<t_1} = \emptyset, I\right) = \frac{1}{p} \quad \forall Q_i \in \mathcal{Q}_M \, \forall q_i \in \mathbb{F}_p, \tag{5.44}$$

where $h_{<t_1} = \emptyset$ denotes we know nothing about the system before time t_1.

In other words, the information of any question is zero unit, the information of system is also zero,

$$I(Q_i|h_{<t_1}) = 0 \;\; \forall Q_i \in \mathcal{Q}_M, \qquad I_{system}(h_{<t_1}) = 0. \tag{5.45}$$

At time t_1 we conduct an interrogation of Q_1 and obtain an outcome m. After this interrogation we obtain 1 unit information about Q_1:

$$\Pr\left(\text{“}Q_1, m', t_2\text{”}|\text{“}Q_1, m, t_1\text{”}, h_{<t_1}, I\right) = \delta_{m,m'} \quad I(Q_1|\text{“}Q_1, m, t_1\text{”}, h_{<t_1}) = 1. \tag{5.46}$$

All questions in $\mathcal{Q}_M$ are mutually complementary, hence the outcome probabilities of all

other questions are remained as uniform distributions,

$$\begin{aligned} \Pr\left(\text{“}Q_i, q_i, t_2\text{”}\middle|\text{“}Q_1, m, t_1\text{”}, I\right) &= \frac{1}{p} \quad \forall Q_i \in \mathcal{Q}_M \; Q_i \neq Q_1 \; \forall q_i \in \mathbb{F}_p, \\ I(Q_i|\text{“}Q_1, m, t_1\text{”}) &= 0 \quad \forall Q_i \in \mathcal{Q}_M \; Q_i \neq Q_1. \end{aligned} \tag{5.47}$$

The information of all other questions are just zero, when calculating the information of the system, we needn't to exclude any questions since there is only one non-zero term. The information of system is just one unit:

$$I_{system}(\text{“}Q_1, m, t_1\text{”}, h_{<t_1}) = \sum_{Q_i \in \mathcal{Q}_M} I(Q_i|\text{“}Q_1, m, t_1\text{”}, h_{<t_1}) = 1. \tag{5.48}$$

At time t_2 we take another interrogation of Q_2 with outcome n. After this interrogation we obtain 1 unit information about Q_2:

$$\begin{aligned} \Pr\left(\text{“}Q_2, n', t_3\text{”}\middle|\text{“}Q_2, n, t_2\text{”}, \text{“}Q_1, m, t_1\text{”}, h_{<t_1}, I\right) &= \delta_{n,n'} \\ I(Q_2|\text{“}Q_2, n, t_2\text{”}, \text{“}Q_1, m, t_1\text{”}, h_{<t_1}, I) &= 1. \end{aligned} \tag{5.49}$$

The information of Q_1 is lost according to the second interrogation, and information of all other questions are still zero:

$$\begin{aligned} \Pr\left(\text{“}Q_i, q_i, t_3\text{”}\middle|\text{“}Q_2, n, t_2\text{”}, \text{“}Q_1, m, t_1\text{”}, h_{<t_1}, I\right) &= \frac{1}{p} \quad \forall Q_i \in \mathcal{Q}_M \; s.t. \; Q_i \neq Q_2 \; \forall q_i \in \mathbb{F}_p, \\ I(Q_i|\text{“}Q_2, n, t_2\text{”}, \text{“}Q_1, m, t_1\text{”}, h_{<t_1}) &= 0 \quad \forall Q_i \in \mathcal{Q}_M \; s.t. \; Q_i \neq Q_2. \end{aligned} \tag{5.50}$$

Similarly the information of the system is still 1 unit, it didn't exceed the upper bound,

$$I_{system}(\text{“}Q_2, n, t_2\text{”}, \text{“}Q_1, m, t_1\text{”}, h_{<t_1}) = \sum_{Q_i \in \mathcal{Q}_M} I(Q_i|\text{“}Q_2, n, t_2\text{”}, \text{“}Q_1, m, t_1\text{”}, h_{<t_1}) = 1 \tag{5.51}$$

Quantum Mechanics Scenario

At time t_1, we take measurement $\hat{X}$ on a single p-dimensional system with an outcome ω_p^m. After this measurement, the state of the system is ensured, which is the eigenstate

of $\hat{X}$:

$$|\psi\rangle_{t>t_1} = |\tilde{m}\rangle\,. \tag{5.52}$$

We gain 1 unit information about $\hat{X}$ and zero information about all other measurements,

$$\Pr\left(\text{“}\hat{X}, \omega_p^{m'}, t_2\text{”}\middle|\text{“}\hat{X}, \omega_p^{m}, t_1\text{”}, I\right) = \delta_{m,m'}, \quad I(\hat{X}|\text{“}\hat{X}, \omega_p^{m}, t_1\text{”}) = 1. \tag{5.53}$$

The information of the system is 1 unit, achieving the upper bound. As expected, when information of the system achieving the upper bound we could ensure the state.

At time t_2, we take another measurement $\hat{Z}$ with an outcome ω_p^n. After the second measurement, the state of the system is now changed to the eigenstate of $\hat{Z}$:

$$|\psi\rangle_{t>t_2} = |n\rangle\,. \tag{5.54}$$

We now gain 1 unit information about $\hat{Z}$ and the information of $\hat{X}$ is lost. Information of all other measurements in $\mathcal{Q}_M$ remain the same. The information of the system is still 1 unit.

5.4.1.3 Interpretation of the size of $\mathcal{Q}_M$

The density matrix of a quantum system, which lies in $\mathbb{C}^p$, has a degree of freedom of $p^2 - 1$ due to hermiticity and normalization. In the question set structure, every question in $\mathcal{Q}_M$ can be represented as a p-tuple probability distribution, resulting in every question having $p - 1$ degrees of freedom. As deduced above, for a single p-ary question set system, there are a total of $p + 1$ mutually independent questions, and the total degree of freedom of $\mathcal{Q}_M$ will be exactly $(p-1)(p+1) = p^2 - 1$, which is equal to the degree of freedom of the density matrix of a p-dimensional system. This shouldn't be surprising, as each value in a p-tuple probability distribution corresponds to the probability of a projection.

This is consistent with our assumption that $\mathcal{Q}_M$ contains the smallest number of p-outcome questions that determine the state of the system, as well as the probability distributions of questions in the set $\mathcal{Q} \setminus \mathcal{Q}_M$.

5.4.2 Two-body composite system

5.4.2.1 Additional relations for composite systems

Based on the connections above and the discussion of logical gates for a p-ary question set structure, we propose a set of connections between a two-body p-ary system under the question set structure and quantum mechanics in $\mathbb{C}^p \otimes \mathbb{C}^p$.

p-ary two-body system in quantum question structure	Quantum mechanics in $\mathbb{C}^p \otimes \mathbb{C}^p$
Question Q_a with p different outcomes	Unitary operator $\hat{U}_a$ with p different eigenvalues
The probabilities of set $\mathcal{Q}_M$ determines state of system	Probabilities of projections of MUBs determines density matrix
Composite question is in the form of $Q_a *_i Q_b$	Composite operator in the form of $\hat{U}_a \otimes \hat{U}_b^i$

Table 5.9: ***Comparison between Question Set Structure and Quantum Mechanics on Two-body System***

1. For a two-body composite system, composite question $Q_a *_i Q_b$ is related to the composite operator $\hat{U}_a \otimes \hat{U}_b^i$ in $\mathcal{L}(\mathbb{C}^p \otimes \mathbb{C}^p)$. $\hat{U}_a \otimes \hat{U}_b^i$ has p distinct eigenvalues $\{\omega_p^0, \omega_p^1, \cdots, \omega_p^{p-1}\}$, each eigenvalue has degeneracy p.

2. $\mathcal{Q}_M$ for composite system contains both complementary and compatible questions. Two questions are compatible if and only if their corresponding unitary operators commute.

5.4.2.2 Example of information change when interrogating compatible questions

Here is an example of interrogations on composite compatible questions on 5-dimension, consider a family of commuting operators, $\{\hat{X} \otimes \hat{X}, \hat{Z} \otimes \hat{Z}^4, \hat{X}\hat{Z} \otimes \hat{X}\hat{Z}^4, \hat{X}\hat{Z}^2 \otimes \hat{X}\hat{Z}^3, \hat{X}\hat{Z}^3 \otimes \hat{X}\hat{Z}^2, \hat{X}\hat{Z}^4 \otimes \hat{X}\hat{Z}\}$, and label them in the form of composite questions, $\{Q_1 *_1 Q_1, Q_2 *_4 Q_2, Q_3 *_1 Q_6, Q_4 *_1 Q_5, Q_5 *_1 Q_4, Q_6 *_1 Q_3\}$.

Quantum Question Scenario

At time t_1 if we take an interrogation of question $Q_1 *_1 Q_1$ and obtain an outcome m, we could immediately write down the following two outcome distributions:

$$\begin{aligned}
&\Pr\left(\text{“}Q_1 *_1 Q_1, m', t_2\text{”}\middle|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}, I\right) = \delta_{m,m'}, \\
&\Pr\left(\text{“}Q, q, t_2\text{”}\middle|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}, I\right) = \frac{1}{5} \quad \forall Q \neq Q_1 *_1 Q_1 \quad \forall q \in \mathbb{F}_5.
\end{aligned} \tag{5.55}$$

This suggests that we gain 1 unit information about $Q_1 *_1 Q_1$,

$$I(Q_1 *_1 Q_1|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}) = 1, \tag{5.56}$$

but 0 unit information of all other questions,

$$I(Q|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}) = 0 \quad \forall Q \neq Q_1 *_1 Q_1. \tag{5.57}$$

The information of the system is 1 unit:

$$\begin{aligned}
I_{system}(\text{“}Q_1 *_1 Q_1, m, t_1\text{”}) &= \sum_{Q \in \mathcal{Q}_M} I(Q|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}) \\
&= I(Q_1 *_1 Q_1|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}) + \sum_{\substack{Q \in \mathcal{Q}_M \\ Q \neq Q_1 *_1 Q_1}} I(Q|\text{“}Q_1 *_1 Q_1, m, t_1\text{”}) \\
&= 1 + 0 + 0 + \cdots = 1.
\end{aligned} \tag{5.58}$$

Here when calculating the information of the system, we could sum the information of all questions in $\mathcal{Q}_M$ without excluding any questions. The only non-zero information we obtain is on question $Q_1 *_1 Q_1$, and all other questions are independent of it, nothing else could be derived from the outcome of $Q_1 *_1 Q_1$.

We then take an interrogation of question $Q_2 *_4 Q_2$ at time t_2 and obtain an outcome n,

$$\Pr\left(\text{“}Q_2 *_4 Q_2, n', t_3\text{”}\middle|\text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}, I\right) = \delta_{n,n'}. \tag{5.59}$$

After the second interrogation, we got 1 unit information about $Q_2 *_4 Q_2$:

$$I(Q_2 *_4 Q_2|\text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}) = 1. \tag{5.60}$$

The information of $Q_1 *_1 Q_1$ is not "lost", since $Q_1 *_1 Q_1$ is compatible with $Q_2 *_4 Q_2$,

$$\begin{aligned} \Pr(\text{“}Q_1 *_1 Q_1, m', t_3\text{”} | \text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}, I) = \delta_{m,m'}, \\ I(Q_1 *_1 Q_1 | \text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}) = 1. \end{aligned} \tag{5.61}$$

The information of $Q_1 *_1 Q_1$ and $Q_2 *_4 Q_2$ are independent to each other, the outcomes of both questions are not related. Therefore the information of the system must contain at least these two questions:

$$\begin{aligned} I_{system}(\text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}) \geq I(Q_1 *_1 Q_1 | \text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}) + \\ I(Q_2 *_4 Q_2 | \text{“}Q_2 *_4 Q_2, n, t_2\text{”}, \text{“}Q_1 *_1 Q_1, m, t_1\text{”}) = 2. \end{aligned} \tag{5.62}$$

The information of the system hits the upper bound, 2 units. Even if we can still gain information about other four compatible composite questions, $\{Q_3 *_1 Q_6, Q_4 *_1 Q_5, Q_5 *_1 Q_4, Q_6 *_1 Q_3\}$, they are not independent and can be derived from the information of $Q_1 *_1 Q_1$ and $Q_2 *_4 Q_2$. In fact from Corollary 5, the outcomes of the other four compatible composite questions must be a function of outcomes of $Q_1 *_1 Q_1$ and $Q_2 *_4 Q_2$:

$$\begin{aligned} Q_3 *_1 Q_6 =& (Q_1 *_1 Q_1) *_1 (Q_2 *_4 Q_2) = m *_1 n = m + n, \\ Q_4 *_1 Q_5 =& (Q_1 *_1 Q_1) *_2 (Q_2 *_4 Q_2) = m *_2 n = m + 2n, \\ Q_5 *_1 Q_4 =& (Q_1 *_1 Q_1) *_3 (Q_2 *_4 Q_2) = m *_3 n = m + 3n, \\ Q_6 *_1 Q_3 =& (Q_1 *_1 Q_1) *_4 (Q_2 *_4 Q_2) = m *_4 n = m + 4n. \end{aligned} \tag{5.63}$$

Quantum Mechanics Scenario

At time t_1, we take measurement of $\hat{X} \otimes \hat{X}$ and obtain an outcome ω_5^m. After this measurement, we cannot write down the state of system, since $\hat{X} \otimes \hat{X}$ is degenerate and all we can know is that the state of system lies in the eigensubspace of $\hat{X} \otimes \hat{X}$. The outcome probability of all other measurements remains unknown,

$$|\psi\rangle_{t>t_1} \in E(\omega_5^m, \hat{X} \otimes \hat{X}). \tag{5.64}$$

We gain 1 unit information about $\hat{X} \otimes \hat{X}$, but 0 unit information of all other measure-

ments. The information of the system is 1 unit.

We then take measurement $\hat{Z} \otimes \hat{Z}^4$ with outcome ω_5^n at time t_2. Based on these two measurements, we can then write down the state of the system, which is a maximally entangled state:

$$\begin{aligned} |\psi\rangle_{t>t_2} = \frac{1}{\sqrt{5}} \Bigg(|0\rangle \left|\frac{m}{4}\right\rangle + \omega_5^{4n} |1\rangle \left|\frac{m-1}{4}\right\rangle + \omega_5^{3n} |2\rangle \left|\frac{m-2}{4}\right\rangle \\ + \omega_5^{2n} |3\rangle \left|\frac{m-3}{4}\right\rangle + \omega_5^{n} |4\rangle \left|\frac{m-4}{4}\right\rangle \Bigg) . \end{aligned} \tag{5.65}$$

After the second interrogation, we got 1 unit information about $\hat{Z} \otimes \hat{Z}^4$. The information of $\hat{X} \otimes \hat{X}$ is retained, since $\hat{X} \otimes \hat{X}$ and $\hat{Z} \otimes \hat{Z}^4$ commute. The outcome probability of the other four commuting composite operators, $\{\hat{X}\hat{Z} \otimes \hat{X}\hat{Z}^4, \hat{X}\hat{Z}^2 \otimes \hat{X}\hat{Z}^3, \hat{X}\hat{Z}^3 \otimes \hat{X}\hat{Z}^2, \hat{X}\hat{Z}^4 \otimes \hat{X}\hat{Z}\}$, are ensured:

$$\begin{aligned} \hat{X}\hat{Z} \otimes \hat{X}\hat{Z}^4 |\psi\rangle_{t>t_2} &= \omega_5^{m+n} |\psi\rangle , \\ \hat{X}\hat{Z}^2 \otimes \hat{X}\hat{Z}^3 |\psi\rangle_{t>t_2} &= \omega_5^{m+2n} |\psi\rangle , \\ \hat{X}\hat{Z}^3 \otimes \hat{X}\hat{Z}^2 |\psi\rangle_{t>t_2} &= \omega_5^{m+3n} |\psi\rangle , \\ \hat{X}\hat{Z}^4 \otimes \hat{X}\hat{Z} |\psi\rangle_{t>t_2} &= \omega_5^{m+4n} |\psi\rangle . \end{aligned} \tag{5.66}$$

This means the information of those four operators are derived from the information of $\hat{X} \otimes \hat{X}$ and $\hat{Z} \otimes \hat{Z}^4$. When calculating the information of the system, we need to exclude those four operators. Moreover, based on this state, the outcome probability of all other non-commuting composite operators will be just the uniform distribution. The information of the system will be 2 units, as expected, since the state of the system is already ensured.

5.4.2.3 Size of $\mathcal{Q}_M$ and degrees of freedom of density matrix

For a single system, we find that the degree of freedom of $\mathcal{Q}_M$ is equal to the degree of freedom of the density matrix. In fact, this relation also holds for a two-body composite system. Since all composite questions also have p outcomes, every question in $\mathcal{Q}_M$ can be represented as a p-tuple probability distribution with $p-1$ degrees of freedom.

As for the size of $\mathcal{Q}_M$, consider a composite system $\mathcal{Q}_{AB}$ that contains two individual system $\mathbb{Q}_A$ and $\mathbb{Q}_B$, the number of elements in $\mathcal{Q}_{M_{AB}}$ can be derived from its structure.

Recall that from assumption 2, $\mathcal{Q}_{M_{AB}} = \mathcal{Q}_{M_A} \cup \mathcal{Q}_{M_B} \cup \tilde{\mathcal{Q}}_{M_{AB}}$ and $\tilde{\mathcal{Q}}_{M_{AB}} = \{Q_a *_i Q_b | Q_a \in \mathcal{Q}_{M_A}, Q_b \in \mathcal{Q}_{M_B}, i \in \mathbb{F}_p^*\}$, therefore we could calculate the cardinality of $\mathcal{Q}_{M_{AB}}$:

$$\begin{aligned} |\mathcal{Q}_{M_{AB}}| &= |\mathcal{Q}_{M_A}| + |\mathcal{Q}_{M_B}| + \left|\tilde{\mathcal{Q}}_{M_{AB}}\right| \\ &= (p+1) + (p+1) + (p+1)(p+1)(p-1) \\ &= (p+1)(p^2+1). \end{aligned} \tag{5.67}$$

The total degree of freedom of $\mathcal{Q}_{M_{AB}}$ is now equal to the product of $|\mathcal{Q}_{M_{AB}}|$ and $p-1$, which is equal to $p^4 - 1$, the same number of degrees of freedom as the density matrix of a two-body p-dimensional system in quantum mechanics.

The agreement of degrees of freedom can be generalized to an N-body p-ary/p-dimensional system. In this situation, $\mathcal{Q}_M$ will contain composite questions from single systems, two-body systems, and so on, up to N-body systems. The total number of questions in $\mathcal{Q}_M$ is given by:

$$\begin{aligned} &\binom{N}{1}(p+1) + \binom{N}{2}(p+1)^2(p-1) + \binom{N}{3}(p+1)^3(p-1)^2 + \cdots \\ &\qquad + \binom{N}{N}(p+1)^N(p-1)^{N-1} \\ &\qquad = \frac{p^{2N}-1}{p-1}. \end{aligned} \tag{5.68}$$

The total degree of freedom of $\mathcal{Q}_M$ is $p^{2N} - 1$, which is the same as the degree of freedom of the density matrix for an N-body p-dimensional quantum system.

5.5 Methodology

5.5.1 Abstracting Observables as Questions

In this study, we abstract the concept of an observable in quantum mechanics as a question. An observable $\hat{O}$ is represented as a question Q. This abstraction facilitates the reformation of quantum measurement processes using information theory, offering a novel perspective on understanding quantum systems. The motivation behind this abstraction is to provide an intuitive framework that aligns with the statistical nature of quantum mechanics

and leverages the well-established principles of information theory.

5.5.2 Characterization of Questions

Each question Q is characterized by a collection of outcome probabilities under a given context. This can be mathematically expressed as:

$$Q: \{\Pr(\text{outcome}|\text{context})\} \tag{5.69}$$

Here, $\Pr(\text{outcome}|\text{context})$ denotes the probability of a particular outcome given the context. By framing measurements as questions, we encapsulate the essence of quantum uncertainty and the probabilistic interpretation of measurement outcomes. This characterization allows us to systematically analyze the informational content of different measurements and their implications on the system's state.

5.5.3 Information Content of a Question

The information content of a question under a given context is defined as a function of the outcome probabilities. We utilize the Shannon entropy H to quantify this information:

$$I(Q|\text{context}) = H(\{\Pr(\text{outcome}|\text{context})\}) \tag{5.70}$$

Shannon entropy, which measures the uncertainty associated with the outcomes, provides a robust framework for quantifying the information gained from a measurement. This approach enables us to assess the value of different measurements in terms of the knowledge they provide about the quantum system.

5.5.4 Information of the System

To quantify the total information content of the system under a given context, we sum the information content of a selected set of questions. Thus, the total information of the system is expressed as:

$$I_{\text{system}}(\text{context}) = \sum I(Q|\text{context}) \tag{5.71}$$

This summation approach allows us to comprehensively capture the informational content of a quantum system by aggregating the contributions from individual measurements. It provides a holistic view of the system's informational structure, reflecting the collective impact of multiple measurements.

5.5.5 Measurement Processes

Single System Measurements When investigating a single quantum system, we perform various measurements to extract information. Each measurement outcome provides information about the system's state post-measurement. For instance, if a measurement $\hat{A}$ is performed at time t_1 and results in an outcome λ_A, the system collapses to the eigenstate corresponding to λ_A. The information gained from this measurement can be represented as:

$$\Delta I_{t<t_1 \to t>t_1}(\hat{A}) = I_{t>t_1}(\hat{A}) - I_{t<t_1}(\hat{A}) \tag{5.72}$$

Assuming we initially know nothing about the system and the possible outcome of measurement $\hat{A}$, we assign a uniform prior probability distribution:

$$\Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}\right)|_{t<t_1} = \Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}|I\right) = \left(\frac{1}{N}, \frac{1}{N}, \cdots, \frac{1}{N}\right) \tag{5.73}$$

For most information measures, the peak probability distribution will achieve a maximal value, $\Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}\right)_{t>t_1} = \Pr\left(\text{“}\hat{A}, \vec{\lambda}_A, t\text{”}|\text{“}\hat{A}, \lambda_A, t_1\text{”}, I\right)$, indicating maximal information gain:

$$\Delta I_{t<t_1 \to t>t_1}(\hat{A}) = H\left(\Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}\right)_{t>t_1}\right) - H\left(\Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}\right)_{t<t_1}\right) = 1 - 0 = 1 \tag{5.74}$$

Multiple System Measurements In quantum state tomography, we perform different measurements on many identical systems. By statistically analyzing the outcomes, we can infer the original prepared state. The information gained from the sequence of measurements D_n is updated as:

$$I_{D_n} = H(\Pr\left(\{q_1, q_2, \cdots, q_s\}|D_n, I\right)) \tag{5.75}$$

where $\{q_1, q_2, \cdots, q_s\}$ are parameters uniquely determining the system's state. This approach leverages Bayesian inference to update our knowledge based on measurement data, ensuring

a consistent framework for state estimation.

5.5.6 Evolution of Information About Single Observable

For an isolated system, the information about a measurement $\hat{A}$ remains unchanged if no further actions are taken. However, performing another measurement $\hat{B}$ at a later time t_2 can affect our knowledge about $\hat{A}$. If $\hat{A}$ and $\hat{B}$ are pairwise independent, the information about $\hat{B}$ can be calculated similarly:

$$\Delta I_{t<t_2 \to t>t_2}(\hat{B}) = I_{t>t_2}(\hat{B}) - I_{t<t_2}(\hat{B}) = 1 - 0 = 1 \tag{5.76}$$

However, if $\hat{A}$ and $\hat{B}$ do not commute, the knowledge gained from $\hat{A}$ may be partially or fully lost after t_2:

$$\Delta I_{t<t_2 \to t>t_2}(\hat{A}) = H\left(\Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}\right)_{t>t_2}\right) - H\left(\Pr\left(\text{“}\hat{A}, t, \vec{\lambda}_A\text{”}\right)_{t<t_2}\right) < 0 \tag{5.77}$$

5.5.7 Composite Systems

Non-Interacting Subsystems For a composite system with no interactions between subsystems, the total information is the sum of the information of the individual subsystems:

$$I_{\text{composite}} = I_{\text{subsystem 1}} + I_{\text{subsystem 2}} \tag{5.78}$$

This linear addition reflects the independence of the subsystems and provides an upper bound for the composite system's information.

Interacting Subsystems When interactions are present, additional terms representing correlations between subsystems are included in the set of probability distributions. These correlations are constructed to ensure they have the same number of outputs as the probability distributions of the subsystems, reflecting the dependency between outcomes of measurements on different subsystems. For example, a correlation between measurements A and B on different subsystems could be:

Correlation: “sum (modulo n) of their outcomes”

This correlation is only meaningful in the presence of interactions.

5.5.8 Information Upper Bound

We propose that the information content of a quantum system should have an upper bound, justified by the gain/loss cycle observed during repeated measurements. For a qubit system, choosing three mutually unbiased measurements (e.g., $\hat{\sigma}_x$, $\hat{\sigma}_y$, $\hat{\sigma}_z$) ensures that the information content remains finite and consistent. The total information is the sum of the information of these measurements:

$$I_{\text{qubit}} = I(\hat{\sigma}_x) + I(\hat{\sigma}_y) + I(\hat{\sigma}_z) \tag{5.79}$$

This approach ensures a finite and manageable upper bound for the information content of quantum systems.

5.6 Conclusion

Summary In this chapter, we present a new way to describe finite-dimensional quantum systems without relying on the language of linear spaces.

We begin with an ideal finite-dimensional quantum system, where every measurement yields the same and a finite number of outcomes. We can abstract each measurement as posing a question to the system. We introduce a new set of assumptions, motivated by quantum tomography and information theory, to deduce the relationships between these questions. Among these assumptions, we use classical logical gates to represent joint measurements in composite quantum systems. By rewriting the logical gate operations in terms of truth tables, we find that all feasible logical gates can be connected through a specific orthogonal array, which exhibits special properties in prime-number-dimensional cases.

The information of a single question is used to quantify our knowledge of a particular question based on a given background. It is defined as the generic entropy of the probability distribution of all possible outcomes for that question. The information of the system is used to quantify our knowledge of the entire system based on a specific background. It is defined as the sum of the information from selected questions, where these selected questions are

assumed to characterize the state of the whole system. The constraints on logical gates and the assumptions regarding the information of the system lead us to derive detailed relations among these selected questions. These selected questions are mutually complementary, akin to the complementarity between Pauli matrices. The number of selected questions is also determined, which is $p + 1$ for a p-ary system. Furthermore, the degree of freedom of the system is found to be the same as the degree of freedom of the density matrix of the corresponding quantum system.

We have also established a connection between this new structure and the conventional quantum system. This bridge is built on the concept of mutually unbiased bases, which share similarities in construction and properties with orthogonal arrays. Each concept we introduce in this new framework has a direct correspondence with concepts in the conventional quantum system. Furthermore, through derivations within this new structure, we have uncovered new insights into conventional quantum systems with the assistance of this connection.

Comparison to prior work As mentioned in the introduction, our work is built upon a lineage of research that originated with Rovelli and was continued by Brukner, Zeillinger, and Höhn. All of these works aim to employ information to describe quantum theory. There are two significant differences between our work and the work of others. The first difference is our emphasis on arbitrary prime-number-dimensional systems, rather than exclusively binary systems. The binary case, as we have demonstrated in the section on the representation of logical gates, is somewhat fortuitous. The second difference is our attempt to alleviate confusion regarding the definition of information, encompassing both the information of a question and the information of a system. Ideally, we seek to represent the information of a system as a combination of information from specific questions. It is essential to determine the types of questions required to characterize the system. In table 5.10 we show the comparison between our work and three selected work.

	Rovelli	Z & B	Höhn	Our work
Basic carrier of unit information	Binary outcome question	Binary outcome question	Binary outcome question	**p-ary** outcome question
Finiteness of information of system	Yes	Yes	Yes	Yes
Information of single system	Sum of complete questions	Sum of all complementary questions	Sum of independent bits	Sum of all questions in $\mathcal{Q}_M$
Information of composite system		Sum of selected correlation questions	Sum of independent bits	Sum of **rel. ind. questions** in $\mathcal{Q}_M$ compare to context

Table 5.10: ***Comparison between the Four Ideas that Describing Information of System*** (1)"p-ary outcome question" denotes to an arbitrary prime number dimensional system. We extend the type of discussed system from binary case to higher dimensional case. (2) We provide a criterion to choose the selected questions for characterizing composite system, that is, the mutually independent questions and the choice will be changed under different background.

We would like to emphasize that the most crucial aspect of incorporating information theory into quantum systems is to clarify the notion and definition of the concept of information, as well as the interpretation of probability distributions. Since most information measures are based on probabilities, we contend that the majority of the confusions surrounding information stem from an unclear understanding of probability distributions.

Discussion on unsolved problems In this chapter, we define the information of a single measurement, similar to Shannon entropy, although we do not provide a detailed expression. At present, we have not identified a compelling reason to choose a specific

measure, and the extreme case of this measure suffices for our purposes. We leave this measure in its general form, which may be useful for deriving other ideas. We also point out the challenge of defining the information of a system, for which we do not present a complete solution but suggest a way to improve it.

It is worth noting that both logical gates and MUBs have special properties in prime-number dimensions. This is why we focus exclusively on prime-number dimensional systems and why we were motivated to establish a connection between the question formalism and conventional quantum mechanics. Such a coincidence may have a profound mathematical explanation. Unfortunately, dealing with non-prime number dimensional cases is considerably more challenging. The construction of orthogonal arrays and MUBs is mathematically difficult in arbitrary finite dimensions, and whether this difficulty has a physical significance remains to be determined.

CHAPTER 6

Conclusion

6.1 Key Findings and Contributions

The first study focused on the operational perspective of information gain from quantum measurements, introducing the Principle of Information Increase. This principle was used to determine the most appropriate measure for quantifying information gain, leading to the identification of differential information gain as the most physically meaningful metric. The study highlighted the significance of the Jeffreys binomial prior, demonstrating its optimal characteristics in maximizing information communication.

One of the most remarkable findings was the equivalence of the expected values of differential and relative information gain for any prior and for any n-outcome probabilistic source. This unexpected result may provide a unified perspective that can simplify the analysis of information in quantum systems. By proving that the differential information gain is more physically meaningful, this study extends previous theoretical frameworks and offers a more nuanced understanding of information gain.

Furthermore, within the family of symmetric beta distributions, the Jeffreys binomial prior was shown to exhibit notable characteristics that enhance the intuitive notion that more data from measurements leads to more knowledge about the system. This result aligns with Summhammer's work and confirms that the Jeffreys prior enables maximal information communication. Additionally, the novel concept of robustness was introduced and applied to the Jeffreys binomial prior, suggesting that it exhibits maximal robustness within the set of symmetric beta distributions. This work raises intriguing questions about the potential extension of this feature to the Jeffreys multinomial prior and other probabilistic distributions.

The second study proposed a new framework for describing finite-dimensional quantum systems without relying on linear spaces. Traditional reconstructions of quantum theory often focus on two-dimensional systems, while this work extends the discussion to higher-

dimensional systems, especially prime-number dimensional cases. By conceptualizing quantum measurements as questions posed to the system, the study introduced classical logical gates to represent joint measurements in composite quantum systems. This approach revealed special properties in prime-number dimensional cases, leading to a deeper understanding of the structure of quantum measurements.

The information of a single question was used to quantify knowledge of a particular question based on a given background, defined as the generic entropy of the probability distribution of all possible outcomes for that question. This approach allowed for the introduction of constraints on logical gates and assumptions regarding the information of the system, which led to detailed relations among selected questions. These selected questions were found to be mutually complementary, similar to the complementarity between Pauli matrices, and their number was determined to be $p+1$ for a p-ary system, where p is a prime number. The degree of freedom of the system was found to be the same as the degree of freedom of the density matrix of the corresponding quantum system.

Moreover, the study established connections between this new framework and conventional quantum mechanics through the concept of mutually unbiased bases (MUBs), which share similarities in construction and properties with orthogonal arrays. Each concept introduced in this new framework has a direct correspondence with concepts in conventional quantum systems. This bridge provides new insights into conventional quantum systems and enhances our comprehension of information flow during measurements on maximally entangled systems.

6.2 Significance and Implications

The findings of these studies have profound implications for both theoretical and practical aspects of quantum information theory. The first study's introduction of the Principle of Information Increase and the identification of differential information gain provide a robust framework for optimizing quantum communication and state tomography. By establishing that the expected values of differential and relative information gain are equivalent, the study offers a unified perspective that can simplify and enhance the analysis of information in quantum systems.

The Jeffreys binomial prior, shown to exhibit maximal robustness within the set of symmetric beta distributions, provides a valuable tool for quantum state tomography and communication. The study's findings on the characteristics of this prior are particularly significant for scenarios involving the communication of maximal information, such as in quantum cryptography and quantum computing, where understanding and optimizing information gain is crucial.

The second study's framework for describing quantum systems through questions and logical gates opens new avenues for understanding the foundational structure of quantum theory. By extending information theory to prime-number dimensional quantum systems, the study reveals new relationships between classical logical operations and quantum measurements. This approach not only bridges the gap between classical and quantum information theories but also provides a novel way to explore the complex relationships between measurements and quantum states.

The insights gained from this study can enhance the development of quantum algorithms and improve the design of quantum information protocols. For example, understanding the special properties of prime-number dimensional systems can lead to more efficient algorithms for quantum computing and more robust protocols for quantum communication. Furthermore, the framework's ability to connect new theoretical constructs with conventional quantum mechanics provides a deeper understanding of how information flows and is processed in quantum systems, which is fundamental for advancements in quantum technologies.

6.3 Limitations and Future Research Directions

Despite the significant contributions of this thesis, there are several limitations that warrant further investigation. The analysis in the first study was restricted to symmetric beta distributions. Future research should explore the applicability of the principle and concept of robustness to other priors and probabilistic sources, particularly the robustness of Jeffreys multinomial prior in n-outcome probabilistic sources. Additionally, a deeper understanding of the robustness of the Jeffreys prior remains an open question that could reveal further insights into information measures.

Expanding the scope of priors and distributions studied will help validate the robustness concept and potentially uncover new priors that exhibit similar or superior characteristics. Such research could involve empirical studies or theoretical analyses to test the robustness of different priors in various quantum information scenarios. Understanding how these priors perform in practice will be crucial for developing more effective quantum state tomography techniques and optimizing quantum communication protocols.

The second study focused exclusively on prime-number-dimensional systems, highlighting the special properties of logical gates and mutually unbiased bases in these dimensions. However, extending this framework to non-prime number dimensions poses significant mathematical challenges. Future research should address these challenges and investigate whether the difficulties have physical significance. For instance, developing methods to construct orthogonal arrays and MUBs in arbitrary finite dimensions could significantly advance our understanding of quantum measurements and information theory.

Moreover, defining the information measure function remains an open problem. While this study suggested ways to improve the definition, a comprehensive solution is still needed. Future research could explore alternative definitions and measures of information that are consistent across different quantum systems and dimensions. Investigating how these measures relate to existing concepts in quantum mechanics and information theory could lead to a more unified and comprehensive understanding of information in quantum systems.

Additionally, the use of logical entropy as an alternative choice of information measure function presents an interesting direction for future research. Logical entropy, while offering a different perspective on measuring uncertainty, lacks the intuitive informational relationship with quantum theory that measures like Shannon or von Neumann entropy provide. Exploring the potential applications and limitations of logical entropy in quantum contexts could yield new insights and help refine our understanding of informational measures in quantum systems.

APPENDIX A

Appendix for Principle of Information Increase

A.1 Derivation of Differential Information Gain

The posterior is determined by T_N and prior. For the sake of simplicity we would set the prior belongs to the family of beta distributions:

$$\Pr(p|I) = \frac{p^\alpha (1-p)^\alpha}{B(\alpha+1, \alpha+1)} \tag{A.1}$$

where $\alpha > -1, B(x, y)$ is the beta function.

Given N, there are 2^N different T_N. However, we may not need to calculate all the 2^N sequences. Suppose every toss is independent, this happens in quantum mechanics, then this coin tossing model would become a binomial distribution. Let h_N be the number of heads inside T_N, the posterior $\Pr(p|N, T_N, I)$ is equivalent to $\Pr(p|N, h_N, I)$ and likelihood will be

$$\Pr(h_N|N, p, I) = \binom{N}{h_N} p^{h_N} (1-p)^{N-h_N} \tag{A.2}$$

hence the posterior after N tosses

$$\begin{aligned} \Pr(p|N, h_N, I) &= \frac{\Pr(h_N|N, p, I)\Pr(p|I)}{\int \Pr(h_N|N, p, I)\Pr(p|I)dp} \\ &= \frac{p^{h_N+\alpha}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+1, N-h_N+\alpha+1)} \end{aligned} \tag{A.3}$$

The information gain in the $N+1$th toss would be

$$I_{diff} = D_{KL}(\Pr(p|N+1, \{T_N, t_{N+1}\}, I)||\Pr(p|I)) - D_{KL}(\Pr(p|N, h_N, I)||\Pr(p|I)) \tag{A.4}$$

I_{diff} is determined by h_N, prior and the result of $N+1$th toss t_{N+1}. t_{N+1} could be

either "Head" or "Tail", then posterior after $N+1$ tosses could be

$$\Pr(p|N+1,\{T_N,t_{N+1}=\text{“}Head\text{”}\},I)=\frac{p^{h_N+\alpha+1}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+2,N-h_N+\alpha+1)} \tag{A.5}$$

$$\Pr(p|N+1,\{T_N,t_{N+1}=\text{“}Tail\text{”}\},I)=\frac{p^{h_N+\alpha}(1-p)^{N-h_N+\alpha+1}}{B(h_N+\alpha+1,N-h_N+\alpha+2)} \tag{A.6}$$

Taking $t_{N+1}=$ "$Head$",the first term in (A.4) would become

$$\begin{aligned}
&D_{KL}(\Pr(p|N+1,\{T_N,t_{N+1}=\text{“}Head\text{”}\},I)||\Pr(p|I))\\
&=\int_0^1 \Pr(p|N+1,h_N+1,I)ln\frac{\Pr(p|N+1,h_N+1,I)}{\Pr(p|I)}dp\\
&=\int_0^1 \frac{p^{h_N+\alpha+1}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+2,N-h_N+\alpha+1)}ln\frac{p^{h_N+1}(1-p)^{N-h_N}B(\alpha+1,\alpha+1)}{B(h_N+\alpha+2,N-h_N+\alpha+1)}dp\\
&=\int_0^1 \frac{p^{h_N+\alpha+1}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+2,N-h_N+\alpha+1)}\{ln[p^{h_N+1}(1-p)^{N-h_N}]+ln\frac{B(\alpha+1,\alpha+1)}{B(h_N+\alpha+2,N-h_N+\alpha+1)}\}dp\\
&=\int_0^1 \frac{p^{h_N+\alpha+1}(1-p)^{n-h_N+\alpha}}{B(h_N+\alpha+1,n-h_N+\alpha+1)}ln[p^{h_N+1}(1-p)^{N-h_N}]dp+ln\frac{B(\alpha+1,\alpha+1)}{B(h_N+\alpha+2,n-h_N+\alpha+1)}
\end{aligned} \tag{A.7}$$

By using the integral

$$\int_0^1 x^a(1-x)^b ln(x)dx=B(a+1,b+1)[\psi(a+1)-\psi(a+b+2)] \tag{A.8}$$

where $\psi(x)$ is the digamma function[9], we can obtain the following result

$$\begin{aligned}
&D_{KL}(\Pr(p|N+1,\{T_N,t_{N+1}=\text{“}Head\text{”}\},I)||\Pr(p|I))\\
=&(h_N+1)\psi(h_N+\alpha+2)+(N-h_N)\psi(N-h_N+\alpha+1)-(N+1)\psi(N+2\alpha+3)\\
&+ln\frac{B(\alpha+1,\alpha+1)}{B(h_N+\alpha+2,n-h_N+\alpha+1)}
\end{aligned} \tag{A.9}$$

[9]The digamma function can be defined in terms of gamma function: $\psi(x)=\frac{\Gamma'(x)}{\Gamma(x)}$

The second term in (A.4) would become

$$
\begin{aligned}
& D_{KL}(\Pr(p|N,h_N,I)||\Pr(p|I)) \\
=& \int_0^1 \Pr(p|N,h_N,I) ln \frac{\Pr(p|N,h_N,I)}{\Pr(p|I)} dp \\
=& \int_0^1 \frac{p^{h_N+\alpha}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+1,N-h_N+\alpha+1)} ln \frac{p^{h_N}(1-p)^{N-h_N}B(\alpha+1,\alpha+1)}{B(h_N+\alpha+1,N-h_N+\alpha+1)} dp \\
=& \int_0^1 \frac{p^{h_N+\alpha}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+1,N-h_N+\alpha+1)} \{ln[p^{h_N}(1-p)^{N-h_N}] + ln \frac{B(\alpha+1,\alpha+1)}{B(h_N+\alpha+1,N-h_N+\alpha+1)}\} dp \\
=& \int_0^1 \frac{p^{h_N+\alpha}(1-p)^{n-h_N+\alpha}}{B(h_N+\alpha+1,n-h_N+\alpha+1)} ln[p^{h_N}(1-p)^{N-h_N}] dp + ln \frac{B(\alpha+1,\alpha+1)}{B(h_N+\alpha+1,n-h_N+\alpha+1)} \\
=& h_N \psi(h_N+\alpha+1) + (N-h_N)\psi(N-h_N+\alpha+1) - N\psi(N+2\alpha+2) \\
& + ln \frac{B(\alpha+1,\alpha+1)}{B(h_N+\alpha+1,n-h_N+\alpha+1)}
\end{aligned}
\tag{A.10}
$$

Now we obtain the final expression of (A.4)

$$
\begin{aligned}
I_{diff}(t_{N+1} = \text{“}Head\text{”}) &= D_{KL}(\Pr(p|N+1,\{T_N, t_{N+1} = \text{“}Head\text{”}\},I)||\Pr(p|I)) \\
&\quad - D_{KL}(\Pr(p|N,h_N,I)||\Pr(p|I)) \\
&= \psi(h_N+\alpha+2) - \psi(N+2\alpha+3) + \\
&\quad \frac{h_N}{h_N+\alpha+1} - \frac{N}{N+2\alpha+2} + \ln \frac{N+2\alpha+2}{h_N+\alpha+1}
\end{aligned}
\tag{A.11}
$$

Similarly we can obtain the I_{diff} when $t_{N+1} =$ “$Tail$”

$$
\begin{aligned}
I_{diff}(t_{N+1} = \text{“}Tail\text{”}) &= \psi(N-h_N+\alpha+2) - \psi(N-h_N+2\alpha+3) + \\
&\quad \frac{N-h_N}{N-h_N+\alpha+1} - \frac{N}{N+2\alpha+2} + \ln \frac{N+2\alpha+2}{N-h_N+\alpha+1}
\end{aligned}
\tag{A.12}
$$

This suggests that for fixed N and α, $I_{diff}(t_{N+1} =$ “$Head$”) and $I_{diff}(t_{N+1} =$ “$Tail$”) are symmetric since h_N is ranging from 0 to N.

A.2 Derivation of Relative Information Gain

From Appendix A we know that the posterior after N tosses is

$$\Pr(p|N, T_N, I) = \Pr(p|N, h_N, I) = \frac{p^{h_N+\alpha}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+1, N-h_N+\alpha+1)} \tag{A.13}$$

Therefore the posterior after $N+1$ tosses would be

$$\Pr(p|N+1, T_{N+1}, I) = \frac{\Pr(h_N, T_{N+1}|p, N+1, I)\Pr(p|I)}{\int_0^1 \Pr(h_N, T_{N+1}|p, N+1, I)\Pr(p|I)dp} \tag{A.14}$$

Depends on different results of t_{N+1}, the posterior after $N+1$ tosses would be

$$\Pr(p|N+1, \{T_N, t_{N+1} = \text{“}Head\text{”}\}, I) = \frac{p^{h_N+\alpha+1}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+2, N-h_N+\alpha+1)} \tag{A.15}$$

$$\Pr(p|N+1, \{T_N, t_{N+1} = \text{“}Tail\text{”}\}, I) = \frac{p^{h_N+\alpha}(1-p)^{N-h_N+\alpha+1}}{B(h_N+\alpha+1, N-h_N+\alpha+2)} \tag{A.16}$$

And the corresponding relative information gain would be

$$\begin{aligned}
&I_{rel}(t_{N+1} = \text{“}Head\text{”}) \\
&= D_{KL}(\Pr(p|N+1, \{T_N, t_{N+1} = \text{“}Head\text{”}\}, I)||\Pr(p|N, h_N, I)) \\
&= \int_0^1 \Pr(p|N+1, \{T_N, t_{N+1} = \text{“}Head\text{”}\}, I) \ln \frac{\Pr(p|N+1, \{T_N, t_{N+1} = \text{“}Head\text{”}\}, I)}{\Pr(p|N, h_N, I)} dp \\
&= \int_0^1 \frac{p^{h_N+\alpha+1}(1-p)^{N-h_N+\alpha}}{B(h_N+\alpha+2, N-h_N+\alpha+1)} \ln \frac{pB(h_N+\alpha+1, N-h_N+\alpha+1)}{B(h_N+\alpha+2, N-h_N+\alpha+1)} dp \\
&= \psi(h_N+\alpha+2) - \psi(N+2\alpha+3) + \ln \frac{N+2\alpha+2}{h_N+\alpha+1}
\end{aligned} \tag{A.17}$$

$$I_{rel}(t_{N+1} = \text{“}Tail\text{”}) = \psi(N-h_N+\alpha+2) - \psi(N-h_N+2\alpha+3) + \ln \frac{N+2\alpha+2}{N-h_N+\alpha+1} \tag{A.18}$$

A.3 Equivalence of Expected Differential Information Gain and Expected Relative Information Gain

In a n-outcome model, the probability of each outcome is p_i, and

$$p_1 + p_2 + \cdots + p_n = 1 \tag{A.19}$$

After N "tosses", the data sequence has the form

$$D_N = (f_1, f_2, \cdots, f_n), \quad \sum_{i=1}^{n} f_i = N \tag{A.20}$$

where f_i is the number of ith outcomes in these N tosses.

We may use a tuple $\vec{p} = (p_1, p_2, \cdots, p_n)$ to represent the probabilities of these outcomes. The prior is just $\Pr(\vec{p}|I)$, and the posterior based on the data D_N is $\Pr(\vec{p}|D_N, I)$.

The average value of the ith outcome probability is

$$\langle p_i \rangle = \int p_i \Pr(\vec{p}|D_N, I) dp_1 dp_2 \cdots dp_n \tag{A.21}$$

Assume the $(N+1)$th toss is the ith outcome, and the posterior of these after this additional toss is

$$\begin{aligned} \Pr(\vec{p}|D_N, d_{N+1} = \text{“}i\text{”}, I) &= \frac{p_i \Pr(\vec{p}|D_N, I)}{\int p_i \Pr(\vec{p}|D_N, I) dp_1 dp_2 \cdots dp_n} \\ &= \frac{p_i}{\langle p_i \rangle} \Pr(\vec{p}|D_N, I) \end{aligned} \tag{A.22}$$

Then we can write I_{diff} as

$$\begin{aligned} I_{\text{diff}}(d_{N+1} = \text{“}i\text{”}) =& D_{\text{KL}}(\Pr(\vec{p}|D_N, d_{N+1} = \text{“}i\text{”}, I) | \Pr(\vec{p}|I)) - D_{\text{KL}}(\Pr(\vec{p}|D_N, I) | \Pr(\vec{p}|I)) \\ =& \int \frac{p_i}{\langle p_i \rangle} \Pr(\vec{p}|D_N, I) \ln \frac{p_i \Pr(\vec{p}|D_N, I)}{\langle p_i \rangle \Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \\ & - \int \Pr(\vec{p}|D_N, I) \ln \frac{\Pr(\vec{p}|D_N, I)}{\Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \end{aligned} \tag{A.23}$$

Then the expected differential information gain is given by

$$
\begin{aligned}
\overline{I_{\text{diff}}} &= \sum_{i=1}^{n} \langle p_i \rangle \, I_{\text{diff}}(d_{N+1} = \text{“}i\text{”}) \\
&= \sum_{i=1}^{n} \int p_i \Pr(\vec{p}|D_N, I) \ln \frac{p_i \Pr(\vec{p}|D_N, I)}{\langle p_i \rangle \Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \\
&\quad - \sum_{i=1}^{n} \langle p_i \rangle \int \Pr(\vec{p}|D_N, I) \ln \frac{\Pr(\vec{p}|D_N, I)}{\Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \\
&= \sum_{i=1}^{n} \left[\int p_i \Pr(\vec{p}|D_N, I) \ln \frac{p_i}{\langle p_i \rangle} dp_1 dp_2 \cdots dp_n + \int p_i \Pr(\vec{p}|D_N, I) \ln \frac{\Pr(\vec{p}|D_N, I)}{\Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \right] \\
&\quad - \int \Pr(\vec{p}|D_N, I) \ln \frac{\Pr(\vec{p}|D_N, I)}{\Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \\
&= \sum_{i=1}^{n} \int p_i \Pr(\vec{p}|D_N, I) \ln \frac{p_i}{\langle p_i \rangle} dp_1 dp_2 \cdots dp_n + \int \sum_{i=1}^{n} p_i \Pr(\vec{p}|D_N, I) \ln \frac{\Pr(\vec{p}|D_N, I)}{\Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \\
&\quad - \int \Pr(\vec{p}|D_N, I) \ln \frac{\Pr(\vec{p}|D_N, I)}{\Pr(\vec{p}|I)} dp_1 dp_2 \cdots dp_n \\
&= \sum_{i=1}^{n} \int p_i \Pr(\vec{p}|D_N, I) \ln \frac{p_i}{\langle p_i \rangle} dp_1 dp_2 \cdots dp_n
\end{aligned}
\tag{A.24}
$$

Similarly, I_{rel} can be written as

$$
\begin{aligned}
I_{\text{rel}}(d_{N+1} = \text{“}i\text{”}) &= D_{\text{KL}}(\Pr(\vec{p}|D_N, d_{N+1} = \text{“}i\text{”}, I) | \Pr(\vec{p}|D_N, I)) \\
&= \int \frac{p_i}{\langle p_i \rangle} \Pr(\vec{p}|D_N, I) \ln \frac{p_i \Pr(\vec{p}|D_N, I)}{\langle p_i \rangle \Pr(\vec{p}|D_N, I)} dp_1 dp_2 \cdots dp_n \\
&= \int \frac{p_i}{\langle p_i \rangle} \Pr(\vec{p}|D_N, I) \ln \frac{p_i}{\langle p_i \rangle} dp_1 dp_2 \cdots dp_n
\end{aligned}
\tag{A.25}
$$

Then the expected relative information gain is, accordingly,

$$
\overline{I_{\text{rel}}} = \sum_{i=1}^{n} \langle p_i \rangle \, I_{\text{rel}}(d_{N+1} = \text{“}i\text{”}) = \sum_{i=1}^{n} \int p_i \Pr(\vec{p}|D_N, I) \ln \frac{p_i}{\langle p_i \rangle} dp_1 dp_2 \cdots dp_n \tag{A.26}
$$

From (A.24) and (A.26), we can see that in this n-outcome model, the expected differential information gain $\overline{I_{\text{diff}}}$ and expected relative information gain $\overline{I_{\text{rel}}}$ are equal, irrespective of the choice of prior.

APPENDIX B

Appendix for Quantum Questions

B.1 Correspondence of Corollary 6 in quantum mechanics

Lemma 1. In $\mathbb{C}^p \otimes \mathbb{C}^p$, among composite operators in the form of $\hat{A} \otimes (\hat{B})^k$ where $\hat{A}, \hat{B} \in \{\hat{X}, \hat{Z}, \hat{X}\hat{Z}, \hat{X}\hat{Z}^2, \cdots, \hat{X}\hat{Z}^{p-1}\}$ and $k \in \{1, 2, \cdots, p-1\}$, there are at most $(p+1)$ different mutually commuting composite operators.

Proof. By contradiction, assume there are at least $(p+2)$ different mutually composite operators, say $\hat{A}_1 \otimes (\hat{B}_1)^{k_1}, \hat{A}_2 \otimes (\hat{B}_2)^{k_2}, \cdots, \hat{A}_{p+2} \otimes (\hat{B}_{p+2})^{k_{p+2}}$.

Since the choice of $\hat{B}_i$ is limited, there will be n and m such that $1 \leq n < m \leq p+1$ and $\hat{B}_n = \hat{B}_m$. This leads to three situations:

1. If $\hat{A}_n \neq \hat{A}_m$, then this leads to contradiction since $\left[\hat{A}_n \otimes (\hat{B}_n)^{k_n}, \hat{A}_m \otimes (\hat{B}_n)^{k_m}\right] \neq 0$.

$$\begin{aligned}\left[\hat{A}_n \otimes (\hat{B}_n)^{k_n}, \hat{A}_m \otimes (\hat{B}_m)^{k_m}\right] =& \hat{A}_n\hat{A}_m \otimes (\hat{B}_n)^{k_n}(\hat{B}_n)^{k_m} - \hat{A}_m\hat{A}_n \otimes (\hat{B}_n)^{k_m}(\hat{B}_n)^{k_n} \\ =& \left(\hat{A}_n\hat{A}_m - \hat{A}_m\hat{A}_n\right) \otimes \hat{B}_n^{k_n+k_m}\end{aligned} \tag{B.1}$$

   Without loss of generality, assume $\hat{A}_n = \hat{X}^{i_n}\hat{Z}^{j_n}, \hat{A}_m = \hat{X}^{i_m}\hat{Z}^{j_m}$, where $i_n, i_m \in \{0, 1\}$ $j_n = `\delta_{i_n,0} + n\delta_{i_n,1}$ $j_m = \delta_{i_m,0} + m\delta_{i_m,1}$ $n, m \in \mathbb{F}_p^*$. By using Weyl commutation relation, we have the following result:

$$\begin{aligned}\hat{A}_n\hat{A}_m - \hat{A}_m\hat{A}_n =& \hat{X}^{i_n}\hat{Z}^{j_n}\hat{X}^{i_m}\hat{Z}^{j_m} - \hat{X}^{i_m}\hat{Z}^{j_m}\hat{X}^{i_n}\hat{Z}^{j_n} \\ =& \hat{X}^{i_n}\hat{Z}^{j_n}\hat{X}^{i_m}\hat{Z}^{j_m} - \omega_p^{i_nj_m - i_mj_n}\hat{X}^{i_n}\hat{Z}^{j_n}\hat{X}^{i_m}\hat{Z}^{j_m}\end{aligned} \tag{B.2}$$

   However, $i_nj_m - i_mj_n \neq 0 \pmod p$ since $\hat{A}_n \neq \hat{A}_m$. This means $\omega_p^{i_mj_n - i_nj_m} \neq 1$ and $\left[\hat{A}_n \otimes (\hat{B}_n)^{k_n}, \hat{A}_n \otimes (\hat{B}_m)^{k_m}\right] \neq 0$.

2. If $\hat{A}_n = \hat{A}_m$ and $k_n \neq k_m$, then the common eigensubspace of these two operators is ensured. Let $|a\rangle$ be the eigenstate of $\hat{A}_n$ and $\hat{b}$ be the eigenstate of $\hat{B}_n$ such that $\hat{A}_n|a\rangle = \omega^a|a\rangle$ and $\hat{B}_n|b\rangle = \omega^b|b\rangle$. The eigenspace of $\hat{A}_n \otimes (\hat{B}_n)^{k_n}$ and $\hat{A}_n \otimes (\hat{B}_n)^{k_m}$ can then

be expressed in terms of $|a\rangle \otimes |b\rangle$:

$$\begin{aligned} E(\omega_p^n, \hat{A}_n \otimes (\hat{B}_n)^{k_n}) &= Span(|a_1\rangle \otimes |b_1\rangle, |a_2\rangle \otimes |b_2\rangle, \cdots, |a_p\rangle \otimes |b_p\rangle) \qquad a_i + k_n b_i = n \\ E(\omega_p^m, \hat{A}_n \otimes (\hat{B}_n)^{k_m}) &= Span(|a_1\rangle \otimes |b_1\rangle, |a_2\rangle \otimes |b_2\rangle, \cdots, |a_p\rangle \otimes |b_p\rangle) \qquad a_i + k_m b_i = m \end{aligned} \tag{B.3}$$

The common eigenspace of $\hat{A}_n \otimes (\hat{B}_n)^{k_n}$ and $\hat{A}_n \otimes (\hat{B}_n)^{k_m}$ can then be determined.

$$E(\omega_p^n, \hat{A}_n \otimes (\hat{B}_n)^{k_n}) \cap E(\omega_p^m, \hat{A}_n \otimes (\hat{B}_n)^{k_m}) = span(|a_{n,m}\rangle \otimes |b_{n,m}\rangle) \tag{B.4}$$

where $a_{n,m} + k_n b_{n,m} = n \pmod p$ and $a_{n,m} + k_m b_{n,m} = m \pmod p$.

$\forall \hat{A}_l, \hat{B}_l \in \{\hat{X}, \hat{Z}, \hat{X}\hat{Z}, \hat{X}\hat{Z}^2, \cdots, \hat{X}\hat{Z}^{p-1}\}$, we have[3]:

$$\begin{aligned} \hat{A}_l |a_{n,m}\rangle &= |a_{n,m} \oplus a_l\rangle \\ \hat{B}_l |b_{n,m}\rangle &= |b_{n,m} \oplus b_l\rangle \end{aligned} \tag{B.5}$$

where $\oplus$ is the addition in $\mathbb{F}_p$ and $a_l, b_l \in \mathbb{F}_p$. $a_l = 0$ if and only if $\hat{A}_l = \hat{A}_n$, $b_l = 0$ if and only if $\hat{B}_l = \hat{B}_n$.

$$\hat{A}_l \otimes (\hat{B}_l)^{k_l} |a_{n,m}\rangle \otimes |b_{n,m}\rangle = |a_{n,m} \oplus a_l\rangle \otimes |b_{n,m} \oplus k_l b_l\rangle \tag{B.6}$$

This suggests that $|a_{n,m}\rangle \otimes |b_{n,m}\rangle$ cannot be the eigenstate of $\hat{A}_l \otimes (\hat{B}_l)^{k_l}$ if $\hat{A}_n \neq \hat{A}_l$ or $\hat{B}_n \neq \hat{B}_l$.

$|a_{n,m}\rangle \otimes |b_{n,m}\rangle$ cannot be the eigenstate of composite operators other than the members of $\{\hat{A}_n \otimes (\hat{B}_n)^1, \hat{A}_n \otimes (\hat{B}_n)^2, \cdots, \hat{A}_n \otimes (\hat{B}_n)^{p-1}\}$. There is no more composite operators that $|a_{n,m}\rangle \otimes |b_{n,m}\rangle$ is one of its eigenstates, which means there are at most $(p-1)$ mutually commuting composite operators and this contradicts to the assumption.

3. If $\hat{A}_n = \hat{A}_m$ and $k_n = k_m$, then $\hat{A}_n \otimes (\hat{B}_n)^{k_n} = \hat{A}_m \otimes (\hat{B}_m)^{k_m}$ and this contradicts the assumption.

□

B.2 Correspondence of Assumption 5 in quantum mechanics

Lemma 2. In $\mathbb{C}^p \otimes \mathbb{C}^p$, $\forall \hat{A}, \hat{B}, \hat{C}, \hat{D} \in \{\hat{X}, \hat{Z}, \hat{X}\hat{Z}, \hat{X}\hat{Z}^2, \cdots, \hat{X}\hat{Z}^{p-1}\}$, if $\hat{A} \neq \hat{C}, \hat{B} \neq \hat{D}$ then $\forall m \in \mathbb{F}_p^*$, $\exists! n \in \mathbb{F}_p^*$ such that $[\hat{A} \otimes \hat{B}^m, \hat{C} \otimes \hat{D}^n] = 0$.

Proof. Let $\hat{A} = \hat{X}^{i_1}\hat{Z}^{i_2}, \hat{B} = \hat{X}^{j_1}\hat{Z}^{j_2}, \hat{C} = \hat{X}^{k_1}\hat{Z}^{k_2}, \hat{D} = \hat{X}^{l_1}\hat{Z}^{l_2}$

$$\begin{aligned}\left[\hat{A} \otimes \hat{B}^m, \hat{C} \otimes \hat{D}^n\right] &= \hat{A}\hat{C} \otimes \hat{B}^m\hat{D}^n - \hat{C}\hat{A} \otimes \hat{D}^n\hat{B}^m \\ &= \hat{X}^{i_1}\hat{Z}^{i_2}\hat{X}^{k_1}\hat{Z}^{k_2} \otimes (\hat{X}^{j_1}\hat{Z}^{j_2})^m(\hat{X}^{l_1}\hat{Z}^{l_2})^n - \hat{X}^{k_1}\hat{Z}^{k_2}\hat{X}^{i_1}\hat{Z}^{i_2} \otimes (\hat{X}^{l_1}\hat{Z}^{l_2})^n(\hat{X}^{j_1}\hat{Z}^{j_2})^m\end{aligned} \tag{B.7}$$

By using Weyl commutation relation, we can obtain the following relation:

$$\hat{X}^{i_1}\hat{Z}^{i_2}\hat{X}^{k_1}\hat{Z}^{k_2} = w_p^{i_2k_1 - i_1k_2}\hat{X}^{k_1}\hat{Z}^{k_2}\hat{X}^{i_1}\hat{Z}^{i_2}$$

$$(\hat{X}^{j_1}\hat{Z}^{j_2})^m(\hat{X}^{l_1}\hat{Z}^{l_2})^n = w_p^{mn(j_2l_1 - j_1l_2)}(\hat{X}^{l_1}\hat{Z}^{l_2})^n(\hat{X}^{j_1}\hat{Z}^{j_2})^m$$

In order to let the commutation relation $[\hat{A} \otimes \hat{B}^m, \hat{C} \otimes \hat{D}^n] = 0$ holds, we must have

$$w_p^{i_2k_1 - i_1k_2} w_p^{mn(j_2l_1 - j_1l_2)} = 1$$

$$n = \frac{i_1k_2 - i_2k_1}{m(j_2l_1 - j_1l_2)}$$

If $\hat{A} \neq \hat{C}, \hat{B} \neq \hat{D}$ then both numerator and denominator cannot be zero. Moreover both numerator and denominator are elements in $\mathbb{F}_p^*$. This means n is a unique element in $\mathbb{F}_p^*$. □

BIBLIOGRAPHY

[1] S Aravinda, R Srikanth, and Anirban Pathak, *On the origin of nonclassicality in single systems*, Journal of Physics A: Mathematical and Theoretical **50** (2017), no. 46, 465303.

[2] John Stufken A.S. Hedayat, N. J. A. Sloane, *Orthogonal arrays: Theory and applications*, 1 ed., Springer series in statistics, Springer-Verlag New York, 1999.

[3] Somshubhro Bandyopadhyay, P. Oscar Boykin, Vwani Roychowdhury, and Farrokh Vatan, *A new proof for the existence of mutually unbiased bases*, 2001, arXiv:quant-ph/0103162.

[4] Charles H. Bennett, *The thermodynamics of computation—a review*, International Journal of Theoretical Physics **21** (1982), no. 12, 905–940.

[5] James O. Berger and Jose M. Bernardo, *Ordered group reference priors with application to the multinomial problem*, Biometrika **79** (1992), no. 1, 25–37.

[6] L. Brillouin, *Maxwell's demon cannot operate: Information and entropy. I*, Journal of Applied Physics **22** (1951), no. 3, 334–337.

[7] Časlav Brukner and Anton Zeilinger, *Conceptual inadequacy of the shannon information in quantum measurements*, Phys. Rev. A **63** (2001), 022113.

[8] ———, *Information and fundamental elements of the structure of quantum theory*, 2002, arXiv:quant-ph/0212084.

[9] ———, *Information invariance and quantum probabilities*, Foundations of Physics **39** (2009), 677–689.

[10] Ariel Caticha, *Entropic dynamics, time and quantum theory*, Journal of Physics A: Mathematical and Theoretical **44** (2011), no. 22, 225303.

[11] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack, *Quantum probabilities as bayesian probabilities*, Phys. Rev. A **65** (2002), 022305.

[12] Jing-Ling Chen, Libin Fu, Abraham A. Ungar, and Xian-Geng Zhao, *Degree of entanglement for two qubits*, Phys. Rev. A **65** (2002), 044303.

[13] Giulio Chiribella, *Agents, subsystems, and the conservation of information*, Entropy **20** (2018), no. 5.

[14] R. Clifton, J. Bub, and H. Halvorson, *Characterizing quantum theory in terms of information-theoretic constraints*, Foundations of Physics **33** (2003), 1561–1591.

[15] R. T. Cox, *Probability, frequency and reasonable expectation*, American Journal of Physics **14** (1946), no. 1, 1–13.

[16] L. Czekaj, M. Horodecki, P. Horodecki, and R. Horodecki, *Information content of systems as a physical principle*, Phys. Rev. A **95** (2017), 022119.

[17] H. De Raedt, M. I. Katsnelson, and K. Michielsen, *Quantum theory as plausible reasoning applied to data obtained by robust experiments*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **374** (2016), no. 2068, 20150233.

[18] David Deutsch, *Quantum theory, the church–turing principle and the universal quantum computer*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **400** (1985), 97–117.

[19] ———, *It from qubit*, Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity (John D. Barrow, Paul C. W. Davies, and Jr. Harper, Charles L., eds.), Cambridge University Press, Cambridge, 2004, pp. 90–102.

[20] D. Dieks, *Communication by EPR devices*, Physics Letters A **92** (1982), no. 6, 271–272.

[21] T. Durt, B. Englert, I. Bengtsson, and K. Zyczkowski, *On mutually unbiased bases*, International Journal of Quantum Information **08** (2010), no. 04, 535–640.

[22] David Ellerman, *Introduction to logical entropy and its relationship to shannon entropy*, 2021, arXiv:2112.01966.

[23] U. Fano, *Description of states in quantum mechanics by density matrix and operator techniques*, Rev. Mod. Phys. **29** (1957), 74–93.

[24] Christopher A. Fuchs, N. David Mermin, and Rüdiger Schack, *An introduction to QBism with an application to the locality of quantum mechanics*, American Journal of Physics **82** (2014), no. 8, 749–754.

[25] Christopher A. Fuchs and Rüdiger Schack, *QBism and the greeks: why a quantum state does not represent an element of physical reality*, Physica Scripta **90** (2015), no. 1, 015104, Published 31 December 2014.

[26] Philip Goyal, *Prior probabilities: An information-theoretic approach*, AIP Conference Proceedings **803** (2005), no. 1, 366–373.

[27] Philip Goyal, Kevin H. Knuth, and John Skilling, *Origin of complex quantum amplitudes and Feynman's rules*, Phys. Rev. A **81** (2010), 022109.

[28] Alexei Grinbaum, *Elements of information-theoretic derivation of the formalism of quantum theory*, International Journal of Quantum Information **1** (2003), no. 3, 289–300.

[29] Rishabh Gupta, Rongxin Xia, Raphael D. Levine, and Sabre Kais, *Maximal entropy approach for quantum state tomography*, PRX Quantum **2** (2021), 010318.

[30] Philipp Andres Höhn, *Quantum theory from rules on information acquisition*, Entropy **19** (2017), no. 3.

[31] ———, *Toolbox for reconstructing quantum theory from rules on information acquisition*, Quantum **1** (2017), 38.

[32] I D Ivonovic, *Geometrical description of quantal state determination*, Journal of Physics A: Mathematical and General **14** (1981), no. 12, 3241.

[33] Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White, *Measurement of qubits*, Phys. Rev. A **64** (2001), 052312.

[34] E. T. Jaynes, *Information theory and statistical mechanics I*, Phys. Rev. **106** (1957), 620–630.

[35] ———, *Information theory and statistical mechanics II*, Phys. Rev. **108** (1957), 171–190.

[36] ———, *Information theory and statistical mechanics*, W. A. Benjamin, Inc., 1963.

[37] Andrei B. Klimov, Denis Sych, Luis L. Sánchez-Soto, and Gerd Leuchs, *Mutually unbiased bases and generalized bell states*, Phys. Rev. A **79** (2009), 052101.

[38] L. D. Landau and E. M. Lifshitz, *Statistical Physics : Volume 5*, 3rd. ed., Butterworth-Heinemann, 2013 (eng).

[39] R. Landauer, *Irreversibility and heat generation in the computing process*, IBM Journal of Research and Development **5** (1961), no. 3, 183–191.

[40] Chun-Wang Ma and Yu-Gang Ma, *Shannon information entropy in heavy-ion collisions*, Progress in Particle and Nuclear Physics **99** (2018), 120–158.

[41] Vaibhav Madhok, Carlos A. Riofrío, Shohini Ghose, and Ivan H. Deutsch, *Information gain in tomography–a quantum signature of chaos*, Physical Review Letters **112** (2014), 014102.

[42] Lluís Masanes, Markus P. Müller, Remigiusz Augusiak, and David Pérez-García, *Existence of an information unit as a postulate of quantum theory*, Proceedings of the National Academy of Sciences **110** (2013), no. 41, 16373–16377.

[43] Robert D. McMichael, Sergey Dushenko, and Sean M. Blakley, *Sequential Bayesian experiment design for adaptive Ramsey sequence measurements*, Journal of Applied Physics **130** (2021), no. 14, 144401.

[44] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, *Quantum-process tomography: Resource analysis of different strategies*, Phys. Rev. A **77** (2008), 032322.

[45] M. K. Patra, *Quantum state determination: estimates for information gain and some exact calculations*, Journal of Physics A: Mathematical and Theoretical **40** (2007), no. 35, 10887–10902.

[46] Ben Placek, Daniel Angerhausen, and Kevin H. Knuth, *Analyzing exoplanet phase curve information content: Toward optimized observing strategies*, The Astronomical Journal **154** (2017), no. 4, 154.

[47] John Preskill, *Chapter 10. Quantum Shannon Theory*, Quantum Computation Course Notes, 2022, Available online.

[48] Yihui Quek, Stanislav Fort, and Hui Khoon Ng, *Adaptive quantum state tomography with neural networks*, npj Quantum Information **7** (2021), no. 105.

[49] C. Radhakrishna Rao, *Orthogonal arrays*, Scholarpedia **4** (2009), no. 7, 9076, revision #137077.

[50] Alfréd Rényi, *On measures of entropy and information*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, vol. 4, University of California Press, 1961, pp. 547–562.

[51] Carlo Rovelli, *Relational quantum mechanics*, International Journal of Theoretical Physics **35** (1996), 1637.

[52] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), no. 3, 379–423.

[53] Johann Summhammer, *Maximum predictive power and the superposition principle*, International Journal of Theoretical Physics **33** (1994), 171–178.

[54] _____, *Maximum predictive power and the superposition principle*, 1999, arXiv:quant-ph/9910039.

[55] John Archibald Wheeler, *Information, physics, quantum: The search for links*, Proceedings III International Symposium on Foundations of Quantum Mechanics, 1989, pp. 354–358.

[56] W. K. Wootters and W. H. Zurek, *A single quantum cannot be cloned*, Nature **299** (1982), no. 5886, 802–803.

[57] William K. Wootters, *Communicating through probabilities: Does quantum theoryoptimize the transfer of information?*, Entropy **15** (2013), no. 8, 3130–3147.

[58] William K Wootters and Brian D Fields, *Optimal state-determination by mutually unbiased measurements*, Annals of Physics **191** (1989), no. 2, 363–381.

[59] Juan Yin, Yuan Cao, Hai-Lin Yong, Ji-Gang Ren, Hao Liang, Sheng-Kai Liao, Fei Zhou, Chang Liu, Yu-Ping Wu, Ge-Sheng Pan, Li Li, Nai-Le Liu, Qiang Zhang, Cheng-Zhi

Peng, and Jian-Wei Pan, *Lower bound on the speed of nonlocal correlations without locality and measurement choice loopholes*, Phys. Rev. Lett. **110** (2013), 260407.