(Yet Another) Reconstruction of Quantum Theory

Bharath Ron

The physical systems of interest to us are those whose observed phenomenon can be described by the notion of an 'observable'. The starting idea here is that the world can be described by sentences of the form 'the observable R has the value R_i'. The next step is to articulate this idea in the mathematical language. This means that we need to construct an appropriately predictive mathematical theory where sentences like the one above can be represented. There have been mainly two successful approaches for this.

In classical mechanics, the approach is to think of observables as continuous maps from a manifold Ω, called the phase space to the set of real numbers

$$A : \Omega \longrightarrow \mathbb{R}.$$

As we will discuss later, this model of an observable fails when the application domain is expanded to simultaneously include both discrete and continuous case. Although quantum theory has the same idea, it is formulated in a different way that makes it compatible with the extended application domain. To understand how the quantum approach is different from the classical approach we will first try to understand what is meant by an 'observable'.

1 | Generalised Probabilistic Theories

The discussion here will be closest to the Ludwig school [**?**]-[4], with some stuff borrowed from the quantum logic literature [9]-[10]. This can be thought of as a modern formulation of Heisenberg's original idea. Hans Primas in §4 [5] characterises approaches to formulation of quantum theory into three main categories, the quantum logic approach, algebraic approach, and the convex state-space approaches.

The convex state-space approach was developed by Ludwig and his peers. This approach starts with measuring instruments and preparation instruments and is the most general approach among the three. A physical theory is in some sense interpreted from outside in terms of pre-theories not belonging to the theory in question itself. Usually when one tries

to formulate quantum theory one starts with a pre-theory such as classical mechanics and then 'quantizes' the theory. This makes the theory messy and the underlying physical ideas hidden and unclear. In Ludwig's approach the pre-theory is the theory of preparation of and measuring instruments. The assumed structure in this instrumentalist view of physics can be motivated with simple thought experiments and hence we will adopt this approach to the formulation of quantum theory. The construction and behavior of instruments will not be of interest to us. Any changes occurring in the instruments during 'measurements' will be accepted as objective events. In this point of view, the fundamental notions of quantum theory have to be defined operationally in terms of macroscopic instruments and prescriptions for their application. Quantum mechanics is then interpreted entirely in terms of such instruments and events which are the changes occurring to instruments. These instruments and events are our links to 'objective reality'.

1.1 | Effects & Ensembles

From this instrumentalist or operational point of view, the notion of 'state' can be defined in terms of the preparation procedure. A preparation procedure is characterized by the kind of system it prepares. The measuring instrument that is capable of undergoing changes when an experiment is performed for evaluating a collection of possible events. The possible results of such an experiment are called outcomes of the experiment. The observable change in the instrument is called an effect.

To simplify the procedure consider instruments that record 'hits'. These instruments perform simple 'yes-no' measurements. Any measurement can be interpreted as a combination of 'yes-no' measurements. These 'yes-no' instruments can be used to build any general instrument. Suppose we have such a measuring instrument, label its measuring instrument by R_i. If the experiment is conducted a lot of times, we get a relative frequency of occurrence of 'yes'. Here 'yes' is an observable change in the instrument. It is hence an observable effect. To every preparation procedure ρ and measuring instrument R_i there exists a probability $\mu(\rho|R_i)$ of the occurrence of 'yes' associated with the pair.

$$(\rho, R_i) \longrightarrow \mu(\rho|R_i).$$

The numbers $\mu(\rho|R_i)$ are called operational statistics.

Two completely different preparation procedures may give the same operational statistics, that is, they may give same probabilities for every experiments. Such preparation procedures must be considered equivalent operationally. An equivalence class of preparation procedures yielding the same operational statistics for experiments is called an ensemble. Similarly, there may be two measuring instruments that have the same chances of undergoing a change for similarly prepared systems. Such measuring instruments must be operationally considered to be equivalent. An equivalence class of change for measuring instruments is called an effect. An effect is the equivalence class of all instruments that undergo a change for the same possible outcome. By considering equivalence classes we have obtained the structure of sets. Without introducing any new physical law we have obtained the basic mathematical structure for modelling preparation procedures and measuring instruments in terms of the ensembles and effects they describe. Since we are considering equivalence classes, the construction and behavior of the instruments are irrelevant.

Denote the set of ensembles by $\mathcal{S}$ and the set of effects by $\mathcal{E}$. The maps of interest to us are the following,

$$\mathcal{S} \times \mathcal{E} \xrightarrow{\mu} [0, 1].$$

For a possible outcome R_i, we will denote the corresponding effect by E_{R_i}. Each effect acts on the ensembles of the system, and each ensemble acts on the effects of the system to yield the corresponding operational statistic

$$E_{R_i} : \rho \mapsto \mu(\rho|R_i), \quad \mu_\rho : R_i \mapsto \mu(\rho|R_i).$$

In general, an experiment can measure a collection of possible outcomes, hence the measuring instrument used for the experiment can be described by a collection of outcomes it can measure. For each such possible outcome we can assign its corresponding effect. If $R \equiv \{R_i\}$ is a measuring instrument, then each outcome R_i of the measuring instrument R corresponds to an effect E_{R_i} called the effect of R_i. Hence each measuring instrument $R \equiv \{R_i\}$ can be modelled as an effect valued map.

Similarly, each ensemble fixes the operational statistics for possible outcomes, and hence we have,

$$\mu^R_\rho : R_i \mapsto \mu^R_\rho(R_i) = \mu(\rho|R_i).$$

$\mathcal{E}$ and $\mathcal{S}$ together with the pairing $\mu : \mathcal{E} \times \mathcal{S} \to [0, 1]$ is called an operational theory.

Since any two preparations giving the same result on every effect represent the same ensemble and two measurement procedures that cannot distinguish ensembles represent the same effect, ensembles and effects are mutually separating with respect to the pairing $\mu(\cdot|\cdot)$. Hence studying the mathematical structures of one of these sets also already tells us a lot about the mathematical structure of the other. It is hence sufficient to focus our attention on understanding the structure of either the space of effects or the space of ensembles. At this stage note that ensembles are not the most primitive concepts in this framework. What measuring instruments an experimenter is allowed to apply to a preparation procedure already constrains the collection of allowed preparation procedures. So the space of effects which model the measuring instruments is a more primitive concept than the space of ensembles. Although this might seem like a trivial reason, the two paths diverge significantly.

On a historic note, this is where we see theoretical physics branch off into two different groups. The first group does in the direction of understanding the structure of the space of ensembles, following Schrödinger, Dirac, Feynman, and others; Since state spaces in classical theories are also closely related to the topology and geometry this group has been able to make a lot of progress in developing the standard approach to quantum field theories behind the celebrated standard model by utilising the geometric intuition behind classical theories and by understanding and generalising gauge theoretic ideas of Maxwell and others. The other group branched out in the direction of understanding the structure of the space of effects, following Heisenberg, Jordon, Born, von Neumann, Haag, Kastler, Araki, Borchers and others. This group has been able to make much less progress comparitively possibly because the space of effects is very strange and unfamiliar compared ro the more familiarity and intuitive classical theories. This lack of pre-theories such as classical theory has left this group stranded in the middle of nowhere, and the ones who did pursue this new path are forced to build up starting from almost nothing.

Most physicists avoid the second path possibly due to unfamiliarity. We will however follow the second path due its stronger foundations and embrace the risk of losing the centuries old, and possibly false intuition, for classical theories.[1]

[1] The need for quantum theory already indicates that atleast some of the ideas behind classical theories are not compatible with observed phenomena. Keeping that in mind we feel it is justified to not blindly trust the intuition we have developed for classical theories.

1.1.1 | Convex Embeddings

In order to be able to work with effects and ensembles we now embed them inside a set with mathematical structure that respects the expected operational relations and also allows us to do mathematics with. In our case the operational requirement is that we should be able to make sense of taking mixtures of effects and ensembles. Accounting for the fact that preparation procedures can be combined to produce a mixed ensemble, the set of ensembles should be closed under forming of mixtures.

The notion of mixing corresponds mathematically to the notion of convex combination, see [6], we expect the set $\mathcal{S}$ to have structure that enables us to take convex combinations and we expect each functional E_{R_i} preserves the convex structure since we expect the operational statistics to preserve this convexity. To make sense of taking convex linear combinations we need to be able make sense of linearity combinations. Hence we must embed the set of effects and the set of ensembles in a vector space with respect to a field that is atleast as large as the real numbers. It is convinient to choose the field to be the field of complex numbers since the field of complex numbers is algebraically closed and allows us sufficient use of analytic tools. It is important to note that preparation and measuring instruments producing the same ensembles and effects are not equal, in fact, the notion of equality will not even make sense. The transition from preparation and measuring instruments to ensembles and effects is a transition from the real world to the abstract mathematical world. It should also be noted that it does not make sense to 'prepare' closed systems, one has to assume such systems start off in some state a priori.

A generalized probabilistic theory (GPT) is the embedding of $\mathcal{E}$ and $\mathcal{S}$ inside vector spaces, such that the ensembles and effects are uniquely determined by the operational statistics they produce. This uniqueness of operational statistics is called the principle of tomography in quantum foundations and quantum information literature. To construct a generalised probabilistic theory for $\mathcal{E}$ and $\mathcal{S}$ we start by thinking of effects and ensembles as linear functionals on a suitable space. Denote by $\mathcal{A}_*$ the set of maps

$$\omega(R) = \textstyle\sum_{i\in I} \alpha_i \mu(\rho_i|R), \quad \forall R \in \mathcal{E},$$

where ρ_i are ensembles and α_i are scalars with I a finite set. Since any linear combination of such maps also belongs to $\mathcal{A}_*$, it follows that $\mathcal{A}_*$ is a complex vector space. Denote by $\mathcal{A}$ the set of maps,

$$A(\rho) = \textstyle\sum_{j\in J} \beta_j \mu(\rho|R_j), \quad \forall \rho \in \mathcal{S},$$

where R_j are effects and β_j are scalars with J a finite set. It follows that $\mathcal{A}$ is also a complex vector spaces. We can embed ensembles inside $\mathcal{A}_*$ with the map,

$$\rho \mapsto \mu_\rho,$$

and similarly embed effects inside $\mathcal{A}$ with the map,

$$R_i \mapsto E_{R_i}.$$

Abusing notation we will denote the images of $\mathcal{E}$ and $\mathcal{S}$ by the same. The elements of the convex subsets $\mathcal{E}$ of $\mathcal{A}$ and $\mathcal{S}$ of $\mathcal{A}_*$ are called effects and states respectively.

The vector space structure of $\mathcal{A}_*$ and $\mathcal{A}$ allows us to develop algebraic tools for studying effects and ensembles. So we embedded an operational theory by viewing the effects and ensembles as linear functionals on each other and took 'quotients' with operational equivalences. equip $\mathcal{A}$ and $\mathcal{A}_*$ with mathematical structure sufficient for their study.[2]

[2]Quantum mechanics can be done in more complicated fields such as the field of quarternions. We will however avoid discussions on such choice as it will digress too much from the main goal of this thesis.

Since we expect the ensembles and effects to be determined by the probabilities they produce we must also expect the vector spaces $\mathcal{A}$ and $\mathcal{A}_*$ to inherit a relation between each other from the pairing $\mu(\cdot|\cdot)$ of $\mathcal{E}$ and $\mathcal{S}$. So we must have a pairing

$$\langle\cdot|\cdot\rangle : \mathcal{A}_* \times \mathcal{A} \to \mathbb{C}$$

that coincides with μ for effects and ensembles, that is, $\langle E_{R_i}|\mu_\rho\rangle = \mu(R_i|\rho)$. We require the scaling of both effects and ensembles by a unit length complex number to give same pairing as the unscaled pairing. This requirement along with real linearity in each argument forces the pairing $\langle\cdot|\cdot\rangle$ to be sesquilinear.

1.1.2 | The Banach *-Structure

The operational requirements force the space of effects and ensembles to be vector spaces. This is however insufficient for analysis. To bring any predictivity to the theory we need more structure than mere vector space structure. Since we can make sense of one measurement after another, the space of effects should come equipped with an algebra structure. Similarly, the notion of accuracy of instruments gives rise to a notion of nearness, and gives rise to a topology. We intend to understand the relation between the algebraic and the topological structures.

Construction of a Norm

We can have two measuring instruments for an event where one is more accurate than the other. In such a case, the less accurate measuring instrument will readily undergo a change compared to the more accurate instrument. If the instrument is too accurate, it becomes difficult to observe any changes occuring in the measuring instrument. In this sense, the more accurate instrument should be 'closer' to the instrument which never undergoes any changes. So, there needs to be a way of quantifying the ease of noticing the changes occuring to the instruments. The instrument which always undergoes a change should be the easiest to notice changes and the instrument which never undergoes any changes should be the hardest.

Let $\|\cdot\|_\mathcal{E}$ be the overall likelihood of undergoing change. Suppose we have two measuring instruments which undergo changes for outcomes R_i and R_j, and let E_{R_i} and E_{R_j} be the corresponding effects. Suppose the two instruments could undergo changes under some similar situations, then it means that we have more redundancy. The more redundancies, the easier it is to undergo changes, and more likely it is to notice changes. Hence we must have,

$$\|E_{R_i} + E_{R_j}\|_\mathcal{E} \leq \|E_{R_i}\|_\mathcal{E} + \|E_{R_j}\|_\mathcal{E}.$$

If the accuracy of either of the instruments is changed, the accuracy of the combined instrument should also change, and this change should be proportional to the change in accuracy of the instrument. Hence we expect $\|\cdot\|_\mathcal{E}$ to be continuous under addition. Hence $\|\cdot\|_\mathcal{E}$ satisfies the triangle inequality.

When a measuring instrument is applied to any preparation procedure it can potentially undergo a change. Even though this gives us an operational interpretation of $\|\cdot\|_\mathcal{E}$ in terms of accuracy of the measuring instruments it is more convinient to define it as a supremum. Since $\|\cdot\|_\mathcal{E}$ represents the overall likelihood of undergoing a change it must correspond to the highest chance of undergoing over all preparation procedures. Hence for every $F \in \mathcal{E}$, $\|F\|_\mathcal{E}$ can be defined as the supremum,

$$\big\|E\big\|_\mathcal{E} \coloneqq \sup_{\mu_\rho \in \mathcal{S}} \big|\big\langle E|\mu_\rho\big\rangle\big|,$$

where the supremum is taken over all preparation procedures which is represented by the convex embedded subset $\mathcal{S}$ of $\mathcal{E}_*$. By the assumption that there will always exist some preparation procedure with non-trivial operational statistics, we deduce that $\|\cdot\|_{\mathcal{E}}$ is non-trivial, that is, $\|E\|_{\mathcal{E}} = 0$ if and only if $E \equiv 0$. By definition it also follows that

$$\big\|\lambda E\big\|_{\mathcal{E}} = |\lambda|\big\|E\big\|_{\mathcal{E}}, \quad \forall \lambda \in \mathbb{C}.$$

Since $\|\cdot\|_{\mathcal{E}}$ satisfies triangle inequality, and satisfies the above scaling condition it defines a norm on $\mathcal{E}$. Similarly, $\mathcal{E}_*$ also inherits a norm from $\mathcal{E}$ such that for any $\sigma \in \mathcal{E}_*$,

$$\big\|\sigma\big\|_{\mathcal{E}_*} \coloneqq \sup_E \big|\big\langle E|\sigma\big\rangle\big|$$

where the supremum is taken over effects which is represented by the convex embedded subset $\mathcal{P}$ of $\mathcal{E}$. We can hence require both $\mathcal{E}$ and $\mathcal{E}_*$ to be a complex, normed spaces. By the boundedness of the functionals

$$E_{R_i} \mapsto \big\langle E_{R_i}|\mu_\rho\big\rangle,$$

we note that every ensemble gives rise to a continuous linear functional on $\mathcal{E}$ with respect to the topology induced by the norm $\|\cdot\|_{\mathcal{E}}$ and similarly for $\mathcal{E}_*$. Under the idealisation that there exist instruments of every level of accuracy, we can assume $\mathcal{E}$ to be closed with respect to taking limits in the above defined norms. Hence we will assume the spaces $\mathcal{E}_*$ and $\mathcal{E}$ to be Banach spaces. This allows us to talk about limits and allows us to do mathematical analysis with ensembles and effects. We assume the existence of a unique element $1 \in \mathcal{E}$ which corresponds to the trivial instrument which is always true for any preparation procedure the uniqueness is assumed because if there are two instruments which are always 'yes' they are operationally indistinguishable and are hence operationally equivalent. By definition of the norm we have $\|1\|_{\mathcal{E}} = 1$. In this sense, the element $1 \in \mathcal{E}$ is the 'existence' element for the system. Similarly we will assume the existence of the unique 0 element which outputs 'no' for every preparation procedure.

Since $\mathcal{E}$ and $\mathcal{E}_*$ inherit topologies from each other we expect $\mathcal{E}$ and $\mathcal{E}_*$ to be related to each other topologically also. Since $\mathcal{S}$ and $\mathcal{P}$ separate each other with respect to the pairing $\langle\cdot|\cdot\rangle$, we will assume $\langle\mathcal{E}|\mathcal{E}_*\rangle$ is a dual pair.

Topological vs Algebraic Data

Let the outcome corresponding to measuring the outcome R_i after R_j be $R_i \wedge R_j$, the corresponding effect is denoted by $E_{R_i}E_{R_j}$. If we vary the accuracy of either of these measuring instruments we expect the accuracy of the combined effect to vary accordingly, that is to say the map

$$\big(E_{R_i}, E_{R_j}\big) \mapsto E_{R_i}E_{R_j}$$

is continuous. This is condition imposes conditions on $\mathcal{E}$, making it a Banach algebra. We make this heuristic argument precise following [**?**].

We need to show the product structure is compatible with the norm $\|\cdot\|_{\mathcal{E}}$ on $\mathcal{E}$. The idea that $\mathcal{E}$ has a natural action of the space of effects and we can embed effects inside the bounded operators on $\mathcal{B}(\mathcal{E})$, which comes equipped with an algebra structure given by composition of operators and the sup-norm which we will denote by $\|\cdot\|$. So, we should expect the Banach space $\mathcal{E}$ to be an algebra such that $\mathcal{L}_E F = EF$ is continuous, and similarly, $\mathcal{R}_F E = EF = \mathcal{L}_E F$ is continuous, so $\mathcal{L}_E$ and $\mathcal{R}_F$ are both bounded.

By continuity of $\mathcal{R}_F$ we have,

$$\big\|\mathcal{R}_E F\big\|_{\mathcal{E}} = \big\|\mathcal{L}_F E\big\|_{\mathcal{E}} \leq \big\|\mathcal{L}_F\big\|_{\mathcal{E}}\,\big\|E\big\|_{\mathcal{E}}.$$

Hence $\{\mathcal{R}_E F\}_{E\in\mathcal{E}}$ is a bounded set for every F. By uniform boundedness principle, pointwise boundedness implies that the set $\{\mathcal{R}_E\}_{E\in\mathcal{E}}$ is uniformly bounded, with the uniform bound $|||\mathcal{R}|||$. Hence for every $E \in \mathcal{E}$,

$$\|\mathcal{R}_E\| \leq |||\mathcal{R}|||\, \|E\|_{\mathcal{E}},$$

Hence the map

$$\mathcal{R} : E \to \mathcal{R}_E$$

defines a bounded and hence continuous linear map from $\mathcal{E}$ to $\mathcal{B}(\mathcal{E})$ of bounded linear maps on $\mathcal{E}$. $\mathcal{R}$ is a continuous algebra homomorphism. Assuming $\|1\|_{\mathcal{E}} = 1$, we have,

$$\|E\|_{\mathcal{E}} = \|\mathcal{R}_E(1)\|_{\mathcal{E}} \leqq \|\mathcal{R}_E\| \leqq |||\mathcal{R}|||\, \|E\|_{\mathcal{E}}.$$

$\mathcal{R}(\mathcal{E})$ is a norm closed subalgebra of $\mathcal{B}(\mathcal{E})$. $\mathcal{R}$ is an algebraic isomorphism from $\mathcal{E}$ to $\mathcal{R}(\mathcal{E})$. Since for any two bounded linear operators S and T on $\mathcal{E}$ we have, $\|TS\| \leq \|T\|\, \|S\|$, $\mathcal{R}(\mathcal{E})$ inherits this property and we have,

$$\|EF\|_{\mathcal{E}} \leq \|E\|_{\mathcal{E}} \|F\|_{\mathcal{E}}$$

Since $\|\mathcal{R}_1\| = 1$ we have, $\|\mathcal{R}_E(1)\|_{\mathcal{E}} = \|E\|_{\mathcal{E}} \leq \|\mathcal{R}_E\| \leq \|\mathcal{R}_1\|\, \|E\|_{\mathcal{E}} = \|E\|_{\mathcal{E}}$. Hence $\mathcal{R}$ is an isometric isomorphism and $\mathcal{R}(\mathcal{E})$ is a Banach algebra. Hence abusing notation we will denote $\mathcal{R}(\mathcal{E})$ by $\mathcal{E}$ itself.

Every preparation procedure can be followed up by a measurement. The preparation procedure together with the measurement can itself be thought of as a preparation procedure, let us denote such an operation on the ensembles by

$$(E^\dagger \mu_\rho)(F) = \mu_\rho(FE).$$

The action of E on $\mathcal{E}$ corresponds to performing the experiment corresponding to the effect E before, and the action of $E^\dagger$ on ensembles corresponds to the composite preparation where one performs the experiment for E after initial preparation. Hence $\mathcal{E}_*$ has an action of $\mathcal{E}$ given by,

$$(\lambda E + \mu F)^\dagger \mu_\rho = \lambda E^\dagger \mu_\rho + \mu F^\dagger \mu_\rho$$

So we can think of $\mathcal{E}_*$ as a $\mathcal{E}$-module.

The structure of the state space and its relation with the space of effects gives extra structure on the space of effects. For every linear map f on $\mathcal{E}$ the sesquilinear pairing $\langle\cdot|\cdot\rangle$ gives us a map on $\mathcal{E}_*$ defined by

$$\left\langle E_{R_i} | f^\dagger \mu_\rho \right\rangle = \left\langle f E_{R_i} | \mu_\rho \right\rangle.$$

For any linear operators f, g on $\mathcal{E}$ we must have $(f\circ g)^\dagger = g^\dagger \circ f^\dagger$ and $(\lambda f)^\dagger = \overline{\lambda} f^\dagger$. This gives us a map on the space of effects, $\dagger : \mathcal{E} \to \mathcal{E}$.[3] Since the norm is given by $\|f\| = \sup\left\{\left|\langle fE|\mu_\rho\rangle\right|\right\}$, we have,

$$\left\|f^\dagger f\right\| = \sup\left\{\left|\left\langle f^\dagger f E|\mu_\rho\right\rangle\right|\right\} = \sup\left\{\left|\left\langle fE|f\mu_\rho\right\rangle\right|\right\}.$$

For effects we have, $\langle EF|E^\dagger\mu_\rho\rangle = \mu_\rho(E^2F)$.

[3]Since we are looking at effects as operators, we can also obtain the †-structure by starting with $\perp$ operation on the space of effects, and use the equality between kernel of an operator and $\perp$ of image of the adjoint of the operator as the starting point. We have avoided this path since it involves more steps to relate projection to a subspace corresponding to the kernel of an operator with the operator.

[[Something Fishy here]]

Hence we expect the norm $\|\cdot\|$ on $\mathcal{E}$ to be such that $\|f^\dagger f\| = \|f\|^2$. An algebra $\mathcal{A}$ together with a norm $\|\cdot\|$ is said to be a Banach algebra if it is complete with respect to the norm $\|\cdot\|$ and satisfies

$$\|AB\| \leq \|A\|\|B\|, \quad \forall A, B \in \mathcal{A}$$

for every $A, B \in \mathcal{A}$. For Banach algebras it follows by triangle inequality that for fixed B, for any ϵ, we can find A_i and A_j with $\|A_i - A_j\| < \epsilon/\|B\|$, and hence it follows that $\|A_iB - A_jB\| \leq \|A_i - A_j\|\|B\| < \epsilon$. Hence it follows that multiplication is jointly continuous with respect to the norm topology.

If $\mathcal{A}$ also has an involution $*$, it is called a Banach *-algebra. A C^*-algebra is a Banach *-algebra $\mathcal{A}$ which satisfies the C^*-identity, that is,

$$\|A^*A\| = \|A\|^2, \quad \forall A \in \mathcal{A}$$

If $\mathcal{A}$ is a C^*-algebra we have $\|A^*A\| = \|A\|^2 \leq \|A\|\|A^*\|$ and similarly by symmetry we will also have $\|A^*\|^2 \leq \|A\|\|A^*\|$, and hence we have

$$\|A\| \leq \|A^*\| \leq \|A\|.$$

Hence the $*$-operation preserves the norm, that is $\|A\| = \|A^*\|$ for every $A \in \mathcal{A}$ and must be continuous with respect to the topology induced by the norm.

From now on we will denote the algebra that can be used to model the space of effects by $\mathcal{A}$, called the algebra of observables which will be assumed to be a C^*-algebra,

$$\mathcal{E} \equiv \mathcal{A}.$$

From the discussion earlier we expect every 'yes-no' effects E to satisfy $E^2 = E$. Hence the collection of all 'yes-no' effects corresponds to the collection of all projection operators on $\mathcal{A}$ denoted by $\mathcal{P}(\mathcal{A})$, and we have $\mathcal{P} \equiv \mathcal{P}(\mathcal{A}) \subset \mathcal{A}$. Since we expect that every experiment can be decomposed in terms of more elementary 'yes-no' instruments, we must expect $\mathcal{A}$ to be generated by $\mathcal{P}(\mathcal{A})$.

We expect two effects to undergo changes more often for different preparation procedures if they were closer to each other. Hence for every preparation procedure and nearby effects we expect the corresponding operational statistics to also be nearby. Any linear functionals satisfying this continuity requirement is called normal. The collection of all normal linear functionals is itself a Banach space equipped with sup-norm. We will denote the space of normal linear functionals on $\mathcal{A}$ by $\mathcal{A}_*$. Hence we have,

$$\mathcal{E}_* \equiv \mathcal{A}_*.$$

Since an ensemble assigns to each effects the corresponding operational statistics, we expect ensembles to assign numbers in the interval $[0, 1]$ for every projection in $\mathcal{A}$ and by assumption must assign the value 1 for the element 1, the norm of such functionals must be 1. Hence the state space corresponds to the space $\mathcal{S}(\mathcal{A})$ of positive norm 1 normal functionals on $\mathcal{A}$. We have, $\mathcal{S} \equiv \mathcal{S}(\mathcal{A}) \subset \mathcal{A}_*$. The elements of the state space $\mathcal{S}(\mathcal{A})$ generate $\mathcal{A}_*$.

We have a natural pairing of $\mathcal{A}$ and linear functionals on $\mathcal{A}$ given by

$$\langle\cdot|\cdot\rangle : (A, \omega) \mapsto \omega(A),$$

where $A \in \mathcal{A}$ and ω is a linear functional on $\mathcal{A}$. From the discussion earlier, we require $\langle \mathcal{A} | \mathcal{A}_* \rangle$ to be a dual pair. Hence $\mathcal{A}$ is the topological dual space of the Banach space $\mathcal{A}_*$ of normal linear functionals on $\mathcal{A}$.

A C^*-algebra $\mathcal{A}$ which is the dual space of a Banach space $\mathcal{A}_*$ is called a von Neumann algebra or a W^*-algebra. $\mathcal{A}_*$ is called the predual of $\mathcal{A}$ and we will prove later the uniqueness of predual. If we denote the space of continuous linear functionals on a Banach space V by V^*, for a von Neumann algebra $\mathcal{A}$ we have

$$(\mathcal{A}_*)^* = \mathcal{A}.$$

The weak*-topology defined by the semi-norm on $\mathcal{A}$ by taking its modulus of linear functionals in $\mathcal{A}_*$ is called the weak topology on the von Neumann algebra $\mathcal{A}$ denoted by $\sigma(\mathcal{A}, \mathcal{A}_*)$.

1.2 | What are Observables?

In general a measuring instrument can simultaneously measure a collection of possible outcomes. The mathematical representatives of such generalised measuring instruments are called observables. We now extrapolate the algebraic structures for collection of effects for such possible outcomes by studying the expected properties of such measuring instruments.

1.2.1 | Boolean Algebras & Maps

Consider a collection of possible outcomes $R \equiv \{R_i\}$ that can be measured simultaneously using the same instrument. In this case a single experiment is performed. Let $\{E_{R_i}\}$ be the collection of effects corresponding to the possible outcomes $\{R_i\}$. We are interested in understanding the algebraic structure that should be expected of such effects. To understand the algebraic structure of $\{E_{R_i}\}$, we can ask what are the queries that can be answered after an experimenter performs a measurement with the same instrument. We should obviously expect whether the outcome was R_i for each $R_i \in R$. The experimenter can also ask if the outcome was not R_i. So, if $\neg R_i$ denotes the outcome corresponding to the experimental outcome not being R_i, then we should expect $E_{\neg R_i} \in \{E_{R_i}\}$. We will denote such elements by the notation

$$E_{\neg R_i} \equiv E_{R_i}^{\perp}.$$

Similarly, the experimenter can ask if the outcome was R_i or R_j, and if the outcome was R_i and R_j for $R_i, R_j \in R$. This tells us about the algebraic structure of effects when they can be measured by a single instrument. The effects of a measuring instrument measured by a single instrument should be expected to have classical logical operations such as meet, join and not, corresponding to whether the measurement detects outcome R_i *or* R_j, R_i *and* R_j, the outcome is *not* R_i. Denote these operations by $R_i \vee R_j$, $R_i \wedge R_j$, and $\neg R_i$ respectively. Clearly we should have for any $R_i \in R$,

$$R_i \wedge R_i = R_i$$
$$R_i \vee R_i = R_i.$$

The composite operations involving $\vee$, $\wedge$, and $\neg$ can be figured out by simple thought. We list these properties below without attempting to be minimal. We expect the operations $\wedge$

and $\vee$ to be commutative and associative, that is, for every $R_i, R_j \in R$,

$$\begin{aligned} R_i \vee R_j &= R_j \vee R_i \\ R_i \wedge R_j &= R_j \wedge R_j \\ R_i \wedge (R_i \wedge R_k) &= (R_i \wedge R_j) \wedge R_k \\ R_i \vee (R_i \vee R_k) &= (R_i \vee R_j) \vee R_k. \end{aligned}$$

$\wedge$ and $\vee$ satisfy the absorption property, that is,

$$\begin{aligned} R_i \vee (R_i \wedge R_j) &= R_i \\ R_i \wedge (R_i \vee R_j) &= R_i. \end{aligned}$$

We also expect distributivity between $\wedge$ and $\vee$, that is, for every $R_i, R_j, R_k \in R$

$$\begin{aligned} R_i \vee (R_j \wedge R_k) &= (R_i \vee R_j) \wedge (R_i \vee R_k) \\ R_i \wedge (R_j \vee R_k) &= (R_i \wedge R_j) \vee (R_i \wedge R_k). \end{aligned}$$

Let $\mathbb{I}$ denote the outcome which is always true when the measurement is performed. Similarly let 0 denote the outcome which never undergoes a change. $\mathbb{I}$ and 0 are called the top and bottom elements respectively. We say $\{R_i\}$ is exhaustive if $\vee_i R_i = \mathbb{I}$. We say that R_i and R_j are disjoint if $R_i \wedge R_j = 0$. Then we expect

$$\begin{aligned} R_i \wedge \mathbb{I} &= R_i \\ R_i \vee 0 &= R_i \\ R_i \vee \neg R_i &= \mathbb{I} \\ R_i \wedge \neg R_i &= 0. \end{aligned}$$

If R is closed under $\wedge$, $\vee$ and $\neg$ and contains the elements $\mathbb{I}$ and 0 it is called a Boolean algebra, denoted by Σ_R. These axioms of Boolean algebra forces the complements to be unique. Associativity ensures that we can make sense of $\wedge_k R_{i_k}$ and $\vee_k R_{i_k}$ for any finite collection $\{R_{i_k}\}$. Σ_R is said to be a complete Boolean algebra if we can make sense of $\vee_k R_{i_k}$ and $\wedge_k R_{i_k}$ for arbitrary collections.[4]

The effects of a measuring instrument corresponding to a single measuring instrument should also inherit Boolean algebra structure, so the maps of interest to us are those which preserve the Boolean algebra structure. If Σ_R and Σ_S are Boolean algebras, a mapping

$$E_R : \Sigma_R \to \Sigma_S$$

[4]The notion of competion can be described intuitively by viewing Boolean algebras in lattice theoretically. We denote

$$R_i \leq R_j \Leftrightarrow R_i \wedge R_j = R_j$$

An upper bound of a finite collection $\{R_{i_k}\}$ is an element R_j such that $R_{i_k} \leq R_j$ for all k. The supremum $\sup_k\{R_{i_k}\}$ for the collection is the lowest upper bound for the collection, that is, for any U with $R_{i_k} \leq U$ for all k implies $\sup_k\{R_{i_k}\} \leq U$. We have $\sup_k\{R_{i_k}\} = \vee_k R_{i_k}$ for any finite collection.

A Boolean algebra is said to be complete if every collection has a supremum with respect to the above order. For complete Boolean algebras the existence of supremum allows us to use the notation

$$\sup_k\{R_{i_k}\} \equiv \vee_k R_{i_k}.$$

for arbitrary collection $\{R_{i_k}\}$. We can similarly make define lower bounds and infimum for a collection $\{R_{i_k}\}$, and the infimum can be defined as $\inf_k\{R_{i_k}\} = \wedge_k R_{i_k}$. Existence of supremums combined with Boolean algebra requirements guarantees existence of infimums and vice-versa.

is called a Boolean map if it preserves the Boolean operations.

An observable is a collection of effects corresponding to such measuring instruments. So, each observable should correspond to a map from a Boolean algebra to the space of effects that respects the Boolean algebra structure. An observable is a pair (Σ_R, E_R), where Σ_R is a Boolean algebra and E_R is a map

$$E_R : \Sigma_R \to \mathcal{A}.$$

which is a Boolean map to its image $E_R(\Sigma_R) \subseteq \mathcal{A}$. In particular $E_R(\Sigma_R)$ will have to be a commutative subalgebra of $\mathcal{A}$. We will denote an observable by its corresponding Boolean map E_R. We will use complete Boolean algebra structure as an idealization of outcomes measurable by a single measuring instrument, since the labels used for possible outcomes in physical theories is the set of real numbers which has this completeness property.

1.2.2 | Stone Duality

The definition of an observable is closely related to the algebraic structures on Boolean algebras. Understand the structure of Boolean algebras can be useful for understanding the mathematical modelling of observables. We now prove Stone's characterisation of Boolean algebras, and describe the notion of a Stonean topological space.

The relation between Stonean spaces and von Neumann algebras, in §**??**, gives us a characterisation of observables in quantum theories.

Stone Space of Σ

Let Ω be a set. Equipped with union, intersection and complements, the powerset $\mathcal{P}(\Omega)$ of Ω is a Boolean algebra, where the empty set and the whole set Ω are the bottom and top elements respectively. Stone's characterisation of Boolean algebras says that every Boolean algebra can be thought of as a Boolean algebra of subsets of a set.

We now start with an abstract completel Boolean algebra Σ. Which for the sake of heuristics and intuition maybe assumed to be the Boolean algebra Σ_R for simultaneously measurable outcomes $R \equiv \{R_i\}$ for heuristics and intuition.

An ideal of Σ is a set $I \subset \Sigma$ containing the bottom element of Σ, and closed under the operation $\vee$, such that

$$R_i \wedge R_j \in I, \quad \forall R_i \in I,\, R_j \in \Sigma.$$

If the Boolean algebra Σ_R corresponds to a collection of simultaneously measurable outcomes $R \equiv \{R_i\}$, and we can think of each R_i as an instrument in itself. The outcome $R_i \vee R_j$ describes the outcome where a prepared system is passed through R_i and R_j in parallel. The closure under $\vee$ says that all such parallel instruments also represent an outcome in I_R.

Similarly, $R_i \wedge R_j$ describes the outcome in which both R_i and R_j are true. Hence the condition $R_i \in I_R$, $R_j \in \Sigma_R$ implies $R_i \wedge R_j \in I_R$, says that I_R contains all such elementary outcomes, describing all of the most accurate instruments.

The Space of Maximal Ideals

The notion of an ideal of a Boolean algebra says that outcomes belonging to the ideal can be expressed by a combination of most accurate instruments. Since the ideal need not be the whole Boolean algebra, it might be the case that the collection does not include all such 'most accurate instruments'.

The notion of maximal ideals helps us distinguish such instruments. An ideal of $I \subset \Sigma$ is called a proper ideal if $I \neq \Sigma$. Maximal ideal is a proper ideal which is not contained in any other proper ideal. Clearly proper ideals cannot contain the top element $\mathbb{I}$, because that would mean all the most accurate instruments would be contained in the ideal. if $R_i \in I$ then I cannot contain $\neg R_i$, because otherwise

$$R_i \vee \neg R_i = \mathbb{I} \notin I$$

By Zorn's lemma we may assume every proper ideal is contained in a maximal ideal.

Theorem 1.1. *$I \subset \Sigma$ is a maximal ideal iff $\forall R_i \in \Sigma$, either $R_i \in I$ or $\neg R_i \in I$.*

Lemma 1.2. *Every proper ideal is contained in a maximal ideal.*

The idea for construction of $S(\Sigma)$ is to view the elements of Σ as functions. Let **2** be the set $\{0, 1\}$.

Similarly, A field of a set is a collection of subsets of the set which is closed under union, intersections and complements. Stone's representation theorem or Stone duality for Boolean algebras suggests that every Boolean algebra Σ is a field of sets of a space, called the Stone space $S(\Sigma)$ for Σ.

The completion properties of Σ are closely related to the topological properties of $S(\Sigma)$

Lemma 1.3. *Every maximal ideal is the kernel of a Boolean homomorphism.*

A topological space is called totally disconnected if every open set is the union of all its subsets that are simultaneously open and closed. A Hausdorff topological space Ω that is compact and totally disconnected is called a Stonean space.

1.3 | Heisenberg's Approach

By the end of the nineteenth century, it was clear that certain elementary processes obeyed 'discontinuous' laws. That is to say, there exist observables whose collection of effects forms discrete sets, and also observables whose collection of effects forms a continuous set.

If the collection of effects of an observable Q can be labeled by a discrete set, the observable E_Q corresponds to a map, which assigns to each collection of values of the observable $Q_F \equiv \{Q_i\}_{i\in F\subseteq\mathbb{N}}$ the effect $\sum_{i\in F} E_{Q_i}$ for the measuring the value to be in the set $\{Q_i\}_{i\in F}$. Hence the observable is a map,

$$E_Q : \mathcal{B}(\mathbb{N}) \to \mathcal{A}.$$

Where $\mathcal{B}(\mathbb{N})$ consists of the subsets of $\mathbb{N}$. Each ensemble μ_ρ is a map $\mu_\rho : \mathcal{A} \to \mathbb{R}$. The composite map is

$$\mu_\rho \circ E_Q \equiv \mu_\rho^Q : \mathcal{B}(\mathbb{N}) \longrightarrow \mathbb{R}.$$

Which assigns to the possible event Q_F its probabilty of occurance $\mu_\rho(Q_F)$ as described above. If the observable is measured, the sum total probability of occurrence of atleast one of the values should be 1. This means that

$$\sum_{i\in\mathbb{N}} \mu_\rho^Q(Q_i) = 1$$

where $\mu_\rho^Q(Q_i)$ is the probability the observable has a value Q_i when the state is μ_ρ. In particular, the sum should make sense. The expected value of the observable R for the state μ_ρ is given by, $\langle Q\rangle_\rho = \sum_{i\in\mathbb{N}} Q_i \mu_\rho^Q(Q_i)$.

On the other hand if the collection of effects of an observable can be labeled by a continuous set, that is, the measurements are labelled by countable unions of open intervals of $\mathbb{R}$, such an observable corresponds to a map, which assigns to intervals R_i of $\mathbb{R}$ the effect E_{R_i} where R_i is the interval in which the value of the observable R lies. Each ensemble μ_ρ corresponds to a function, $\mu_\rho : \mathcal{A} \to \mathbb{R}$. The gives us a composite map,

$$\mu_\rho^R : \mathcal{B}(\mathbb{R}) \longrightarrow \mathbb{R}.$$

Which assigns to the interval R_i a probabilty $\mu_\rho^R(R_i)\lambda(R_i)$ where λ is the standard Lebesgue measure on $\mathbb{R}$. Again as before, if the observable is measured, the sum total probability of occurrence of one of the values should be 1. This means that

$$\int_{\mathbb{R}} \mu_\rho^R(x) d\lambda(x) = 1$$

In particular the integral should make sense. Each state corresponds to the measure $\mu_\rho^R d\lambda$. The expected value of the observable R for the state μ_ρ is given by, $\langle R\rangle_\rho = \int_{\mathbb{R}} x\mu_\rho^R(x) d\lambda(x)$.

1.3.1 | Observables in Classical Theories

Classical theories model observables using continuous functions on a manifold. It is sufficient to assume integrability of the functions for this model. If the manifold is Ω, observables in classical physics correspond to integrable functions of the form $R : \Omega \to \mathbb{R}$. The state in classical physics corresponding to the ensemble ρ is modeled by a positive Radon measure μ_ρ on Ω. The probability measure the state μ_ρ associates with the observable R is

$$\mu_\rho^R : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$$

This is the composite map,

$$\mathcal{B}(\mathbb{R}) \xrightarrow{R^{-1}} \mathcal{B}(\Omega) \xrightarrow{\mu_\rho} \mathbb{R}.$$

Which assigns to the effect E_{R_i} the probabilty $\mu_\rho^R(R_i) = \mu_\rho(R^{-1}(R_i))$. Note that $R^{-1}(R_i)$ is a measurable set in Ω and corresponds to the set of all points in Ω such that the evaluation of the function R lies in the interval R_i. Since the total probability should be 1 we have

$$\int_{R^{-1}(\mathbb{R})} d\mu_\rho(x) = 1.$$

The expected value of the observable R for an ensemble ρ is given by

$$\langle R \rangle_\rho = \int_{R^{-1}(\mathbb{R})} R(x) d\mu_\rho(R^{-1}(x))$$

The characteristic functions which assign to each Borel set in Ω the value 1 correspond to effects in this model. The collection of all characteristic functions on Ω is itself an observable and since every integrable function can be approximated as combinations of simple function, every observable in the classical model coexists with each other.

The space of all measures $\mathcal{M}(\Omega)$ forms a vector space. The set of measures with the above property of which ensures the integral is 1 ensures that it is a convex subset of $\mathcal{M}(\Omega)$. The extreme points of this convex set are the delta distributions which correspond to the points of the manifold Ω. If we denote the convex set of regular probability measures on Ω by $\mathcal{S}(\Omega)$ the state space of the classical theory is

$$\mathcal{S} \equiv \mathcal{S}(\Omega).$$

Since preparation procedures are independent of which measuring instrument is applied the states must assign a probability distribution to each observable. The collection of all integrable functions $L^1(\Omega)$ acts as the space of effects.

This is a good model as long as all the observables take continuum of values. The problem starts when there exists a discrete observable simultaneously alongside a continuous observable. Suppose there exists a discrete observable, then each state $\mu_\rho \in \mathcal{S}(\Omega)$ should also correspond to a probability distribution,

$$\mu_\rho^Q : \mathcal{B}(\mathbb{N}) \to \mathbb{R}$$

such that,

$$\sum_{i \in \mathbb{N}} \mu_\rho^Q(Q_i) = 1$$

where $\mu_\rho^Q(Q_i) = \mu_\rho(Q^{-1}(Q_i))$. Such functions correspond to summable sequences denoted by $l^1(\Omega)$. The coexistence is not possible because the space of integrable functions $L^1(\Omega)$ is not isomorphic to the space of summable sequences $l^1(\Omega)$.[5]

The conclusion we can draw is that if the observables are modeled as maps from a manifold, the state space of the physical system cannot give a probability distribution for each observable. There does not exist a common state space that can accommodate both discrete and continuous observables. So, the problem with classical theory is that it fails to consider some of the observed phenomena. So the physical ideas behind classical theory are too constraining. We now have to figure out what this hidden extra idea is and construct a new mathematically model for observables that does not take this extra idea into account.

[5]One way to prove this is via the so called Schur's property, which says that if a sequence is weakly convergent then it is also convergent in the norm. $l^1(\Omega)$ has the Schur's property and $L^1(\Omega)$ will not, proving that they cannot be isomorphic.

1.3.2 | Hilbert Spaces

Heisenberg's radical solution to the problem of mathematical modeling of observables was to think of observables as operators on a vector space. The spectrum of the operator is to be thought of as values of the observable. Compared to classical theories, this is an extremely abstract and radical change. The classical model of an observable was geometric and hence intuitive. Von Neumann, Hilbert, and others were able to figure out the underlying idea and reformulated it clearly in terms of 'Hilbert spaces'. The key to von Neumann's articulation of Heisenberg's model lies in the isomorphism between the space of square summable functions and the space of square-integrable functions as Hilbert spaces, due to Riesz & Fischer.

As noted before what we want is a convex state space $\mathcal{S}$ that can provide a probability function for both discrete and continuous variables. Each element $\mu_\rho \in \mathcal{S}$ should give rise to functions, $\mu_\rho^Q : \mathcal{B}(\mathbb{N}) \to \mathbb{R}$, and $\mu_\rho^R : \mathcal{B}(\mathbb{R}) \longrightarrow \mathbb{R}$, such that,

$$\sum_{i\in\mathbb{N}} \mu_\rho^Q(Q_i) = 1,$$

and

$$\int_{\mathbb{R}} \mu_\rho^R(x) d\lambda(x) = 1.$$

At this stage, von Neumann defined the notion of a Hilbert space which provides the appropriate mathematical language to get all the ingredients together. Abstractly a Hilbert space is a pair $(\mathcal{H}, \langle\cdot|\cdot\rangle_{\mathcal{H}})$ where $\mathcal{H}$ is a vector space and $\langle\cdot|\cdot\rangle_{\mathcal{H}}$ an inner product on it, and is topologically a complete normed space. Every Hilbert space has an orthonormal basis and an element of the Hilbert space can be uniquely specified by its coordinates with respect to a complete orthonormal system. A Hilbert space is said to be separable if it has a countable basis, and any two separable Hilbert spaces are isomorphic.

A sequence of complex numbers or a function from natural numbers to complex numbers is said to be square summable if

$$\sum_{i\in\mathbb{N}} \big|f(i)\big|^2 < \infty.$$

With pointwise addition and scalar multiplication, the set of all square summable sequences is a complex vector space. It can be endowed with an inner product,

$$\big\langle f|g\big\rangle_{l^2} = \sum_{i\in\mathbb{N}} \overline{f(i)} g(i).$$

Together with this inner product the space of square summable sequences of complex numbers is a Hilbert space denoted by $l^2(\mathbb{N})$. A measurable function from the real line to complex numbers is called square integrable if,

$$\int_{\mathbb{R}} \big|f(x)\big|^2 d\lambda(x) < \infty,$$

where λ is the standard Lebesgue measure on $\mathbb{R}$. The collection of all square-integrable functions is a complex vector space. This vector space can be endowed with an inner product,

$$\big\langle f|g\big\rangle_{L^2} = \int_{\mathbb{R}} \overline{f(x)} g(x) d\lambda(x).$$

Two square integrable functions are equivalent if they are same almost everywhere with respect to λ. The collection of equivalence classes of square integrable functions inherits a

vector space structure from the space of square integrable functions. Together with the inner product $\langle\cdot|\cdot\rangle_{L^2}$ the space of equivalence classes of square-integrable functions is a Hilbert space, denoted by $L^2(\mathbb{R})$.

Theorem 1.4. (Riesz-Fischer) *$L^2(\mathbb{R}) \cong l^2(\mathbb{N})$ as Hilbert spaces.*

This isomorphism provides us with a state space that allows for coexistence of discrete and continuous observables. See §1.4 [10] for a proof of Riesz-Fischer theorem. This isomorphism acts as the starting point for von Neumann's reformulation of Heisenberg's model. Von Neumann's approach was to compare the the space of functions on discrete and continuous spaces instead of comparing discrete space and continuous space themselves. The isomorphism as Hilbert spaces of the space of square integrable functions and the space of square summable sequences allows us to develop a unified mathematical model where coexistence of both discrete and continuous observables is possible. If observables are treated as self-adjoint operators on a separable Hilbert space $(\mathcal{H}, \langle\cdot|\cdot\rangle_{\mathcal{H}})$, then the collection of all unit length vectors can be used as the extreme points of a state space common to both discrete and continuous observables. The values of the observable correspond to the spectrum of the operator. The interpretation is that the self-adjoint operator via spectral theorem, which we will discuss later on, gives rise to a collection of projection operators which correspond to the collection of effects for the observable, and the length of the projection for a state corresponds to the probability of the state undergoing a change for that effect.

2 | Quantum Theory in Hilbert Spaces

The discussion of the previous section suggests that if we assume the coexistence of continuous and discrete observables, the algebra structure on the space of effects can be mathematically modeled using the structure of a von Neumann algebra. We now describe the special case of quantum theory in the Hilbert space formalism, and describe the space of effects and the space of ensembles. The necessary structure for the abstract mathematical framework of quantum theory can be found in Hilbert spaces and operator algebras on Hilbert spaces.

2.1 | The Lattice of Closed Subspaces

We need to study the algebra of operators to understand the structure of the space of effects $\mathcal{E}$. Let $(\mathcal{H}, \langle\cdot|\cdot\rangle)$ be the complex separable Hilbert space. $\mathcal{P}(\mathcal{H})$ denote the set of all closed subspaces. Denote $\mathcal{H}_i \leq \mathcal{H}_j$ if and only if $\mathcal{H}_i \subseteq \mathcal{H}_j$. The relation $\leq$ is a partial ordering in $\mathcal{P}(\mathcal{H})$. Join $\vee$ of a family $\{\mathcal{H}_i\}_{i\in I}$ is the linear span of the family denoted $\vee_i \mathcal{H}_i$. Meet $\wedge$ of a family $\{\mathcal{H}_i\}_{i\in I}$ is the intersection of the family, denoted $\wedge_i \mathcal{H}_i$. The orthocomplement of $\mathcal{H}_i$ in $\mathcal{P}(\mathcal{H})$ denoted by $\mathcal{H}_i^{\perp}$ is the closed subspace of vectors $\varphi \in \mathcal{H}$ such that $\langle\varphi|\mathcal{H}_i\rangle = 0$. Since there is a bijection between closed subspaces of the Hilbert space $\mathcal{H}$ and projection operators acting on the Hilbert space, the set of all projection operators on the Hilbert space inherits a lattice structure from the lattice of closed subspaces. Abusing notation, we will denote the projection operators on $\mathcal{H}$ by $\mathcal{P}(\mathcal{H})$. The orthocomplement of the projection E is the projection onto the orthogonal complement of the subspace corresponding to the projection operator E and is denoted by $E^{\perp}$. It follows that

$$E^{\perp} = I - E.$$

The lattice structure of $\mathcal{P}(\mathcal{H})$ coming from the above relations gives us the necessary structure to get the mathematical representatives of physical observables.[6] The non-Boolean lattice $\mathcal{P}(\mathcal{H})$ of projections should act as the space of effects

$$\mathcal{E} \equiv \mathcal{P}(\mathcal{H}).$$

Let R_i be the outcome of an measuring equipment. Let the corresponding effect be the projection operator given by E_{R_i}. We can think of the equipment not undergoing any changes as an outcome itself, and this corresponds to the outcome $\neg R_i$. The effect corresponding to the outcome $\neg R_i$ is the projection operator to the complement of the subspace spanned by E_{R_i},

$$E_{\neg R_i} = E_{R_i}^{\perp}.$$

If R_i and R_j are outcomes of two measuring equipments then the effect corresponding to when either of the two equipments undergo a change will correspond to the projection to the closed subspace spanned by the two closed subspaces.

$$E_{R_i \vee R_j} = E_{R_i} + E_{R_j}.$$

We can treat a system prepared in an initial state undergoing a measurement for R_i to be a measuring equipment by itself. If this preparation undergoes a change for the outcome R_j we have a composite outcome, and corresponds to the product,

$$E_{R_j R_i} = E_{R_j} E_{R_i}.$$

The order of experiment is important. In the definition of an observable the order was unimportant because an observable consists of a collection of effects which can be measured simultaneuously since they are measured by the same measuring equipment. In this general case the effects might correspond to multiple measuring equipments. Since multiple measuring equipments cannot be simultaneuously applied, we cannot expect such a commutative structure.

For a family of projection operators to represent an observable, we should make sure that the family forms a Boolean algebra. A quantum mechanical observable is an additive measure of the form,

$$E_A : \Sigma_A \to \mathcal{P}(\mathcal{H}).$$

Such maps are called a projection valued function. In physical experiments, the statements usually are of the form 'the value of the observable lies in some set ϵ_i of real numbers'. To accommodate the fact that the measurement scale is consisting of real numbers, we identify Σ_A with the Boolean algebra of Borel sets of $\mathbb{R}$.[7]

The quantum observables are analogous to classical random variables, namely, that of a projection valued measure,

$$E_A : \mathcal{B}(\mathbb{R}) \to \mathcal{P}(\mathcal{H}).$$

[6] In the quantum logic literature the central objects that model possible events are the so called orthomodular lattices. Orthomodular lattices are studied abstractly in [12].

[7] It should be noted that the observables need not be real, the physics community has historically decided to use real numbers to label the outcomes of experiments. Any other labeling should work equally well. Döring and Isham have done an interesting generalization of this scheme [19]. Their idea is to replace the Boolean structure in Σ_A with a more general propositional language system and question if values of the system should be more general than 'real'. Though I find this to be a beautiful generalization for the future of quantum theory I do not think this is the part needing fixing for solving the foundational problems in quantum theory. I believe we can get a lot of work done with real measurement scales themselves.

This generalizes the classical case, for which mathematical representatives were the measure space $(\Omega, \Sigma(\Omega), \mu)$, where the σ-algebra, $\Sigma(\Omega)$ is a class of subsets of the set Ω which correspond to events and μ is a probability measure. A classical random variable is defined as a map $X : \Omega \to \mathbb{R}$. The map doing the work in assigning necessary probabilities is its pre-image, considered as a set map,

$$X^{-1} : \mathcal{B}(\mathbb{R}) \to \Sigma(\Omega).$$

A spectral measure is a projection operator-valued function E defined on the sets of $\mathbb{R}$ such that, $E(\mathbb{R}) = I$ and $E(\coprod_{i\in\mathbb{N}} \epsilon_i) = \sum_{i\in\mathbb{N}} E(\epsilon_i)$, where ϵ_is are disjoint Borel sets of $\mathbb{R}$. The spectral theorem says that every self-adjoint operator A corresponds to a spectral measure E_A such that,

$$A = \int_{\mathbb{R}} \lambda \, dE_A(\lambda),$$

and conversely, every spectral measure corresponds to a self-adjoint operator. In the finite-dimensional case this reduces to $A = \sum_i \lambda_i E_i$ where E_i is the projections onto eigenspaces of the eigenvalues λ_i. The special feature of quantum theory is that, unlike classical mechanics, the space of effects is a non-commutative algebra.

2.1.1 | Gleason's Theorem

States are linear functionals on the space of effects. In case of quantum theory the space of effects is identified with projection operators on $\mathcal{H}$ and hence the mathematical representatives of the physical states in the quantum case are the maps,

$$\omega : \mathcal{P}(\mathcal{H}) \to \mathbb{R}$$

such that $\omega(0) = 0$, $\omega(E^\perp) = 1 - \omega(E)$ and $\omega(\sum_i E_i) = \sum_i \omega(E_i)$, for mutually orthogonal E_i. They naturally extend linearly to the space of bounded operators $\mathcal{B}(\mathcal{H})$ on $\mathcal{H}$.

A positive linear functional on the space of bounded operators $\mathcal{B}(\mathcal{H})$ on the Hilbert space $\mathcal{H}$ is called a non-commutative probability measure or a normal state if for every projection operators $E \in \mathcal{P}(\mathcal{H})$,

$$\omega(E^\perp) = 1 - \omega(E)$$

and for a countable collection of mutually orthogonal projection operators $\{E_i\}_{i\in\mathbb{N}}$,

$$\omega\big(\textstyle\sum_{i\in\mathbb{N}} E_i\big) = \displaystyle\sum_{i\in\mathbb{N}} \omega(E_i).$$

When restricted to commutative subalgebras, non-commutative probability measures correspond to probability measures. For each observables corresponding to a spectral measure $E_A : \mathcal{B}(\mathbb{R}) \to \mathcal{P}(\mathcal{H})$, the composition,

$$\mathcal{B}(\mathbb{R}) \xrightarrow{E_A} \mathcal{P}(\mathcal{H}) \xrightarrow{\omega} [0, 1],$$

satisfies the conditions of a probability measure. So, each state ρ assigns to an observable E_A a probability measure, $\mu^A_\omega \equiv \omega \circ E_A : \mathcal{B}(\mathbb{R}) \to [0, 1]$. $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu^A_\omega)$ is a classical probability space, where μ^A_ω is a probability measure. The Gleason's theorem characterises non-commutative probability measures on Hilbert spaces.

Theorem 2.1. (Gleason) *If the complex separable Hilbert spaces $\mathcal{H}$ of dimension greater than 2, then every non-commutative measure ω is of the form*

$$\omega(E) = Tr(\rho E),$$

where ρ is a positive semidefinite self-adjoint operator of unit trace.

Sketch of Proof

The idea of Gleason's proof is to reduce the general case to the case of two dimensions, where one can use relatively simpler tools. The starting point is the definition of the so called frame functions. A frame function f of weight W on the separable Hilbert space $\mathcal{H}$ is a real valued function on the unit sphere of $\mathcal{H}$ such that for any orthonormal basis, $\{|\varkappa_i\rangle\}_{i\in\mathbb{N}}$, $\sum_i f(|\varkappa_i\rangle) = W$. A frame function is regular if there exists a self-adjoint operator ρ such that for every unit vector $|\varkappa\rangle \in \mathcal{H}$,

$$f(|\varkappa\rangle) = \langle\varkappa|\rho\varkappa\rangle$$

Gleason proves that every frame function on two dimensional Hilbert spaces is regular. For Hilbert spaces of dimension greater than three the result holds for the restriction to every two dimensional subspace. Then the continuity of frame functions is proved. Every non-negative frame function on a Hilbert space of dimension greater than three is regular. Much of the hard work lies in this part. We will skip this hard and technical part. See Gleason's original paper [20] for details.

Suppose $\omega : \mathcal{P}(\mathcal{H}) \to [0,1]$ is a function such that $\omega(0) = 0$, $\omega(E^\perp) = 1 - \omega(E)$ and $\omega(\vee_i E_i) = \sum_i \omega(E_i)$ for mutually orthogonal E_i. Let E_φ be the projection onto the subspace spanned by the unit vector $|\varphi\rangle$. Then $f(|\varphi\rangle) = \omega(E_{|\varphi\rangle})$ defines a non-negative frame function. By regularity there exists a self-adjoint operator ρ such that,

$$f(|\varphi\rangle) = \langle\varphi|\rho\varphi\rangle.$$

Since this holds for all unit vectors, ρ is positive semi-definite. Denote by $E_\mathcal{H}$ the projection onto the whole Hilbert space which is the identity operator. Given an orthonormal basis $\{|\varphi_i\rangle\}_{i\in\mathbb{N}}$ of $\mathcal{H}$ we have,

$$\omega(E_\mathcal{H}) = \sum_i \omega(E_{|\varphi_i\rangle}) = \sum_i \langle\varphi_i|\rho\varphi_i\rangle = Tr(\rho).$$

For any closed subspace $\mathcal{K} \subset \mathcal{H}$ denote by $E_\mathcal{K}$ the corresponding projection operator. Take an orthonormal basis $\{|\varkappa_i\rangle\}_{i\in I}$ for $\mathcal{K}$ and extend it to one of $\mathcal{H}$. Then we can write

$$\omega(E_\mathcal{K}) = \sum_{i\in I} \omega(|\varkappa_i\rangle) = \sum_{i\in I} \langle E_\mathcal{K}\varkappa_i|\rho\varkappa_i\rangle = Tr(\rho E_\mathcal{K}).$$

So for all projection operators $E \in \mathcal{P}(\mathcal{H})$,

$$\omega(E) = Tr(\rho E).$$

□

Positive semidefinite self-adjoint operator of unit trace are called density matrices. Note that the converse of the theorem clearly holds, that is, every density matrix determines a state as defined in the above formula. More general quantum experiments correspond to positive operator-valued measures. The effects E are given by positive operators, $O \leq E \leq I$. Since

these should sum to 1 for an experiment, it will be a resolution of identity $\sum_i E_{A_i} = I$, where the E_{A_i}'s are effects. The resolution of identity $E_A : A_i \to E_{A_i}$ is called positive operator-valued measure (POVM). General quantum mechanical experiments are represented by pairs (ρ, E_A). For Gleason's theorem in this setting see [17].

We call ω satisfying the equivalent conditions of Gleason's theorem a Gleason measure. Every state corresponds to a positive semidefinite self-adjoint operator of unit trace. We denote the set of all states on the Hilbert space $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$. The mathematical representatives of ensembles are states,

$$\mathcal{S} \equiv \mathcal{S}(\mathcal{H})$$

This is a convex set. The topology on $\mathcal{S}(\mathcal{H})$ comes from the inner product on $\mathcal{H}$.

The Krein-Milman theorem says that a compact convex subset of a Hausdorff locally convex topological vector space is equal to the closed convex hull of its extreme points. The extreme points of this convex set are called pure states, pure states satisfy $\rho = \rho^2$ and correspond to some vector $|\varphi\rangle$ in the Hilbert space $\mathcal{H}$ and ρ is the projection onto the subspace generated by $|\varphi\rangle$. Such a state is denoted by $|\varphi\rangle\langle\varphi|$.

For an observable with an associated self-adjoint operator A the probability that the observable takes a value lying in the interval ϵ is given by,

$$\mu_\rho^A(\epsilon) = Tr(\rho E_A(\epsilon)).$$

The expectation value of the observable will be,

$$\langle A \rangle = \int_{\mathbb{R}} \lambda \, d\mu_\rho^A(\lambda) = Tr(\rho A).$$

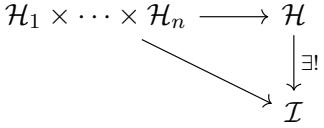
All Hilbert spaces will be assumed to be separable and complex.

Tensor Product

Given a finite number of Hilbert spaces $\mathcal{H}_i$, for n quantum systems, the problem is to describe the Hilbert space appropriate to the 'product' system. Let $\mathcal{I}$ denote a possible solution to this problem: that is the states of $\mathcal{I}$ are supposed to be states for the product system. Then, at the very least, some of the preparation procedures for the product system should be obtainable by arranging in some manner the preparation procedures on the individual systems. We should be able to construct a certain function,

$$\mathcal{H}_1 \times \cdots \times \mathcal{H}_n \to \mathcal{I}.$$

The interpretation of this map is that it introduces a component from each individual system into the product system. Accounting for the superpositions, the product system should inherit the structure from the components. The map above must be linear for each component. The universal solution $\mathcal{H}$ to this problem is the algebraic tensor product. It is the vector space $\mathcal{H}$ together with an n-linear map such that, for any n-linear map $\mathcal{H}_1 \times \cdots \times \mathcal{H}_n \to \mathcal{I}$, there exists a unique linear map : $\mathcal{H} \to \mathcal{I}$,



This vector space inherits a canonical inner product from the component Hilbert spaces. The completion of

$$\mathcal{H} = \bigotimes_{n \geq i \geq 1} \mathcal{H}_i$$

under the canonical inner product will serve as the Hilbert spaces for product systems. Such a topology is justified because if two states are near to each other in a component, then we expect them to be near to each other after this subsystem is introduced into the product system.

References

[1] G Ludwig, Foundations of Quantum Mechanics I. Springer-Verlag, 1983

[2] G Ludwig, Foundations of Quantum Mechanics II. Springer-Verlag, 1985

[3] G Ludwig, An Axiomatic Basis for Quantum Mechanics Volume 1. Springer-Verlag, 1985

[4] K Kraus, States, Effects, and Operations. Lectures Notes in Physics, Springer-Verlag, 1983

[5] H Primas, Chemistry, Quantum Mechanics and Reductionism. Springer-Verlag, 1983

[6] S Gudder, Convexity and Mixtures. SIAM Review, Vol. 19, No. 2, April 1977

[7] J M Jauch, Foundations of Quantum Mechanics. Addison-Wesley Publishing Company, 1968

[8] V Moretti, Fundamental Mathematical Structures of Quantum Theory. Springer, 2018

[9] M Redei, Quantum Logic in Algebraic Approach. Springer, 1998

[10] J von Neumann, Mathematical Foundations of Quantum Theory. Princeton University Press, 1955

[11] V S Varadarajan, Geometry of Quantum Theory. Van Nostrand, New York, 1970

[12] G Kalmbach, Orthomodular Lattices. Academic Press, 1983

[13] P Busch, M Grabowski, P Lahti, Operational Quantum Physics. Springer, 1995

[14] A M Gleason, Measures on the closed subspaces of a Hilbert space. Journal of Mathematics and Mechanics, vol. 6, no. 6, pp. 885–893, 1957

[15] H Granström, Gleason's theorem. Master's thesis, Stockholm University, `http://www.kiko.fysik.su.se/en/thesis/helena-master.pdf`, 2006

[16] A Peres, Quantum Theory: Concepts and Methods, Kluwer Academic Publishers, 2002

[17] P Busch, Quantum states and generalized observables: A simple proof of Gleason's theorem. Physical Review Letters 91(12):120403. `http://arxiv.org/abs/quant-ph/9909073`, 2003

[18] J S Lomont, P Mendelson, The Wigner Unitarity-Antiunitarity Theorem. Annals of Mathematics, Second Series, Vol. 78, No. 3, pp. 548-559, 1963

[19] A Döring, C Isham, What is a thing? `https://arxiv.org/abs/0803.0417v1`, 2008

[20] A M Gleason, Measures on the closed subspaces of a Hilbert space. Journal of Mathematics and Mechanics, vol. 6, no. 6, pp. 885–893, 1957

[21] H Granström, Gleason's theorem. Master's thesis, Stockholm University, `http://www.kiko.fysik.su.se/en/thesis/helena-master.pdf`, 2006