On the Foundations of Experimental Statistical Sciences

Written July 1979

George Svetlichny[1]

[1]Departamento de Matmática, Pontifícia Universidade Católica, Rio de Janeiro, Brazil
svetlich@mat.puc-rio.br http://mat.puc-rio.br/~svetlich

Contents

PREFACE

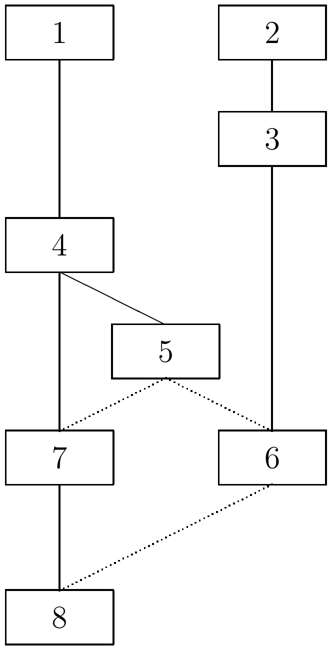
The many attempts at trying to explain the extreme regularity of the set of propositions referring to physical systems, exemplified by the Boolean algebra of subsets of phase space of classical systems, and the orthomodular lattice of projections in a Hilbert space of quantum systems, has resulted in various axiomatic foundations [1, 2, 3]. To a large extent, these have not been totally satisfactory seeing that in all of them some of the axioms seem very *ad hoc* and removed from any operational or phenomenological interpretation. Rather than attempt yet a different formulation, we investigate the properties that can be expected of a completely general statistical theory without trying to impose assumptions that would lead to one of the standard physical examples. The hope is that by understanding better the general theories, we can perhaps see better what is so special about the special ones. Following one line of investigation, we find that statistical theories can be manipulated rather freely, being objects of a bicomplete monoidally closed category. In a sense this provides more machinery than we use but it's unsuspected machinery with much potential. Another line of study leads to what is probably the main result of this work: the identification of those measuring procedures that cannot be interpreted as involving ad hoc interferences by the experimenter. This leads to the result that theories with only two pure states can be imbedded in Boolean theories with restriction on state production, adding yet another example of statistical theories with a natural associated lattice of propositions, a Boolean algebra in this case.

Our studies were originally motivated by physical examples, especially by the challenging and deceitful guide of the foundations of quantum mechanics, but the formalism is purposefully general, and we interpret many of the objects in terms of other types of experiments. Mathematical modelling in biology and psychology should be quite different from either Boolean or quantum modelling, and it is these sciences that we eventually hope to benefit.

A brief description of the content of each chapter is as follows:

In Chapter 1 we set the stage by presenting mostly already well known material, formalizing the notion of state and yes-no observation in terms of a certain duality of convex sets. Nothing is really new except for some examples and possibly some of the commentary. We follow this in Chapter 2 by a presentation of convex set theory from the point of view of universal algebra. This allows us to view convex sets as objects of a bicomplete monoidally closed category. In Chapter 3 we imbed the convex set duality introduced in the first chapter into a canonical vector space duality following a pattern already established in the literature. Chapter 4 goes beyond the first chapter in introducing into the formalism simultaneous measurements and many-exit operations on states. We follow an axiomatic approach, introducing these as primitive entities and axiomatizing all the properties they should have based on their phenomenological interpretation. This is the essential formalism of this work. The short Chapter 5 briefly deals with the introduction of ideal objects in statistical theories, that is, objects that are not physically realizable, but arbitrarily well approximated by physically realizable ones. We use this notion mainly to be able to assume in subsequent chapters that all convex sets are algebraically closed, although the notion of idealization naturally appears in various other places. Chapter 6 returns to the simple statistical theories introduced in Chapter 1 and using the categorical machinery of the second chapter shows that these theories form a bicomplete monoidally closed category. This is a calculus of statistical theories. We also introduce another category of theories, with not so many nice properties, but one more realistically reflecting the intuitive notion of strength of theories and providing for the formalization of localization. Chapter 7 contains the most important results. we identify, by an analysis of the complexity of states, those measuring procedures that cannot be interpreted as containing interferences by the experimenter that would produce spurious complexity in the observed data. In parallel with this we explore the notion of an entropy function, stopping somewhat short however of a fully satisfying treatment. The chapter also contains various passages related to quantum mechanical models, pointing out the special nature of these. The last Chapter 8 takes up the theme of lattices of propositions using the concepts of the previous chapter. Scratching only the surface, we discover a Boolean algebra of propositions in theories with only two pure states, and include some commentary on the general case.

We diagram below the logical interdependence of the chapters. Dotted lines mean minimal dependence that can be successfully handled by recourse to the indices of symbols and definitions provided at the end.



Many people contributed to this work. Special thanks go to Paul Otterson who provided the maze example of Chapter 8, and who contributed with helpful discussions and observations, to Carlos Malamut, whose master's thesis [4] incorporated some of the material, and without whose interest the intricacies of foundation studies might have overpowered the original motivation for this project, and to Stuart Turner for help in the proof of Theorem 7.7.

This work was supported financially by the Financiadora de Estudos e Projetos (FINEP) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), agencies of the Brazilian Government which deserve our gratitude.

Chapter 1

Preliminaries

In this work we consider the formal foundations of statistical experimental sciences. Such a science involves the following type of experimentation: by certain well defined procedures, we prepare a state of affairs for study, and at some later time, according to other well defined procedures, we observe whether the so obtained state of affairs possesses or not a gives property. What interests a statistical science is the asymptotic frequency with which repeated preparations yield a given observation. We shall not here go into the problem of bow we identify repeated executions of experimental procedures as being a repetition of the same experiment. Such an identification is necessary in order to make valid sense of the frequency of occurrence of some property and we suppose this problem is solved in some manner. A science may not be wholly statistical. Psychology for example builds models of behavior on other than statistical grounds, yet part of psychology can be handled by a statistical theory. It has only to be recalled that the multivariate statistical approach known as factor analysis was developed primarily by psychologists of human behavior. In animal psychology also, the frequency of a particular type of behavior is often the subject of research, as in learning for example. Thus a given non-statistical science can have a statistical theory associated with it and a quantitative model can be built for this part. Certain properties of this statistical part can be ascertained directly from experimental data, independently of what model one uses for the presumed mechanisms by which the data is produced. Such experimentally determinable properties should shed light not only on the problem of model construction but also on the non-statistical aspects as well. It is these properties that we wish to investigate.

At the minimum, we can say that any formalization of a statistical science must begin by introducing two sets $\mathcal{S}$ and $\mathcal{O}$ and a partial function $\langle\cdot,\cdot\rangle$ with domain $D \subset \mathcal{S} \times \mathcal{O}$ and with values in the interval $[0,1]$. Each $\sigma \in \mathcal{S}$ is interpreted as a preparation procedure and each $p \in \mathcal{O}$ as an observation procedure. If $(\sigma, p) \in D$ then p is applicable to σ and $\langle\sigma, p\rangle$ is the asymptotic frequency of times the given property is observed under repeated preparations of σ and observations of p. While strictly speaking, and according to this formalism, not every observation procedure is applicable to every preparation procedure (confronting a hungry rat with an optical collimator rather than a T maze is not likely to be considered a legitimate experiment), we shall immediately sidestep this issue by formally setting $\langle\sigma, p\rangle = 0$ if $(\sigma, p) \notin D$, thus not bothering to formalize the notion of applicability. From now on $\langle\cdot,\cdot\rangle$ is a total function.

We call $\mathcal{S}$ the *set of states* and $\mathcal{O}$ the *set of observations*. It should be stressed from the outset that every state should be considered as representing an *ensemble*, that is, a hypothetical population that we can realize practically only in part by repeated preparations. In what follows a preparation procedure shall be, at a convenient point in the discussion, identified with the ensemble it is hypothetically capable of generating, identifying thus procedures that generate the same ensemble. The state of affairs resulting from each particular preparation we shall call a *copy* of the state. Similar statements hold for the observations. Each particular realization of a preparation of a state a followed by a realization of an observation p, we call an *execution* of the *experiment* (σ, p). Only by a repetition of a large number of such executions can the number $\langle\sigma, p\rangle$ be computed since with a single execution the result is simply *yes* or *no*. Each triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle$ we shall call a *statistical system* or *system* for short.

It may well happen that different procedures yield the same state of affairs. In such cases it must be true that if σ_1 and σ_2 really produce the same state of affairs, then within experimental error $\langle\sigma_1, p\rangle = \langle\sigma_2, p\rangle$ for all $p \in \mathcal{O}$. Similarly different observation procedures can be connected with the same property, and if p_1, and p_2 are two such, then again within experimental error, $\langle\sigma, p_1\rangle = \langle\sigma, p_2\rangle$ for all $\sigma \in \mathcal{S}$. The notion of same state and property generally come from extrastatistical grounds. Thus in flipping a coin there is only one preparation: flip; there are two observations: heads (H) and tails (T). Thus $\mathcal{S}$ can be taken as the one point set $\{*\}$ and $\mathcal{O}$ as the two point set $\{\mathrm{H}, \mathrm{T}\}$ and if the coin is fair we have $\langle *, \mathrm{T}\rangle = \langle *, \mathrm{H}\rangle = 1/2$. Statistically heads and tails are the same, but phenomenologically different. At this stage of the discussion however, we shall identify statistically equivalent elements, for the simplicity so obtained compensates for the loss of certain distinctions which however we shall later to a large extent recover by a more thorough formalism. In any case if $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ is a system as introduced above, we define an equivalence relation in $\mathcal{S}$ and $\mathcal{O}$ as follows:

$$\begin{aligned} \sigma \sim \sigma' \quad &\Leftrightarrow \quad \forall p \in \mathcal{O}, \quad \langle \sigma, p\rangle = \langle \sigma', p\rangle; \\ p \sim p' \quad &\Leftrightarrow \quad \forall \sigma \in \mathcal{S}, \quad \langle \sigma, p\rangle = \langle \sigma, p'\rangle. \end{aligned}$$

Of course if $\mathcal{S}$, $\mathcal{O}$, and $\langle\cdot,\cdot\rangle$ are experimentally determinable objects, such equalities hold only up to a certain error. We do not enter here into the investigation of how to test hypotheses of the form $\langle \sigma, p\rangle = \langle \sigma', p\rangle$. In any case, once these relations are introduced, we can pass to the quotient sets $\mathcal{S}/\!\sim$, $\mathcal{O}/\!\sim$ and the corresponding function, still denoted by the same symbol, $\langle\cdot,\cdot\rangle : \mathcal{S}/\!\sim \times \mathcal{O}/\!\sim \to [0,1]$ defined by $\langle [\sigma], [p]\rangle = \langle \sigma, p\rangle$ where from now on we use the bracket to denote equivalence classes of elements. We call $(\mathcal{S}/\!\sim, \mathcal{O}/\!\sim, \langle\cdot,\cdot\rangle)$ the *reduced system* of $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$. We say that a statistical system is *separated* if the families of functions $\{\langle\cdot, p\rangle \mid p \in \mathcal{O}\}$ and $\{\langle \sigma, \cdot\rangle \mid \sigma \in \mathcal{S}\}$ separate points of $\mathcal{S}$ and $\mathcal{O}$ respectively. The reduced system is always separated. In a separated system states and observations are determined entirely by their statistical behavior. Phenomenologically this is not always the case as the coin tossing example shows. For rich systems with complicated $\mathcal{S}$ and $\mathcal{O}$ we may encounter empirical separated systems. This is believed to be the case of quantum mechanics for example, where in principle only statistical behavior is observable. However, we have no means of knowing when this should be the case, and should not make a principle of it.

Once we are working with a separated system it is advantageous to introduce further structure in the sets $\mathcal{S}$ and $\mathcal{O}$. It makes sense to assume they are convex in the sense that if $\sigma_1, \sigma_2 \in \mathcal{S}$ and $0 < \lambda < 1$ we can talk of the element $\lambda\sigma_1 + (1-\lambda)\sigma_2 \in \mathcal{S}$. Given procedures for the preparation of σ_1 and σ_2 this new state is prepared by the following procedure: choose a number ξ randomly in $[0,1]$ from a uniformly distributed population (by a random number table say); if it turns out that $\xi < \lambda$ prepare σ_1 if $\xi > \lambda$ prepare σ_2. Certainly this is a practical procedure and we can assume $\mathcal{S}$ is closed under convex combinations. As before, we here gloss over some facts: in practice λ is a rational number while we are assuming the above can be carried out for any λ. Similar convex combinations can be introduced in $\mathcal{O}$ and we assume they have been. Remembering that we are in a reduced triple we should note that what is of practical importance is to able to identify a given state a as being a mixture $\lambda\sigma_1 + (1-\lambda)\sigma_2$, $0 \neq \lambda \neq 1$ even though the preparation leading to σ may not involve such deliberate mixing procedures. An example of this is furnished by physics of molecular beams. A beam of molecular hydrogen consists of a certain fraction f of orthohydrogen and a fraction $1-f$ of parahydrogen . Hence we can write $\sigma = f\sigma_{\text{ortho}} + (1-f)\sigma_{\text{para}}$, yet the preparation of σ involved no random number table contingent decisions. Of course it's possible to build apparatus to prepare pure orthohydrogen and pure parahydrogen and these can be used along with a random number generator to recreate a state having identical statistical properties as σ above. This is not the main point, the main point is the importance of identifying mixed states, as this implies a simplification of the study of their statistical behavior, the problem passing on to the components of the mixture. The present work has to a large extent been influenced by the prototype of the statistical behavior of beams of physical particles applying the ideas however to far removed contexts. A "beam" then becomes simply an ensemble of prepared states be they physical, biological, psychological or otherwise. In such a way we propose to unify the study of statistical sciences, physics becoming simply a particular case.

It should now be noted that by the frequentistic interpretation of $\langle\cdot,\cdot\rangle$ this function is biaffine. That is:

$$\begin{aligned}\langle\lambda\sigma_1 + (1-\lambda)\sigma_2, p\rangle &= \lambda\langle\sigma_1, p\rangle + (1-\lambda)\langle\sigma_2, p\rangle,\\ \langle\sigma, \lambda p_1 + (1-\lambda)p_2\rangle &= \lambda\langle\sigma, p_1\rangle + (1-\lambda)\langle\sigma, p_2\rangle\end{aligned}$$

for all λ, σ, σ_1, σ_2, p_1, p_2. To see this, consider the first expression. Suppose that we have made a large number N of repetitions of the experiment $(\lambda\sigma_1 + (1-\lambda)\sigma_2, p)$. In terms of mean expected behavior of these N repetitions, λN correspond to copies of σ_1 and $(1-\lambda)N$ to copies of σ_2. Of these λN instances of σ_1 we obtain $\lambda N\langle\sigma_1, p\rangle$ positive observations of p, and of the $(1-\lambda)N$ instances of σ_2 we obtain $(1-\lambda)N\langle\sigma_2, p\rangle$ positive observations of p. Thus the total number of positive responses is $\lambda N\langle\sigma_1, p\rangle + (1-\lambda)N\langle\sigma_2, p\rangle$ meaning that the frequency is $\lambda\langle\sigma_1, p\rangle + (1-\lambda)\langle\sigma_2, p\rangle$. By the frequentistic interpretation this frequency is precisely $\langle\lambda\sigma_1 + (1-\lambda)\sigma_2, p\rangle$ demonstrating the first equality. A similar argument works for the second equality.

In addition to the convex structure we can assume further additional structure for $\mathcal{O}$. Since each $p \in \mathcal{O}$ corresponds to a *yes-no* observation we can consider the observation $\neg p$ obtained by interchanging *yes* and *no*. We thus assume that $\mathcal{O}$ is provided with the map $\neg : \mathcal{O} \to \mathcal{O}$. By the frequentistic interpretation we find, using the same methods as in the previous paragraph, that

$$\begin{aligned} \langle \sigma, \neg p \rangle &= 1 - \langle \sigma, p \rangle, \\ \neg(\lambda p_1 + (1-\lambda)p_2) &= \lambda \neg p_1 + (1-\lambda)\neg p_2 \end{aligned}$$

for all λ, σ, p, p_1, p_2. From this it is also clear using the fact that the system is separated that

$$\neg\neg p = p.$$

Thus $\neg$, which is called *negation*, is an affine involution of $\mathcal{O}$.

We further assume that $\mathcal{O}$ is provided with a distinguished element $\mathbf{1}$ having the property that $\langle \sigma, \mathbf{1} \rangle = 1$ for all σ. We interpret $\mathbf{1}$ as being the observation of a tautological property, or simply the observation that a state has been prepared. The negation $\neg\mathbf{1}$ of $\mathbf{1}$ we denote by $\mathbf{0}$ and this corresponds to an observation of a contradictory property, and as an element of $\mathcal{O}$ satisfies $\langle \sigma, \mathbf{0} \rangle = 0$ for all σ.

We are thus able at this point to make a formal definition:

Definition 1.1 *A* statistical triple *is a separated statistical system* $(\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ *where* $\mathcal{S}$ *and* $\mathcal{O}$ *are convex sets and* $\langle \cdot, \cdot \rangle : \mathcal{S} \times \mathcal{O} \to [0, 1]$ *is a biaffine map, such that the following axioms are satisfied:*

1. $\forall p \in \mathcal{O}, \quad \exists \neg p \in \mathcal{O}$ *such that* $\forall \sigma \in \mathcal{S}, \quad \langle \sigma, \neg p \rangle = 1 - \langle \sigma, p \rangle,$
2. $\exists \mathbf{1} \in \mathcal{O}$ *such that* $\forall \sigma \in \mathcal{S}, \quad \langle \sigma, \mathbf{1} \rangle = 1.$

We note that an easy consequence of (1) and separatedness is that $\neg p$ is unique and that the map $\neg : \mathcal{O} \to \mathcal{O}$ is affine. We likewise deduce, as has already been mentioned, the existence of the element $\mathbf{0} \in \mathcal{O}$ such that $\forall \sigma \in \mathcal{S}, \langle \sigma, \mathbf{0} \rangle = 0$.

We have deliberately avoided mentioning the vector spaces in which $\mathcal{S}$ and $\mathcal{O}$ are to be considered as convex subsets. For our immediate purposes it doesn't matter, and since we shall in the next chapter firstly dispense with them by an intrinsic axiomatization of convex sets and secondly reinstitute them in a canonical manner, we gloss over the problem here, pausing only to introduce the following terminology.

The extreme points of $\mathcal{S}$ are called *pure states*, and the extreme points of $\mathcal{O}$, *pure observations*. States and observations that are not pure are called *mixed*. We shall also at times refer to S as the *state figure* and to $\mathcal{O}$ as the *observation figure*.

We now present a few examples of statistical triples that have traditionally been considered and a few that haven't.

Example 1.1 *Kolmogorov Probability*

Let (Ω, Σ) be any measurable space, where Σ is a σ-algebra of subsets of Ω. Let $\mathcal{S}$ be the set of all probability measures in (Ω, Σ) and $\mathcal{O}$ be the set of all measurable functions $f : \Omega \to [0, 1]$. We define $\langle \cdot, \cdot \rangle$ by $\langle \mu, f \rangle = \mu(f) = \int f(w) d\mu(w)$. The set of extreme points of $\mathcal{S}$ are the *dichotomous measures*, that is, such that for all $A \in \Sigma$, $\mu(A)$ is always either 0 or 1. There is a map which associates to each point $\omega \in \Omega$ an extreme point of $\mathcal{S}$ given by $\omega \mapsto \delta_\omega$ where δ_ω is the *Dirac measure* defined by $\langle \delta_\omega, f \rangle = f(\omega)$. In general this map is neither injective nor surjective. In many important cases though it is bijective. The pure observations of any Kolmogorov triple are precisely the characteristic functions χ_A of measurable subsets. The elements $\mathbf{0}$ and $\mathbf{1}$ are the constant functions 0 and 1 respectively, and $\neg f = 1 - f$.

If μ is a pure state then $\langle \mu, \chi_A \rangle = \mu(A)$ is either 0 or 1; in other words, a pure state has a pure property either certainly or certainly not. This is the intuitive content of sharply defined states and sharply defined measurements.

In mechanics, Ω is taken to be the phase space of the physical system. Statistical mechanics deals with certain canonically defined measures in Ω such as the Gibbs and the microcanonical ensembles, classical mechanics restricts itself to Dirac states, other measures are introduced only as auxiliary objects.

An important difference between our viewpoint and the usual one is the introduction of mixed observations. We have been motivated by reference [5] in this regard.

if $\Omega = \{\omega_1, \ldots, \omega_k\}$ is finite, then we can take for Σ the algebra of all subsets of Ω. Now every probability measure can be written as $\mu = \sum_i m_i \delta_{\omega_i}$ where $m_i = \mu(\{\omega_i\}) \geq 0$ and $\sum m_i = 1$. Thus S can be identified with the $n - 1$ simplex $\{m \in \mathbb{R}^n \,|\, m_i \geq 0, \sum m_i = 1\}$ and identifying $\langle \cdot, \cdot \rangle$ with the usual Euclidean inner product in $\mathbb{R}^n$ we see that $\mathcal{O}$ is identifiable with the cube $\{p \in \mathbb{R}^n \,|\, 0 \leq p_i \leq 1\}$. In this case $\mathbf{1}$ is the point $(1, 1, \ldots, 1)$ and $\mathbf{0}$ is the origin; negation is given by $\neg(p_1, p_2, \ldots, p_k) = (1 - p_1, 1 - p_2, \ldots, 1 - p_k)$.

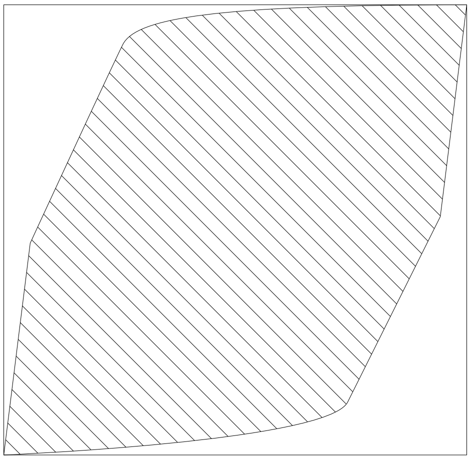
We can think of this system as that of infinite messages in an alphabet on k letters $\sigma_1, \sigma_2, \ldots, \sigma_k$ which can be identified as copies of the pure states. A pure observation consists of verifying whether an instance of one of the letters falls within a predetermined subset of letters. We assume the asymptotic frequency of such events exists. Such messages are produced by the outcomes of Bernoulli trials. As every cryptanalyst knows, natural languages approximate such behavior, even though natural languages are not Bernoulli processes. In natural languages, symbols at different positions are correlated, but for this example, such correlations are not part of the observation procedures.

Example 1.2 *Boolean triples*

This is a system closely related to Kolmogorov probability, but with the measure-theoretic troubles disregarded. Let $\mathbb{B}$ be a Boolean algebra. By Stone's representation theorem [6], $\mathbb{B}$ is identifiable with the algebra of closed-open subsets of a completely disconnected compact Hausdorff space X. We let $\mathcal{S}$ be the set of all finitely additive measures on $\mathbb{B}$. We.let $\mathcal{O}$ be the set of continuous functions $f : X \to [0,1]$ and we set $\langle \mu, f \rangle = \int f \, d\mu$. This is a finitely additive integral and so perhaps deserves some explanation. Consider the set $\mathcal{F}$ of *simple functions* $f = \sum_i c_i \chi_{A_i}$ where the c_i are real and A_i are closed-open sets. Such functions are continuous, and by the Stone-Weierstrass theorem are dense in $\mathcal{C}(X)$ considered as a Banach space with the supremum norm. For $f \in \mathcal{F}$ we can define the integral by $\int f \, d\mu = \sum_i c_i \mu(A_i)$. It can be easily checked that this defines a norm continuous linear functional on cF and so can be extended to the closure $\mathcal{C}(X)$. Hence $\langle \cdot, \cdot \rangle$ is well defined. The extreme points of $\mathcal{S}$ are Dirac measures and the extreme points of $\mathcal{O}$ are characteristic functions of closed-open sets. Negation is given by $\neg f = 1 - f$ and $\mathbf{0}$ and $\mathbf{1}$ are the constant functions 0 and 1 respectively. When $\mathbb{B}$ is a finite Boolean algebra with k atoms, this triple is isomorphic to the Kolmogorov triple with Ω being a set with k points.

Example 1.3 *Two dimensional triples*

This is the simplest non-trivial triple, that is one in which $\mathcal{S}$ is spanned by two pure states σ_0 and σ_1. Thus $\mathcal{S}$ is a line segment $\{\sigma_\lambda = \lambda\sigma_1 + (1-\lambda)\sigma_0 \mid 0 \le \lambda \le 1\}$ which we identify with the line segment between $(0,1)$ and $(1,0)$ in $\mathbb{R}^2$. Hence now a $\sigma_\lambda = (\lambda, 1-\lambda)$. We consider $\mathcal{O} \subset \mathbb{R}^2$ and if we use the normal duality, $\mathcal{O}$ must be a subset of the square $[0,1]\times[0,1]$. The existence of negation means that $\mathcal{O}$ is symmetric under inversion through the point $(1/2, 1/2)$ and the separation requirement means that besides the segment $[\mathbf{0},\mathbf{1}]$ there must be at least one more point in $\mathcal{O}$ (and by convexity, many more). An example of such a set is given by the figure below:



The boundary of $\mathcal{O}$ consists of two parts, the *upper boundary* that lies above the segment $[\mathbf{0},\mathbf{1}]$ and the *lower boundary* that lies below. We include $\mathbf{0}$ and $\mathbf{1}$ in both parts. In case the upper boundary is the graph of a function we denote this function by b. This occurs if and only if $\mathbf{0}$ is the only point of the form $(0, y)$ on the boundary. If $\mathcal{O}$ is all of the square then we have a Kolmogorov triple on a two point set, or equivalently a Boolean triple with $\mathbb{B}$ having two atoms.

Example 1.4 *Hilbert space quantum mechanical triple*

Let $\mathcal{H}$ be a complex Hilbert space with inner product $(\cdot,\cdot)$ We let $\mathcal{S}$ be the set of positive trace class operators ρ of trace 1, and let $\mathcal{O}$ be the set of operators A such that $0 \le A \le I$. We define $\langle \rho, A\rangle = \mathrm{Tr}(\rho A)$. The extreme points of $\mathcal{S}$ correspond to pure quantum mechanical states, and these are operators of the form $(\phi,\cdot)\phi$ where $||\phi|| = 1$, $\phi \in \mathcal{H}$. Such a ϕ is defined only up to a multiplicative constant of modulus 1. The extreme points of $\mathcal{O}$ are orthogonal projections. Negation is given by $\neg A = 1 - A$, and $\mathbf{0}$ and $\mathbf{1}$ are the operators 0 and 1 respectively. It is customary to call elements of $\mathcal{S}$ *density matrices*. Just as for the Kolmogorov triple in mechanics, this formalism differs from the usual one in the admission of mixed observations.

Example 1.5 *Stochastic triples*

Let $(X, \mathcal{X})$ and (Ω, Σ) be measurable spaces with σ-algebras $\mathcal{X}$ and Σ respectively. Let P_x be a family of probability measures in Ω indexed by X and such that for all $A \in \Sigma$, $x \mapsto P_x(A)$ is a $\mathcal{X}$-measurable function on X. We let $\tilde{\mathcal{S}}$ be the space of probability measures on X, and $\tilde{\mathcal{O}}$ be the set of measurable functions $f : \Omega \to [0, 1]$. We define $\langle \mu, f \rangle = \int (\int f(\omega) P_x(d\omega))\, d\mu(x)$. The system $(\tilde{\mathcal{S}}, \tilde{\mathcal{O}}, \langle \cdot, \cdot \rangle)$ may not be separated; the stochastic triple is the corresponding reduced system. This construction corresponds to stochastic processes which of course have additional structure. In such cases Ω is the path space, X the coordinate space of the process and μ is the initial distribution.

This triple is closely related to the Kolmogorov triple on (Ω, Σ), and can be obtained from it by restricting the states to be the measures that can be written as $\int_X P_x \, d\mu(x)$.

Example 1.6 *Empirical triple*

The ideas presented here could be used as a basis for procedures in multivariate statistics. The sets $\mathcal{S}$ and $\mathcal{O}$ are experimentally determinable given the results of observations of particular phenomena. Let thus $Y_1, Y_2, \ldots, Y_m$ be a set of dichotomous random variables with possible values 0 and 1. Let there be a sample of N observations of the m-tuple $(Y_1, Y_2, \ldots, Y_m)$. We thus have in fact a matrix of observations Y_{ij}, $i = 1, \ldots, m$; $j = 1, \ldots, N$. We wish to interpret the data according to the model that we are dealing with a statistical triple $(\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ with the Y_i corresponding to points in $\mathcal{O}$ and each empirical observation being performed upon a certain state in $\mathcal{S}$. Although we have not developed the general statistical procedures to make such an analysis, there is nothing in principle to prevent any set of data from being treated in this way. Certain existing procedures can already be applied if we make some *a-priori* assumptions about the data. If the sample can be divided into K sub-samples of n_k observations, $k = 1, \ldots, K$, each one of which can be considered as corresponding to a single state σ_k, then by averaging the variables within the subsamples we obtain the following new sample variables:

$$\hat{Y}_{ik} = \frac{1}{n_k} \sum_j Y_{ij}$$

where j runs over the k-th sample. We now consider the $\hat{Y}_{ik}$ as components of a vector $\hat{Y}_{i\cdot}$. On the basis of these we can perform an oblique axis factor analysis [7] and fit a model of the form:

$$\hat{Y}_{i\cdot} = \sum_{\ell=1}^{L} a_{i\ell} Z_{\ell\cdot} + b_i$$

where the $Z_{\ell\cdot}$ are picked to be certain of the $\hat{Y}_{i\cdot}$ and b_i is the coefficient of the presumably always present factor **1** of mean 1 and variance 0, that is, the constant 1. Once this is done, consider in $\mathbb{R}^{L+1}$ the convex envelope of the set of vectors

$$\{(1, 0, \ldots, 0), (0, 0, \ldots, 0)\} \cup \{(b_i, a_{i\cdot}), (1 - b_i, -a_{i\cdot}) \,|\, i = 1, \ldots, m\}$$

and also the convex envelope of the set of vectors $\{(1, Z_{\cdot k}) \,|\, k = 1, \ldots, K\}$. These give the empirically determined sets $\mathcal{S}$ and $\mathcal{O}$ respectively. If the factor analysis is well done in that one cannot reduce the number of factors and preserve goodness of fit, the two sets of vectors $\{(b_i, a_{i\cdot}) \,|\, i = 1, \ldots, m\}$ and $\{(1, Z_{\cdot k}) \,|\, k = 1, \ldots, K\}$ are entire in $\mathbb{R}^{L+1}$. The empirically determined inner product that serves to define the duality of the empirical triple is given by:

$$\langle \sum \alpha_k (1, Z_{\cdot k}), \sum \beta_i (b_i, a_{i\cdot}) \rangle = \sum_{i,\ell,k} \beta_i a_{i\ell} Z_{\ell k} \alpha_k.$$

Note that even though the sums on the left hand side involve in general sets that are linearly dependent, the definition is consistent, for if any of the sums vanish, the right hand side vanishes also. In this triple **1** is $(1, 0, \ldots, 0)$, the origin is **0**, and negation is given by $\neg p = \mathbf{1} - p$.

More sophisticated statistical analysis must of course be made if we don't have *a-priori* identifications of which observation is done in which state.

Normal factor analysis techniques usually confine themselves to the elaboration of a factor model, and after appropriate coordinate changes to the identification of the factors with some naturally interpretable variables. Little, if any, attention is paid to the geometrical shapes of the sets of empirical points in the various spaces involved. A study of these shapes should give clues as to the underlying mechanisms that could be involved in producing the observed data. To illustrate these remarks, consider the physics of a light bench. The polarization states of light are well modelled by a Hilbert space triple with $\mathcal{H} = \mathbb{C}^2$. If we use *Pauli spin matrices*:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

we can write any 2×2 matrix m as $m = \alpha I + \vec{\beta} \cdot \vec{\sigma}$. We find that hermitian matrices correspond to α, and $\vec{\beta}$ real. Now $\mathcal{S}$ is defined by $\alpha = 1/2$, $|\vec{\beta}| \leq 1/2$, and $\mathcal{O}$ by $|\vec{\beta}| \leq \min(\alpha, 1-\alpha)$ which is a four dimensional convex region. Furthermore, $\mathrm{Tr}((\alpha I + \vec{\beta} \cdot \vec{\sigma})(\gamma I + \vec{\delta} \cdot \vec{\sigma})) = 2(\alpha\gamma + \vec{\beta} \cdot \vec{\delta})$. Now if the above mentioned factor analysis were done on a large number of optical bench experiments, the above convex figures and duality, up to an affine isomorphism, would be found. With sufficient knowledge on our part, we would recognize them as coming from a Hilbert space model. Empirical data can therefore in some systematic way call for certain model types. Now the freedom to choose our coordinates arbitrarily implies that only the affine invariants of the sets $\mathcal{S}$ and $\mathcal{O}$ and of the pairing $\langle \cdot, \cdot \rangle : \mathcal{S} \times \mathcal{O} \to [0, 1]$ are intrinsic to the data. Based on this observation we now prepare to formulate the study of convex sets independent of their possible imbeddings in linear spaces and of any possible coordinate systems therein.

Chapter 2

The Category of Convex Sets

Because of the role that convex sets play in the previous discussion, we propose here to give a condensed development of convex set theory from an intrinsic viewpoint. We want to consider a convex set as an object existing apart from any vector space in which it may be embedded. To be able to do this the best procedure is to develop convex set theory as an algebraic theory and to investigate the category that is so formed.

Definition 2.1 *An* abstract convex set $\mathcal{C}$ *is a set endowed with the following additional algebraic structure: To each $n \geq 1$, and to each n-tuple $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, of real numbers such that $0 \leq \lambda_i \leq 1$ and $\sum_i \lambda_i = 1$ there is given an n-ary operation $\Phi_\Lambda : \mathcal{C}^n \to \mathcal{C}$ where for intuitive we write $\sum \lambda_i x_i$ instead of $\Phi_\Lambda(x_1, x_2, \ldots, x_n)$. These operations are subject to the following axioms:*

1. $1 \cdot x = x$.
2. $\sum_j \lambda_j (\sum_i \mu_{ji} x_i) = \sum_i (\sum_j \lambda_j \mu_{ji}) x_i$.
3. $\sum_j \lambda_j (\sum_i \mu_{ji} x_{ji}) = \sum_{ij} (\lambda_i \mu_{ji}) x_{ji}$.
4. *If* $0 \neq \lambda \neq 1$ *and* $\lambda x + (1-\lambda) y = \lambda x + (1-\lambda) z$, *then* $y = z$.

Easy consequences of these axioms are: (a) $1 \cdot x + 0 \cdot y = x$ which implies that in any expression of the form $\sum \lambda_i x_i$ we can simply drop any term with $\lambda_i = 0$; and (b) that in order to define the operations Φ_Λ we need only know how to form $\lambda x + (1-\lambda) y$ for any x, y and $0 \leq \lambda \leq 1$.

Given two convex sets $\mathcal{C}$ and $\mathcal{D}$ a morphism $\phi : \mathcal{C} \to \mathcal{D}$ is a set map that is a homomorphism of the algebraic system; that is:

$$\phi(\sum_i \lambda_i x_i) = \sum_i \lambda_i \phi(x_i).$$

We shall call such a ϕ an *affine map*.

We denote by *Conv* the category whose objects are convex sets and whose morphisms are affine maps. The set of morphisms $\phi : \mathcal{C} \to \mathcal{D}$ we denote by $Conv(C, D)$. The empty set and any singleton set are obvious objects of this category.

Given any convex set $\mathcal{C}$ a subset $\mathcal{D} \subset \mathcal{C}$ is called a *convex subset* if $x, y \in \mathcal{D}, 0 \leq \lambda \leq 1 \Rightarrow \lambda x + (1 - \lambda)y \in \mathcal{D}$. The intersection of any family of convex subsets is a convex subset, hence given any subset $S \subset \mathcal{C}$ we can define its convex envelope $\mathrm{conv}(S)$ as being the intersection of all the convex subsets that contain S. This is the smallest convex subset that contains S and consists of points of the form $\sum_i \lambda_i s_i$ with s_i in S.

A subset F of a convex set $\mathcal{C}$ is called a *face* of $\mathcal{C}$ if $x \in F, y, z \in \mathcal{C}, 0 \leq \lambda \leq 1, x = \lambda y + (1 - \lambda)z \Rightarrow y, z \in F$. Clearly $\mathcal{C}$ is a face of $\mathcal{C}$. A point x such that the singleton set $\{x\}$ is a face is called a *pure*, or *extreme point*. All other points are called *mixed*. If $\phi : \mathcal{C} \to \mathcal{D}$ is a morphism and G is a face of $\mathcal{D}$ then $\phi^{-1}(G)$ is a face of $\mathcal{C}$.

The aim of this chapter is to show that *Conv* is a bicomplete monoidally closed category [8]. To achieve this we must first construct an equivalent category which is also of great utility because it interprets each convex set as being canonically imbedded as a base of a cone in a functorially dependent seminormed vector space.

We first note that $Conv(C, D)$ can itself be considered as an object in *Conv* by defining $\sum_i \lambda_i \phi_i$ by the formula:

$$(\sum_i \lambda_i \phi_i)(x) = \sum_i \lambda_i \phi_i(x).$$

We shall write $\hat{Conv}(C, D)$ for this object in *Conv*.

We further note that if *Lin* is the category of real linear spaces with linear maps as morphisms, then there exists a functor *Lin* $\to$ *Conv* which interprets each linear space as a convex set in the obvious manner.

To simplify notation, we give no names to the various identification functors that we are about to start using, letting the context supply the necessary information. Note that if V is a real linear space, then $Conv(C, V)$

can also be thought of as a real linear space; namely, define $a\phi + b\psi$ by $(a\phi + b\psi)(x) = a\phi(x) + b\psi(x)$ for a, b real.

Let C now be a nonempty convex set. We show that $Conv(C, \mathbb{R})$ separates points in C; that is, given x, $y \in C$, $x \neq y$ there is an affine map $\phi : C \to \mathbb{R}$ such that $\phi(x) \neq \phi(y)$. To show this we first imbed C by a rather laborious construction into a real vector space V_C in a way that the convex structure of C becomes part of the linear structure of V_C In this way any affine map $C \to \mathbb{R}$ will have an extension to a linear map $V_C \to \mathbb{R}$, and conversely, the restriction of any such linear map to the imbedded image of C provides an affine map $C \to \mathbb{R}$. Thus using the fact that $Lin(V_C, \mathbb{R})$ separates points of V_C we will conclude the needed result.

In order to understand the construction that follows, think of V_C as being formed by formal differences of the form $ax - by$; x, $y \in C$; a, $b \in \mathbb{R}$; a, $b > 0$.

Consider the set of all quadruples of the form $(x, y, a, b) \in C \times C \times (0, \infty) \times (0, \infty)$ and for intuitive reasons write such a quadruple as $ax - by$. Introduce now an equivalence relation $ax - by \sim a'x' - b'y'$ by firstly requiring that $a + b' = a' + b$; call this number D; and secondly by requiring

$$\frac{a}{D}x + \frac{b'}{D}y' = \frac{a'}{D}x' + \frac{b}{D}y$$

where of course the symbolic sums now refer to the algebraic operations performed in the convex set C. Let V_C be the set of equivalence classes $[ax - by]$ of these quadruples. We now define a linear structure in V_C. First, we define, choosing any $x_0 \in C$:

$$0 = [1x_0 - 1x_0].$$

Given $r \in \mathbb{R}$, we define scalar multiplication by:

$$r[ax - by] = \begin{cases} [(ra)x - (rb)y] & \text{if} \quad r\rangle 0 \\ 0 & \text{if} \quad r = 0 \\ [(-rb)y - (-ra)x] & \text{if} \quad r\langle 0 \end{cases}$$

and we define addition by:

$$[ax - by] + [a'x' - b'y'] = \\ \left[(a + a')\left(\frac{a}{a + a'}x + \frac{a'}{a + a'}x'\right) - (b + b')\left(\frac{b}{b + b'}y + \frac{b'}{b + b'}y'\right)\right]$$

where the sums $a + a'$ and $b + b'$ are to be interpreted in $\mathbb{R}$, and the other two as standing for the operations in C.

A tedious and totally unenlightening verification shows that V_C as so defined is in fact a real vector space.

We now define an imbedding $j_C : C \to V_C$ by:

$$j_C x = [2x - 1x].$$

To see that j_C is injective suppose that $[2x - 1x] = [2y - 1y]$, then by the definition of equivalence we must have $(2/3)x + (1/3)y = (2/3)y + (1/3)x$ which by the axioms of convex sets is easily shown to be equivalent to $x = y$. A further tedious verification shows that j_C is affine, hence we can identify the convex structure in C as being induced by the linear structure in V_C. Let now $\phi : C \to \mathbb{R}$ be any affine map, we define $V_\phi : V_C \to \mathbb{R}$ by setting $V_\phi[ax - by] = a\phi(x) - b\phi(y)$. We have $\phi = V_\phi \circ j_C$. We conclude finally, using the argument already presented, that $Conv(C, \mathbb{R})$ separates points in C.

Having established this, we can proceed in a more canonical fashion. Consider the set $C^{**} = Conv(\hat{C}onv(C, \mathbb{R}), \mathbb{R})$. There is a natural affine map $j_C : C \to C^{**}$ given by

$$(j_C x)(\phi) = \phi(x)$$

for $\phi \in \hat{C}onv(C, \mathbb{R})$, $x \in C$. By the previous discussion we now know that j_C is injective. In what follows, we drop the subscript in j_C whenever convenient.

Since C^{**} is naturally a vector space, we define $V(C)$ as being the vector space spanned by the image of C by j. An element v of $V(C)$ is therefore of the form, $\sum r_i j x_i$. Let I be the set of indices i such that $r_i > 0$, and let K be the set of indices k such that $r_k < 0$. Let $D^+ = 1 + \sum\{r_i \mid i \in I\}$, and $D^- = 1 - \sum\{r_k \mid k \in K\}$. Let $x_0 \in C$ be arbitrary. We have:

$$v = D^+ \left(\frac{1}{D^+} j x_0 + \sum_I \frac{r_i}{D^+} j x_i \right) - D^- \left(\frac{1}{D^-} j x_0 + \sum_K \frac{-r_k}{D^-} j x_k \right)$$

which being of the form $D^+ j x^+ - D^- j x^-$; $D^+, D^- > 0$; x^+, $x^- \in C$, is now identifiable with the element $[D^+ x^+ - D^- x^-]$ of the vector space V_C constructed earlier. We thus recover our previous construction. We furthermore note that any element of $V(C)$ can be written as $a j x_1 - b j x_2$; x_1, $x_2 \in C$ and a, $b > 0$. We can clearly allow a or b to be zero, though in the above manipulations we avoided this to be able to relate $V(C)$ to V_C.

The space $V(C)$ has considerable additional structure. Let $K(C)$ be the cone of elements of the form ajx, $a \geq 0$, $x \in C$; thus jC is identified with the base of this cone and we have $V(C) = K(C) - K(C)$. Given $v = ajx - bjy \in V(C)$ we can define the linear functional $\tau_C(v) = a - b$. Note that τ is positive on $K(C)$. We now define the functional

$$p_C(v) = \inf\{a + b \,|\, v = ajx - bjy;\, a, b \geq 0\}$$

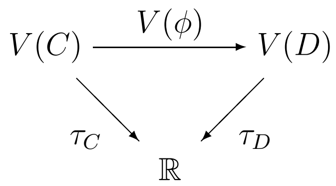
or better

$$p_C(v) = \inf\{\tau(v_1) + \tau(v_2) \,|\, v = v_1 - v_2,\, v_i \in K(C)\}.$$

One verifies that p_C is a seminorm that coincides with τ on $K(C)$; furthermore $\tau_C \circ j_C(x) = 1$. Let now $\phi : C \to D$ be an affine map. Define $V(\phi) : V(C) \to V(D)$ by:

$$V(\phi)(ajx - bjy) = aj\phi(x) - bj\phi(y).$$

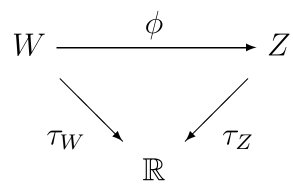
One checks that $V(\phi \circ \psi) = V(\phi) \circ V(\psi)$, $V(1_C) = 1_{V(C)}$, and that the diagram



commutes.

For the omitted case $C = \emptyset$ we set $V(C) = \{0\}$, the zero dimensional vector space. The rest of the structure can now be easily identified.

We are now ready to define *Bsn*, the category of based seminormed linear spaces. The objects of this category are real vector spaces W endowed with the following additional structure: a proper convex cone K such that $W = K - K$ (proper means that $-K \cap K = \{0\}$) and a positive linear functional τ such that $x \in K$, $\tau(x) = 0 \Rightarrow x = 0$. Morphisms in this category are positive linear maps $\phi : W \to Z$ such that the diagram



commutes (positive means $\phi(K(W)) \subset K(Z)$). Now *Conv* and *Bsn* are equivalent categories. We have already exhibited a functor $C \mapsto V(C)$; $\phi \mapsto V(\phi)$ from *Conv* to *Bsn*. The functor in the other direction which establishes the equivalence is: $W \mapsto \{v \in K(W) \mid \tau(v) = 1\}$; $\phi \mapsto \phi|W$.

Just as for the category *Conv* we shall identify $Bsn(W, Z)$ with an object of *Bsn*. Now $Bsn(W, Z)$ is a convex subset of $Lin(W, Z)$ since $\tau \circ \phi = \tau$ and $\tau \circ \psi = \tau$ imply $\tau \circ (\lambda\phi + (1-\lambda)\psi) = \tau$ and ϕ, ψ positive imply $\lambda\phi + (1-\lambda)\psi$ positive. We thus define $\hat{Bsn}(W, Z) = V(Bsn(W, Z))$ thinking of $Bsn(W, Z)$ as a convex set. Now by equivalence we have a natural bijection $Bsn(V(C), V(D)) \simeq Conv(C, D)$ which is readily seen to be an isomorphism of convex sets. Thus $V(\hat{Conv}(C, D)) \simeq V(Bsn(V(C), V(D)) \simeq \hat{Bsn}(V(C), V(D))$ a fact that will be useful later so we dignify it to be a lemma.

Lemma 2.1 *There is a naturalism Bsn isomorphism:*

$$V(\hat{Conv}(C, D)) \simeq \hat{Bsn}(V(C), V(D)).$$

We also need the following technical result in a subsequent demonstration.

Lemma 2.2 *Let $\phi \in \hat{Bsn}(W, Z)$ then we have the equality*

$$\tau_Z(\phi(w)) = \tau_{\hat{Bsn}(W,Z)}(\phi) \cdot \tau_W(w).$$

Proof: Suppose $\phi = a\alpha - b\beta$; a, $b \geq 0$; α, $\beta \in \hat{Bsn}(W, Z)$, then $\tau_Z(\phi(w)) = \tau_Z(a\alpha(w) - b\beta(w)) = a\tau_Z \circ \alpha(w) - b\tau_Z \circ \beta(w) = a\tau_W(w) - b\tau_W(w) = (a - b)\tau_W(w) = \tau_{\hat{Bsn}(W,Z)}(\phi) \cdot \tau_W(w)$. Q.E.D

The advantage of using *Bsn* instead of *Conv* is that we can use the linear structure to manipulate the convex structure.

Theorem 2.1 *Conv is a bicomplete monoidally closed category.*

The proof will proceed by stages. To show bicompleteness we need to prove the existence of products, equalizers, coproducts, and coequalizers. To prove monoidal closure we will exhibit a tensor product $C \otimes D$ of convex sets and a natural bijection $Conv(C \otimes D, E) \simeq Conv(C, \hat{Conv}(D, E))$.

1. *Existence of products*

   Let C_α be a family of convex sets and let $C = \prod C_\alpha$ be the set theoretic product and $\pi_\beta : C \to C_\beta$ be the set theoretic projections. We furnish C with the following convex structure: $\lambda(x_\alpha) + (1-\lambda)(y_\alpha) = (\lambda x_\alpha + (1-\lambda)y_\alpha)$. To show that C is the product in *Conv* is now trivial.

2. Existence of equalizers

   Let ϕ_1, $\phi_2 : C \to D$ be two affine maps. Consider the subset $E = \{x \in C \mid \phi_1(x) = \phi_2(x)\}$. Now E is a convex subset and the inclusion map $\iota : E \to C$ is the set theoretic equalizer, which being affine is easily seen to be the equalizer in *Conv.*

3. Existence of coproducts

   Let C_α, be a family of convex sets. Consider the family $V(C_\alpha)$ of the corresponding vector spaces. Let $V = \prod V(C_\alpha)$ be the vector space coproduct, that is, the subspace of the vector space product defined by $(v_\alpha) \in V \Leftrightarrow v_\alpha \in V(C_\alpha)$ and $v_\alpha = 0$ for all but a finite number of indices. We define $\rho_\alpha : C_\alpha \to V$ by: $\rho_\alpha(x) = (y_\beta)$ where

   $$y_\beta = \begin{cases} j_{C_\alpha} x & \text{if } \beta = \alpha \\ 0 & \text{if } \beta \neq \alpha. \end{cases}$$

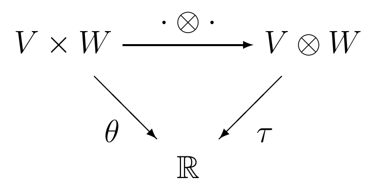
   Let $C = \mathrm{conv}(\bigcup_\alpha \rho_\alpha(C_\alpha))$ be the convex envelope of the union of the images of all the C_α, by the maps ρ_α. Continue to call by ρ_α the injections $C_\alpha \to C$. We show that C is the coproduct in *Conv.* Clearly each ρ_α is affine. Let $\phi_\alpha : C_\alpha \to D$ be a family of affine maps. Now each $x \in C$ can be uniquely written as $x = \sum\{\lambda_\alpha \rho_\alpha(x_\alpha) \mid \alpha \in F\}$ where F is a finite set of indices, $\lambda_\alpha \geq 0$, $\sum \lambda_a = 1$, and $x_\alpha \in C_\alpha$. We wish to establish the existence of a unique affine map $\psi : C \to D$ such that $\psi \circ \rho_\alpha, = \phi_\alpha$ but by affinity and the factorization requirement we see that ψ must be given by $\psi(x) = \sum_F \lambda_\alpha \phi_\alpha(x_\alpha)$ which is readily shown to satisfy the requirements.

4. Existence of coequalizers

   Let ϕ_1, $\phi_2 : C \to D$ be two morphisms in *Conv.* Consider the convex set $D \times D$. An equivalence relation $R \subset D \times D$ is said to be *convex* if it is a convex subset of $D \times D$. Seeing that the total relation $D \times D$ is a convex equivalence relation, and that an intersection of an arbitrary family of convex equivalence relations is also a convex equivalence relation, we conclude that the set of convex equivalence relations is a complete lattice. The smallest element is equality $E = \{(x, y) \mid x = y\}$. Thus, let R be the smallest convex equivalence relation containing all pairs of the form $(\phi_1(x), \phi_2(x))$, $x \in C$. The passage to the quotient space D/R is compatible with the convex structure in D; and we can introduce a convex structure in the quotient by setting $\lambda[x] + (1-\lambda)[y] = [\lambda x + (1-\lambda)y]$. The set theoretic canonical map $\kappa : D \to D/R$ is easily seen to be affine. Let now $\psi : D \to F$ be a morphism such that $\psi \circ \phi_1 = \psi \circ \phi_2$; we note that $(\psi \times \psi)^{-1}(E)$ is a convex equivalence relation in D, and since it contains all pairs of the form $(\phi_1(x), \phi_2(x))$, it contains R. Thus ψ is constant on the equivalence classes of R and so there is a unique factorization $\psi = \hat{\psi} \circ \kappa$; $\hat{\psi}$ is easily shown to be affine showing that K is the coequalizer.

5. Existence of tensor product

   To show that *Conv* has a tensor product, its enough to show that the equivalent category *Bsn* has a tensor product. Let V and W be based seminormed vector spaces and consider the linear space tensor product $V \otimes W$. Define a cone in this space as being the convex cone K generated by elements of the form $v \otimes w$, $v \in K(V)$, $w \in K(W)$. Since any element of $V \otimes W$ has a representation $\sum x_i \otimes y_i$ and since we can write $x_i = v_i - v_i'$; $y_i = w_i - w_i'$; v_i, $v_i' \in K(V)$; w_i, $w_i' \in K(W)$; we conclude that $V \otimes W = K - K$. Define $\tau : V \otimes W \to R$ by the universal diagram:



   where $\theta(x, y) = \tau_V(x)\tau_W(y)$. If $\tau(v) = 0$ for $v \in K$, then $v = \sum x_i \otimes y_i$; $x_i \in K(V)$, $y_i \in K(W)$ and $\sum \tau(x_i)\tau(y_i) = 0$. Since each term is positive we have $\tau(x_i)\tau(y_i) = 0$ for all i, that is, for all i, either $x_i = 0$ or $y_i = 0$, hence $v = 0$. We conclude that K is a proper cone and that $V \otimes W$ has a natural *Bsn* structure. We now show there is a natural bijection
$$Bsn(V, \hat{Bsn}(W, Z)) \simeq Bsn(V \otimes W, Z).$$

A *Bsn* morphism $\phi : V \to \hat{Bsn}(W,Z)$ coincides with a *Lin* morphism $\hat{\phi} : V \to Lin(W,Z)$ which in turn defines a unique *Lin* morphism $\phi^{\S} : V \otimes W \to Z$. We need show that $\phi^{\S}$ is a *Bsn* morphism. Let $x \otimes y \in V \otimes W$ then $\phi^{\S}(x \otimes y) = (\hat{\phi}x)(y)$ so $\tau_Z \phi^{\S}(x \otimes y) = \tau_Z((\hat{\phi}x)(y))$ which by Lemma 2.2 is equal to $\tau_{\hat{Bsn}}(W,Z)(\hat{\phi}x)\tau_W(y)$ which in turn is equal to $\tau_V(x)\tau_W(y)$ since $\hat{\phi}$ is actually a *Bsn* morphism. As this last number is $\tau_{V\otimes W}(x \otimes y)$ we conclude that $\tau_Z \circ \phi^{\S} = \tau_{V\otimes W}$. Furthermore, if $x \in K(V)$, $y \in K(W)$, then $\phi^{\S}(x \otimes y) = \phi(x)(y)$, but since $\phi(x) \in Bsn(W,Z)$ we have $\phi(x)(y) \in K(Z)$, showing that $\phi^{\S}$ is positive. We conclude from all this that $\phi^{\S}$ is a *Bsn* morphism. We have established one side of the bijection however, using the same equations the argument is reversible and we conclude that *Bsn* has a tensor product. Returning to *Conv* we define $C \otimes D$ to be the base of the cone in the *Bsn* tensor product $V(C) \otimes V(D)$, establishing thus a natural isomorphism $V(C \otimes D) \simeq V(C) \otimes V(D)$. Therefore $C \otimes D$ is the convex envelope of the set of points of the form $j_C x \otimes j_D y$; $x \in C$, $y \in D$. We now have using Lemma 2.1, $Conv(C, \hat{Conv}(D,E)) \simeq Bsn(V(C), \hat{Bsn}(V(D),V(E))) \simeq Bsn(V(C) \otimes V(D), V(E)) \simeq Bsn(V(C \otimes D), V(E)) \simeq Conv(C \otimes D, E)$ where each bijection is natural. Thus *Conv* is a monoidally closed category and we conclude the proof of our theorem. Q.E.D.

The tensor product of two convex sets is generally hard to compute. For example $[0,1] \otimes [0,1]$ is affinely isomorphic to the tetrahedron, but this requires some work to establish.

As a final observation on the theorem we note that the initial object in *Conv* is the empty set $\emptyset$ where the unique initial morphism $\emptyset \to C$ is the empty map. As has already been noted $V(\emptyset) = \{0\}$ is the initial object in *Bsn*. The final object in *Conv* is any singleton set $\{*\}$. Note that $V(\{*\}) \simeq \mathbb{R}$ where $j* = 1$.

Chapter 3

The Canonical Vector Space Duality of a Statistical Triple

By a *statistical pairing*, or *pairing* for short, we shall mean a system $(C, D, \langle \cdot, \cdot \rangle)$ where C and D are convex sets and $\langle \cdot, \cdot \rangle : C \times D \to [0, 1]$ is biaffine. Given a pairing, we can imbed C in its *Bsn* space $V(C)$. We can now think of each $d \in D$ as defining a linear map $V(C) \to \mathbb{R}$ by $\alpha jc - \beta jc' \mapsto \alpha\langle c, d\rangle - \beta\langle c', d\rangle$. Thus D is mapped to a subset of the algebraic dual of $V(C)$. Let $V_0(D)$ be the subspace of this dual spanned by the image of D. Thus we have a pairing of $V(C)$ with $V_0(D)$ and we continue to write $\langle \cdot, \cdot \rangle$ for the bilinear form defining this pairing. Note that if $P = (C, D, \langle \cdot, \cdot \rangle)$ is a pairing, then so is $JP = (D, C, \langle \cdot, \cdot \rangle_J)$ where $\langle d, c\rangle_J = \langle c, d\rangle$. Hence we can also pair $V(D)$ with $V_0(C)$ defined analogously.

In case a pairing is separated, we call it a *statistical duality*, or *duality* for short. In this case the correspondences $d \mapsto \langle \cdot, d\rangle$ and $c \mapsto \langle c, \cdot\rangle$ imbed D as a subset of $Conv(C, [0, 1])$ and C as a subset of $Conv(D, [0, 1])$, identifications that we will normally use without mention.

A statistical triple $(S, \mathcal{O}, \langle \cdot, \cdot \rangle)$ is clearly a duality, and in this case we note that $\mathbf{0}$ is mapped to the origin of $V_0(\mathcal{O})$. We have in this case:

Theorem 3.1 *The pairing of $V(S)$ with $V_0(\mathcal{O})$ is separated.*

Proof: Suppose first that $\langle aj\sigma - bj\sigma', r\rangle = 0$ for all $r \in V_0(\mathcal{O})$, then in particular for $r = \mathbf{1}$ we conclude $a = b$. Thus if $a \neq 0$ we have that for all $p \in \mathcal{O}$, $\langle j\sigma - j\sigma', p\rangle = 0$, or $\langle \sigma, p\rangle = \langle \sigma', p\rangle$ and hence $\sigma = \sigma'$ since triples are separated. In any case therefore $aj\sigma - bj\sigma' = 0$ and the pairing separates points of $V(S)$.

Suppose now $\langle j\sigma, \sum r_i p_i \rangle = 0$ for all $\sigma \in S$, then as before we can write $\sum r_i p_i = ap - bq$; p, $q \in \mathcal{O}$, a, $b > 0$. Thus we have $a\langle \sigma, p \rangle = b\langle \sigma, q \rangle$. One of the ratios a/b or b/a must lie in $[0, 1]$, say b/a. Hence $\langle \sigma, p \rangle = (b/a)\langle \sigma, q \rangle = \langle \sigma, (b/a)p + (1 - b/a)\mathbf{0} \rangle$. Since triples are separated we have $p = (b/a)q + (1 - b/a)\mathbf{0}$ but then in $V_0(\mathcal{O})$, $ap - bq = (a - b)\mathbf{0} = 0$, and so the pairing separates points of $V_0(\mathcal{O})$. Q.E.D

We note that in five of the examples in the first chapter, the statistical triples were thought of as being already within their respective linear dualities. For example, in quantum mechanics, $V(S)$ is the space of all trace class matrices, and $V_0(\mathcal{O})$ the space of all self adjoint bounded operators with the pairing given by $\mathrm{Tr}(PA)$ where P is trace class and A is bounded and self adjoint.

Chapter 4

A More Extensive Formalism

Various experiences in trying to develop the formal foundations of quantum mechanics [9] have demonstrated that the scheme so far introduced is too restrictive in two aspects. We rarely perform simply a single yes-no observation, we generally observe several properties at the same time, or use instruments with scales whose response is some real number and is not restricted to merely registering yes-no alternatives. Furthermore, the act of observation does not normally mean the end of experimentation; observation could be part of preparation of the state. That is, some state is prepared, then observed, and if it satisfies certain conditions it's kept for further observations, and if not, it's rejected and the state preparation procedure is aborted and considered a failure. An example of such a procedure would be an initial screening of subjects for a psychology experiment by administering certain tests.

In previous works these two amplifications of the formalism has been done within an already existing statistical triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$. That is, instruments and operations were defined within an already mounted formalism and then their properties studied.

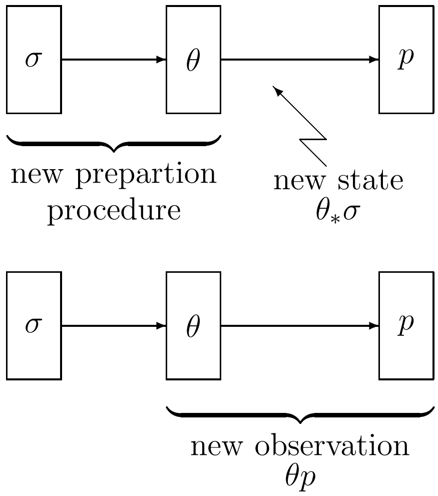
The point of view we shall follow here is to introduce these notions as primitive, to formalize the properties that we intuitively feel they have in their concrete realizations in experiments, and to show that in doing so, we can, under certain phenomenological assumptions, construct in the end a statistical triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ within which the given structure can be represented.

Consider a measuring instrument with a scale, such as a voltmeter, a balance, a yardstick, or an apparatus for measuring visual thresholds. Such an instrument does not provide us with a yes-no alternative but with a family of such. We think of the scale as an abstract set X such that each point $x \in X$ represents a possible reading of the instrument. If $A \subset X$ is in some sense a sufficiently regular subset, we can associate to A a yes-no observation described by: the reading of the instrument fell within A. If it is meaningful to attach such yes-no observations to the sets A, $B \subset X$, then it's equally meaningful to attach them to $X \setminus A$, $A \cup B$, and $A \cap B$. Thus we can assume these subsets form a Boolean algebra.

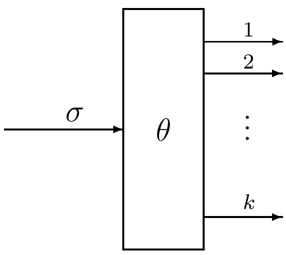
Without further ado we introduce a set $\mathcal{I}$ of *instruments*. To each $I \in \mathcal{I}$ is associated a certain Boolean algebra $\mathbb{B}_I$ the algebra of events in the *scale* of the instrument. (For the scale we can take the Stone representation space of $\mathbb{B}_I$) We still assume that we have a convex set $\mathcal{S}$ of states. Given $\sigma \in S$, $I \in \mathcal{I}$, and $A \in \mathbb{B}_I$ we denote by $\langle \sigma, I(A) \rangle$ the asymptotic frequency that in the state σ the instrument has a reading in the event A of the scale. We must assume by the frequentistic interpretation that for each $\sigma \in \mathcal{S}$, $A \mapsto \langle \sigma, I(A) \rangle$ is a measure on $\mathbb{B}_I$ Of special interest are the instruments in which $\mathbb{B}$ is finite, say a Boolean algebra with n atoms. In practice, due to technical limits on resolution of observations, we can always assume the scale of an instrument to be divided into a finite number of mutually disjoint subsets. Other types of scales are idealizations introduced for ease in theoretical investigations. It's easier in physical theories for example, to talk about an abstract length with values in $[0, \infty)$ instead of actual metersticks, microcalipers, interferometers, planetary orbits, and other length measuring instruments and methods as mathematical objects formalized within the theory. For foundation studies however, these generalizations to infinite algebras are cumbersome to handle, and we shall therefore assume for the rest of this work that $\mathbb{B}$ is always a finite Boolean algebra. If now $\mathbb{B}$ has n atoms, we can assume it to be the Boolean algebra of subsets of the set $\{1, 2, \ldots, n\}$. This latter set we denote by $\mathbf{n}$. and the associated Boolean algebra by $\mathbb{B}_n$. The trivial algebra $\{\emptyset\}$ we denote by $\mathbb{B}_0$. For ease of notation, we denote the atom $\{k\}$ of $\mathbb{B}_n$ simply by k. We can now think of I as being the simultaneous observation of n mutually exclusive properties. We shall refer to these properties as the *atomic observations*, or simply *atoms* of I. A copy of a state can posses only at most one of these properties, which one however, can vary from copy to copy. In these circumstances we can also introduce the property which is the conjunction of the negations of the given set of properties. That is, we can add to $\mathbb{B}$ one more atom and now assume that in addition to having a set of mutually exclusive properties we also have a set of mutually exhaustive ones. This procedure we call *exhaustion* of the instrument I. If we denote by \$ the largest element of $\mathbb{B}$ then we say that I is *exhaustive* if $\langle \sigma, I(\$)\rangle = 1$ for all $\sigma \in S$. The exhaustion of any instrument is clearly exhaustive. A final convenient way of intuitively thinking of I with $\mathbb{B}$ finite with n atoms is the following: think of I as an apparatus with a panel of n lights, when applied to a copy of the state σ any one of the lights may register, but never more than one. When I is exhaustive then there is always some light that does register. Every event $A \subset \mathbb{B}$ is a subset of lights, and $\langle \sigma, I(A)\rangle$ is the asymptotic frequency with which a light registers within the subset A when applied to a state σ. Let now $A_l, A_2, \ldots, A_k$ be a set of k mutually disjoint subsets of the atoms of $\mathbb{B}_n$. We can now consider the instrument J on $\mathbb{B}_k$ defined as follows: if $S \subset \mathbf{k}$, then $\langle \sigma, J(S)\rangle = \sum\{\langle \sigma, I(A_j)\rangle \mid j \in S\}$. We call J a *condensation* of I and I a *refinement* of J. In terms of properties, the j-th atomic property of J is the logical disjunction of the atomic properties of I that lie in the subset A_j. In terms of a panel of lights we consider the event of any light registering in the subset A_j as triggering of the registration of the j-th light on the panel of J. For each A_j that is empty we therefore have a dummy light on the panel of J that never registers. Note that the union of the A_j is not necessarily all the atoms of $\mathbb{B}_n$, certain atoms may thus be simply ignored. Such a situation may turn an exhaustive instrument into one which is not.

Least the reader be led terribly astray, we must warn him or her that the above notion of simultaneous observations is rather different from the same called notion as is commonly considered in quantum mechanics. The actual physical mechanism that leads to the assignment of one of the properties of an instrument to a copy of a state could involve interactions with the copy over a prolonged time interval. What is essential is that only one of the properties is affirmed of the copy in the end. We shall explore in Chapter 7 the relationship that our notion has to the one of the usual quantum mechanical models.

Consider now an *operation*, that is, an observation followed by either rejection or an acceptance as a new state. Given such an operation θ and a further observation p, we can consider the situation from two points of view, as diagrammed below (note that p is not the observation involved in the mechanism of the operation, it is a subsequent one):

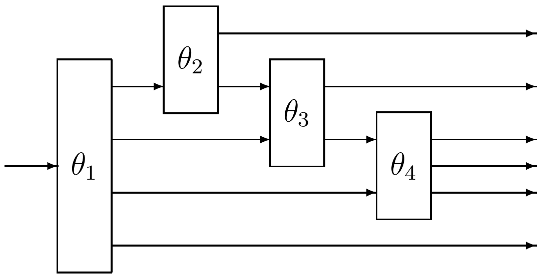


The two viewpoints are equivalent, though for certain purposes the second is more convenient, for if θ rejects every copy of σ, then $\theta_*\sigma$ strictly speaking doesn't exist, yet θp continues to make sense. Now this type of reasoning, though extensively treated in the literature on the foundations of quantum mechanics, suffers from the same defects that lead to the consideration of instruments instead of isolated observations. If we now admit that the observation needed to define an operation can be performed by an instrument, then instead of rejecting or accepting, we can classify according to the response of the instrument. A more adequate reflection of actual practice would be diagrammed as follows:



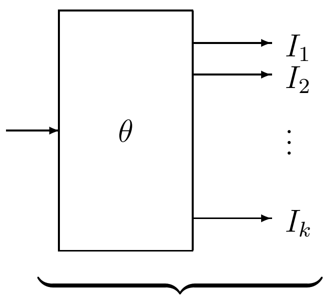
That is, a copy of a state is prepared and observed with an instrument; on the basis of the result, the copy is either rejected or is classified as belonging to one of k possible new states. We say the operation has k *exits*. In terms of beam physics parlance, θ is a beam splitter, for other situations however this is merely a metaphor. In a psychological experiment for instance, the subject can be given a preliminary test, and on the basis of the result he or she is classified within any one of k categories. Note that we are not assuming that after such a classification the subject continues to be a copy of the original state; the fact of having been tested could certainly influence subsequent behavior, and such influences must be part of the formalism.

Note that outputs of operations can be inputs to others so that a complicated array can be constructed. For example:



In terms of a collection of tests, we can consider such an array as a system comprising an initial test and contingent subsequent tests by which a subject is prepared for final observations in any one of the six categories at the end. Furthermore, anywhere along the line any one of the operations could, depending on its functioning, reject the subject from the system. Note that in the above diagram, the double entries in θ_3 and θ_4 are merely apparent, since a given copy of the state makes its way only along one route, and by the very interpretation of the formalism, there is never any simultaneous entries or exits. We can think of the whole array above as itself being a single operation with six exits.

Given an operation θ with k exits, we can place a different instrument at each exit:



new instrument
$\theta(I_1, I_2, \ldots, I_k)$

If each I_i is based on $\mathbb{B}_{n_i}$, the combined system is now a new instrument $\theta(I_1, I_2, \ldots, I_k)$ based on $\mathbb{B}_n$, $n = \sum n_i$. Of course even if each I_i is exhaustive, the new instrument may not be, since θ may reject some copies.

We must now introduce into our scheme something that would correspond to the convex combinations of observations introduced in the first chapter. Let $\Lambda = (\lambda_l, \lambda_2, \ldots, \lambda_n)$, $\lambda_i \geq 0$, $\sum \lambda_i = 1$ be a probability measure on n points. We define an operation θ_Λ that functions according to the following prescription: Divide the unit interval $[0, 1]$ into the following subintervals $[0, \lambda_1)$, $[\lambda_1, \lambda_1 + \lambda_2)$, $[\lambda_1 + \lambda_2 + \cdots + \lambda_j, \lambda_1 + \lambda_2 + \cdots + \lambda_{j+1})$, $[\lambda_1 + \lambda_2 + \cdots + \lambda_{n-1}, \lambda_1 + \lambda_2 + \cdots + \lambda_n]$ There are n such intervals, some of which may be empty. Whenever a copy of a state σ is presented to θ_Λ pick a number randomly in $[0, 1]$ from a uniform distribution there and classify the copy of σ as belonging to class i if this number fell within the i-th interval. Since the operation θ_Λ does not involve any interaction with the copy of σ, and in fact the result can be decided even before that particular copy is prepared, we must admit that at each one of the exits of θ_Λ the given copy continues to be a copy of the same state. In other words θ_Λ simply classifies a-priori in a random manner a fraction λ_i of the population σ as belonging to class i without basing the classification on any characteristics of σ at all. We call θ_Λ a *stochastic splitter*.

There is yet another sort of operation that must be introduced into any formalism because it can always be performed in practice. Let I be an instrument with n atoms and $\tau_1, \tau_2, \ldots, \tau_n \in \mathcal{S}$. We can define and operation as follows: Observe a state σ with I, if I does not register, destroy the copy of the state, if the k-th property of I registers, destroy the copy of the state and prepare state τ_k considering this as the outcome of the operation. Notice that here the transformed copy of the state τ_k could have absolutely nothing to do with the original copy of σ since the preparation can be performed independently of anything to do with σ. We call this operation a *substitution operation* and denote it by θ^I_τ where $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$. It could better be called a *sleight of hand operation* being a type of operation by which a magician puts a bottle of champagne in a hat and then pulls out a rabbit. We use both names in this work.

Now except for sleight of hand operations, we shall usually in what follows abandon all mention or investigation of the instrument needed in the mechanism of the operation. All that matters are the results of applying the operation to a state σ. In fact in many operations, such as light filters for example, the instrument is not made accessible to the experimenter, the operation performs its classifications automatically, making the said instruments something that is only virtually present.

We now formally define a statistical theory:

Definition 4.1 *A* statistical theory *consists of the following:*

1. *A convex set $\mathcal{S}$ called the* set of states*.*

2. *A set $\mathcal{I}$ called the* set of instruments*. This set is a disjoint union of subsets $\mathcal{I}_n$, $n = 1, 2\ldots$ where if $I \in \mathcal{I}_n$ we say n is the sort of I and call it the number of its* atomic observation*, or* atoms *for short.*

3. *A set $\mathcal{R}$ called the* set of operations*. This set is a disjoint union of subsets $\mathcal{R}_n$, $n = 1, 2\ldots$ where if $\theta \in \mathcal{R}_n$ we say n is the* sort *of θ and call it the number of* exits *of the operation.*

These objects are subject to the following laws:

Axiom 4.1 *Each $I \in \mathcal{I}_n$is a function $I(\cdot) : \mathbb{B}_n \to Conv(\mathcal{S}, [0, 1])$. We call $I(A)$ the property of the reading of I to fall within A. The number $\langle\sigma, I(A)\rangle$ has the frequentistic interpretation of being the asymptotic frequency with which the event A occurs in the state σ when viewed with I. According to this interpretation we must assume that $\langle a, I(\cdot)\rangle$ is a measure on $\mathbb{B}_n$ for all σ.*

Axiom 4.2 *Each $\theta \in \mathcal{R}_k$ is a map $\theta : \mathcal{I}^k - \mathcal{I}$. We call $\theta(I_1, I_2, \ldots, I_k)$ the instrument constructed by placing I_j on the j-th exit of θ. If n_j is the sort of I_j, then the sort of $\theta(I_1, I_2, \ldots, I_k)$ is $n = \sum n_j$. We identify $\mathbb{B}_n$ with $\oplus\mathbb{B}_{n_j}$ via the injections $\mathbb{B}_{n_j} \to \mathbb{B}_n$ that sends the atom m of $\mathbb{B}_{n_j}$ into the atom $m + \sum_{i=1}^{j-1} n_i$ of $\mathbb{B}_n$.*

Axiom 4.3 *For any $k \geq 1$ and any k-tuple $\Lambda = (\lambda_1, \ldots, \lambda_k)$ of numbers such that $\lambda_i \geq 0$, $\sum \lambda_i = 1$; there is an operation $\theta_\lambda \in \mathcal{R}_k$ such that $\langle\sigma, \theta_\Lambda(I_1, I_2, \ldots, I_k)(A)\rangle = \sum \lambda_i\langle\sigma, I_i(A \cap \mathbb{B}_{n_i})\rangle$ where $A \cap \mathbb{B}_{n_i}$ is the projection of $A \in \mathbb{B}_n$ onto the summand $\mathbb{B}_{n_i}$. We call θ_Λ a* stochastic splitter*. For $k = 1$ we have $\theta_1(I) = I$ and we call θ_1 by Id being the identity map $\mathcal{I} \to \mathcal{I}$.*

Axiom 4.4 *For any* $I \in \mathcal{I}_k$ *and any* $\tau = (\tau_1, \tau_2, \ldots, \tau_k) \in S^k$ *there is an operation* θ^I_τ *such that* $\langle \sigma, \theta^I_\tau(I_1, \ldots, I_k)(A) \rangle = \sum_i \langle \sigma, I(\{i\}) \rangle \langle \tau_i, I_i(A \cap \mathbb{B}_{n_i}) \rangle$ *with the same notation as in axiom 4.3. We call* θ^I_τ *a* substitution *or* sleight of hand operation*.*

Axiom 4.5 *The set* $\mathcal{R}$ *is closed under* compositions*. That is, given* $\theta \in \mathcal{R}_m$ *and* $\psi_1, \ldots, \psi_m$ *in* $\mathcal{R}_{n_1}, \ldots, \mathcal{R}_{n_m}$ *respectively, there is an operation* $\theta\{\psi_1, \ldots, \psi_m\}$ *in* $\mathcal{R}_{n_1 + \cdots + n_m}$ *which satisfies:*

$$\theta\{\psi_1, \ldots, \psi_m\}(I_1^{(1)}, \ldots, I_{n_1}^{(1)}, I_1^{(2)}, \ldots, I_{n_2}^{(2)}, \ldots, I_1^{(m)}, \ldots, I_{n_m}^{(m)}) = \theta(\psi_1(I_1^{(1)}, \ldots, I_{n_1}^{(1)}), \psi_2(I_1^{(2)}, \ldots, I_{n_1}^{(2)}), \ldots, \psi_m(I_1^{(m)}, \ldots, I_{n_m}^{(m)})).$$

This allows us to form systems of operations. For example, the system shown on page 37 is

$$\theta_1\{\theta_2\{\mathrm{Id}, \theta_3\{\mathrm{Id}, \theta_4\}\}, \theta_3\{\mathrm{Id}, \theta_4\}, \theta_4, \mathrm{Id}\}.$$

Axiom 4.6 *We assume* $\mathcal{I}$ *is closed under* condensation*. That is, if* $I \in \mathcal{I}_m$ *and* $\phi : \mathbb{B}_n \to \mathbb{B}_m$ *is a* lattice *morphism, then* $I \circ \phi \in I_n$*. We call* $I \circ \phi$ *the condensation of* I *along* ϕ*. The condensation defined by the map* $\iota : \mathbb{B}_1 \to \mathbb{B}_m$ *which sends the only atom* 1 *of* $\mathbb{B}_1$ *into the largest element* \$ *of* $\mathbb{B}_m$ *is called a* total *or* full condensation *of* I*. It corresponds to forming the conjunction of all the atomic properties of* I*.*

Axiom 4.7 *We assume* $\mathcal{R}$ *is closed under* condensations of exits*, that is any subset of exits can be considered as a single exit. Let* $\theta \in \mathcal{R}_n$ *and* $\phi : \mathbf{n} \to \mathbf{m}$ *be any map. Then we must have* $\phi_* \theta \in \mathcal{R}_m$ *where*

$$\phi_* \theta(I_1, \ldots, I_m) = \theta(I_{\phi(1)}, \ldots, I_{\phi(n)}) \circ \phi^{\S}$$

and where $\phi^{\S} : \oplus_{j=1}^m \mathbb{B}_{n_j} \to \oplus_{i=1}^n \mathbb{B}_{n_{\phi(i)}}$ *is the lattice morphism defined by* $\phi^{\S}(A \cap \mathbb{B}_{n_j}) = \bigcup_i \{A \cap \mathbb{B}_{n_{\phi(i)}} \mid j = \phi(i)\}$*.*

Hence to condense the first two exits of a three exit operation we must define what would happen if we place an instrument I on the condensation of the two exits an instrument J on the third. Place therefore a copy of I on the first two exits, J on the third, and condense the corresponding atoms of the two copies of I.

We also include under this axiom the act of *ignoring* an exit, which we consider as a type of condensation. Thus we assume that if $\theta \in \mathcal{R}_n$ then $\kappa\theta \in \mathcal{R}_{n-1}$ where $\kappa\theta(I_1, I_2, \ldots, I_{n-1})$ is that condensation of $\theta(I_1, I_2, \ldots, I_n)$ which ignores all the atoms of I_n. Operationally this corresponds to rejecting every copy of a state that leaves by the last exit.

Axiom 4.8 *We assume that $\mathcal{I}$ is closed under* exhaustion. *If $I \in \mathcal{I}_n$, then its exhaustion $\hat{I} \in \mathcal{I}_{n+1}$ is defined on say $\mathbb{B}_n \oplus \mathbb{B}_1$ by:*

$$\langle \sigma, \hat{I}(A) \rangle = \begin{cases} \langle \sigma, I(A) \rangle \text{ if } A \cap \mathbb{B}_1 = \emptyset \\ \langle \sigma, I(A \cap \mathbb{B}_n) \rangle + 1 - \langle \sigma, I(\$) \rangle \text{ otherwise.} \end{cases}$$

For any instrument I consider now the total condensation of $\hat{I}$ This is an instrument $\mathbf{1}$ of sort 1 with the property that $\langle \sigma, \mathbf{1} \rangle = 1$ for all σ. Consider now $\hat{\mathbf{1}}$ and condense by ignoring the first atom; this is now an instrument $\mathbf{0}$ of sort 1 such that $\langle \sigma, \mathbf{0} \rangle = 0$ for all σ. With these two special instruments we can now proceed with the axiomatization.

Axiom 4.9 *We assume that the exits of an operation can be viewed as new state preparations if when applied to a state σ, a positive fraction of times the particular exit is used. Let $\theta \in \mathbb{R}_n$ and consider those i for which $\nu_i(\sigma) = \langle \sigma, \theta(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{1}, \mathbf{0}, \ldots, \mathbf{0}) \neq 0$ where $\mathbf{1}$ stands in the i-th place. We must now allow that there exists a state $\theta_*^{(i)}\sigma$ such that for all $I_1, I_2, \ldots, I_n$;*

$$\langle \sigma, \theta(I_1, \ldots, I_n)(A) \rangle = \sum \{ \nu_i(\sigma) \langle \theta_*^{(i)} \sigma, I_i(A \cap \mathbb{B}_{n_i}) \rangle \mid \nu_i(\sigma) \neq 0 \}.$$

Note that $\nu_i(\sigma)$ is the fraction of times a new state exits through i when θ is applied to σ. By convention we will write sums of the form above ranging over all i having it understood that if $\nu_i(\sigma) = 0$, the corresponding term is not present since no state leaves by that exit.

We have in particular for $\theta = \theta_\tau^I$ that $\nu_i(\sigma) = \langle \sigma, I(\{i\}) \rangle$ and $\theta_*^{(i)}\sigma = \tau_i$, and for $\theta = \theta_\Lambda$ that $\nu_i(\sigma) = \lambda_i$, $\theta_*^{(i)}\sigma = \sigma$.

Axiom 4.10 *Instruments separate states. That is, if $\langle \sigma, I(A) \rangle = \langle \sigma', I(A) \rangle$ for all I and A, then $\sigma = \sigma'$.*

We note that defining $I \in \mathcal{I}_n$ as a map $\mathbb{B}_n \to Conv(\mathcal{S}, [0,1])$ and $\theta \in \mathcal{R}_n$ as a map $\mathcal{I}^n \to \mathcal{I}$ we are a already identifying as being n identical any two realizable instruments or operations that have identical statistical behavior.

This is similar to the assumption previously made about the sets $\mathcal{S}$ and $\mathcal{O}$, but now that we have a greater variety of mathematical objects we recover certain distinctions that previously we lost. For example, for tossing a balanced coin, we can introduce the instrument that looks at the two exclusive possibilities of heads and tails, though we still can't tell which is which.

The complexity of the above formalization is its major drawback, fortunately some simplification is possible. We proceed to reformulate the theory to make it more manageable.

Having a statistical theory, we define $\mathcal{O}$ as being a subset of $Conv(\mathcal{S}, [0,1])$ consisting of the union of all images of $I \in \mathcal{I}$:

$$\mathcal{O} = \bigcup \{I(A) \mid A \in \mathbb{B}_I, I \in \mathcal{I}\}.$$

Proposition 4.1 *$\mathcal{O}$ is identifiable with $\mathcal{I}_1$ that is with the instruments of sort* 1.

Proof: If $I \in \mathcal{I}_1$ then we identify I with the image $I(1)$ of the only atom 1 of $\mathbb{B}_1$. If now $p = I(A)$, $I \in \mathcal{I}_n$, $A \in \mathbb{B}_n$, then condensing along the map $\phi : \mathbb{B}_1 \to \mathbb{B}_n$ defined by $1 \mapsto A$ we obtain an instrument of sort 1 which corresponds exactly to p. Q.E.D

Proposition 4.2 *$\mathcal{O}$ is a convex subset of $Conv(S, [O,1])$.*

Proof: Let p_1, $p_2 \in \mathcal{I}_1$ be two elements of $\mathcal{O}$ and $0 \leq \lambda \leq 1$. Consider the stochastic splitter $\theta_{(\lambda,1-\lambda)}$ and let q be the total condensation of $\theta_{(\lambda,1-\lambda)}(p_1, p_2)$. Then $q = \lambda p_1 + (1-\lambda)p_2$ Q.E.D

Proposition 4.3 *$\mathcal{O}$ is closed under* negation*, that is, if $p \in \mathcal{O}$ then there is a $\neg p \in \mathcal{O}$ such that $\langle \sigma, \neg p \rangle = 1 - \langle \sigma, p \rangle$ for all σ.*

Proof: Let $p \in \mathcal{I}_1$ and let I be the exhaustion of p; then $\neg p$ is the image of the new atom of I. Q.E.D

We have already shown on page 41 the following:

Proposition 4.4 *$\mathcal{O}$ contains an element* **1** *such that for all σ, $\langle \sigma, \mathbf{1} \rangle = 1$.*

We have thus shown that $(\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ is a separated statistical triple where $\langle \cdot, \cdot \rangle$ is the restriction of the natural duality between $\mathcal{S}$ and $Conv(\mathcal{S}, [0,1])$ to $\mathcal{S} \times \mathcal{O}$.

We can now consider each $I \in \mathcal{I}_n$ as being a map $I : \mathbb{B}_n \to \mathcal{O}$ and thus interpret each instrument within the triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ as is usually done.

There is still a more cogent way of viewing $\mathcal{I}_n$; it can be interpreted as a subset of $\mathcal{O}^n$ by the identification $I \mapsto (I(1), I(2), \ldots, I(n))$.

Proposition 4.5 *$\mathcal{I}_n$ is a convex subset of $\mathcal{O}^n$.*

Proof: Let $0 \leq \lambda \leq 1$; I, $J \in \mathcal{I}_n$. As an element of $\mathcal{O}^n$, $\lambda I + (1-\lambda)J$ is equal to the condensation of $\theta_{(\lambda,1-\lambda)}(I,J)$ which identifies corresponding atoms of I and J. Q.E.D

Let now $\phi : \mathbf{n} \to \mathbf{m}$ be any partial map and consider the induced map $\phi_* : V_0(\mathcal{O})^n \to V_0(\mathcal{O})^m$ given by $\phi_*(a_1, \ldots, a_n) = (b_1, \ldots, b_m)$ where $b_j = \sum\{a_i \,|\, \phi(i) = j\}$. The condensation axiom for instruments now becomes:

Proposition 4.6 $\phi_*(\mathcal{I}_n) \subset \mathcal{I}_m$.

Here of course we identify $\mathcal{I}_n$with its corresponding subset of $V_0(\mathcal{O})^n$.

Consider now the map $e : V_0(\mathcal{O})^n \to V_0(\mathcal{O})^{n+1}$ given by $e(a_1, \ldots, a_n) = (a_1, \ldots, a_n, \mathbf{1} - a_1 - a_2 + \cdots - a_n)$ The exhaustion axiom for instruments now becomes:

Proposition 4.7 $e(\mathcal{I}_n) \subset \mathcal{I}_{n+1}$.

We now proceed to interpret each operation θ within the triple. Let $\theta \in \mathcal{R}_n$ and define $\theta^{(i)} : \mathcal{O} \to \mathcal{O}$ by the following formula: given p

$$\langle\sigma, \theta^{(i)}p\rangle = \nu_i(\sigma)\langle\theta_*^{(i)}\sigma, p\rangle$$

where in case $\nu_i(\sigma) = 0$, the right hand side is by convention zero. We first need to show that $\theta^{(i)}p \in \mathcal{O}$. But we have from axiom 4.9 that $\theta^{(i)}p$ is the total condensation of $\theta(\mathbf{0}, \ldots, \mathbf{0}, p, \mathbf{0}, \ldots, \mathbf{0})$ where p is in the i-th place. Furthermore, axiom 4.9 shows that $\mathcal{O}$ is completely determined by the $\theta^{(i)}$ since the equation there can be written as:

$$\langle\sigma, \theta(I_1, \ldots, I_k)(A)\rangle = \sum\langle\sigma, \theta^{(i)}I_i(A \cap \mathbb{B}_{n_i})\rangle.$$

Let $\mathcal{Q}$ be the set of all $\theta^{(i)}$ for all $\theta \in \mathcal{R}$.

Proposition 4.8 *$\mathcal{Q}$ is identifiable with $\mathcal{R}_1$.*

Proof: Clearly if $\theta \in \mathcal{R}_1$, then θ is immediately identifiable with $\theta^{(1)} \in \mathcal{Q}$. Let therefore $\theta \in \mathcal{R}_n$, $n > 1$, and consider the condensation of θ obtained by rejecting all exits other than i. As an element of $\mathcal{R}_1$ this operation is now identifiable with $\theta^{(i)}$. Q.E.D

With obvious changes in notation we proceed.

Proposition 4.9 *Each $\theta \in \mathcal{Q}$ is an affine map.*

Proof: For each $\sigma \in \mathcal{S}$, $\langle \sigma, \theta(\lambda p + (1-\lambda)q)\rangle = \nu(\sigma)\langle \theta_*\sigma, \lambda p + (1-\lambda)q\rangle = \lambda\nu(\sigma)\langle \theta_*\sigma, p\rangle + (1-\lambda)\nu(\sigma)\langle \theta_*\sigma, q\rangle = \lambda\langle \sigma, \theta p\rangle + (1-\lambda)\langle \sigma, \theta q\rangle = \langle \sigma, \lambda\theta p + (1-\lambda)\theta q\rangle$. Hence $\theta(\lambda p + (1-\lambda)q) = \lambda\theta p + (1-\lambda)\theta q$ Q.E.D

We have immediately from the definitions.

Proposition 4.10 $\nu(\sigma) = \langle \sigma, \theta\mathbf{1}\rangle$.

In particular we see that in all cases $\langle \sigma, \theta p\rangle \leq \langle \sigma, \theta\mathbf{1}\rangle$.

If now θ, $\theta' \in \mathcal{R}_1$ then $\theta\theta'$ corresponds exactly to the operation $\theta\{\theta'\}$. Hence we have:

Proposition 4.11 *If θ, $\theta' \in \mathcal{Q}$, then $\theta\theta' \in \mathcal{Q}$.*

We thus see that $\mathcal{Q}$ is a semigroup; in fact it is a monoid with zero, since Id is the identity and the map $p \mapsto \mathbf{0}$, the zero, can be identified with the condensation of the stochastic splitter $\theta_{(0,1)}$ obtained by ignoring the second exit.

Proposition 4.12 *$\mathcal{Q}$ is a convex set.*

Proof: If θ_1, $\theta_2 \in \mathcal{Q}$ and $0 \leq \lambda \leq 1$, then $\lambda\theta_1 + (1-\lambda)\theta_2$ is the full condensation of $\theta_{(\lambda,1-\lambda)}\{\theta_1, \theta_2\}$. Q.E.D

We can now identify $\mathcal{R}_k$ with a subset of $\mathcal{Q}^k$ by the map $\theta \mapsto (\theta^{(1)}, \ldots, \theta^{(k)})$ Since every $\theta \in \mathcal{Q}$ has an obvious extension, still called θ, to a linear map $V_0(\mathcal{O}) \to V_0(\mathcal{O})$ we can thus also think of $\mathcal{Q}^k$ as a subset of $(Lin(V_0(\mathcal{O}), V_0(\mathcal{O})))^k$.

Proposition 4.13 *$\mathcal{R}_k$ is a convex subset of $\mathcal{Q}^k$.*

Proof: If θ, $\theta' \in \mathcal{R}_k$ and $0 \leq \lambda \leq 1$, then $\lambda\theta + (1-\lambda)\theta'$ can be identified with the condensation of $\theta_{(\lambda,1-\lambda)}\{\theta, \theta'\}$ which identifies corresponding atoms of θ and θ'. Q.E.D

The existence of stochastic splatters now becomes:

Proposition 4.14 *If* $0 \leq \lambda_i$, $\sum_{i=1}^{k} \lambda_i = 1$, *then* $(\lambda_1 Id, \lambda_2 Id, \ldots, \lambda_k Id) \in \mathcal{R}_k$.

The existence of substitution operations now becomes:

Proposition 4.15 *Let* $I = (p_1, \ldots, p_k) \in \mathcal{I}_k$ *and* $\tau = (\tau_1, \ldots, \tau_k) \in \mathcal{S}^k$, *then* $(\langle \tau_1, \cdot \rangle p_1, \ldots, \langle \tau_k, \cdot \rangle p_k) \in \mathcal{R}_k$

The axiom of composition now becomes:

Proposition 4.16

If $(\theta_1, \ldots, \theta_m) \in \mathcal{R}_m$ and $({\psi^j}_1, \ldots, {\psi^j}_{n_j}) \in \mathcal{R}_{n_j}$; $j = 1, \ldots, m$ then

$$(\theta_1 \psi_1^1, \theta_1 \psi_2^1, \ldots, \theta_1 \psi_{n_1}^1, \theta_2 \psi_1^2, \ldots, \theta_2 \psi_{n_2}^2, \ldots, \theta_m \psi_1^m, \ldots, \theta_m \psi_{n_m}^m) \in \mathcal{R}_n$$

where $n = n_1 + \cdots + n_m$.

Let now $\phi : \mathbf{n} \to \mathbf{m}$ be any partial map and consider the induced map $\phi_0 : (Lin(V_0(\mathcal{O}), V_0(\mathcal{O})))^n \to (Lin(V_0(\mathcal{O}), V_0(\mathcal{O})))^m$ given by $\phi_0(A_1, \ldots, A_n) = (B_1, \ldots, B_m)$ where $B_j = \sum \{A_i \mid \phi(i) = j\}$. The condensation axiom for operations now becomes:

Proposition 4.17 $\phi_0(\mathcal{R}_n) \subset \mathcal{R}_m$.

The fact that operations act on instruments to produce other instruments is now expressed by:

Proposition 4.18 *If* $(\theta_1, \ldots, \theta_k) \in \mathcal{R}_k$ *and* $I_j = ({p^j}_1, \ldots, {p^j}_{n_j}) \in \mathcal{I}_{n_j}$, $j = 1, \ldots, k$ *then* $(\theta_1 p_1^1, \ldots, \theta_1 p_{n_1}^1, \ldots, \theta_k p_1^k, \ldots, \theta_k p_{n_k}^k) \in \mathcal{I}_n$ *where* $n = n_1 + \cdots + n_k$.

In view of the above propositions we now redefine a statistical theory as follows:

Definition 4.1′ *A statistical theory* T *is a statistical triple* $(\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ *endowed with the following additional structure:*

1. For each $n \geq 0$ there is given a convex subset $\mathcal{I}_n \subset \mathcal{O}^n$.
2. There is given a convex subsemigroup $\mathcal{Q}$ of $\hat{C}onv(\mathcal{O}, \mathcal{O})$ containing Id and 0.
3. For each $n \geq 1$ there is given a convex subset $\mathcal{R}_n \subset \mathcal{Q}^n$.

This collection of objects satisfies the following axioms:

Axiom 4.2′ *The statement of Proposition 4.18*

Axiom 4.3′ *The statement of Proposition 4.14*

Axiom 4.4′ *The statement of Proposition 4.15*

Axiom 4.5′ *The statement of Proposition 4.16*

Axiom 4.6′ *The statement of Proposition 4.6*

Axiom 4.7′ *The statement of Proposition 4.17*

Axiom 4.8′ *The statement of Proposition 4.7*

Axiom 4.9′ *Given $\sigma \in \mathcal{S}$ and $\theta \in \mathcal{Q}$, then:*

1. *If $\langle \sigma, \theta \mathbf{1} \rangle \neq 0$, $\exists\, \theta_* \sigma \in \mathcal{S}$ such that $\forall\, p \in \mathcal{O}$, $\langle \sigma, \theta p \rangle = \langle \sigma, \theta \mathbf{1} \rangle \langle \theta_* \sigma, p \rangle$.*
2. *If $\langle \sigma, \theta \mathbf{1} \rangle = 0$ then $\forall\, p \in \mathcal{O}$, $\langle \sigma, \theta p \rangle = 0$.*

We note that there is no axiom correspondent to Axiom 4.1 since the content of that axiom is already implied in the definition; likewise for Axiom 4.10.

Let $(\theta_1, \theta_2, \ldots, \theta_k) \in \mathcal{Q}^k$ be any k-tuple. Given instruments $I_1, I_2, \ldots, I_k$, with I_i based on $\mathbb{B}_n$, consider the map $I : \oplus \mathbb{B}_{n_i} \to Conv(\mathcal{S}, [0,1])$ given by

$$\langle \sigma, I(A) \rangle = \sum \langle \sigma, \theta_i I_i(A \cap \mathbb{B}_{n_i}) \rangle. \tag{4.1}$$

This map may or may not be an instrument. If for every set of instruments $I_1, I_2, \ldots, I_k$; I is an instrument, we say that $(\theta_1, \ldots, \theta_k)$ is *consistent*.

The question now arises whether given a consistent k-tuple $(\theta_1, \theta_2, \ldots, \theta_k)$ if there is an operation $\theta \in \mathcal{R}_k$ such that $\theta_i = \theta^{(i)}$. Given instruments $(I_1, \ldots, I_k)$ the existence of the instrument I above means operationally that it can be physically realized to any degree of approximation. The ability to do so by means other than the existence of an operation θ with $\theta_i = \theta^{(i)}$ is hard to imagine but is not to be excluded. We thus propose to introduce a simplifying assumption at this level of our formalization, noting however that it is a phenomenological assumption.

We say that a statistical theory is *operational* if every consistent k-tuple $(\theta_1, \theta_2, \ldots, \theta_k)$ is of the form $\theta_i = \theta^{(i)}$ for some $\theta \in \mathcal{R}_k$. Noting that for any such θ the k-tuple $(\theta^{(1)}, \ldots, \theta^{(k)})$ is always consistent we see that in an operational theory the sets $\mathcal{R}_k$ are completely determined by the sets $\mathcal{I}$ and $\mathcal{Q}$.

We note that the basic phenomenological assumption which distinguishes operational theories from more general ones is the following: If instruments exist whose behavior in relation to other existing instruments and to existing operations is given by formula (4.1), then it's possible to construct a many exit operation such that placing $I_1, \ldots, I_k$ on the exits we obtain the instrument I. The introduction of this hypothesis rests on certain conceptual distinctions. We can conceive of two types of operations, those that are immediately followed by observations at each exit and which are so tightly bound to their observing instruments that it doesn't make sense to detach them, and those in which we can effect further operations at the exits by detaching the instruments. We can conceive that the first type be possible and the second type not. The first type is not essentially distinguishable from just another type of instrument and it only effectively acts as an operation with instruments on the exits. Saying that an actual operation exists means that we allow for the possibility of removing some of the exit instruments and substituting them by others or by further operations. Of course, there is no reason why a statistical science be operational in these terms. It seems to us though that operational statistical theories is the next simplest level of formalism beyond statistical triples.

We close this chapter with the demonstration that a statistical triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ has associated to it a canonical operational statistical theory that in a certain sense admits the maximum family of instruments and operations.

Let $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ be given. We define $\mathcal{I}_n$ as the set of all maps of the form $I : \mathbb{B}_n \to \mathcal{O}$ where $n \geq 1$, and $\langle\sigma, I(\cdot)\rangle$ is a measure on $\mathbb{B}_n$. For further use we enlarge on this. The Boolean algebra $\mathbb{B}_n$ can be viewed as the set of characteristic functions on $\mathbf{n}$. That $\langle\sigma, I(\cdot)\rangle$ is a measure for all σ means that we can extend I by convexity to the set of all functions $f : \mathbf{n} \to [0,1]$ by setting $I(f) = \sum f(i)I(\{i\})$. Thus an instrument can be viewed as an affine map from the set of observations $\mathcal{O}(\mathbb{B}_n)$ of the Boolean triple based on $\mathbb{B}_n$ to $\mathcal{O}$ with the two absurd observations corresponding. We shall thus also consider I as an affine map $\mathcal{O}(\mathbb{B}_n) \to \mathcal{O}$. Let now Q be the set of all affine maps $\theta : \mathcal{O} \to \mathcal{O}$ that satisfy axiom $4.9'$. We reinterpret $\mathcal{I}_n$ now as being a subset of $\mathcal{O}_n$ and one easily shows that $\mathcal{I}_n$ is convex and Axioms 4.6

$'$ and 4.8$'$ are satisfied. We define $\mathcal{R}_n \subset \mathcal{Q}^n$ as consisting of consistent n-tuples $(\theta_1, \ldots, \theta_n)$. We first show that Q is convex and that Axiom 4.9$'$ is satisfied. To do so we first establish a lemma.

Lemma 4.1 *Let $\theta_i \in \mathcal{Q}$; $r_i \geq 0$, $i = 1, \ldots, k$; $\sigma \in \mathcal{S}$. Interpret each θ_i as a linear transformation of $V_0(\mathcal{O})$. Then if $D = \sum r_i \nu_{\theta_i}(\sigma) \neq 0$ there is a state $\sigma_* \in \mathcal{S}$ such that for all $p \in \mathcal{O}$, $\langle \sigma, \sum r_i \theta_i p \rangle = D \langle \sigma_*, p \rangle$.*

Proof: We have: $\langle \sigma, \sum r_i \theta_i p \rangle = \sum r_i \langle \sigma, \theta_i p \rangle = \sum r_i \nu_{\theta_i}(\sigma) \langle \theta_{i*} \sigma, p \rangle$. Hence we can take for σ_* the state $\sum (r_i \nu_{\theta_i}(\sigma)/D) \theta_{i*} \sigma$. Q.E.D

To establish convexity of $\mathcal{Q}$ we need only note that if θ, $\theta' \in \mathcal{Q}$; $0 \leq \lambda \leq 1$; then clearly $\lambda\theta + (1-\lambda)\theta' \in Conv(\mathcal{O}, \mathcal{O})$. Lemma 4.1 now shows that Axiom 4.9$'$ is satisfied by this convex combination.

To establish Axiom 4.7$'$ let $(\theta_1, \ldots, \theta_n)$ be consistent and $\phi : \mathbf{n} \to \mathbf{m}$ be a partial map. We can always assume that ϕ is surjective, since $(\phi_0 \theta)_j = 0$ for j not in the image of ϕ and hence doesn't contribute. Since ϕ can be factored into a product of surjections, each one of which is injective except on a single pair of points, we can reduce the probem by induction and relabelling to showing that $\theta_1 + \theta_2 \in \mathcal{Q}$ and that $(\theta_1 + \theta_2, \theta_3, \ldots, \theta_n)$ is consistent. Now consistency will follow from the axiom on condensations of instruments, which already has been established. We first show therefore that $\theta_1 + \theta_2$ maps $\mathcal{O}$ into $\mathcal{O}$. Let $p \in \mathcal{I}_1$, by definition of consistency, $\theta(p, p, \mathbf{0}, \ldots, \mathbf{0})$ is an instrument, and its total condensation gives $\theta_1 p + \theta_2 p$. Hence $\theta_1 + \theta_2 : \mathcal{O} \to \mathcal{O}$. Lemma 4.1 now establishes Axiom 4.9$'$ and the condensation axiom for operations is proved.

Let θ, $\theta' \in \mathcal{Q}$ and assume $\langle \sigma, \theta \mathbf{1} \rangle \neq 0$. We have $\langle \theta_* \sigma, \theta' p \rangle \langle \sigma, \theta \mathbf{1} \rangle = \langle \sigma, \theta\theta' p \rangle$. Set $p = \mathbf{1}$ and we have

$$\langle \theta_* \sigma, \theta' \mathbf{1} \rangle \langle \sigma, \theta \mathbf{1} \rangle = \langle \sigma, \theta\theta' \mathbf{1} \rangle.$$

Suppose $\langle \sigma, \theta\theta' \mathbf{1} \rangle \neq 0$ then $\langle \theta_* \sigma, \theta' \mathbf{1} \rangle \neq 0$ and so $\langle \theta_* \sigma, \theta' p \rangle = \langle \theta'_* \theta_* \sigma, p \rangle \langle \theta_* \sigma, \theta' \mathbf{1} \rangle$ hence $\langle \theta'_* \theta_* \sigma, p \rangle \langle \theta_* \sigma, \theta' \mathbf{1} \rangle \langle \sigma, \theta \mathbf{1} \rangle = \langle \sigma, \theta\theta' p \rangle$ or $\langle \theta'_* \theta_* \sigma, p \rangle \langle \sigma, \theta\theta' \mathbf{1} \rangle = \langle \sigma, \theta\theta' p \rangle$ which means that in this case $(\theta\theta')_* \sigma = \theta'_* \theta_* \sigma$. Now under the hypothesis that $\langle \sigma, \theta \mathbf{1} \rangle \neq 0$ we see from the displayed equation above that if $\langle \sigma, \theta\theta' \mathbf{1} \rangle = 0$ then $\langle \theta_* \sigma, \theta' \mathbf{1} \rangle = 0$ hence for all p, $\langle \theta_* \sigma, \theta' p \rangle = 0$ which amounts to $\langle \sigma, \theta\theta' p \rangle = 0$. Suppose now $\langle \sigma, \theta \mathbf{1} \rangle = 0$ then by definition $\langle \sigma, \theta\theta' \mathbf{1} \rangle = 0$ and still by definition we have for all p, $\langle \sigma, \theta\theta' p \rangle = 0$. We have thus shown that $\theta\theta' \in \mathcal{Q}$, and so $\mathcal{Q}$ is a semigroup. Clearly the maps Id and $0 : p \mapsto \mathbf{0}$ belong to $\mathcal{Q}$ and these serve as the identity and zero of the semigroup.

The other axioms are now easily established. Axiom 4.3′ is true since if $\lambda_i \geq 0$, $\sum \lambda_i = 1$; $i = 1, \ldots, n$; then $(\lambda_1 \mathrm{Id}, \ldots, \lambda_n \mathrm{Id})$ is consistent. Axiom 4.2′ follows directly from the definition of consistency. To establish Axiom 4.4′, let $I = (p_1, \ldots, p_k)$ be an instrument, $\tau = (\tau_1, \ldots, \tau_k) \in \mathcal{S}^k$, and $I_j = ({p^j}_1, \ldots, {p^j}_{n_j})$ be any instrument for $j = 1, \ldots, k$. We must show that

$$\langle \tau_1, p_1^1 \rangle p_1, \ldots, \langle \tau_1, p_{n_1}^1 \rangle p_1, \langle \tau_2, p_1^2 \rangle p_2, \ldots, \langle \tau_2, p_{n_2}^2 \rangle p_2, \ldots$$

$$\ldots, \langle \tau_k, p_1^k \rangle p_k, \ldots, \langle \tau_k, p_{n_k}^k \rangle p_k$$

are atoms of an instrument. Consider I as an affine map $\mathcal{O}(\mathbb{B}_k) \to \mathcal{O}$ and let $\psi : \mathcal{O}(\oplus \mathbb{B}_{n_i}) \to \mathcal{O}(\mathbb{B}_k)$ be the affine map given by sending the characteristic function of $\{m\}$ into $\langle \tau_a, p_b^a \rangle$ times the characteristic function of $\{a\}$ where $n_1 + \cdots + n_{a-1} < m \leq n_1 + \cdots + n_a$ and $b = m - (n_1 + \cdots + n_{a-1})$. It is readily checked that ψ is indeed well defined and that $I \circ \psi$ defines as an instrument precisely the set of atomic observations displayed above.

To show Axiom 4.5′, let $\theta = (\theta_1, \ldots, \theta_m)$ and $\psi_j = (\psi_1^j, \ldots, \psi_{n_j}^j)$; $j = 1, \ldots, m$ be consistent. Placing any instruments of the exits of the ψ_j we obtain by consistency instruments I_j, placing these now on the exits of θ we obtain by consistency of θ an instrument I which however is the instrument obtained by placing the original instruments on the exits of $\theta\{\psi_1, \ldots, \psi_m\}$ showing the consistency of this last combination.

The statistical theory that we constructed above from a statistical triple, we call the *canonical statistical theory* associated to the triple. It is the most generous theory that a triple can possess.

We have thus introduced the following levels of generality: 1) Statistical triples, 2) Operational statistical theories, and 3) Statistical theories. We have also shown that each type can be taken to be a particular case of the higher type. Each new type introduces distinctions which the earlier one ignores. An operational statistical theory does not admit that any map $I : \mathbb{B}_n \to \mathcal{O}$ for which for all σ, $\langle \sigma, I(\cdot) \rangle$ is a measure, is necessarily a realizable measuring process. It does however admit that the realizability of certain instruments implies the realizability of certain many exit operations. General theories remove this final assumption.

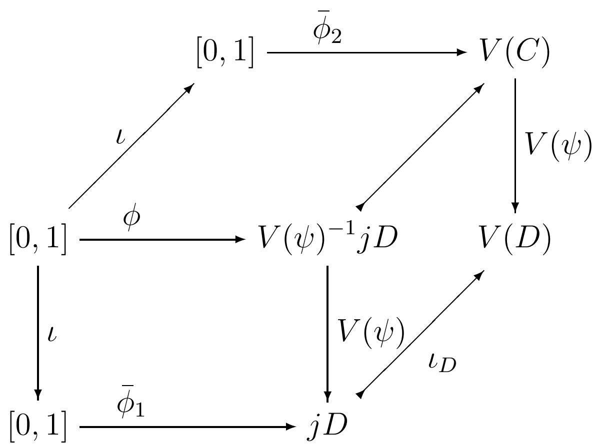
Chapter 5

Idealizations

Mathematical models of phenomena generally have a great difficulty constructing mathematical objects that correspond to actual phenomenal relations. Models generally give simple relations among idealized objects, and these in turn can only be approximated by actual situations. In quantum mechanics, for example, we formalize the possibility of precise measurements, that is we assume that to a certain projection P corresponds a certain observation procedure. Now actually, the physically realizable observation procedure does't correspond to P due to inherent errors and ambiguities in the construction of the apparatus. The projection is an idealized measurement. The actual measuring apparatus would,within a Hilbert space model appear as an operator A, $0 \leq A \leq 1$, but which particular operator this is, is not at all easily determinable. We must admit therefore that a phenomenologically constructed statistical theory could for theoretical reasons be looked upon as beng imbedded within another in which ideal objects can appear, and whose appearance is justified by the relative ease by which they can be treated mathematically. To formalize this notion we first make the following observation: Given a statistical theory, we can place on $\mathcal{S}$ the initial topology with respect to a family of functions of the form: $\sigma \mapsto \langle \sigma, p \rangle$. Dually, we can place on $\mathcal{O}$ the initial topology with respect to some family of functions of the form: $p \mapsto \langle \sigma, p \rangle$. Since each $\mathcal{I}_n$ can be considered as a subset of $\mathcal{O}^n$ we can place on $\mathcal{I}_n$ the topology induced by the product topology on $\mathcal{O}^n$, and $\mathcal{I}$ can be considered as a topological sum of the $\mathcal{I}_n$. Finally on $\mathcal{R}_n$ we can place the topology of pointwise convergence on $\mathcal{I}^n$ seeing that $\mathcal{R}_n$ is a set of maps $\mathcal{I}^n \to \mathcal{I}$. We now view $\mathcal{R}$ as the topological sum of the $\mathcal{R}_n$. Let there now be two statistical theories $T = (\mathcal{S}, \mathcal{I}, \mathcal{R})$ and $T' = (\mathcal{S}', \mathcal{I}', \mathcal{R}')$. we say that T'

is an *idealization* of T if there are given inclusions $\mathcal{S} \subset \mathcal{S}'$, $\mathcal{I} \subset \mathcal{I}'$, $\mathcal{R} \subset \mathcal{R}'$ such that in T' both $\langle \sigma, I(A) \rangle$ and $\theta(I_1, \dots, I_n)$ coincide with their values in T when all objects belong to T, and such that $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{R}$ are dense in $\mathcal{S}'$, $\mathcal{I}'$, and $\mathcal{R}'$ in the following topologies: the topology on $\mathcal{S}'$ is the initial one with respect to the family $\{\sigma' \mapsto \langle \sigma', p \rangle \,|\, p \in \mathcal{O}\}$, and the topology on $\mathcal{O}'$ is the initial one with respect to the family $\{p' \mapsto \langle \sigma, p' \rangle \,|\, \sigma \in \mathcal{S}\}$. The topologies on $\mathcal{I}'$ and $\mathcal{R}'$ are then constructed as explained above. The intuitive content of this density requirement is that ideal objects can be approximated by real ones to any degree of approximation that can be defined by a finite number of real objects. Stronger criteria of approximation can of course be considered. The study of the class of all idealizations of a given theory is an important undertaking since it should provide us with a way of conceiving all possible ideal theoretical frameworks for a given empirical situation. we don't pursue this question further, but point out that except for the explicitly empirical triple discussed in the first chapter, all of the examples so far considered should be thought of as already being idealizations of some unspecified but more phenomenological theories.

One idealization that can always be carried out is to assume each of the convex sets $\mathcal{S}$, $\mathcal{I}_n$, $\mathcal{R}_n$ to be algebraically closed. We call a convex set C *algebraically closed* if each morphism $\phi : [0,1) \to C$ can be extended to a morphism $\bar{\phi} : [0,1] \to C$. Each convex set is contained in an algebraically closed set $\bar{C}$ such that if D is any algebraically closed set, and $\psi : C \to D$ is a morphism, then there is a unique extension $\bar{\psi} : \bar{C} \to D$. To see this we note that the *Conv* inclusion $\iota : [0,1) \subset [0,1]$ is a bimorphism, for as a *Set* map we can identify $V(\iota) : V([0,1)) \to V([0,1])$ with $\mathbb{R}^2 \simeq \mathbb{R}^2$ which certainly gives a bimorphism in *Bsn*. This means that any two extensions of $\phi : [0,1) \to C$ to $[0,1]$ are equal. From this we can conclude that the intersection of any family of algebraically closed sets is algebraically closed. Thus let $\bar{C}$ be the intersection of all algebraically closed sets containing jC in $V(C)$. We see that $\bar{C} \subset \{x \in V(C) \,|\, \tau(x) = 1\}$ since the latter hyperplane is algebraically closed. If $\psi : C \to D$ is a morphism into an algebraically closed convex set D we have by the fact that $V(C)$ is algebraically closed that given any morphism $\phi : [0,1) \to V(\psi)^{-1} jD$ there is a commutative diagram:



Now $V(\psi) \circ \bar{\phi}_2 \circ \iota = \iota_D \circ \bar{\phi}_1 \circ \iota \Rightarrow V(\psi) \circ \bar{\phi}_2 = \iota_D \circ \bar{\phi}_1$ since ι is an epimorphism. This shows that $\bar{\phi}_2([0,1]) \subset V(\psi)^{-1}jD$, that is, that $V(\psi)^{-1}jD$ is algebraically closed, hence $\bar{C} \subset V(\psi)^{-1}jD$ and $V(\psi)\bar{C} \subset jD$, proving the claim.

The reader can now straightforwardly, though laboriously, check that replacing $\mathcal{S}$, $\mathcal{I}_n$, and $\mathcal{R}_n$ by their algebraic closures we can uniquely extend all of the defining morphisms of a statistical theory to act within the algebraic closures, thus passing to an idealization where all of the defining convex sets are algebraically closed.

Chapter 6

The Category of Statistical Triples

Any statistical theory embodies only a partial knowledge of the world. As we pursue our investigation of phenomena, we may embody our knowledge in various separate statistical theories. Assuming a certain fundamental unity in nature, we expect that this fragmented knowledge can be combined into a unified viewpoint. Thus we expect a certain calculus of statistical theories by which they can be related and combined. The main result of this chapter is that statistical triples form a certain bicomplete monoidally closed category, providing us therefore with some of the necessary operations for such a calculus. Because of the complexities involved, we've not investigated whether statistical theories as defined in Chapter 4 form a similar category in any natural way. If so, it would be a strong argument for the adequacy of the concepts, and if not it should be cause for a search for modifications.

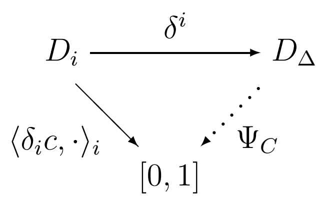
To prove our result we first establish a more general one and then specialize to statistical triples.

Let $P = (C, D, \langle\cdot,\cdot\rangle)$ and $P' = (C', D', \langle\cdot,\cdot\rangle')$ be two dualities and $f : C \to C'$ and affine map. We then have an induced morphism $f^* : D' \to \hat{Conv}(C, [0, 1])$ given by $(f^*d')c = \langle fc, d'\rangle'$. In case $f^*D' \subset D$ we say f is a *morphism of dualities* and we have: $\langle fc, d'\rangle' = \langle c, f^*d'\rangle$. We denote by *Dual* the category whose objects are dualities and whose morphisms are the maps just described. Note that a *Dual* morphism $f : P \to P'$ is also a *Conv* morphism $f : C \to C'$ and also a *Set* morphism of the underlying sets. The context will generally supply the needed interpretation. We also have two functors C, D : *Dual* $\to$ *Conv* which using the above notation are defined as follows: $\mathrm{C}(P) = C$, $\mathrm{C}(f) = f$; $\mathrm{D}(P) = D$, $\mathrm{D}(f) = f^*$. Of these functors C is covariant and D is contravariant. We also have the contravariant involution $J : \mathit{Dual} \to \mathit{Dual}$ already defined on page 31; clearly $J(f) = f^*$.

We denote by *Trip* the full subcategory of *Dual* of statistical triples.

Theorem 6.1 *Dual is a bicomplete monoidally closed category.*

Proof: Let Δ be a diagram in *Dual* with a typical arrow $\delta_{ij} : P_i \to P_j$. We are then provided with two diagrams $\mathrm{C}\Delta$ and $\mathrm{D}\Delta$ in *Conv* with typical arrows $\delta_{ij} : C_i \to C_j$ and $\delta^*_{ij} : D_j \to D_i$ respectively. Since *Conv* is bicomplete consider the limit C of $\mathrm{C}\Delta$ with projections $\delta_i : C \to C_i$ and the colimit D_Δ of $\mathrm{D}\Delta$ with injections $\delta^i : D_i \to D_\Delta$ Let now $c \in C$ and define the morphism $\Psi_C : D_\Delta \to [0,1]$ by universality from the following diagram:



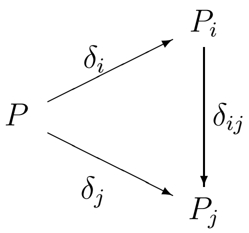
and let $\langle \cdot, \cdot \rangle_\Delta : C \times D_\Delta \to [0,1]$ be given by $\langle c, d \rangle_\Delta = \Psi_c(d)$. We must show that $\langle \cdot, \cdot \rangle_\Delta$ is biaffine. The affinity in d is trivial, we show affinity in c: $\Psi_{\lambda c_1 + (1-\lambda) c_2} \circ \delta^i = \langle \delta_i(\lambda c_1 + (1-\lambda)c_2), \cdot \rangle_i = \lambda \langle \delta_i c_1, \cdot \rangle_i + (1-\lambda)\langle \delta_i c_2, \cdot \rangle = \lambda \Psi_{c_1} \circ \delta^i + (1-\lambda)\Psi_{c_2} \circ \delta^i = (\lambda \Psi_{c_1} + (1-\lambda)\Psi_{c_2}) \circ \delta^i$ which by universality implies $\Psi_{\lambda c_1 + (1-\lambda)c_2} = \lambda \Psi_{c_1} + (1-\lambda)\Psi_{c_2}$

We now make a few observations. Since $\delta_i : C \to C_i$, we have a map $\delta_\pi : C \to \prod C_i$. As this map is injective as a map of sets whenever C is a product of equalizer, we see that it's injective for any diagram. Likewise from $\delta^i : D_i \to D_\Delta$ we obtain a map $\delta^{\amalg} : \coprod D_i \to D_\Delta$ and as this is surjective as a map of sets when D_Δ is a coproduct or coequalizer, it's surjective for any diagram.

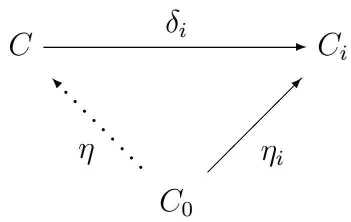
Assume now that $\Psi_{c_1} = \Psi_{c_2}$ thus $\Psi_{c_1} \circ \delta^i = \langle \delta_i c_1, \cdot \rangle_i = \langle \delta_i c_2, \cdot \rangle_i = \Psi_{c_2} \circ \delta^i$. Since each P_i is separated we have $\delta_i c_1 = \delta_i c_2$ for all i and hence $\delta_\pi c_1 = \delta_\pi c_2 \Rightarrow c_1 = c_2$ since δ_π is injective.

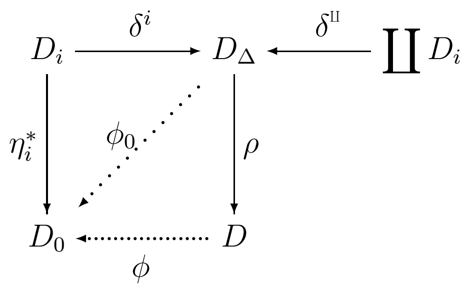
On the other hand $\langle \cdot, \cdot \rangle_\Delta$ may not separate points of D_Δ. Let $P = (C, D, \langle \cdot, \cdot \rangle)$ be the reduced system associated to $(C, D_\Delta, \langle \cdot, \cdot \rangle_\Delta)$. We show that P is the limit of Δ. Let $\rho : D_\Delta \to D$ be the canonical morphism and define $\gamma^i = \rho \circ \delta^i$. We first show that $\gamma^i = \delta^*_i$; this is equivalent to $\langle \delta_i c, d_i \rangle_i = \langle c, \gamma^i d_i \rangle$, but $\gamma^i d_i = [\delta^i d_i]$ so $\langle c, \gamma^i d_i \rangle = \langle c, [\delta^i d_i] \rangle = \langle c, \delta^i d_i \rangle_\Delta =$

$\Psi_c(\delta^i d_i) = \langle \delta_i c, d_i \rangle_i$. Thus $\delta_i : P \to P_i$ is a morphism in *Dual* with $\delta_i^* = \gamma^i$. By construction the diagram below commutes

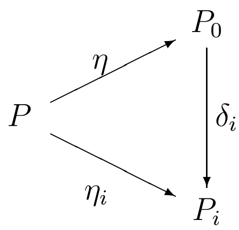


Let now $\eta_i : P_0 \to P_i$ be any family of morphisms compatible with Δ. We have the diagrams:





By universality η and ϕ_0 exist and are unique. To show the existence and uniqueness of ϕ we must show that if $d_1 \sim d_2$ in D_Δ then $\phi_0(d_1) = \phi_0(d_2)$ and thus defining $\phi[d]$ by $\phi_0(d)$. Since $\delta^{\amalg}$ is surjective, we have for $d \in D_\Delta$ a representation $d = \sum \lambda_i \delta^i d_i$; thus, $\phi_0(d) = \sum \lambda_i \phi_0 \delta^i d_i = \sum \lambda_i \eta_i^* d_i$. For $c_0 \in C_0$ then $\langle \eta c_0, d \rangle_\Delta = \sum \lambda_i \langle \eta c_0, \delta^i d_i \rangle_\Delta = \sum \lambda_i \langle \delta_i \eta c_0, d_i \rangle_i = \sum \lambda_i \langle \eta_i c_0, d_i \rangle_i = \sum \lambda_i \langle c_0, \eta_i^* d_i \rangle_0 = \langle c_0, \phi_0 d \rangle_0$. On the other hand if $d_1 \sim d_2$ have $\langle \eta c_0, d_1 \rangle_\Delta = \langle \eta c_0, d_2 \rangle_\Delta \Rightarrow \langle c_0, \phi_0 d_1 \rangle_0 = \langle c_0, \phi_0 d_2 \rangle_0 \Rightarrow \phi_0 d_1 = \phi_0 d_2$ since P_0 is separated. Hence ϕ exists and is unique. We now have $\langle \eta c_0, [d] \rangle = \langle \eta c_0, d \rangle_\Delta = \langle c_0, \phi_0 d \rangle_0 = \langle c_0, \phi[d] \rangle_0$ establishing that $\phi = \eta^*$. Clearly by construction the diagram



commutes and so P is indeed the limit of Δ.

The colimit of Δ is now clearly $J \lim J\Delta$ and so *Dual* is bicomplete.

We now proceed to monoidal closure. Consider the set $Dual(P, P')$, this is canonically identified with a subset of $Conv(C, C')$. Let now f_1, $f_2 \in Dual(P, P')$, then it's easy to see that $(\lambda f_1 + (1-\lambda) f_2)^* = \lambda f_1^* + (1-\lambda) f_2^*$, hence by convexity of D, $(\lambda f_1^* + (1-\lambda) f_2^*)(D') \subset D$ or in other words $Dual(P, P')$ as a subset of $\hat{Conv}(C, C')$ is convex. Let $f \in Dual(P, P')$ and $(c, d') \in C \times D'$; then we have a map $C \times D' \to [0,1]$ given by $(c, d') \mapsto \langle fc, d' \rangle' = \langle c, f^* d' \rangle$. Being biaffine it defines a morphism $\Psi_f : C \otimes D' \to [0,1]$ such that $c \otimes d' \mapsto \langle fc, d' \rangle$. The map $\langle \cdot, \cdot \rangle_\times : Dual(P, P') \times (C \otimes D') \to [0,1]$ given by $\langle f, r \rangle_\times = \Psi_f(r)$ is biaffine and defines a pairing $(Dual(P, P'), C \otimes D', \langle \cdot, \cdot \rangle_\times)$. Now $\langle \cdot, \cdot \rangle_\times$ clearly separates points of $Dual(P, P')$ but may not separate points of $C \otimes D'$. We let $\hat{Dual}(P, P')$ be the associated reduced system $(Dual(P, P'), C \otimes D'/\sim, \langle \cdot, \cdot \rangle_\otimes)$. It's easy to see that it defines a contra-covariant functor $Dual \times Dual \to Dual$. We now exhibit a tensor product $\otimes$ in *Dual* and a natural bijection

$$Dual(P_1, \hat{Dual}(P_2, P_3)) \simeq Dual(P_1 \otimes P_2, P_3).$$

we thus define $C = \mathrm{C}(P_1 \otimes P_2)$ to be $C_1 \otimes C_2$ and $D = \mathrm{D}(P_1 \otimes P_2)$ to be the set of those morphisms $f : C_1 \otimes C_2 \to [0,1]$ such that for all $c_1 \in C_1$, and all $c_2 \in C_2$ we have $f(\cdot \otimes c_2) \in D_1$, and $f(c_1 \otimes \cdot) \in D_2$; of course $\langle r, f \rangle = f(r)$. Now $\langle \cdot, \cdot \rangle$ separates points of D by definition. To show separation in C it is enough to note that there is a canonical injection $D_1 \otimes D_2 \to D$ defined by means of: $\langle c_1 \otimes c_2, d_1 \otimes d_2 \rangle = \langle c_1, d_1 \rangle \langle c_2, d_2 \rangle$. Now $Dual(P_1, \hat{Dual}(P_2, P_3)) \subset Conv(C_1, \hat{Conv}(C_2, C_3)) = Conv(C_1 \otimes C_2, C_3)$ so we have a natural injection $U : Dual(P_1, \hat{Dual}(P_2, P_3)) \to Conv(C_1 \otimes C_2, C_3)$. If now $\phi \in Dual(P_1, \hat{Dual}(P_2, P_3))$ we must show that $U\phi \in Dual(P_1 \otimes P_2, P_3)$ which amounts to showing that if $d \in D_3$ then $(U\phi)^* d \in D$. We have $\langle c_1 \otimes c_2, (U\phi)^* d \rangle = \langle (U\phi) c_1 \otimes c_2, d \rangle = \langle (\phi c_1)(c_2), d \rangle = \langle c_2, (\phi c_1)^* d \rangle$, also

$\langle(\phi c_1)(c_2), d\rangle = \langle \phi c_1, [c_2 \otimes d]\rangle = \langle c_1, \phi^*[c_2 \otimes d]\rangle$. From these equations we see that $((U\phi)^*d)(c_1 \otimes \cdot) = \langle \cdot, (\phi c_1)^* d\rangle \in D_2$ and $((U\phi)^*d)(\cdot \otimes c_2) = \langle \cdot, (\phi)^*[c_2 \otimes d]\rangle \in D_1$ hence $U\phi \in \mathit{Dual}(P_1 \otimes P_2, P_3)$. Reciprocally consider now $\psi \in \mathit{Dual}(P_1 \otimes P_2, P_3)$ thus $\psi : C_1 \otimes C_2 \to C_3$ and this defines canonically $(W^*\psi) : C_1 \to \hat{Conv}(C_2, C_3)$ by $((W\psi)(c_1))(c_2) = \psi(c_1 \otimes c_2)$. We must show that $(W\psi)(C_1) \subset \mathit{Dual}(P_2, P_3)$ and that $(W\psi) \in \mathit{Dual}(P_1, \hat{Dual}(P_2, P_3))$. Let $c \in C_1$, then $\langle c_2, ((W\psi)(c_1))^* d\rangle = \langle ((W\psi)(c_1))c_2, d\rangle = \langle \psi(c_1 \otimes c_2), d\rangle = \langle c_1 \otimes c_2, \psi^* d\rangle = (\psi^* d)(c_1 \otimes c_2)$. Since $\psi^* d \in \mathrm{D}(P_1 \otimes P_2)$ then $(\psi^* d)(c_1 \otimes \cdot) \in D_2$ which means that $((W\psi)(c_1))^* d \in D_2$ and so $(W\psi)(c_1) \in \mathit{Dual}(P_2, P_3)$. Let now $[c' \otimes d] \in \mathrm{D}(\hat{Dual}(P_2, P_3))$; we have $\langle c, (W\psi)^*[c' \otimes d]\rangle = \langle ((W\psi)c)c', d\rangle = \langle \psi(c \otimes c'), d\rangle = \langle c \otimes c', \psi^* d\rangle$. Since $(\psi^* d)(\cdot \otimes c') \in D_1$ we have $(W\psi)^*[c' \otimes d] \in D_1$ and so $(W\psi) \in \mathit{Dual}(P_1, \hat{Dual}(P_2, P_3))$ establishing finally a bijection and proving the theorem. Q.E.D

Theorem 6.2 *Trip is a bicomplete monoidally closed category.*

Proof: Let Δ be a nonempty diagram in *Trip*; it has a limit and colimit in *Dual*. It only remains to show that these objects belong to *Trip*. To do so we must exhibit a negation for the observations and the existence of an absurd observation $\mathbf{0}$. We borrow notation from the proof of the previous theorem. Let $f : T \to T'$ be a morphism in *Trip* where $T = (\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ and $T' = (\mathcal{S}', \mathcal{O}', \langle \cdot, \cdot \rangle')$. We have $\langle \sigma, f^* \neg p\rangle = \langle f\sigma, \neg p\rangle' = 1 - \langle f\sigma, p\rangle' = 1 - \langle \sigma, f^* p\rangle = \langle \sigma, \neg f^* p\rangle$ hence f^* commutes with negation; furthermore, $\langle \sigma, f^* \mathbf{0}'\rangle = \langle f\sigma, \mathbf{0}'\rangle' = 0$ hence $f^* \mathbf{0}' = \mathbf{0}$. Let now $T = (\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ be the limit of Δ in *Dual*. In this case $\mathcal{O} = \mathcal{O}_\Delta / \sim$ where $\mathcal{O}_\Delta$ is the colimit of $\mathrm{D}\Delta$. Since δ^*_{ij} commutes with negation we have by universality a unique morphism $\neg : \mathcal{O}_\Delta \to \mathcal{O}_\Delta$ which renders commutative the diagram

$$\begin{array}{ccccc}
\mathcal{O}_i & \xrightarrow{\delta^i} & \mathcal{O}_\Delta & \xrightarrow{\rho} & \mathcal{O} \\
\neg \downarrow & & \neg \downarrow & & \\
\mathcal{O}_i & \xrightarrow{\delta^i} & \mathcal{O}_\Delta & \xrightarrow{\rho} & \mathcal{O}
\end{array}$$

We show that $\neg$ can be defined in $\mathcal{O}$. If $p \in \mathcal{O}_\Delta$ then by surjectivity of $\delta^{\amalg}$ we can write $p = \sum \lambda_i \delta^i p_i$ and by the diagram $\neg p = \sum \lambda_i \delta^i \neg p_i$. Thus $\Psi_\sigma(p) = \sum \lambda_i \langle \delta_i \sigma, p_i\rangle_i$ and so $\Psi_\sigma(\neg p) = \sum \lambda_i \langle \delta_i \sigma, \neg p_i\rangle_i = 1 - \sum \lambda_i \langle \delta_i \sigma, p_i\rangle_i = 1 - \Psi_\sigma(\neg p)$. From this we conclude that if p is equivalent to q, then so is $\neg p$ to $\neg q$, and $\neg$ can be defined by $\neg[p] = [\neg p]$, and furthermore $\langle \sigma, \neg[p]\rangle =$

$\langle\sigma,[\neg p]\rangle = \langle\sigma,\neg p\rangle_\Delta = \Psi_\sigma(\neg p) = 1-\Psi_\sigma(p) = 1-\langle\sigma,p\rangle_\Delta = 1-\langle\sigma,[p]\rangle$ showing that $\neg$ is indeed a negation.

To show the existence of $\mathbf{0}$ in $\mathcal{O}$, let $\mathbf{0}_i \in \mathcal{O}_i$ be its absurd observation. Then $\gamma^i\mathbf{0}_i \in \mathcal{O}$ and for $\sigma \in \mathcal{S}$ we have $\langle\sigma,\gamma^i\mathbf{0}_i\rangle = \langle\sigma,\delta^i\mathbf{0}_i\rangle_\Delta = \langle\delta_i\sigma,\mathbf{0}_i\rangle_i = 0$. Hence any $\gamma^i\mathbf{0}_i$ is the absurd observation of $\mathcal{O}$.

Remembering now that the colimit is computed as $J\lim J$, and using when necessary an appropriate notation, we now assume that $T = (\mathcal{S},\mathcal{O},\langle\cdot,\cdot\rangle)$ is the colimit in *Dual* of Δ with injections $\delta^i : \mathcal{O}_i \to \mathcal{O}$. Since δ^*_{ij} commutes with negation, we are given by universality an affine isomorphism $\neg : \mathcal{O} \to \mathcal{O}$ commuting with the δ^{i*}. Now $\Psi_{\neg p}\circ\delta^i = \langle\cdot,\delta^*_i\neg p\rangle_i = \langle\cdot,\neg\delta^*_i p\rangle = 1-\langle\cdot,\delta^*_i p\rangle = 1-\Psi_p\circ\delta^i = (1-\Psi_p)\circ\delta^i$ and so by universality $\Psi_{\neg p} = 1-\Psi_p$. Since $\mathcal{S} = \mathcal{S}_\Delta/\!\sim$, $\langle[\sigma],\neg p\rangle = \langle\sigma,\neg p\rangle_\Delta = \Psi_{\neg p}(\sigma) = 1-\Psi_p(\sigma) = 1-\langle\sigma,p\rangle_\Delta = 1-\langle[\sigma],p\rangle$ and so $\neg$ is indeed a negation. Consider now the morphism $\omega_i : \{*\} \to \mathcal{O}_i$ given by $* \mapsto \mathbf{0}_i$. Universality provides a map $\omega : \{*\} \to \mathcal{O}$ which determines an element $\omega(*) = \mathbf{0}$. Since $\Psi_{\mathbf{0}}\circ\delta^i = \langle\delta^i\cdot,\mathbf{0}_i\rangle_i$. we see by universality that $\mathbf{0}$ is the absurd observation of $\mathcal{O}$.

To complete the proof of bicompleteness we must exhibit a final and initial object which correspond to the limit and colimit of an empty diagram. It's immediately apparent that the initial object is $(\emptyset,\{*\},\emptyset)$ where $\emptyset : \emptyset\times\{*\} \to [0,1]$ is the empty morphism. The final object is $(\{*\},[0,1],\langle\cdot,\cdot\rangle)$ where $\langle *,\lambda\rangle = \lambda$. Note that this is different from the final object in *Dual* which is $(\{*\},\emptyset,\emptyset)$.

We now proceed to monoidal closure.

Consider now $\hat{Dual}(T,T')$ for two statistical triples T, T'. We have the map $\mathbf{1}_S\times\neg : \mathcal{S}\times\mathcal{O}' \to \mathcal{S}\times\mathcal{O}'$ which being an isomorphism induces an isomorphism $\neg : \mathcal{S}\otimes\mathcal{O}' \to \mathcal{S}\otimes\mathcal{O}'$, such that $\neg(\sigma\otimes p') = \sigma\otimes\neg p'$. One easily sees that this map is compatible with the equivalence relation in $\mathcal{S}\otimes\mathcal{O}'$ and since $\langle f,\neg(\sigma\otimes p')\rangle_\times = \langle f,(\sigma\otimes\neg p')\rangle_\times = \langle f\sigma,\neg p'\rangle' = 1-\langle f\sigma,p'\rangle' = 1-\langle f,(\sigma\otimes p')\rangle_\times$ the map is a negation. Since $\langle f,[\sigma\otimes\mathbf{0}']\rangle = \langle f,\sigma\otimes\mathbf{0}'\rangle_\times = \langle f\sigma,\mathbf{0}'\rangle_\times = 0$ we see that $\mathbf{0} = [\sigma\otimes\mathbf{0}']$ is the absurd observation. Thus $\hat{Dual}(T,T')$ is a statistical triple and we can thus identify it with an object in *Trip* which we denote by $\hat{Trip}(T,T')$.

Consider now the tensor product in *Dual* of two statistical triples: $T\otimes T'$. If $f \in \mathrm{D}(T\otimes T')$, then for all $\sigma\in\mathcal{S}$, $\sigma'\in\mathcal{S}'$; $f(\cdot\otimes\sigma')\in\mathcal{O}$ and $f(\sigma\otimes\cdot)\in\mathcal{O}'$; thus $1-f(\cdot\otimes\sigma')\in\mathcal{O}$ and $1-f(\sigma\otimes\cdot)\in\mathcal{O}'$, and so defining $\neg f$ by $\langle r,\neg f\rangle = 1-\langle r,f\rangle$ we obtain $\neg f \in \mathrm{D}(T\otimes T')$ and so $\neg$ is a negation. Clearly the map that sends $\mathcal{S}\otimes\mathcal{S}'$ to 0 in $[0,1]$ belongs to $\mathrm{D}(T\otimes T')$ and is the absurd observation of $\mathrm{D}(T\otimes T')$. We can now consider $T\otimes T'$ as an object of

Trip. Since *Trip* is a full subcategory of *Dual* and since the tensor products coincide in *Dual* and *Trip*, and $\hat{Dual}(T,T')$ coincides with $\hat{Trip}(T,T')$ we still maintain a natural bijection $Trip(T_1, \hat{Trip}(T_2,T_3)) = Trip(T_1 \otimes T_2, T_3)$ showing that *Trip* is monoidally closed. Q.E.D

For the sake of gaining some familiarity we cite a few examples

1. *Product* $T \times T' = (\mathcal{S} \times \mathcal{S}', (\mathcal{O} \amalg \mathcal{O}')/\sim, \langle \cdot, \cdot \rangle))$ where $\langle (\sigma, \sigma'), [\lambda p + (1-\lambda)p'] \rangle = \lambda \langle \sigma, p \rangle + (1-\lambda)\langle \sigma', p' \rangle'$ and we have $\lambda p + (1-\lambda)p' \sim \mu q + (1-\mu)q'$ if and only if in $V_0(\mathcal{O} \amalg \mathcal{O}')$ of the pairing $(\mathcal{S} \times \mathcal{S}', \mathcal{O} \amalg \mathcal{O}', \langle \cdot, \cdot \rangle)_\Delta)$ they differ by $\alpha \mathbf{1} - \alpha \mathbf{1}'$ for a real α. Operationally, the experiment $((\sigma, \sigma'), [\lambda p + (1-\lambda)p'])$ is performed as follows: prepare simultaneously copies of σ and σ' in a way that they don't interfere with each other; toss a coin with head-tail probabilities $(\lambda, 1-\lambda)$, if the result is heads, observe p on the copy of σ, if tails, observe p' on the copy of σ'.

   The product is an operation that has not been usually considered. For example, the product of two Kolmogorov, or two quantum triples is not one of these again, nor readily related to one.

2. *Coproduct* $T \amalg T' = (\mathcal{S} \amalg \mathcal{S}', \mathcal{O} \times \mathcal{O}', \langle \cdot, \cdot \rangle))$ where $\langle \lambda \sigma + (1-\lambda)\sigma', (p, p') \rangle = \lambda \langle \sigma, p \rangle + (1-\lambda)\langle \sigma', p' \rangle'$. The experiment $(\lambda \sigma + (1-\lambda)\sigma', (p, p'))$ is performed as follows: toss a coin with head-tail probabilities $(\lambda, 1-\lambda)$; if the result is heads, perform the experiment (σ, p), if tails, perform (σ', p').

   If T and T' are two quantum triples within Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ respectively, then we can identify $\mathcal{S} \amalg \mathcal{S}'$ as being those density matrices that are of the form $\lambda \rho \oplus (1-\lambda)\rho'$, $\rho \in \mathcal{S}, \rho' \in \mathcal{S}'$; and $\mathcal{O} \times \mathcal{O}'$ with the set of operators of the form $A \oplus A'$, $A \in \mathcal{O}$, $A' \in \mathcal{O}'$. This means that $T \amalg T'$ is a quantum theory with $\mathcal{H}$ and $\mathcal{H}'$ as superselection sectors. A coproduct decomposition of a theory can therefore be viewed as generalizing the notion of decomposition into superselection sectors.

   For two Kolmogorov triples based on measurable spaces (X, Σ) and (X', Σ'), the coproduct is the Kolmogorov triple based on $(X \amalg X', \Sigma \amalg \Sigma')$.

3. *Tensor product* Consider the canonical map $\mathcal{O} \times \mathcal{O}' \to \mathrm{D}(T \otimes T')$ defined by $\langle \sigma \otimes \sigma', p \otimes p' \rangle = \langle \sigma, p \rangle \langle \sigma', p' \rangle$ To perform the experiment $(\sigma \otimes \sigma', p \otimes p')$ we must simultaneously prepare copies of σ and σ' in a way that they do not interfere with each other, and observe p on σ and

p' on σ' with the final result being the conjunction of the two separate observational results. The generic experiment in the tensor product is in general an idealized one.

For two Kolmogorov triples in measurable spaces (X, Σ), (X', Σ'), the Kolmogorov triple in $(X \times X', \Sigma \times \Sigma')$ is an idealization of $T \otimes T'$. Likewise for two quantum triples in $\mathcal{H}$ and $\mathcal{H}'$, the quantum triple in $\mathcal{H} \otimes \mathcal{H}'$ is an idealization of $T \otimes T'$. In the first case this idealization becomes identity if X or X' is finite and in the second case if $\mathcal{H}$ or $\mathcal{H}'$ is finite dimensional.

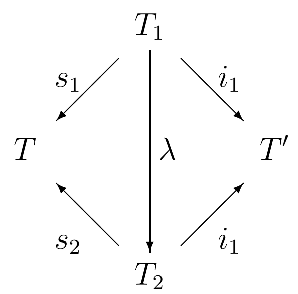
Though we have established the existence of a natural category of statistical triples, it must be noted that the morphisms of this category don't represent in a totally satisfactory way our notions of comparing two systematizations of partial knowledge. Each statistical triple T systematizes a set of procedures for the preparation of copies of states, and procedures for observations of properties. Suppose on T' we embody more descriptions of state preparations than in T, but both theories embody the same procedures for observation of properties. Then every operation in T is the result of restricting ones in T' to the fewer states of T. Hence we have an inclusion $i : \mathcal{S} \subset \mathcal{S}'$ with $i^*\mathcal{O}' \subset \mathcal{O}$ and the possibility that different observations in T' can become identified in T. On the other hand if T' embodies more procedures for observation or properties than T but both theories embody the same procedures for state preparation, then one must have an inclusion $j^* : \mathcal{O} \subset \mathcal{O}'$ and a corresponding surjection $j : \mathcal{S}' \to \mathcal{S}$. In this case state preparations statistically indistinguishable in T may become distinguishable in T'. Thus in both cases T' should be considered a stronger theory than T but in the first case we have $i : T \to T'$ and in the second $j : T' \to T$ which are arrows with different directions. Under more complicated conditions the situation cannot be represented by a morphism in *Trip*. Consider for instance when T' is an idealization of T, then $\mathcal{S} \subset \mathcal{S}'$ and $\mathcal{O} \subset \mathcal{O}'$ and under most circumstances this corresponds to neither a morphism from T to T' nor to one from T' to T.

To analyze these more complex situations consider the following parable. Suppose a master experimenter E' who systematizes the world according to T' hires an apprentice E whose presumably weaker power of systematization is given by a triple T. If E prepares a copy of a state σ for E', the latter notices, by his greater power of observation that somewhat different preparations, though for E would be just as good for preparing σ would in fact prepare different states for E'. We must thus assume that a subset $\mathcal{S}_0 \subset \mathcal{S}'$ corresponds in a many-one fashion to S. On the other hand anything E can observe about S, E' can also, hence to every $p \in \mathcal{O}$ there must be a $p' \in \mathcal{O}'$ which gives $\langle \sigma, p \rangle$ on any element of S_0 that corresponds to σ. Thus O' restricted to S_0 must be able to embody $\mathcal{O}$. We must thus posit a diagram

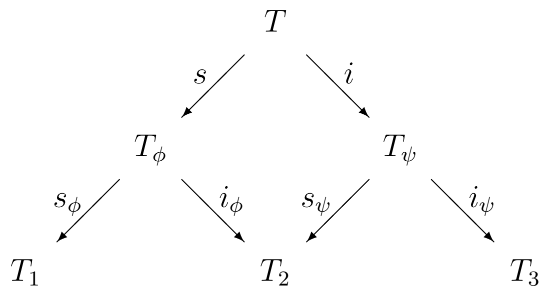
$$T \xleftarrow{\ s\ } T_0 \xrightarrow{\ i\ } T'$$

where T_0 is the reduced triple of $(\mathcal{S}_0, \mathcal{O}', \langle \cdot, \cdot \rangle')$ and as maps of sets $s : \mathcal{S}_0 \to \mathcal{S}$ is a surjection and $i : \mathcal{S}_0 \to \mathcal{S}'$ is an injection. Furthermore $i^*\mathcal{O}' \supset s^*\mathcal{O}$. Reciprocally given such a diagram with s and i satisfying the stated properties we see that $\mathcal{S}_0$ can be considered a subset of $\mathcal{S}'$ that corresponds to a set theoretic epimorphic image of $\mathcal{S}$ and since one readily checks that $s* : \mathcal{O} \to \mathcal{O}_0$ is a monomorphism, $\mathcal{O}$ is embedded in $\mathcal{O}_0$, but $s^*\mathcal{O} \subset i^*\mathcal{O}'$ hence $\mathcal{O}$ can be executed in T'.

Let us now define a new category *Sta* whose object class is that of *Trip* but whose morphisms $\phi : T \to T'$ are equivalence classes of ordered triples (s, T_0, i) where the *Trip* morphism $s : T_0 \to T$ is a set epimorphism, the *Trip* morphism $i : T_0 \to T'$ is a set injection, and $s^*\mathcal{O} \subset i^*\mathcal{O}'$. The equivalence relation $(s_1, T_1, i_1) \sim (s_2, T_2, i_2)$ holds when there is a *Trip* isomorphism $\lambda : T_1 \to T_2$ such that the diagram below commutes:

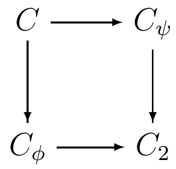


We must show that an associative composition law can be defined for these morphisms. Let therefore $\phi : T_1 \to T_2$ and $\psi : T_2 \to T_3$ be two such morphisms represented by (s_ϕ, T_ϕ, i_ϕ) and (s_ψ, T_ψ, i_ψ) respectively. We then have the diagram:

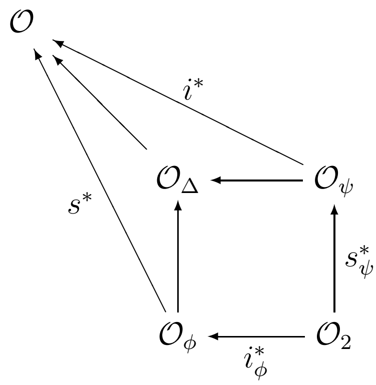


where the square is defined to be a pull-back. We now show that $(s_\phi s, T, i_\psi i)$ represents a morphism from T_1 to T_3 that we call $\psi\phi$.

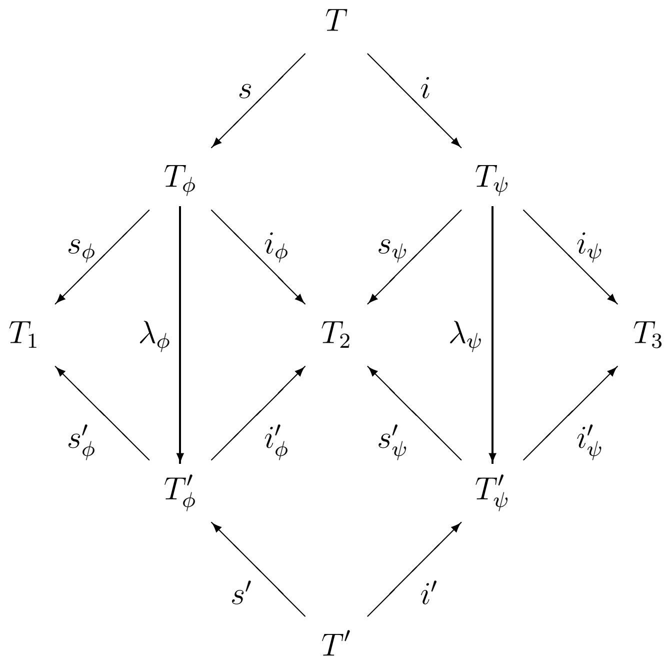
We must have in *Conv* a pull-back



and since *Conv* limits and *Set* limits coincide in the underlying sets and morphisms, we can take $C = \{(c_\phi, c_\psi) \in C_\phi \times C_\psi \mid i_\phi c_\phi = s_\psi c_\psi\}$ with s and i as the restrictions of the corresponding canonical projections. It's immediate that s is a surjection and i an injection. But then $s_\phi s$ is a surjection and $i_\psi i$ an injection. We also have in *Conv* the diagram

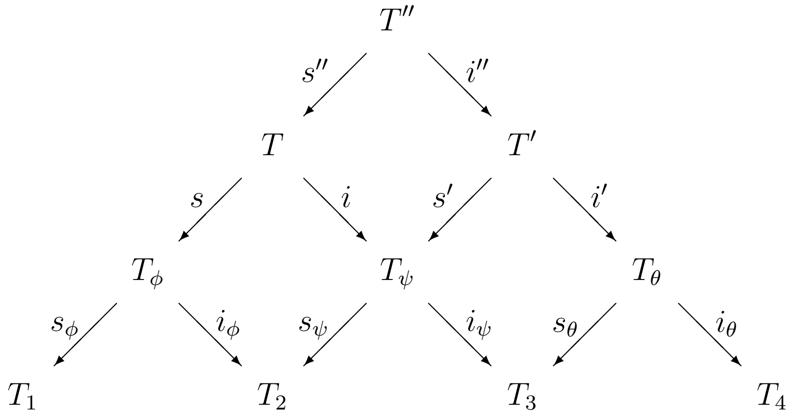


where the square is a push-out. Now $s_\phi^*\mathcal{O}_1 \subset i_\phi^*\mathcal{O}_2$ hence $s^*s_\phi^*\mathcal{O}_1 \subset s^*i_\phi^*\mathcal{O}_2 \Rightarrow s^*s\phi^*\mathcal{O}_1 \subset i^*s_\psi^*\mathcal{O}_2$. On the other hand $s_\psi^*\mathcal{O}_2 \subset i_\psi^*\mathcal{O}_3$ and thus $s^*s_\phi^*\mathcal{O}_1 \subset i^*i_\psi^*\mathcal{O}_3$ and so $(s_\phi s, T, i_\psi i)$ is indeed a morphism representative. We must now show that the morphism is well defined; that is, if we choose other representatives $(s'_\phi, T'_\phi, i'_\phi)$ and $(s'_\psi, T'_\psi, i'_\psi)$ of ϕ and ψ respectively, with isomorphisms $\lambda_\phi : T_\phi \to T'_\phi$, $\lambda_\psi : T_\psi \to T'_\psi$ then the resulting composition lies within the same equivalence class as the previous one. We have the diagram:

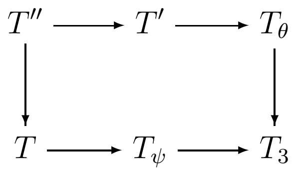


Since T' is a pullback, we have a unique map $\lambda : T \to T'$ such that $s' \circ \lambda = \lambda_\phi \circ s$ and $i' \circ \lambda = \lambda_\psi \circ i$. Now using the same argument seeing that T is also a pullback we get a map in the opposite direction which by universality must be inverse to λ. Thus λ is an isomorphism and by commutativity of the diagram we must have $s'_\phi \circ s' \circ \lambda = s_\phi \circ s$, $i'_\psi \circ i' \circ \lambda = i_\psi \circ i$ and so the composition is well defined.

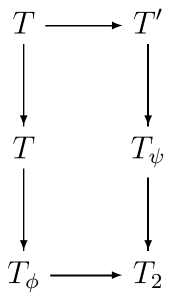
To prove associativity let $\phi : T_1 \to T_2$, $\psi : T_2 \to T_3$ and $\theta : T_3 \to T_4$ be represented by (s_ϕ, T_ϕ, i_ϕ), (s_ψ, T_ψ, i_ψ) and $(s_\theta, T_\theta, i_\theta)$ respectively. Consider the diagram



where each square is a pull-back. By general category theory, the two rectangles



and



are both pullbacks. Hence we see that $(s_\phi s s'', T'', i_\theta i' i'')$ represents both $\theta(\psi\phi)$ and $(\theta\psi)\phi$ and the associative law is proved.

A morphism $\phi : T \to T'$ in *Sta* represents a way of subsuming the knowledge systematized in T within T'; hence T' must be a stronger theory. Note now that an idealization is given by a *Sta* morphism, for if $i : \mathcal{S} \subset \mathcal{S}'$, $j : \mathcal{O} \subset \mathcal{O}'$ we can take for T_0 the reduced triple of $(\mathcal{S}, \mathcal{O}', \langle\cdot,\cdot\rangle')$ and so define the morphism $(1_{\mathcal{S}}, T_0, j) : T \to T'$. Our scientific life would be greatly simplified if *Sta* were a cocomplete category, for this would mean we could always by universal systematic procedures combine fragmented knowledge into a coherent whole. However this is not the case and unfortunately not too much can be said about this category. One can view now however certain existing formalisms in this light. Consider for instance local quantum theory in the C^* algebra formulation. Here to each bounded open region G of space time we have a C^* algebra $A(G)$ which can be assumed to be a subalgebra of the algebra of all bounded operators for some fixed Hilbert space. If $G \subset G'$ we have an inclusion $A(G) \subset A(G')$. The global theory has the C^* algebra A given by the norm closure of $\cup A(G)$. Let $\mathcal{S}$ be the set of states of A and let $\mathcal{O}$ be the elements of A such that $0 \leq a \leq 1$. To each G we associate the local statistical triple $T(G)$ which is the reduced triple of $(\mathcal{S}, \mathcal{O} \cap A(G), \langle \cdot, \cdot \rangle)$. Each restriction morphism of $T = (\mathcal{S}, \mathcal{O}, \langle \cdot, \cdot \rangle)$ to $T(G)$ therefore corresponds to a *Sta* morphism $T(G) \to T$ and T is an idealization of the *Trip* limit of the $T(G)$ and this *Trip* limit coincides with a colimit in that subcategory of *Sta* where i is an isomorphism. The notion of localizability assumes that the global theory is loosely speaking a colimit of its localizations, implying thus a categorical idea in its very conception.

Chapter 7

On the Complexity of States

Consider a state a of a general statistical theory. Intuitively we expect to be able to define a measure of the unpredictability of the state, a measure of the complexity of its repertoire of behavior. Thus if we view a state with an exhaustive instrument $I = (p_1, p_2, \ldots, p_n)$ the classical entropy

$$h(\sigma, I) = -\sum \langle \sigma, p_i \rangle \log \langle \sigma, p_i \rangle$$

is a measure of the unpredictability of observing the state by the instrument I. There are many difficulties however in attributing this entropy wholly to the state. Consider the instrument $((1/2)\mathbf{1}, (1/2)\mathbf{1})$; every state has entropy 1 bit. One could have obtained the same observational result by observing $\mathbf{1}$ and then flipping a coin considering heads and tails as our observational outcomes. Thus the entropy we are seeing here can be wholly interpreted as being due to a stochastic mechanism in the observation process itself rather than in the complexity of the state observed. Suppose more generally that we are observing with an instrument $J = (q_1, q_2, \ldots, q_m)$ and then follow this observation with a stochastic process with n possible outcomes where the process is defined by giving transition probabilities $\mathbf{p}(i, j)$; $i = 1, \ldots, n$; $j = 1, \ldots, m$ that the i-th outcome of the process will be realized if the j-th property of J was observed. The matrix $\mathbf{p}$ is *stochastic* in that $\sum_{i=1}^{n} \mathbf{p}(i, j) = 1$ for all j. The probability that the i-th outcome of the process be realized in a state σ is $\sum_{j=1}^{n} \mathbf{p}(i, j) \langle \sigma, q_j \rangle$.

Lemma 7.1 *Given a stochastic matrix and an instrument J as above, if we set $I = (p_1, p_2, \ldots, p_n)$ where $p_i = \sum_{j=1}^{n} \mathbf{p}(i, j) q_j$ then I is an instrument.*

Proof: Represent each probability measure $\mathbf{p}(\cdot, j)$ by the partition

$$0, \mathbf{p}(1,j), \mathbf{p}(1,j) + \mathbf{p}(2,j), \ldots, \mathbf{p}(1,j) + \cdots + \mathbf{p}(n-1,j), 1$$

of $[0,1]$. The union over j of all of all of these partitions results in a partition $s_0, s_1, \ldots, s_k$ with $s_0 = 0 < s_1 < s_2 < \cdots < s_{k-1} < 1 = s_k$ Let $\lambda_r = s_r - s_{r-1}$; $r = 1, \ldots, k$; then $\lambda_r \geq 0$ and $\sum \lambda_r = 1$. We can express each $\mathbf{p}(i,j)$ as a sum of the λ_r: $\mathbf{p}(i,j) = \sum\{\lambda_r \,|, r \in L_{ij}\}$ where $L_{ij} \subset \{1, \ldots, k\}$ and for each j the L_{ij}; $i = 1, \ldots, n$ form a partition of $\{1, \ldots, k\}$. Consider the instrument $K = \theta_{(\lambda_1, \lambda_2, \ldots, \lambda_k)}(J, \ldots, J)$. To obtain p_i we include in a condensation the j-th atom in the r-th copy of J if and only if $r \in L_{ij}$. Since the L_{ij} form a partition as described above, no atom gets included in more than one condensation and so I is in fact a condensation of K and so by axiorm $4.6'$ is an instrument. Q.E.D

Lemma 7.2 *Given a stochastic matrix and an instrument as in the previous lemma, let* $\pi_i = \{Q_{i,1}, \ldots, Q_{i,\nu(i)}\}$ *be an arbitrary partition of* $\{1, \ldots, m\}$ *for each* i. *Define* $p_{ik} = \sum\{\mathbf{p}(i,j)q_j \,|\, j \in Q_{ik}\}$, *then the* p_{ik}; $i = 1, \ldots, n$; $k = 1 \ldots \nu(i)$ *are atoms of an instrument.*

Proof: We define a new stochastic matrix $\mathbf{t}$ by

$$\mathbf{t}(i,k;j) = \begin{cases} \mathbf{p}(i,j) & \text{if } j \in Q_{ik} \\ 0 & \text{otherwise} \end{cases}$$

where the rows are labelled by the double index (i,k), $i = 1, \ldots, n$; $k = 1 \ldots \nu(i)$. Now apply Lemma 7.1. Q.E.D

The instrument obtained by this last lemma we call a *stochastic condensation* of J by means of $\mathbf{p}$ and the π_i or simply a stochastic condensation. Ordinary condensations that ignore no atoms are particular cases. Stochastic condensations given by Lemma 7.1 we call *full*; they correspond to $\pi_i = \{\{1, \ldots, m\}\}$ for all i.

We shall at a certain point also consider similar constructions, still called stochastic condensations, but with *substochastic* matrices, that is matrices $\mathbf{s}(i,j)$ with $\mathbf{s}(i,j) > 0$, and $\sum_{i=1}^{n} \mathbf{s}(i,j) \leq 1$.

Lemma 7.3 *Without altering the conclusions, one can replace in Lemmas 7.1 and 7.2 the stochastic matrix* $\mathbf{p}$ *with a substochastic matrix* $\mathbf{s}$.

Proof: We only show that the conclusion of Lemma 7.1 is unchanged, since the proof of the altered Lemma 7.2 follows almost verbatim the proof of the original. Introduce now a new row in the matrix $\mathbf{s}$, labelled by 0 and with elements

$$\tilde{\mathbf{s}}(0,j) = 1 - \sum_{i=1}^{n} \mathbf{s}(i,j)$$

where $\tilde{\mathbf{s}}$ is the matrix $\mathbf{s}$ with this added row. We see that $\tilde{\mathbf{s}}$ is stochastic, and the instrument obtained by a full stochastic condensation of J by means of $\tilde{\mathbf{s}}$ is

$$\tilde{I} = (\mathbf{1} - \sum_{i=1}^{n} p_i, p_1, p_2, \ldots, p_n)$$

where $p_i = \sum_{j=1}^{m} \mathbf{s}(i,j) q_j$. Condensing $\tilde{I}$ by eliminating the first atom we obtain an instrument $I = (p_1, \ldots, p_n)$. Q.E.D

Lemma 7.4 *Let* $J = (q_1, \ldots, q_m)$ *be an instrument and* $\mathbf{s}$ *a substochastic matrix. If* $I = (p_1, \ldots, p_n)$ *where* $p_i = \sum_{j=1}^{m} \mathbf{s}(i,j) q_j$ *is an exhaustive instrument then for all* j *either* $q_j = 0$ *or* $\sum_{i=1}^{n} \mathbf{s}(i,j) = 1$

Proof: Since $\sum p_i = 1$ we have $\sum_{i,j} \mathbf{s}(i,j) q_j = \mathbf{1}$ Let $\omega_j = \sum_{i=1}^{n} \mathbf{s}(i,j) \leq 1$. Thus $\sum \omega_j q_j = \mathbf{1}$. Since J is an instrument this is only possible if $\sum q_j = \mathbf{1}$ and $q_j = \mathbf{0}$ whenever $\omega_j \neq 1$. In fact if $\sum q_j \neq \mathbf{1}$, there is a state σ such that $1 > \langle \sigma, \sum q_j \rangle = \sum \langle \sigma, q_j \rangle$ but then as $\omega_j \leq 1$ we have $1 > \sum \omega_j \langle \sigma, q_j \rangle = \langle \sigma, \sum \omega_j q_j \rangle$ which contradicts $\sum \omega_j q_j = \mathbf{1}$. If now $q_a \neq \mathbf{0}$ and $\omega_a \neq 1$ we have a state σ such that $\langle \sigma, q_a \rangle \neq 0$ and we have $1 = \langle \sigma, \sum q_j \rangle \rangle \sum_{j \neq a} \langle \sigma, q_a \rangle + \omega_a \langle \sigma, q_a \rangle \geq \sum \omega_j \langle \sigma, q_j \rangle = \langle \sigma, \sum \omega_j q_j \rangle = 1$, an absurdity. Q.E.D

This last lemma shows that whenever we obtain an exhaustive instrument I by means of a fill stochastic condensation with a substochastic matrix $\mathbf{s}$ from an instrument J, then eliminating the zero atoms of J and the corresponding columns of $\mathbf{s}$ we obtain a stochastic matrix and I as a full stochastic condensation by means of this stochastic matrix from an instrument with nonzero atoms.

For a stochastic matrix $\mathbf{p}(i,j)$, each $\mathbf{p}(\cdot,j)$ is a probability measure on a finite set and we can consider its classical entropy $h(\mathbf{p}(\cdot,j))$ and we set $h(\mathbf{p}) = \sup_j h(\mathbf{p}(\cdot,j))$. If now $J = (q_1, \ldots, q_m)$ is an instrument with nonzero atoms, and I is the stochastic condensation of J by means of $\mathbf{p}$ with $h(\mathbf{p}) > 0$, then we can attribute the entropy $h(\sigma, I)$ of a state o as being partially due to the stochastic process described by $\mathbf{p}$ and thus not completely intrinsic to the state. We say that an exhaustive instrument is *stochastically factorizable* if it can be written as a full stochastic condensation of an exhaustive instrument J with nonzero atoms by means of a stochastic matrix with positive entropy. We must eliminate such instruments from any computation of entropies of states.

This however doesn't finish the story. Consider an instrument $I = \theta_{(\lambda,1-\lambda)}(I_1, I_2)$ with I_i exhaustive. This is a stochastic splitter followed by instruments at each exit. The entropy of this arrangement in the state a is easily computed to be

$$h(\sigma, I) = \lambda h(\sigma, I_1) + (1-\lambda)h(\sigma, I_2) - \lambda \log \lambda - (1-\lambda)\log(1-\lambda).$$

Here the quantity $-\lambda \log \lambda - (1-\lambda)\log(1-\lambda)$ can be interpreted as the entropy due to the action of the stochastic splitter and must not be counted as belonging to the state, but attributed to the intervention of the experimenter. We now generalize this. Given an operation $\Psi = (\psi_1, \ldots, \psi_m)$ and a stochastic matrix $\mathbf{p}$ we can consider the n-tuple $\Theta = (\theta_1, \ldots, \theta_n)$ with $\theta_i = \sum \mathbf{p}(i,j)\psi_j$. By an argument exactly parallel to the one used in Lemma 7.1 θ is a condensation of an operation of the form $\theta_{(\lambda_1,\ldots,\lambda_k)}\{\Psi, \ldots, \Psi\}$ and is thus an operation. It is also easy to see that θ is a condensation of $\Psi\{\theta_{\mathbf{p}(\cdot,1)}, \ldots, \theta_{\mathbf{p}(\cdot,m)}\}$ giving an independent proof that it is an operation. As before we say θ is a *full stochastic condensation* of Ψ. We call an operation *nondestructive* if $\sum \theta_i \mathbf{1} = \mathbf{1}$ (no copy is rejected) and we say that a nondestructive operation is *stochastically factorizable* if it is a full stochastic condensation of a nondestructive operation Ψ with no dummy exits by means of a stochastic matrix $\mathbf{p}$ with $h(p) > 0$. Now if an exhaustive instrument I is stochastically factorizable being a stochastic condensation of J, then we can interpret the action of I as being that of observing J *followed* by a stochastic process applied to the result and independent of the state being observed. On the other hand if I is of the form $\theta(J_1, \ldots, J_m)$ with θ stochastically factorizable being a stochastic condensation of Ψ its action can be interpreted as operating with Ψ, applying a stochastic process at the exits by which states get reclassified into new categories independently of the actual copies that may appear and *prior* to observing with the array $(J_1, \ldots, J_m)$ in a way conditioned to the outcomes of the process. Again, this introduces entropy due solely to the experimenter and this type of instrument must also be excluded from entropy calculations.

Under a more careful analysis we may say that an instrument I should not be allowed in entropy calculations if it can be interpreted as a construction using the means supplied to us by the axioms of Chapter 4 and which contains stochastic elements interpretable as being due to the experimenter. There is a difficulty in this, for consider the instrument obtained by condensing corresponding atoms of the two copies of I in $\theta_{(1/2,1/2)}(I, I)$. This gives us I again but constructed with the obvious stochastic element $\theta_{(1/2,1/2)}$. What happened is that the condensation eliminated the influence of this stochastic element. Essentially $\theta_{(1/2,1/2)}$ is not there and in fact we can change $(1/2, 1/2)$ to any other two point probability $(\lambda, 1 - \lambda)$ without changing the result.

Consider therefore an instrument $I \in \mathcal{I}_n$ that can be written as a construction in terms of operations and other instruments using the means given us by the axioms of Chapter 4. Assume that in this construction there is an occurrence of a stochastic splitter θ_{Λ_0} with $h(\Lambda_0) > 0$. Let now I_Λ be the instrument resulting from replacing that occurrence of θ_{Λ_0} by θ_Λ where Λ is a measure absolutely continuous with respect to Λ_0 (If it were not, new operative exits will be created for which no provision has been taken). We say that the given occurrence of the splitter is *inessential* if $I_\Lambda = I_{\Lambda_0} = I$. Now the map $\Lambda \mapsto I_\Lambda$ is affine and so in the case of essentiality this map is not constant on $\{\Lambda \,|\, \Lambda \ll \Lambda_0\}$, and since Λ_0 is an interior point of this set, I cannot be an extreme point of $\mathcal{I}_n$. We are thus led to consider that pure exhaustive instruments are the ones that cannot be interpreted as having state independent experimenter determined positive entropy stochastic processes in their construction. That is, the entropy they see would be intrinsic to the state. Reciprocally, suppose $I = (1/2)I_1 + (1/2)I_2$, $I_1 \neq I_2$; then we can view I as a condensation of $\theta_{(1/2,1/2)}(I_1, I_2)$ and the essentiality of θ is interpreted as saying that the condensations did not completely eliminate the effects of the splitter. We now show that this point of view is consistent with our previous decision to eliminate instruments interpretable in terms of stochastic factorizability.

Lemma 7.5 *An exhaustive instrument I is stochastically factorizable if and only if there is an observation $q \neq 0$ and two atoms of I which we call p_1 and p_2 after reordering, such that $p_1 - (1/2)q, q, p_2 - (1/2)q, p_3, \ldots, p_n$ is also an instrument.*

Proof: Suppose such q, p_1, p_2 exist, then we see that

$$\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} p_1 - (1/2)q \\ q \\ p_2 - (1/2)q \end{pmatrix}$$

and this can be extended to a stochastic factorization of I. Suppose conversely that I is a full stochastic condensation of an exhaustive instrument $J = (q_1, \ldots, q_m)$, $q_j \neq 0$; $p_i = \sum_j \mathbf{p}(i,j)q_j$; $h(\mathbf{p}) > 0$. Then there are indices a, b, c such that $0 < \mathbf{p}(a,c) < 1$ and $0 < \mathbf{p}(b,c) < 1$. By renumbering we set $a = 1$, $b = 2$. Calling the two numbers above by ν_1 and ν_2 respectively we can perform a stochastic condensation of J as in Lemma 7.2 to get an instrument $p_1 - \nu_1 q_c, \nu_1 q_c, \nu_2 q_c, p_2 - \nu_2 q_c, p_3, \ldots, p_n$. Assume $\nu_1 \leq \nu_2$ so we can write $\nu_1 = \beta\nu_2$, $0 < \beta \leq 1$. We then have

$$\begin{pmatrix} \nu_1 q_c \\ p_2 - \nu_1 q_c \end{pmatrix} = \begin{pmatrix} \beta & 0 \\ 1-\beta & 1 \end{pmatrix} \begin{pmatrix} \nu_2 q_c \\ p_2 - \nu_2 q_c \end{pmatrix}$$

and so by a further stochastic condensation we obtain an instrument $p_1 - \nu_1 q_c, \nu_1 q_c, \nu_1 q_c, p_2 - \nu_1 q_c, p_3, \ldots, p_n$ and setting $q = 2\nu_1 q_c$ a final condensation results in $p_1 - (1/2)q, q, p_2 - (1/2)q, p_3, \ldots, p_n$. Q.E.D

Corollary 7.1 *If I is stochastically factorizable then I is mixed.*

Proof: By Lemma 7.5 there is a $q \in \mathcal{O}$, $q \neq \mathbf{0}$ and two atoms p_i, p_j of I such that $J : q, p_1, \ldots, p_i - (1/2)q, \ldots, p_j - (1/2)q, \ldots, p_n$ is an instrument. By condensing we get two distinct instruments

$$I_1 : p_1, \ldots, p_i + (1/2)q, \ldots, p_j - (1/2)q, \ldots, p_n,$$

$$I_2 : p_1, \ldots, p_i - (1/2)q, \ldots, p_j + (1/2)q, \ldots, p_n,$$

and so $I = (1/2)I_1 + (1/2)I_2$ is mixed. Q.E.D

We observe that we could have proved the corollary without recourse to Lemma 7.5 by arguing that $\mathbf{p} \mapsto$ full stochastic condensation of J, is affine and if $h(\mathbf{p}) > 0$, then $\mathbf{p}$ lies on a segment on which the map is not constant. However, Lemma 7.5 is useful for other purposes.

By an argument totally parallel to the above, we can also prove:

Lemma 7.6 *An operation $\theta = (\theta_1, \ldots, \theta_n)$ is stochastically factorizable if and only if there are two exits, say 1 and 2 after renumbering, and a one exit operations $\psi \neq 0$ such that $\theta_1 - (1/2)\psi, \psi, \theta_2 - (1/2)\psi, \theta_3, \ldots, \theta_n$ is an operation.*

Suppose now that θ is stochastically factorizable and $I = \theta(J_1, \ldots, J_n)$. By condensing the operation $\theta_1 - (1/2)\psi, \psi, \theta_2 - (1/2)\psi, \theta_3, \ldots, \theta_n$ in two ways we obtain the distinct operations

$$\Xi_1 : \theta_1 + (1/2)\psi, \psi, \theta_2 - (1/2)\psi, \theta_3, \ldots, \theta_n,$$

$$\Xi_2 : \theta_1 - (1/2)\psi, \psi, \theta_2 + (1/2)\psi, \theta_3, \ldots, \theta_n,$$

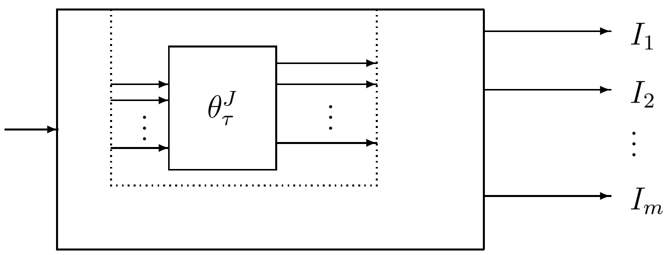
and we find that $I = (1/2)\Xi_1(J_1, \ldots, J_n) + (1/2)\Xi_2(J_1, \ldots, J_n) = (1/2)I_1 + (1/2)I_2$ Now if $I_1 = I_2$ then $(\theta_1 + (1/2)\psi)(J_1) = (\theta_2 - (1/2)\psi)(J_1)$ or, $\psi(J_1) = 0$. Supposing J_1 exhaustive, we have $\psi\mathbf{1} = \mathbf{0}$ which by axiom $4.9'$ implies $\psi = 0$ contradicting the construction. Thus we conclude the following corollary.

Corollary 7.2 *If $I = \theta(J_1, \ldots, J_n)$ with θ stochastically factorizable and J_i exhaustive, then I is mixed.*

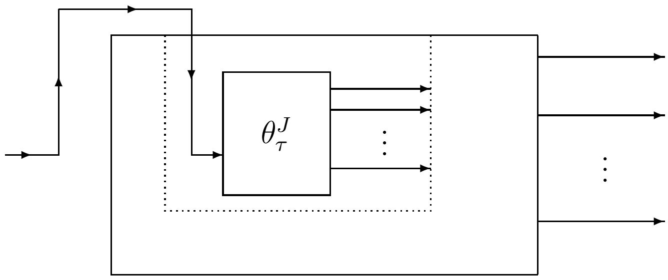
We dignify the following obvious fact to a lemma for future reference:

Lemma 7.7 *If every atom of an instrument is pure, then the instrument is pure.*

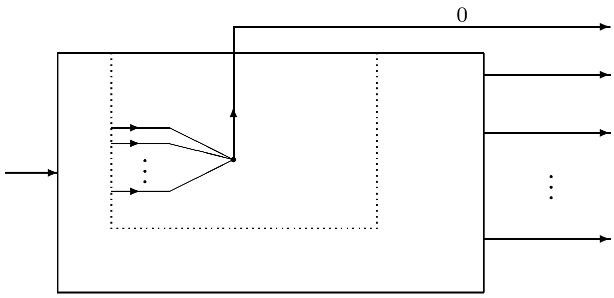
To complete our argument that interprets pure instruments as those free of ad-hoc interventions by the experimenter, we must counter a possible objection that sleight of hand operations must be taken into account. If a given instrument can be constructed by the means of Chapter 4 with the use of substitution operations one can argue that these represent interventions of the experimenter by means of which the original copy of the state is destroyed and arbitrary others are substituted so that the final observations do not entirely refer to the original state but also incorporate the structure of ad hoc states introduced by the experimenter. We now proceed to explain that in case the final instrument is pure, the presence of possible sleight of hand operations is innocuous. Consider schematically such an instrument I which we first assume can be written as $\theta(I_1, \ldots, I_m)$ where θ is itself a construct involving the sleight of hand operation θ^J_τ; $J = (q_1, \ldots, q_n)$, $\tau = (\tau_1, \ldots, \tau_n)$



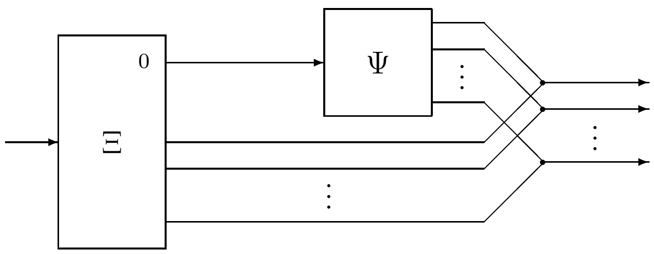
Since the entry of a state into θ^J_τ leads to its observation by J and to a subsequent preparation of one of the states $\tau_1, \ldots, \tau_n$, we can consider all subsequent events as being a stochastic data analysis performed on the outcomes of J. This analysis becomes systematically incorporated into the pattern of final observations by the array $I_1, I_2, \ldots, I_m$. Such incorporation would be innocuous if no real stochastic elements are introduced. Consider now the operation Ψ obtained by direct entry into θ^J_τ bypassing the original entrance:



and the operation Ξ obtained by excising θ^J_τ and leading every state that would normally enter θ^J_τ out to a new exit labelled by 0:



We now see that the instrument I can be given the following schematic description:



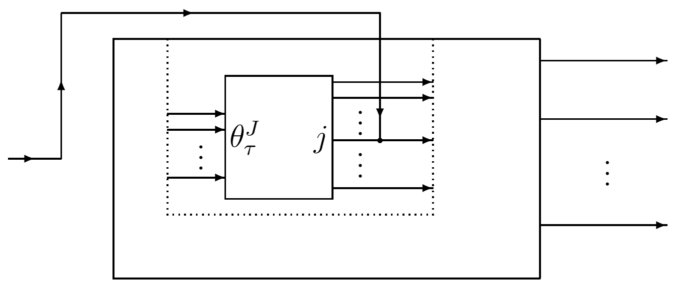
which is a condensation of the instrument

$$\Xi\{\Psi; \mathrm{Id}, \ldots, \mathrm{Id}\}(I_1, \ldots, I_m; I_1, \ldots, I_m).$$

Let us now examine the action of Ψ. Let $p_1, \ldots, p_r$ be the atoms of $(I_1, \ldots, I_m)$. Clearly $\langle \sigma, p_i \rangle$ depend affinely on $\langle \sigma, q_j \rangle$ and we can write

$$\langle \sigma, p_i \rangle = \sum_{j=1}^{n} \mathbf{s}(i,j) \langle \sigma, q_j \rangle + c_i.$$

We proceed to show that $\mathbf{s}(i,j)$ is a substochastic matrix and $c_i = 0$. Consider the operation Υ^j defined by feeding directly into the j-th exit of θ^J_τ in Ψ bypassing the original entrance:



In a state σ we have that $\Upsilon^{j(i)}_* \tau_j$ appears at exit j of Ψ exactly a fraction $\langle \sigma, q_j \rangle \langle \tau_j, \Upsilon^{j(i)} \mathbf{1} \rangle$ times, and so $\langle \sigma, \Psi^{(i)} p \rangle = \sum \langle \sigma, q_j \rangle \langle \tau_j, \Upsilon^{j(i)} \mathbf{1} \rangle \langle \Upsilon^{j(i)}_* \tau_j, p \rangle = \sum \langle \tau_j \Upsilon^{j(i)} p \rangle \langle \sigma, q_j \rangle$ that is,

$$\Psi^{(i)} p = \sum_j \langle \tau_j, \Upsilon^{j(i)} p \rangle q_j.$$

If now p_i corresponds to atom r_{ab} of I_a we have

$$p_i = \sum_j \langle \tau_j, \Upsilon^{j(i)} r_{ab} \rangle q_j$$

and so $c_i = 0$ and

$$\mathbf{s}(i,j) = \langle \tau_j, \Upsilon^{j(i)} r_{ab} \rangle \geq 0.$$

Let $r_a = \sum_b r_{ab}$ we then have

$$\sum_i \mathbf{s}(i,j) = \sum_i \langle \tau_j, \Upsilon^{j(i)} r_{ab} \rangle = \sum_i \langle \tau_j, \Upsilon^{j(i)} r_i \rangle.$$

Now $\Upsilon^{j(1)}r_1, \ldots, \Upsilon^{j(m)}r_m$ is an instrument and so $r^j = \sum_i \Upsilon^{j(i)}r_i \in \mathcal{O}$ and we have $\sum_i \langle \tau_j, \Upsilon^{j(i)}r_i \rangle = \langle \tau_j, r^j \rangle \leq 1$, and $\mathbf{s}$ is substochastic. We now see that I is a condensation of $(\Xi^{(0)}\Psi(I_1, \ldots, I_m), \Xi^{(1)}I_1, \ldots, \Xi^{(m)}I_m)$ and in which $\Xi^{(0)}\Psi(I_1, \ldots, I_m)$ is a stochastic condensation of $\Xi^{(0)}J$ by means of a substochastic matrix. Thus I is a stochastic condensation of $(\Xi^{(0)}J, \Xi^{(1)}I_1, \ldots$ $\ldots, \Xi^{(m)}I_m)$ by means of a substochastic matrix $\tilde{\mathbf{s}}(i,j)$. If now we assume that I is an exhaustive extreme instrument then by Lemma 7.4 and Corollary 7.2 we conclude that $\sum_i \tilde{\mathbf{s}}(i,j) = 1$ unless $\Xi^{(0)}q_j = 0$ and eliminating these atoms of $\Xi^{(0)}J$ we obtain I as a stochastic condensation by means of a stochastic matrix, which by extremity of I must be trivial entropically. Thus the presence of θ^J_τ in the construction of I is innocuous, seeing that the data processing that is effected by subsequent passages of the state is non stochastic in character and does not contribute to the complexity of the final observations. To complete the analysis we remark that any construction of an instrument by the axioms of Chapter 4 is a (possible) exhaustion of a (possible) condensation of an instrument of the form $\theta(I_1, \ldots, I_m)$. Since exhaustions and condensations cannot introduce any new stochastic elements into any previous data analysis, performing the above analysis on $\theta(I_1, \ldots, I_m)$ we reach the same conclusion.

We thus assume the following basic thesis for the rest of this work: *pure instruments are those that cannot be interpreted as incorporating any essential ad hoc interferences of the experimenter, and the complexity they exhibit when applied to a state can be attributed entirely to the state.*

In order to be assured of sufficiently many pure instruments one must generally place the theory within an appropriate idealization. We shall always suppose, according to Chapter 5, that all of the defining convex sets of a statistical theory are algebraically closed. For finite dimensional state figures this is sufficient to assure enough pure instruments that they determine all others.

We can now consider the entropy of a state a as being the supremum of $h(\sigma, I)$ over pure exhaustive instruments, since this roughly speaking is the maximum complexity exhibited by a state that can be intrinsically attributed to it. We shall come to realize that there is still more involved in the notion of entropy, but this quantity is certainly an important one and we call it the *instrumental entropy* of the state and denote it by $H(\sigma)$:

$$H(\sigma) = \sup\{h(\sigma, I) \mid I \text{ pure and exhaustive}\}.$$

Before proceeding with our investigation we need an elementary technical result.

Given two probability measures s and s' on two finite sets X and X' respectively, we say s is a *condensation* of s' (and s' is a *refinement* of s) if there is a map $\phi : X \to X'$ such that $s'(\phi^{-1}(A)) = s(A)$ for all $A \subset X$. The classical entropy of s is of course given by $h(s) = -\sum_X s(x) \log s(x)$.

Lemma 7.8 *Let s' be a refinement of s, then $h(s') \geq h(s)$ and equality holds if and only if s' is also a condensation of s.*

Proof: By induction we need only show that if $0 \leq a$, $0 \leq b$, $0 \leq a+b \leq 1$ then $-a \log a - b \log b \geq -(a+b) \log(a+b)$ with equality holding if and only if at least one of the numbers is zero. The case $a + b = 0$ is trivial and so consider $a + b \neq 0$and the two point probability measure $s_1 = a/(a + b)$, $s_2 = b/(a + b)$. Now $h(s) \geq 0$ with equality holding if and only if one of the s_i is zero. Thus we have

$$-\frac{a}{a+b} \log \frac{a}{a+b} - \frac{b}{a+b} \log \frac{b}{a+b} \geq 0$$

or equivalently, multiplying by a+b and expanding

$$-a \log a - b \log b + (a+b) \log(a+b) \geq 0$$

with equality holding if and only if one of the numbers is zero. This is what we need. Q.E.D

We see in particular that $h(\sigma, I)$ increases with refinements of I, a useful fact for computation.

We proceed to discuss the instrumental entropy function in each of the following theories: (1) Kolmogorov probability (2) Boolean triples (3) Two dimensional triples, and (4) Quantum mechanics.

1. Consider a Kolmogorov triple based on a measurable space (Ω, Σ). We associate with the triple the canonical operational statistical theory.

Theorem 7.1 *In a Kolmogorov triple an exhaustive instrument is stochastically non factorizable if and only if it is pure, which happens if and only if each atom is a characteristic function.*

Proof: Let $I = (f_1, \ldots, f_n)$ and suppose not all f_i are extreme. Renumbering the functions if necessary we can assume $0 < f_1(x_0) < 1$, and $0 < f_2(x_0) < 1$ for some point x_0. Let α be such that $0 < \alpha < f_i(x_0)$, $i = 1, 2$ and set $g = \alpha\chi_{x_0}$. Then $f_1 - (1/2)g, g, f_2 - (1/2)g, f_3, \ldots, f_n$ is an instrument and so by Lemma 7.5 is stochastically factorizable and by Corollary 7.1 is mixed. Conversely, if each f_i is pure then by Lemma 7.7, I is pure. Q.E.D

Theorem 7.2 *Consider a Kolmogorov triple. Let σ be a state that is a sum of Dirac measures: $\sigma = \sum_{i=1}^{\infty} \nu_i \delta_{x_i}$, x_i distinct. Then $H(\sigma) = -\sum \nu_i \log \nu_i$.*

Proof: If $I = (\chi_{A_1}, \ldots, \chi_{A_n})$ is a pure instrument corresponding to the partition $\Omega = \cup A_i$ then $h(\sigma, I) = -\sum \sigma(A_i) \log \sigma(A_i)$ is equal to $-\sum_i (\sum_j \{\nu_j \,|\, x_j \in A_i\}) \log(\sum_k \{\nu_k \,|\, x_k \in A_i\})$. Now any partition can be refined to contain the first N singletons $\{x_i\}$, $i = 1, 2, \ldots, N$. Since $h(\sigma, I)$ increases with refinements we have $H(\sigma) = -\sum \nu_i \log \nu_i$ which is the classical formula. Q.E.D

2. Consider a Boolean triple. Associate with it the canonical operational statistical theory.

Theorem 7.3 *An exhaustive instrument in a Boolean triple is stochastically nonfactorizable if and only if it is pure, and this occurs if and only if each atom is a characteristic function.*

Proof: Let$I = (f_1, \ldots, f_n)$ and suppose as in the proof of Theorem 7.1 that $0 < \alpha < f_i(x_0)$, $i = 1, 2$. The same inequalities can now be affirmed in a closed-open neighborhood U of x_0. Let $g = \alpha\chi_U$, then $f_1 - (1/2)g, g, f_2 - (1/2)g, f_3, \ldots, f_n$ is an instrument and by Lemma 7.5 and Corollary 7.1 is mixed. Conversely if each f_i is pure, by Lemma 7.5, I is pure. Q.E.D

Theorem 7.4 *Consider a Boolean triple. Let σ be a state that is a sum of Dirac measures: $\sigma = \sum_{i=1}^{\infty} \nu_i \delta_{x_i}$, x_i distinct. Then $H(\sigma) = -\sum \nu_i \log \nu_i$.*

Proof: Let the pure instrument I correspond to a partition $X = \cup A_i$ of the Stone's space of the Boolean algebra, into closed open sets. Now any such partition can be refined to a partition by closed open sets such that the first N points x_i, $i = 1, \ldots, N$ lie in different elements of the partition. The conclusion follows as in Theorem 7.2. Q.E.D

3. Let $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ be a two dimensional triple. Associate with it the canonical operational statistical theory,.

For a non zero pair of real numbers $q = (q_1, q_2)$ we define the slope $m(q) = q_2/q_1$.

Lemma 7.9 *Let* $\mathbf{1} = \sum_{i=1}^{n} p_i$ *be a decomposition of* $\mathbf{1} \in \mathcal{O}$ *by non zero elements* p_i *of* $V_0(\mathcal{O})$ *arranged in the order* $m(p_1) \geq m(p_2) \geq \cdots \geq m(p_n)$. *This decomposition defines an instrument if and only if all of the following partial sums belong to* $\mathcal{O}$)*:* $p_1, p_1 + p_2, p_1 + p_2 + p_3, p_1 + p_2 + \cdots + p_{n-1}$.

Before giving the proof we remark that we can identify the above partial sums with the vertices of a convex *polygonal line* obtained by placing the p_i tail to head in the given order.

Proof: Consider the partial sums in increasing slopes: $p_n, p_n + p_{n-1}$, $p_n + p_{n-1} + p_{n-2}, \ldots, p_n + \cdots + p_2$. Each element in these sums is the negation of an element in the original sums. Thus if the first ones lie in $\mathcal{O}$ so do the second ones. By the slope condition, any other partial sum has to lie within the convex figure bounded by the two polygonal lines, hence if the upper line lies in $\mathcal{O}$ so do all the partial sums, and the decomposition is an instrument. Conversely, suppose some partial sum is outside $\mathcal{O}$ then since it is within the convex polygon formed by the upper and lower lines, some point of the boundary of this polygon must lie outside $\mathcal{O}$ and since the upper line is paired to the lower by negation, some point of the upper line lies outside $\mathcal{O}$. By convexity of $\mathcal{O}$ some vertex of the upper line now lies outside $\mathcal{O}$. Q.E.D

We see from this lemma that if we take a convex polygonal line joining $\mathbf{0}$ to $\mathbf{1}$ and lying within $\mathcal{O}$, above or on the segment $[\mathbf{0}, \mathbf{1}]$, then by taking for p_i the sides of this line we obtain an instrument and conversely. Instruments are thus in one to one correspondence with convex polygonal lines joining $\mathbf{0}$ to $\mathbf{1}$, lying within $\mathcal{O}$ and above or on the segment $[\mathbf{0}, \mathbf{1}]$.

Lemma 7.10 *Suppose* $I = (p_1, \ldots, p_n)$*;* $p_i \neq \mathbf{0}$ *is an instrument in a general statistical theory. If some two atoms are proportional, then* I *is stochastically factorizable and hence mixed.*

Proof: Suppose, by reordering if necessary, that $p_1 = \alpha p_2$, $0 < \alpha \leq 1$. Let $J = (p_1 + p_2, p_3, \ldots, p_n)$ be a condensation.

$$\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\alpha}{1+\alpha} \\ \dfrac{1}{1+\alpha} \end{pmatrix} (p_1 + p_2)$$

showing that I is stochastically factorizable and so by Corollary 7.1 is mixed. Q.E.D

Theorem 7.5 *An exhaustive instrumentI* $= (p_1, \ldots, p_n)$ *with* $p_i \neq 0$ *in a two dimensional triple is stochastically non factorizable if and only if it is pure and this occurs if and only if every node of the upper polygonal line of* I *is an extreme point of* $\mathcal{O}$.

Proof: By Lemma 7.10 we can suppose all slopes are distinct. Let $m(p_1) > m(p_2) > \cdots > m(p_n)$ and suppose the first mixed partial sum is $p_1 + p_2 + + \cdots + p_a$. Consider the parallelogram formed by the vertices $p_1 + \cdots + p_{a-1}$, $p_1 + \cdots + p_a$, $p_1 + \cdots + p_{a-1} + p_{a+1}$, $p_1 + \cdots + p_{a+1}$. Call these points r, s, t, u respectively. We argue that there is a line segment lying within $\mathcal{O}$, having s in its interior, and of slope m satisfying $m(p_a) \geq m \geq m(p_{a+1})$. Since s is mixed, there is at least a line segment lying within $\mathcal{O}$ and containing s in its interior. If its slope already satisfies the condition we look no further, if not, then by nature of its slope, the segment is partly in the interior of the parallelogram. Assume the segment is of the form $[x, y]$ with $[x, s)$ outside, and $(s, y]$ within the parallelogram. But now the triangle r, x, u has s in its interior and so contains a segment with the desired property.

Now there is a vector q in $\mathcal{O}$ of slope m since $m(p_a) \geq m \geq m(p_{a+1})$. Furthermore we can choose q small enough that if placed at its midpoint at the vector $s = p_1 + \cdots + p_a$ it would lie wholly in $\mathcal{O}$. Making q yet smaller if need be, we can have $m(p_{a-1}) > m(p_a - (1/2)q) > m(p_a) \geq m \geq m(p_{a+1}) > m(p_{a+1} + (1/2)q) > m(p_{a+2})$. By the correspondence between exhaustive instruments and polygonal lines joining $\mathbf{0}$ to $\mathbf{1}$, we thus have and instrument

$$p_1, \ldots, p_{a-1}, p_a - (1/2)q, q, p_{a+1} + (1/2)q, p_{a+2}, \ldots, p_n$$

and by Lemma 7.5, I is stochastically factorizable and so by Corollary 7.1 is mixed.

Suppose now that every vertex of the upper polygonal line is pure, and suppose $I = (1/2)I_1 + (1/2)I_2$. Now each such vertex is a partial sum of atoms of I and so is a convex combination of partial sums of atoms of I_1 and I_2. By the purity assumption all three of these partial sums coincide. Since the atoms of the instruments can be uniquely recovered from these partial sums, the atoms of all three instruments coincide and I is pure. Q.E.D

Corollary 7.3 *The pure exhaustive instruments for a two dimensional figure form a filtered set under refinement.*

Proof: The pure exhaustive instruments are in one to one correspondence with finite subsets of the extreme points of the upper boundary of $\mathcal{O}$ with $\mathbf{0}$ and $\mathbf{1}$ deleted: we simply identify these with the vertex set of the upper polygonal line of the instrument. For any two such sets, their union gives a common refinement. Q.E.D

Corollary 7.4 *In a two dimensional triple*

$$H(\sigma) = \lim_I \{h(\sigma, I) \mid I \text{ pure, exhaustive}\}.$$

Proof: We use Corollary 7.3 and the remark that the entropy increases with refinements of instruments. Q.E.D

Corollary 7.5 *The two dimensional instrumental entropy is concave:*

$$H(\lambda\sigma + (1-\lambda)\sigma') \geq \lambda H(\sigma) + (1-\lambda)H(\sigma').$$

Proof: We use Corollary 7.4 and the concavity of the classical entropy: $H(\lambda\sigma + (1-\lambda)\sigma') = \lim_I h(\lambda\sigma + (1-\lambda)\sigma', I) \geq \lim_I(\lambda h(\sigma, I) + (1-\lambda)h(\sigma', I)) = \lambda H(\sigma) + (1-\lambda)H(\sigma')$. Q.E.D

Hence if one pure state has infinite instrumental entropy, then so does every mixed state.

We see that if the observation figure is a polygon, then the entropy $H(\sigma)$ is given by $h(\sigma, I_0)$ where I_0 is the instrument corresponding to the upper polygonal boundary of $\mathcal{O}$. In this case it is always finite. The entropy could very well be infinite if $\mathcal{O}$ is not a polygon. Suppose the upper boundary of $\mathcal{O}$ is such that the projection onto the first component of the set of pure points contains a closed interval β. By convexity of $\mathcal{O}$ the same is then true for the projection onto the second component. Let us take a partition of β into n subintervals giving rise to an instrument containing at least the corresponding vectors $(\Delta x_i, \Delta y_i)$. Thus $h(\sigma_1, I) \geq -\sum \Delta x_i \log \Delta x_i$ and if we take the Δx_i all equal to some Δx the right hand side becomes $-n\Delta x \log \Delta x$. As $n \to \infty$, $n\Delta x$ remains constant and so this term becomes infinite, and $h(\sigma_1, I) = \infty$. A similar reasoning shows $h(\sigma_0, I) = \infty$ and by Corollary 7.5 $h(\sigma_\lambda, I) = \infty$.

In such cases it may happen that certain natural constructions can still be finte. Suppose the upper boundary is the graph of a $\mathcal{C}^1$ function b and that all of its points are pure. Let I be a pure exhaustive instrument, and then using the obvious notation we can write $I = ((\Delta x_i, \Delta b_i))_i$. If we now compute $h(\sigma_\lambda, I) - \lambda h(\sigma_1, I) - (1-\lambda)h(\sigma_0, I)$ then after rearrangement we have

$$-\lambda \sum \Delta x_i \log(\lambda + (1-\lambda)\Delta b_i/\Delta x_i) +$$

$$-(1-\lambda)\sum (\Delta b_i/\Delta x_i)\Delta x_i \log(\lambda(\Delta x_i/\Delta b_i) + (1-\lambda))$$

which in the limit goes over to

$$-\lambda \int_0^1 \log(\lambda + (1-\lambda)b'(x))\, dx - (1-\lambda)\int_0^1 b'(x)\log(1-\lambda+\lambda/b'(x))\, dx$$

and this may very well be finite depending of the function $b'(x)$. We may symbolically write the result as $H(\sigma_\lambda) - \lambda H(\sigma_1) - (1-\lambda)H(\sigma_0)$ and this corresponds to the removal of the infinite contributions to the entropy by the pure states.

Under the same hypotheses, the limit of discrimination entropies could also be finite. The discrimination entropy of two probability measures s and r on $\{1, \ldots, n\}$ is defined as $h(s, r) = -\sum s_i \log(s_i/r_i)$. Thus if σ, σ' are two states and I an exhaustive instrument, we can define $H(\sigma, \sigma'; I) = h(\langle \sigma, I(\cdot)\rangle, \langle \sigma', I(\cdot)\rangle)$. Making a calculation similar to the above, we find that in our case $H(\sigma_\lambda, \sigma_\mu; I)$ in the limit approaches

$$-\int_0^1 (\lambda + (1-\lambda)b'(x)) \log \frac{\lambda + (1-\lambda)b'(x)}{\mu + (1-\mu)b'(x)}\, dx$$

which again could be finite depending on $b'(x)$.

That entropies could be infinite in such seemingly simple situations as two dimensional observation figures should not be considered unnatural as the following considerations suggest.

Every polygonal two dimensional observation figure $\mathcal{O}$ can be obtained by restricting from a larger finite dimensional Boolean system. Let the upper polygonal boundary of $\mathcal{Q}$ be represented by the instrument $I = (p_1, \ldots, p_n)$. Consider the n-point Boolean probability triple $(\mathcal{S}_n, \mathcal{O}_n, \langle \cdot, \cdot \rangle)$ and the instrument $J = ((1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, \ldots, 0, 1))$ in it. A state $\sigma \in \mathcal{S}_n$ is given by an n-tuple of numbers $0 \leq s_i \leq 1$, $\sum s_i = 1$. Consider now the two states $\tau_0 = (p_{i,2})$, $\tau_1 = (p_{i,1})$ where $p_{i,2}$ and $p_{i,1}$ are the second and first components of $p_i \in \mathcal{O}$. Restrict now the system to the production only of states in the segment $[\tau_0, \tau_1]$. The reduced system of $([\tau_0, \tau_1], \mathcal{O}_n, \langle\cdot,\cdot\rangle)$ is now precisely the given two dimensional triple with the two instruments above corresponding atom by atom. The high entropy of the pure states for a polygonal observation figure thus can be interpreted as arising from the formal identification of these as highly mixed states in a larger Boolean model. This larger model may or may not have phenomenological significance. The pure components of σ_i in the formal mixture $\tau_0 = \sum p_{i,2}\sigma_i$ could correspond to unrealizable situations. In the case of organisms for example the formal pure states could correspond to conditions under which the organism would be dead, in which case the organism would in a certain sense be always intrinsically mixed though phenomenologically it could be a pure state in the statistical theory adapted to it. If the observation figure is not a polygon, then to interpret a oure state as a formal mixture in a Boolean system we must go to infinite Boolean algebras and hence should not be surprised that the entropy could be infinite. We pursue these ideas further in the next chapter.

4. Consider a quantum triple in a Hilbert space $\mathcal{H}$ of dimension $N < \infty$. Now the canonical operational theory associated to the triple is not what is normally considered as being the statistical theory of quantum mechanics. The canonical theory contains instruments that normally one does not admit. According to the usual interpretation two operators represent observables that can be simultaneously measured if and only if they commute. Thus if $I = (A_1, \ldots, A_n)$ is an instrument we should have $A_iA_j = A_jA_i$ This view, as has already been remarked, does not exactly correspond to our idea of what constitutes an instrument; for example, if I and J are two instruments with some atom of I not commuting with some atom of J, then $\theta_{(1/2,1/2)}(I, J)$ is a perfectly realizable instrument with noncommuting atoms.

Let us consider then the statistical theory $T_c(\mathcal{H})$ generated by using the constructions of Chapter 4 and starting from instruments with commuting atoms. Let $\mathcal{I}_n^c$ be the set of n atom instruments of $T_c(\mathcal{H})$ with commuting atoms.

Lemma 7.11 $T_c(\mathcal{H}) = conv(\mathcal{I}_n^c)$.

Proof: $\mathcal{I}_n(T_c(\mathcal{H}))$ is an algebraically closed subset of a finite dimensional space, thus each point is a finite convex combination of extreme points. It thus suffices to show that each extreme point of $\mathcal{I}_n(T_c(\mathcal{H}))$ belongs to conv($\mathcal{I}_n^c$). Now $\mathcal{I}(T_c(\mathcal{H}))$ can be built up recursively by means of instruments constructed in previous steps and using condensations, exhaustions, stochastic splatters, and sleight of hand operations whose instruments are also constructed in previous recursive steps. The recursion initiates with $\mathcal{I}_n^c$. By analysis made on the effects of the presence of substitution operations on pages 75-78, where $\theta(I_1, \ldots, I_m)$ was replaced by a stochastic condensation of $(\Xi^{(0)}J, \Xi^{(1)}I_1, \ldots, \Xi^{(m)}I_mJ)$ we see that by allowing stochastic condensations we can forgo using sleight of hand operations in the construction. Since stochastic condensations can be effected by stochastic splatters and ordinary condensations, as was done in the proof of Lemma 7.1, we see that in constructing $\mathcal{I}(T_c(\mathcal{H}))$ we can forgo completely the substitution operations. In constructing an extreme instrument I though, each statistical splitter must be inessential, thus replacing each splitter by an appropriate one with only one certain exit and the rest dummy, we see that I must be an exhaustion of a condensation of an element of $\mathcal{I}^c$. But this leads again to an element of $\mathcal{I}^c$. Thus $I \in \mathcal{I}_n^c$. Q.E.D

Lemma 7.12 *The extreme points of $\mathcal{I}_n(T_c(\mathcal{H}))$ are instruments with projections for atoms.*

Proof: Suppose I is an extreme point of $\mathcal{I}_n(T_c(\mathcal{H}))$. Then as was shown in the proof of Lemma 7.11, $I \in \mathcal{I}_n^c$. Thus there is a spectral measure E of a finite set X such that $A_i = \int a_i(x)\, E(dx)$ where $a_i : X \to [0,1]$. Now each instrument $F = (f_1, \ldots, f_n)$ in the Boolean triple based on $\mathbb{B} = \mathcal{P}(X)$ defines and instrument in $\mathcal{I}_n^c$ by $\psi : F \mapsto (B_1, \ldots, B_n)$ where $B_i = \int f_i(x)E(dx)$. Now $\psi^{-1}(I)$ must be a face of $I_n(\mathbb{B})$ seeing that I is extreme. Thus there is an extreme point of $I_n(\mathbb{B})$ whose image by ψ is I. But extreme points of $I_n(\mathbb{B})$ are defined by n-tuples of characteristic functions whose spectral integrals are projections. Thus I has projections as atoms. Q.E.D

Now $T_c(\mathcal{H})$ is a perfectly acceptable fragment of quantum mechanics, though as we shall see it doesn't incorporate all of the normally acceptable ideas.

Theorem 7.6 *The instrumental entropy function in $T_c(\mathcal{H})$ is constant: $H(\rho) = \log N$.*

Proof: We have, by definition,

$$H(\rho) = \sup\{h(\rho, I) \mid I \text{ pure and exhaustive}\}.$$

Now by Lemma 7.12, I has projections for atoms, and as the entropy increases with refinement we can assume that these atoms are one dimensional Projections $(\psi_i, \cdot)\psi_i$ for some orthonormal basis ψ_i, $i = 1, \ldots, N$. Thus $h(\rho, I) = -\sum(\psi_i, \rho\psi_i)\log(\psi_i, \rho\psi_i)$ which is the entropy of the diagonal elements of the matrix of ρ in the orthonormal basis ψ_i, $i = 1, \ldots, N$. Consider now a 2 by 2 hermitian matrix A, then we can write $A = \alpha\mathbf{1} + \vec{\beta} \cdot \vec{\tau}$ where $\vec{\tau}$ is the usual vector of Pauli spin matrices. The diagonal entries of A are $\lambda_1 = \alpha + \beta_3$ and $\lambda_2 = \alpha - \beta_3$. Assume say that $\lambda_1 \geq \lambda_2$. A unitary transform of A corresponds to a rotation of $\vec{\beta}$ and so for any $\mu \leq \lambda_1 - \lambda_2$ the diagonal elements can be changed to $\lambda_1 - \mu$, $\lambda_2 + \mu$ by a unitary change of basis. Applying this reasoning now to the matrix of ρ, we can, by unitary transformations in two dimensional subspaces, bring closer together by any given amount any pair of diagonal entries. Apply now the following procedure: if there is an entry less that $1/N$, then by the fact that $\mathrm{Tr}(\rho) = 1$, there must be one larger that $1/N$. By narrowing the difference between these two entries, one of them can be brought to coincide with $1/N$. After a finite number of steps therefore all the entries are $1/N$ and for this particular basis ψ'_i, and the corresponding instrument $I' = ((\psi'_1, \cdot)\psi'_1, \ldots, (\psi'_N, \cdot)\psi'_N)$ we have $h(\rho, I') = \log N$. Since this number is the maximum that $h(\rho, I)$ can achieve in a pure instrument, we have $H(\rho) = \log N$. Q.E.D

We have here our first major contrast between quantum theory and theories of a more classical nature as in our first three examples. In $T_c(\mathcal{H})$ each state can appear maximally chaotic if the instrument is appropriately chosen. Note that each basis provides a maximal instrument, thus in contrast to finite Boolean schemes and polygonal two dimensional triples, no instrument exists that completely determines the state. A state reveals its true nature only if we compare its behavior with respect to instruments that are incompatible, in the sense that one has an atom that never appears with an atom of the other in yet a third instrument.

To understand the situation better, we must now consider operations. As before, we consider a finite dimensional Hilbert space $\mathcal{H}$ of dimension N. According to the standard interpretation of quantum mechanics, if B is any observable with spectral measure E: $B = \sum_{i=1}^{N} \lambda_i E_i$, then if we observe B the result is one of the numbers λ_i. If before observation the state was ρ, then the new state immediately after having observed the value λ_i is $E_i\rho E_i/\mathrm{Tr}(E_i\rho E_i)$ provided $\mathrm{Tr}(E_i\rho E_i) \neq 0$. The frequency with which λ_i is observed as the value of B is precisely this latter number $\mathrm{Tr}(E_i\rho E_i) = \mathrm{Tr}(\rho E_i)$. Thus B can be thought of as describing an n-exit operation θ_B

in which $\theta^{(i)}_{B^*}\rho = E_i\rho E_i/\mathrm{Tr}(E_i\rho E_i)$, $\nu_i(\rho) = \mathrm{Tr}(E_i\rho E_i)$; but this means that $\theta^{(i)}_B(A) = E_iAE_i$. We now make a slight modification of this idea, namely, we allow for the possibility that observing some of the λ_i may involve a destruction of the state. For instance, in determining linear polarization of a beam of light we place a polarizer in the beam represented by a one dimensional projection P in $\mathbb{C}^2$. If a photon in a state ρ passes through, we have a copy of the the exit state $P\rho P/\mathrm{Tr}(P\rho P)$, but if the photon is absorbed, we do not have access to a copy of the corresponding state $(1-P)\rho(1-P)/\mathrm{Tr}((1-P)\rho(1-P))$. We thus admit the following: A *quantum mechanical statistical theory* is a statistical theory generated by the constructions of Chapter 4 and starting with instruments with commuting atoms and a family F of operations $\theta_{\mathbf{E}}$ where each $\mathbf{E}$ is an n-tuple of orthogonal projections $E_1, \ldots, E_n$ such that $P_{\mathbf{E}} = E_1 + \cdots + E_n \leq 1$, and where $\theta^{(i)}_{\mathbf{E}}(A) = E_iAE_i$. Note that from the inequality we get by multiplying by E_j on the right and left that $\sum_{k\neq j} E_jE_kE_j \leq 0 \Rightarrow E_jE_kE_j = 0 \Rightarrow E_kE_j = 0$ for each $k \neq j$. Thus $\mathbf{E}$ is part of a spectral measure. Let $T_F(\mathcal{H})$ be the statistical theory so defined. Now there is no simple and reasonable way to choose the family F. The usual way to avoid this problem is to accept any $\theta_{\mathbf{E}}$ as a possible operation. In this case we call the resulting theory $T_Q(\mathcal{H})$ and call it the quantum mechanical theory associated to $\mathcal{H}$. We also consider this theory in cases when the dimension of $\mathcal{H}$ is infinite.

In $T_Q(\mathcal{H})$ we have instruments with noncommuting atoms that do not come about via stochastic splitters. Consider the theory of a free particle in one dimension as formalized in $L^2(\mathbb{R})$. Let E and F be the spectral measures of the momentum operator and the position operator respectively. If we first prepare a state that lies in $E(A)\mathcal{H}$, $A \subset \mathbb{R}$, $E(A) \neq 0, 1$, at $t = 0$ and at $t = t_1 > 0$ observe with the instrument $F(B_1), \ldots, F(B_n)$ where the B_i form a partition of $\mathbb{R}$, then in terms of $\mathcal{H}$ we are making a simultaneous observation of properties represented by $A_1, \ldots, A_n$ where $A_i = E(A)U(t_1)F(B_i)U(t_1)^*E(A)$ and U is the unitary group of time translation. Of course, this really involves measurements at different times, yet even for t_1 near zero, $[A_i, A_j]$ is still appreciably different from zero. According to our formalizm there is no way of denying that this is a bona fide instrument, so one can ask just what does the normal interpretation require when it insists on commutativity of the operators that correspond to simultaneous observations. One can argue that what is hidden in the above example is that the $F(B_i)$ in fact commute and it just happens that we are restricting our states to be in $E(A)\mathcal{H}$ which is a proper subspace of $\mathcal{H}$. Thus we've placed our preparations as being part of the observation. One could try to say that only the so called ultimate instruments, those that do not involve previous preparations must be commutative. This idea is in fact somewhat incorporated in the very definition of $T_Q(\mathcal{H})$ whose ultimate instruments come from I^c. Now whether phenomenologically speaking an instrument is ultimate or not is certainly hard to define. It's hard to see how measurements can be done without preparation except in the extreme cases of either not interacting with the state or destroying it instantly. In terms of our formalism, suppose $I = (A_1, \ldots, A_n)$ is an instrument in $T_Q(\mathcal{H})$ with commuting atoms, and for simplicity assume the spectrum of each A_i is finite. Then as before, there is a finite set X, a spectral measure E on X, and functions $a_i : X \to [0,1]$ such that $A_i = \int a_i(x)\, E(dx)$. We can think of the family of functions a_i as an instrument J in the Boolean triple based on $\mathbb{B} = \mathcal{P}(X)$. We can then write J as a convex combination of extreme instruments $J = \sum_1^k \lambda_i J_i$. The spectral integrals of the J_i provide instruments I_i in T_Q with projections for atoms and I is thus a condensation of $\theta_{(\lambda_1,\ldots,\lambda_k)}(I_1, \ldots, I_k)$. We can thus assume ultimate instruments have projections for atoms. If now however $I = (E_1, \ldots, E_n)$ has projections for atoms, then $I = \theta_{\mathbf{E}}(\mathbf{1}, \ldots, \mathbf{1})$ where $\mathbf{E} = (E_1, \ldots, E_n)$ so the only ultimate instrument is $\mathbf{1}$, and the commutativity requirement cannot be explained by any notion of ultimate instrument. Commutativity seems to be the result of attributing self adjoint operators to observables. Now if B is an observable with finite spectrum $\lambda_1, \ldots, \lambda_n$ and E is its spectral measure, then we can consider the result of observing B as that of observing with the instrument $I = (E_1, \ldots, E_n)$ where $E_i = E(\{\lambda_i\})$ and assigning value λ_i to the observation if the i-th property of I is realized. We see in this case that I is an exhaustive instrument with commuting atoms. Suppose now $I = (A_1, \ldots, A_n)$ is any exhaustive instrument of $T_Q(\mathcal{H})$ and suppose each time the i-th property of I is realized we assign the value λ_i to our observation. In general this doesn't correspond to any observable B for then we must have for some spectral measure $(E_1, \ldots, E_n)$ that $\mathrm{Tr}(\rho A_i) = \mathrm{Tr}(\rho E_i)$ for all ρ and so $A_i = E_i$. So I in general cannot be used to construct observables as understood by the standard interpretation, though there is nothing wrong operationally with the procedure just described. This we must understand by realizing that the standard model describes somehow ideal measurements, and so our procedure must correspond to an imprecise observation. We may thus remark in our example of the free particle that the simultaneous observation of the A_i are not ideal position measurements, for before measuring, the instrument projects into the momentum subspace $E(A)\mathcal{H}$. Commutativity seems to involve the notion of ideal measurement.

This raises now two questions of consistency. First of all we had previously considered an operator A with $0 \leq A \leq 1$ as a possibly non-ideal measurement of a yes-no property. By the above discussion this non-ideal measurement should correspond to a two atom instrument $I = (A_1, A_2)$ with 'yes' attributed to the realization of the first property. This is totally consistent if we take $A_1 = A$. It must be emphasized at this point that the atoms of an instrument I must not be treated as observables, even though they are represented by self adjoint operators. Observing with I one of the atoms is realized, that is achieves the value 'yes' and all the others 'no'; we do not observe the eigenvalue of any of them. Each instrument defines any number of not necessarily ideal observables by assigning values to each atom.

The second point of consistency is that we have already argued, by a different route, that ideal observables would correspond to pure instruments, whereas the standard quantum mechanical interpretation would say that they correspond to self adjoint operators, that is instruments with projections as atoms. This is consistent in $T_c(\mathcal{H})$ by Theorem 7.12 but as we shall see shortly in $T_Q(\mathcal{H})$ there are pure instruments with atoms that are not projections. This apparent inconsistency can be resolved by noting that the notion of ideal observable in quantum mechanics means a bit more than purity of the instrument. Ideality in quantum mechanics also means repeatability of the measurement and the preservation of purity of states. Thus observing B again, after having observed the eigenvalue λ leads again to the same eigenvalue now with certainty, and the exit state after the second measurement coincides with the one after the first. Moreover, observing with B, pure states are transformed into pure states. How do we repeat a non-ideal observation? We must determine what happens to a state after a non-ideal measurement. This in principle is not uniquely determined by the corresponding instrument $I = (A_1, \ldots, A_n)$ as occurs in the ideal case. However, in $T_Q(\mathcal{H})$, as we shall briefly show, we can always write I as $\theta(\mathbf{1}, \ldots, \mathbf{1})$ for some operation θ. Now θ is not uniquely determined by I, but if we shift our attention to θ we can talk about its repeatability properties. What we want is that if a state leaves by the i-th exit of θ, then applying θ again, it leaves again by the i-th exit now with certainty and unchanged. We call such a θ *repeatable*. If in addition $\theta_*^{(i)}\sigma$ is pure whenever it is defined and σ is pure, we call θ *ideal*. In our formalism, leaving again by the i-th exit with certainty means

$$\langle \sigma, \theta^{(i)}\mathbf{1}\rangle \neq 0 \Rightarrow \langle \theta_*^{(i)}\sigma, \theta^{(j)}\mathbf{1}\rangle = \delta_{ij},$$

and leaving unchanged means

$$(\theta_*^{(i)})^2\sigma = \theta_*^{(i)}\sigma.$$

This in turn implies $\langle (\theta_*^{(i)})^2\sigma, p\rangle = \langle \theta_*^{(i)}\sigma, p\rangle$ but $\langle (\theta_*^{(i)})^2\sigma, p\rangle\langle \theta_*^{(i)}\sigma, \theta^{(i)}\mathbf{1}\rangle = \langle \theta_*^{(i)}\sigma, \theta^{(i)}p\rangle$ which using the first equation now leads to $\langle \theta_*^{(i)}\sigma, p\rangle = \langle \theta_*^{(i)}\sigma, \theta^{(i)}p\rangle$. This now is equivalent to $\langle \sigma, \theta^{(i)}p\rangle\langle \sigma, \theta^{(i)}\mathbf{1}\rangle = \langle \sigma, (\theta^{(i)})^2 p\rangle\langle \sigma, \theta^{(i)}\mathbf{1}\rangle$ which by Axiom 4.9′ implies $\langle \sigma, \theta^{(i)}p\rangle = \langle \sigma, (\theta^{(i)})^2 p\rangle$ which finally says $(\theta^{(i)})^2 = \theta^{(i)}$. Now for $i \neq j$, $\langle \theta_*^{(i)}\sigma, \theta^{(j)}\mathbf{1}\rangle = 0 \Rightarrow \langle \sigma, \theta^{(i)}\theta^{(j)}\mathbf{1}\rangle\langle \sigma, \theta^{(i)}\mathbf{1}\rangle = 0$ and again by Axiom 4.9′ we conclude that $\theta^{(i)}\theta^{(j)} = 0$. Repeatability is thus equivalent to $\theta^{(i)}\theta^{(j)} = \delta_{ij}\theta^{(i)}$

We now show that every instrument in $T_Q(\mathcal{H})$ is of the form $\theta(\mathbf{1}, \ldots, \mathbf{1})$. By definition of T_Q every instrument is a (possible) exhaustion of a (possible) condensation of a $\theta(I_1, \ldots, I_m)$ where $I_i \in \mathcal{I}^c$. As was already show, each I_i can be written as a convex combination of instruments with projections for atoms and these convex combinations can be obtained by condensing from stochastic splitters with the given instruments with projections for atoms at the exits. An instrument with projections for atoms can be written as $\theta_{\mathbf{E}}(\mathbf{1}, \ldots, \mathbf{1})$. Thus we can replace $\theta(I_1, \ldots, I_m)$ above by a $\theta(\mathbf{1}, \ldots, \mathbf{1})$ where of course we now use θ to mean a different operation. A condensation of such an instrument can be obtained by first condensing exits of θ and again placing $\mathbf{1}$ on the new exits. Any instrument I is therefore a (possible) exhaustion of a $\theta(\mathbf{1}, \ldots, \mathbf{1})$. If now somewhere in the construction of θ, state destroying elements occurred, these are of two possible types: either condensations that ignore exits or operators $\theta_{\mathbf{E}}$ where $\mathbf{E} = (E_1, \ldots, E_k)$ is not a full spectral measure and $p_{\mathbf{E}} = E_1 + \cdots + E_k < 1$. Introduce now a new exit labelled by 0. For each ignored exit in any condensation involved in the construction of θ, allow this ignored exit to lead now to exit 0. For each $\theta_{\mathbf{E}}$ with $p_{\mathbf{E}} < 1$ introduce a $\theta_{\hat{\mathbf{E}}}$ with $\hat{\mathbf{E}} = (I_1, \ldots, I_k; 1 - p_{\mathbf{E}})$ and lead the new exit so obtained to exit 0. The new resulting operation θ' has one more exit and $\theta'(\mathbf{1}, \ldots, \mathbf{1})$ is the exhaustion of $\theta(\mathbf{1}, \ldots, \mathbf{1})$. Hence every instrument in T_Q can be written as $\theta(\mathbf{1}, \ldots, \mathbf{1})$.

For ease in the following computation let us make the inessential assumption that $\dim\mathcal{H} = N < \infty$. We claim that every one exit operation Ψ in T_Q (and subsequently every $\theta^{(i)}$ of an n-exit operation θ) can be written as

$\Psi(A) = \sum \mu_i C_i^* A C_i$. where $\mu_i \geq 0$ We prove this by recursion. Suppose Ψ is an exit of a many exit operation each exit of which has the above form. Condensing with other exits corresponds to summing with other expressions of the same form, which preserves the form. Suppose now we place at the exit of Ψ one of the generating operations: a stochastic splitter θ_Λ, a $\theta_{\mathbf{E}}$ or a sleight of hand operation θ_τ^J Let us examine the j-th exit of the composite; we have:

$$\begin{aligned}
(\Psi\{\theta_\Lambda\})^{(j)}(A) &= \sum_i \lambda_j \mu_i C_i^* A C_i, \\
(\Psi\{\theta_{\mathbf{E}}\})^{(j)}(A) &= \sum_i \mu_i E_j C_i^* A C_i E_j, \\
(\Psi\{\theta_\tau^J\})^{(j)}(A) &= \mathrm{Tr}(\rho_j \sum_i \mu_i C_i^* A C_i) B_j,
\end{aligned}$$

where $J = (B_1, \ldots, B_n)$ and $\tau = (\rho_1, \ldots, \rho_n)$. Of these the first two expressions are certainly of the given form, to see that the third one is also, let us introduce the spectral decompositions of ρ_j and B_j with respect to bases of eigenvectors:

$$\begin{aligned}
\rho_j &= \sum_{k=1}^{N} \lambda_k (\phi_k, \cdot)\phi_k, \\
B_j &= \sum_{m=1}^{N} \beta_m (\psi_m, \cdot)\psi_m.
\end{aligned}$$

Now we see that the third expression above can be written as:

$$\sum_{i,k,m} \lambda_k \mu_i \beta_m ((\psi_m, \cdot)\phi_k)^* C_i^* A C_i ((\psi_m, \cdot)\phi_k)$$

which again is of the claimed form. Since every operation of T_Q is built up using these steps starting from the operation Id, which clearly has the given form, our claim is justified. We thus have

$$\Psi_* \rho = \frac{\sum \mu_i C_i \rho C_i^*}{\mathrm{Tr}(\sum \mu_i C_i \rho C_i^*)}$$

provided the denominator is not zero. The requirement that pure states transform into pure states means that

$$\sum \mu_i C_i (\psi, \cdot)\psi C_i^* = \sum \mu_i (C_i \psi, \cdot) C_i \psi$$

must be of the form $(\phi, \cdot)\phi$ for all ψ. Since each summand is positive, this means that in particular $C_i\psi = \alpha_i\phi$ for each i, hence $C_i = \gamma_i C$ for some fixed C and we have:

$$\Psi(A) = (\sum \mu_i|\gamma_i|^2)C^*AC.$$

Incorporating the square root of the numerical factor into C, we can write $\Psi(A) = C^*AC$. Idempotency now implies that $(C^*)^2AC^2 = C^*AC$ for all A. Now this holding for all self adjoint operators A with $0 \leq A \leq 1$ implies that it holds for all bounded operators and in particular for $(\phi, \cdot)\psi$ which means that $(C^{*2}\phi, \cdot)C^{*2}\psi = (C^*\phi, \cdot)C^*\psi$ which in turn implies that $C^{*2}\psi = \alpha_\psi C^*\psi$ and $\bar{\alpha}_\phi\alpha_\psi = 1$ unless $C^*\phi$ or $C^*\psi$ is zero. This can only be satisfied if α_ψ is independent of ψ whenever $C^*\psi \neq 0$ and we can now write $C^{*2}\psi = \alpha C^*\psi$ with $|\alpha| = 1$. But now $(\alpha C)^{*2} = \bar{\alpha}^2 C^{*2} = \bar{\alpha}^2\alpha C^* = \bar{\alpha}C^* = (\alpha C)^*$ and $(\alpha C)^*$ is an idempotent. Since multiplying C by a phase factor doesn't change Ψ we can write $\Psi(A) = P^*AP$ where P is an idempotent. Taking $A = 1$ we must have $\Psi(1) = P^*P \leq 1 \Rightarrow \|P\| \leq 1$, but a Hilbert space projection with norm not exceeding one is necessarily orthogonal so in fact $\Psi(A) = EAE$ for an orthogonal projection E. An ideal operation therefore has the form $\theta^{(j)}(A) = E_jAE_j$ with $E_iE_j = \delta_{ij}E_i$, that is $\theta = \theta_{\mathbf{E}}$ and its instrument $\theta(\mathbf{1}, \ldots, \mathbf{1})$ has projections for atoms. Instruments corresponding to ideal measurements in T_Q therefore have projections for atoms and thus commute.

We conclude from the above analysis that the condition of commutativity in quantum mechanics corresponds to more than just commensurability, but incorporates in addition repeatability and the preservation of purity of states.

Before exploring further the structure of $T_Q(\mathcal{H})$, we remark that repeatability and preservation of purity of states are rather stringent conditions and a general statistical theory should not have any ideal operations other than the ones that are always present, such as 0, Id, and certain sleight of hand operations. Consider a two dimensional statistical triple and an idempotent purity preserving one exit operation θ, which would be given by a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Now $\theta_*\sigma_\lambda = \sigma_\mu$ where $\mu = (\lambda a+(1-\lambda)c)/(\lambda(a+b)+(1-\lambda)(c+d))$. So $[\sigma_0, \sigma_1]$ gets transformed into $[\sigma_{c/(c+d)}, \sigma_{a/(a+b)}]$ unless $c+d = 0$ or $a+b = 0$. If $(c+d)(a+b) = 0$ but $\theta \neq 0$, then the image interval is degenerate and θ_* is undefined on one of the pure states, if $\theta = 0$ then θ_* is totally undefined. By preservation of purity, there are four possibilities for the image: $[\sigma_0, \sigma_0]$,

$[\sigma_0, \sigma_1]$, $[\sigma_1, \sigma_0]$ and $[\sigma_1, \sigma_1]$ and these correspond to θ of the form

$$\begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix}, \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}, \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix} \text{ and } \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix}$$

respectively. Imposing idempotency results finally in the following four possibilities:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & b \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 \\ c & 0 \end{pmatrix}.$$

Now in general the last two don't exist unless $b = 1$ and $c = 1$, since applying them to $\mathbf{1}$ results in $(b, 1)$ and $(1, c)$ which cannot be assumed to be in $\mathcal{O}$ unless $b = c = 1$. We have however $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = \theta^{\mathbf{1}}_{\sigma_0}$, $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \theta^{\mathbf{1}}_{\sigma_1}$ which certainly must always be present.

This points out yet another essential difference between general theories on the one hand, and quantum mechanics and Boolean theories on the other hand. The latter are completely determined by their ideal operations, whereas general theories have a scarcity of them.

We resume now the study of $T_Q(\mathcal{H})$ and assume now that $\dim \mathcal{H} = \infty$ and let $\mathcal{K} \subset \mathcal{H}$ be a finite dimensional subspace. Suppose now that $A_1, \ldots, A_n$ are positive operators in $\mathcal{K}$ such that $A_1 + \cdots A_n \leq 1_{\mathcal{K}}$. By [10] it is now possible to find a finite dimensional Hilbert space $\mathcal{K}^+ \supset \mathcal{K}$ with projection $P^+ : \mathcal{K}^+ \to \mathcal{K}$ and a spectral measure $E_1^+, \ldots, E_n^+$ on $\mathcal{K}^+$ such that $A_i = P^+ E_i P^+$. Since $\mathcal{K}$ is infinite dimensional we can consider $\mathcal{K}^+ \subset \mathcal{H}$ and so there is a projection P in $\mathcal{H}$ and projections E_i such that $E_1 + \cdots + E_n \leq 1$ and $A_i = P E_i P$. Let $Q = E_1 + \cdots + E_n$, then $P \leq Q$. Now if $\mathbf{E} = \{P\}$ and $I = (E_1, \ldots, E_n)$, then $\theta_{\mathbf{E}}$ and I are objects of T_Q(Ḩ) and we have $\theta_{\mathbf{E}}(I) = (A_1, \ldots, A_n)$. Thus in $T_Q(\mathcal{H})$ with $\mathcal{H}$ infinite dimensional the instruments that act on any finite dimensional subspace can be taken to be those of the canonical operational statistical theory associated to the quantum triple based on the subspace. If our quantum mechanical system is described by an infinite dimensional Hilbert space, and if we admit any self adjoint operator as an observable, then in any finite dimensional subspace $\mathcal{K}$, given any set of operators $A_1, \ldots, A_n$ that are positive and so that $A_1 + \cdots + A_n \leq 1_K$ then there is an ideal operation in $\mathcal{H}$ whose instrument acting on the states in K is represented by $(A_1, \ldots, A_n)$. Admitting $T_Q(\mathcal{H})$ as a model implies admitting at least the instrumental structure of the canonical operational theory associated to a finite dimensional Hilbert space as a realistic model.

We now return to the investigation of the instrumental entropy function.

Theorem 7.7 *Let $T_I(\mathcal{H})$ be a quantum mechanical triple with $\dim\mathcal{H} = N < \infty$ and the set of instruments given by the canonical operational theory. The instrumental entropy function is constant equal to $\log N^2$.*

Before proving the theorem we shall need a few technical results.

Suppose $I = (A_1, \ldots, A_n)$ is not extreme, then $I = (1/2)I_1 + (1/2)I_2$ with $I_j = (A_1^{(j)}, \ldots, A_n^{(j)})$. Let $C_i = A_i^{(2)} - A_i^{(1)}$ then we see that the C_i are self adjoint, not all zero, $\sum C_i = 0$, and $0 \leq A_i + \eta C_i \leq 1$ for $|\eta| \leq 1/2$. Conversely, if such C_i existed, then I could not be extreme since then $\eta \mapsto (A_1 + \eta C_1, \ldots, A_n + \eta C_n)$ would be a nontrivial affine map $[-1/2, 1/2] \to \mathcal{I}_n$ with $0 \mapsto I$.

Suppose we have $0 \leq A + \eta C \leq 1$ for A and C self adjoint and $|\eta| \leq 1/2$. Let $E = E(\{0\}) + E(\{1\})$ be a spectral projection of A and decompose $\mathcal{H}$ into $E\mathcal{H} \oplus (1-E)\mathcal{H}$. In matrix form we have

$$A = \begin{pmatrix} E_1 & 0 \\ 0 & B \end{pmatrix} \quad C = \begin{pmatrix} C_1 & C_2 \\ C_2^* & C_3 \end{pmatrix}$$

where $E_1 = E(\{1\})$ and B, C_1 and C_3 are self adjoint. From the inequality we obtain in particular that $0 \leq E_1 + \eta C_1 \leq 1$ which by the extremality of E_1 implies $C_1 = 0$. Applying now $A + \eta C$ to $x \oplus y$ we get

$$0 \leq \|x\|^2 + 2\eta \mathrm{Re}(x, C_2 y) + (y, (B + C_3)y) \leq \|x\|^2 + \|y\|^2 \Leftrightarrow$$

$$\Leftrightarrow 0 \leq 2\eta \mathrm{Re}(x, C_2 y) + (y, (B + \eta C_3)y) \leq \|y\|^2.$$

If $C_2 \neq 0$ then we can find a y and an x such that $(x, C_2 y)$ is real and different from zero. Picking $\eta \neq 0$ we can then multiply x by a real constant in such a way as to contradict the inequality, thus $C_2 = 0$ and we conclude that the null space of C contains $E\mathcal{H}$. Conversely if C is self adjoint and contains $E\mathcal{H}$ in its null space, then for some $R > 0$, $|C| \leq RE((0,1))$ and so $0 \leq A + \eta(2/R)C \leq 1$ for $|\eta| \leq 1/2$.

Suppose now $I = (A_1, \ldots, A_n)$ is an extreme instrument. By the spectral theorem we can write for each i, $A_i = \sum_j \alpha_{ij}(\phi_{ij}, \cdot)\phi_{ij}$ where $(\phi_{ij})_{j=1,\ldots,N}$ is an orthonormal basis for $\mathcal{H}$. We then have a refinement $J = (\alpha_{ij}(\phi_{ij}, \cdot)\phi_{ij})_{i=1,\ldots,n; j=1,\ldots,N}$ of I. We claim that J is pure. If it weren't, there would exist self adjoint operators C_{ij} not all vanishing with $\sum_{ij} C_{ij} = 0$ and $0 \leq \alpha_{ij}(\phi_{ij}, \cdot)\phi_{ij} + \eta C_{ij} \leq 1$ for $|\eta| \leq 1/2$. By our previous paragraph we conclude that $C_{ij} = \beta_{ij}(\phi_{ij}, \cdot)\phi_{ij}$ with $\beta_{ij} = 0$ if $\alpha_{ij} = 0$ or 1. Let now

$C_i = \sum_j C_{ij} = \sum_j \beta_{ij}(\phi_{ij}, \cdot)\phi_{ij}$ and since some $\beta_{ij} \neq 0$ and each $(\phi_{ij})_{j=1,\ldots,N}$ is an orthonormal basis, not all the C_i vanish. Now $0 \leq A_i + \eta C_i \leq 1$ for $|\eta| \leq 1/2$ and $\sum C_i = 0$ so I is not extreme contrary to the hypothesis.

Since the entropy $h(\rho, I)$ increases with refinements we need only look at instruments of the form where each atom has rank one.

Lemma 7.13 *In $T_I(\mathcal{H})$ an instrument $I = ((\phi_i, \cdot)\phi_i)_{i=1,\ldots,n}$ is extreme if and only if the operators $(\phi_i, \cdot)\phi_i$ for $\|\phi_i\| \neq 0$, 1 are linearly independent as elements of the real vector space of hermitian operators.*

Proof: If I is not extreme then there are real numbers β_i not all zero such that $0 \leq (1 + \eta\beta_i)(\phi_i, \cdot)\phi_i \leq 1$ for $|\eta| \leq 1/2$ and such that $\sum \beta_i(\phi_i, \cdot)\phi_i = 0$. Now the first inequality implies that $\beta_i = 0$ if $\|\phi_i\| = 0$ or 1 and so the $(\phi_i, \cdot)\phi_i$ for $\|\phi_i\| \neq 0$ or 1 are linearly dependent. Following the argument backwards, we find a set of real numbers γ_i not all zero with $\gamma_i = 0$ if $\|\phi_i\| = 0$ or 1 and such that $\sum \gamma_i(\phi_i, \cdot)\phi_i = 0$; setting $\beta_i = \gamma_i/R$ for R sufficiently large, the first inequality is satisfied and we conclude that I is not extreme. Q.E.D

Lemma 7.14 *In $T_I(\mathcal{H})$ if I is extreme with n rank one atoms, then $n \leq N^2$*

Proof: Consider the $(\phi_i, \cdot)\phi_i$ with $\|\phi_i\| = 1$, then their sum is a sum of projections and bounded by 1 and as has already been seen, they must be orthogonal and sum up to a projection E of dimension k, say. Then $\sum\{(\phi_i, \cdot)\phi_i \mid \|\phi_i\| \neq 1\} = 1 - E$ and by the previous lemma these $(\phi_i, \cdot)\phi_i$ are linearly independent operators in $(1 - E)\mathcal{H}$. Since we can have at most $(N - k)^2$ linearly independent hermitian operators in $(1 - E)\mathcal{H}$ we have $n \leq k + (N - k)^2 \leq N^2$. Q.E.D

According to the proof of Theorem 7.6 there is an orthonormal basis $\phi_1, \ldots, \phi_N$ such that $(\phi_i, \rho\phi_i) = 1/N$ where ρ is a given state. Consider now the instrument J with N^2 rank one atoms obtained by repeating each $(1/N)(\phi_i, \cdot)\phi_i$, N times. We find $h(\rho, J) = \log N^2$. Now J is not extreme by Lemma 7.13, however by Lemma 7.14, $\log N^2$ is the maximum value that $h(\rho, I)$ can have in a pure instrument. Our theorem would then be proved once we establish the following result:

Lemma 7.15 *There is a pure instrument arbitrarily close to J.*

Proof: Consider an instrument I with N^2 rank one atoms $((\psi_i, \cdot)\psi_i)_{i=1,\ldots,N^2}$, $\sum(\psi_i, \cdot)\psi_i = 1$. Let ϵ_a, $a = 1, \ldots, N$ be any orthonormal basis and set

$\mu_{ia} = (\psi_i, \epsilon_a)$ The $N^2 \times N$ matrix μ obeys $\sum_i \mu_{ia}\bar{\mu}_{ib} = \delta_{ab}$. Reciprocally, if μ_{ia} is any such matrix then setting $\psi_i = \sum_a \mu_{ia}\epsilon_a$ one gets $1 = \sum(\psi_i, \cdot)\psi_i$. Thus the set of exhaustive instruments with rank one atoms is in one to one correspondence with the set $\mathcal{M}$ of $N^2 \times N$ matrices μ such that $\mu^t\bar{\mu} = 1$. Now $\mathcal{M} \subset \mathbb{C}^{N^3}$ and the condition $\mu^t\bar{\mu} = 1$ imposes $2N^2$ constraints. The real Frechet derivative of $f_{ab}(\mu) = (\mu^t\bar{\mu})_{ab} = \sum_i \mu_{ia}\bar{\mu}_{ib}$ is given by $Df_{ab}(\mu)h = \sum_i(\mu_{ia}\bar{h}_{ib} + \bar{\mu}_{ib}h_{ia})$. Since $\mu^t\bar{\mu} = 1$, the real rank of μ is maximal being $2N$, but this implies that the real rank of the derivative of $f = (f_{ab})$ is likewise maximum being $2N^2$ so that the set $\mathcal{M}$ besides being a real algebraic subvariety of $\mathbb{C}^{N^3}$ is in fact a differentiable submanifold. Let us show that $\mathcal{M}$ is connected. If μ_1 and μ_2 correspond to instruments and $(\xi_i, \cdot)\xi_i$, and $(\eta_i, \cdot)\eta_i$, $i = 1, \ldots, N^2$ respectively, then since $1 = \sum(\xi_i, \cdot)\xi_i = \sum(\eta_i, \cdot)\eta_i$, N of the ξ, say $\xi_1, \ldots, \xi_N$ after renumbering, and N of the η, say $\eta_1, \ldots, \eta_N$ after renumbering, are bases for $\mathcal{H}$. Thus $\xi_i = \sum_{j=1}^N t_{ij}\eta_j$, $i = 1, \ldots, N$. The matrix $T = (t_{ij})$ being invertible possesses a logarithm and defining $W(s) = \exp(s \cdot \log T)$ and setting $\zeta_i(s) = \sum_{j=1}^N W(s)_{ij}\xi_j$ the $\zeta_1(s), \ldots, \zeta_N(s)$ form a base for $\mathcal{H}$ with $\zeta_i(0) = \xi_i$ and $\zeta_i(1) = \eta_i$. For $i > N$ let $\zeta_i(\cdot)$ now be any continuous path from $\zeta_i(0) = \xi_i$ to $\zeta_i(1) = \eta_i$. Let $A(s) = \sum_{i=1}^{N^2}(\zeta_i(s), \cdot)\zeta_i(s)$. We see that $A(s)$ is a positive operator and since $\zeta_1(s), \ldots, \zeta_N(s)$ is a basis for $\mathcal{H}$, $A(s)$ is invertible for all s. We now have $1 = \sum_{i=1}^{N^2}(A(s)^{-1/2}\zeta_i(s), \cdot)A(s)^{-1/2}\zeta_i(s)$ and so provides a path in $\mathcal{M}$ from μ_1 to μ_2. We conclude therefore that $\mathcal{M}$ is an irreducible algebraic variety. Let us now establish a one to one correspondence between $1, 2, \ldots, N^2$ and pairs (a, b); $a, b = 1, \ldots, N$ setting $(a(j), b(j))$ to be the image of j. For the $(\psi_i, \cdot)\psi_i$ to be linearly independent we must have that the determinant of the matrix $F_{ij}(\mu) = \mathrm{Tr}((\psi_i, \cdot)\psi_i((\epsilon_{b(j)}, \cdot)\epsilon_{a(j)} + (\epsilon_{a(j)}, \cdot)\epsilon_{b(j)}))$ be different from zero. We have $F_{ij}(\mu) = \mu_{ia(j)}\bar{\mu}_{ib(j)} + \mu_{ib(j)}\bar{\mu}_{ia(j)}$. We now reason by absurdity. Suppose that for a certain neighborhood V of J we had the implication $\mu \in V \Rightarrow \det F(\mu) = 0$. Now $\det F(\mu) = 0$ defines an algebraic variety $\mathcal{N}$ and if $\mathcal{N} \not\subset \mathcal{M}$ we would have$\mathcal{N} \cap \mathcal{M}$ of codimension at least one in $\mathcal{M}$ since $\mathcal{M}$ is irreducible. But if $V \subset \mathcal{N}$ then $V \cap \mathcal{N} = V$ has codimension zero in $\mathcal{M}$ a contradiction. Thus $V \subset \mathcal{N} \Rightarrow \mathcal{M} \subset \mathcal{N}$. Hence to prove our lemma we need find one single extreme instrument with N^2 rank one atoms. This is the same as finding N^2 linearly independent positive rank one operators whose sum is 1. Assume we already have m operators $(\phi_i, \cdot)\phi_i$ where $m \geq N$ whose sum is 1. For $m = N$ we can take the ϕ_i to form an orthonormal basis. If $m < N^2$ there is a hermitian matrix A linearly independent of the $(\phi_i, \cdot)\phi_i$. Let now $\psi_1, \ldots, \psi_N$ be a basis of $\mathcal{H}$ made up of eigenvectors of A. If each $(\psi_i, \cdot)\psi_i$ were linearly dependent on the $(\phi_i, \cdot)\phi_i$ we can then write $A = \sum \alpha_i(\psi_i, \cdot)\psi_i$ as a linear combination of the $(\phi_i, \cdot)\phi_i$ contradicting the choice. Thus there is a ψ such that the $(\phi_i, \cdot)\phi_i$ and $(\psi, \cdot)\psi$ are all linearly independent. Since the ϕ_i span $\mathcal{H}$ we can write $\psi = \sum c_i\phi_i$. We now want to find numbers δ_i and σ such that $\sum(\phi_i + \delta_i\psi, \cdot)(\phi_i + \delta_i\psi) + \sigma(\psi, \cdot)\psi = 1$. Using the fact that $\sum(\phi_i, \cdot)\phi_i = 1$ we find that we must have

$$\sum \bar{\delta}_i(\psi, \cdot)\phi_i + \sum \delta_i(\phi_i, \cdot)\psi + (\sum |\delta_i|^2 + \sigma)(\psi, \cdot)\psi = 0.$$

Choosing now $\delta_i = \lambda c_i$ we must have $\lambda + \bar{\lambda} + |\lambda|^2 \sum |c_i|^2 + \sigma = 0$ which certainly has a solution for λ, given σ. We note that if σ is made sufficiently small,then λ can also be chosen to be arbitrarily small, but for λ sufficiently small the operators $(\phi_i + \lambda c_i\psi, \cdot)(\phi_i + \lambda c_i\psi)$ and are linearly independent. We have thus found $m + 1$ linearly independent positive rank one operators whose sum is one. Continuing by induction we can finally find N^2 such. We can now conclude that every neighborhood of J contains a pure instrument I. Q.E.D

With this, Theorem 7.7 is also proved.

We have not been able to calculate the instrumental entropy in $T_Q(\mathcal{H})$ due to the difficulty in identifying the pure instruments, although we suspect it also to be constant.

It should by now be fairly convincing that the notion of instrumental entropy is inadequate to give a reasonable justification for the quantum mechanical entropy formula $-\mathrm{Tr}(\rho \log \rho)$ on information theoretic grounds. This expression has more to do with the way a state is analyzed in terms of other states than with intrinsic measures of complexity. We shall not persue this train of thought to any great depth, but end this chapter by introducing a type of complexity that does reproduce the quantum mechanical formula, thus giving credence to the basic idea of this paragraph.

Let us consider representations of states as mixtures of other states: $\sigma = \sum \lambda_i\sigma_i$ suppose the σ_i are distinct. If one of the σ_i is not pure, say a $\sigma_1 = (1/2)\tau_1 + (1/2)\tau_2$, $\tau_1 \neq \tau_2$, then substituting this expression for σ_1 we obtain a more refined decomposition and the original expression would not contain the ultimate ingredients by which σ can be analyzed. Consider now finite dimensional algebraically closed state figures S. Any state can then be written as a mixture of pure states, and these correspond to non refinable decompositions. We define the *decomposition entropy* $D(\sigma)$ of σ as

$$D(\sigma) = \inf\{\sum \lambda_i H(\sigma_i) - \sum \lambda_i \log \lambda_i \,|\, \sigma = \sum \lambda_i \sigma_i;\ \sigma_i \text{ pure}\}.$$

The first term $\sum \lambda_i H(\sigma_i)$ is the mean complexity of the constituents σ_i of σ and the second term $-\sum \lambda_i \log \lambda_i$ is the complexity involved in decomposing into the constituents. The sum can be interpreted therefore as the information theoretic effort involved in understanding σ as a mixture of pure states. Since we want the most economic analysis, we must determine the infimum of the corresponding quantities. We must emphasize however that the decomposition entropy cannot be the best formalization of the information theoretic cost of analyzing a state in terms of others, seeing that it involves only the convex structure of the set $\mathcal{S}$ and the instrumental entropy $\mathcal{H}$, and ignores completely the operational structure of the theory, which by its very nature should have a role in the process of analysis. We compute now some decomposition entropies.

Theorem 7.8 *Given a Boolean triple based on* $\mathbb{B}_n$ *, and* $\sigma = (s_1, \ldots, s_n)$ *a state, then* $D(\sigma) = -\sum s_i \log s_i$*.*

Proof: The decomposition into pure states is unique and so we have $\sigma = \sum s_i \sigma_i$ where $\sigma_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the i-th place. Since $H(\sigma_i) = 0$ the result follows. Q.E.D

Theorem 7.9 *Consider any quantum mechanical theory in a finite dimensional Hilbert space in which the instrumental entropy is a constant* k*. If* ρ *is a state, then* $D(\rho) = -Tr(\rho \log \rho) + k$*.*

Proof: Consider a decomposition $\rho = \sum_{j=1}^{m} r_j(\phi_j, \cdot)\phi_j$ of ρ into pure states. We assume the ϕ_j are different, for if they were not, we could condense the probability measure $r.$ and lower the entropy. Assume none of the r_j are zero. Consider the two dimensional subspace K spanned by ϕ_1 and ϕ_2 and define the state $\rho_1 = (1/(r_1+r_2))(r_1(\phi_1, \cdot)\phi_1 + r_2(\phi_2, \cdot)\phi_2)$. If now $R = r_1 + r_2$ and if $\rho_1 = s_1(\psi_1, \cdot)\psi_1 + s_2(\psi_2, \cdot)\psi_2$ is any other decomposition of ρ_1 in K we have

$$\rho = Rs_1(\psi_1, \cdot)\psi_1 + Rs_2(\psi_2, \cdot)\psi_2 + \sum_{j=3}^{m} r_j(\phi_j, \cdot)\phi_j.$$

The entropy of this new decomposition is found to be

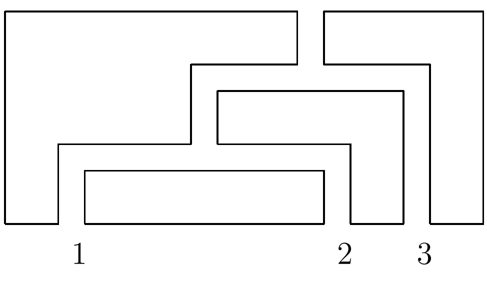
$$-R\log R - R(s_1\log s_1 + s_2\log s_2) - \sum_{j=3}^{m} r_j\log r_j + k.$$

Thus if the original entropy were minimum, the entropy of the given coefficients $(r_1/R, r_2/R)$ of ρ_1 would be minimum. We are thus lead to consider decompositions of minimum entropy of a state in $\mathbb{C}^2$ into two pure states. Now $\mathcal{S}(\mathbb{C}^2)$ is a solid sphere. Given a point in the interior @welve assumed $r_1 \neq 0 \neq r_2$ a decomposition into two pure states corresponds to a line segment through the point and meeting the surface of the sphere in the two pure states. The coefficients of the decomposition are the two ratios into which the line segment is divided by the given point, and the minimal entropy of these coefficients is obtained when the segment passes through the center of the sphere. This corresponds to a decomposition into two orthogonal states. We conclude therefore that the minimum entropy of decomposition is obtained when the ϕ_i are pairwise orthogonal. This corresponds to the case where the r_j are the nonzereo eigenvalues of ρ, counting multiplicities, and so the entropy of such a decomposiotion is $-\mathrm{Tr}(\rho\log\rho) + k$. Q.E.D

Chapter 8

On Lattices of Propositions in Statistical Theories

It is by now notorious that classical mechanics and quantum mechanics each possess well defined lattices of yes-no propositions. Within our formalism we see that in these two cases the propositions correspond to pure observations; characteristic functions for classical mechanics and orthogonal projections for quantum mechanics. The first form a Boolean algebra and the second an orthomodular lattice. One can ask then whether one can identify a lattice structure in a subset of $\mathcal{O}$ in a general statistical theory. We give here a satisfactory answer to this in the case of two dimensional theories, which also hints at the problems to be faced in the general case. Now the first point to establish is that we cannot expect the extreme points of $\mathcal{O}$ to correspond to a lattice of propositions. Consider the following hypothetical system: we take a maze as pictured below



and consider a hypothetical rat which has the following behavior. When placed in a maze and faced with its i-th T intersection it has a probability $(p_i, 1-p_i)$ of a right-left turn where for some θ with $-1 \leq \theta \leq 1$:

$$p_i = \begin{cases} \theta + (1-\theta)p_{i-1} & \text{if the previous turn was right} \\ (1+\theta)p_{i-1} - \theta & \text{if the previous turn was left.} \end{cases}$$

Hence each turn reinforces, positively or negatively, the subsequent turn. Suppose we have a population of these rats. Let $\sigma_{(\theta,\lambda)}$ be the state prepared as follows: place a rat of parameter θ and $p_1 = \lambda$ in the above maze. Let p^1, p^2, and p^3 be the observations of emergence at exits 1, 2, and 3 respectively. A simple calculation shows:

$$\begin{aligned} < \sigma_{(\theta,\lambda)}, p^1 > &= \lambda^2 + (1-\lambda)\theta, \\ < \sigma_{(\theta,\lambda)}, p^2 > &= \lambda(1-\lambda)(1-\theta), \\ < \sigma_{(\theta,\lambda)}, p^3 > &= 1-\lambda. \end{aligned}$$

Consider now a subpopulation with λ fixed. Seeing that the above expressions are affine in θ we see that the states $\sigma_{(\theta,\lambda)}$, λ fixed, lie on a segment of stats with the end points given by the extreme values of θ. We identify this segment with $\mathcal{S}$. Likewise $\mathcal{O}$ is a two dimensional observation figure constructed from p^1, p^2, and p^3 by logical operations and mixtures. We represent the system as a two dimensional statistical triple in the usual manner. Now unless $\lambda = 1$ or 0, p^3 is mixed, yet it should be natural to associate with this system the Boolean algebra of propositions generated by the atomic statements π_i: "the rat appeared at exit i." Mixed observations must therefore in general be considered as candidates for a logic of propositions. On the other hand not every mixed observation should do, only those should be considered that partake in measurements without *ad hoc* interferences by the experimenter. Our discussion of state complexity suggests we look to pure instruments. We note that in the maze example above (p^1, p^2, p^3) is the maximum pure instrument provided that θ varies in some interval of positive length. On the other hand it would still be too restrictive to consider merely atoms of pure instruments, since logical disjunctions of the properties associated to atoms correspond to condensations, and these disjunctions must certainly enter any putative logic of propositions. Let us tentatively call an observation $p \in \mathcal{O}$ a *statistical proposition* if there is a pure refinement of the instrument $(p, \mathbf{1} - p)$. Now even with this point of view, there is little chance of seeing a logic in the set of statistical propositions. To situate the problem better, we consider first a general two- dimensional triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$. In this case the pure instruments are filtered with respect to refinement. Furthermore if I is pure and refines a pure instrument J, there is a unique Boolean algebra monomorphism $\mathbb{B}_J \to \mathbb{B}_I$. We have an increasing filtered family of Boolean algebras and we can form the limit algebra $\mathbb{B}$ with canonical injections $\phi_I : \mathbb{B}_I \to \mathbb{B}$. Let now $(\mathcal{S}_{\mathbb{B}}, \mathcal{O}_{\mathbb{B}}, \langle\cdot,\cdot\rangle_{\mathbb{B}})$ be the Boolean triple associated to $\mathbb{B}$. By universality each $\sigma \in \mathcal{S}$ defines a probability measure on $\mathbb{B}$, thus there is a canonical injection $\phi : \mathcal{S} \to \mathcal{S}_{\mathbb{B}}$. We now prove that $\mathcal{O}$ is recoverable from $\mathcal{O}_{\mathbb{B}}$.

Theorem 8.1 $\phi^* \mathcal{O}_{\mathbb{B}} = \mathcal{O}$.

Proof: Since every element f of $\mathcal{O}_{\mathbb{B}}$ can be uniformly approximated by simple functions $h = \sum c_i \chi_{A_i} \in \mathcal{O}_{\mathbb{B}}$ and since $\mathcal{O}$ is closed in $\mathbb{R}^2$ and ϕ^* continuous, we need only show that these simple functions are mapped into $\mathcal{O}$ to conclude that $\phi^* \mathcal{O}_{\mathbb{B}} \subset \mathcal{O}$. We can assume the A_i disjoint and by construction there is an instrument I such that $A_i \in \phi_I \mathbb{B}_I$; $A_i = \phi_I(D_i)$ say. Since the A_i are disjoint, $0 \leq c_i \leq 1$, and $\phi^* h$ is a sum of atoms of I with coefficient c_i if the atom belongs to D_i and with coefficient 0 otherwise. Clearly this lies in $\mathcal{O}$ being a stochastic condensation of I. Thus $\phi^* \mathcal{O}_{\mathbb{B}} \subset \mathcal{O}$. On the other hand if $p \in \mathcal{O}$ is pure, the instrument $I = (p, \mathbf{1} - p)$ is pure and so $p = \phi^* \chi_A$ where $A = \phi_I(\{1\})$. Thus $\phi^* \mathcal{O}_{\mathbb{B}}$ contains every extreme point of $\mathcal{O}$. Since any point of $\mathcal{O}$ is a convex combination of at most three extreme points, we've proved the claim. Q.E.D

Thus as was mentioned in the previous chapter, we can view any two dimensional theory T as being a Boolean theory in which the production of states is somehow limited to a single interval $[\sigma_0, \sigma_1] \subset \mathcal{S}_{\mathbb{B}}$ and so $\mathcal{O}_{\mathbb{B}}$ collapses to its quotient making the reduces triple of $([\sigma_0, \sigma_1], \mathcal{O}_{\mathbb{B}}, \langle\cdot,\cdot\rangle)$ coincide with T. The Boolean algebra $\mathbb{B}$ should be considered as the lattice of propositions associated to T, but this lattice is not the set of statistical propositions as defined above. To see this, consider a two dimensional triple in which $\mathcal{O}$ is defined by its upper polygonal boundary formed by placing tail to head the following three vectors $p^1 = (0, 1/2)$, $p^2 = (1/2, 1/2)$, $p^3 = (1/2, 0)$. In this case $\mathbb{B} = \mathbb{B}_3$. Now p^1, p^2, p^3 are three statistical propositions of this theory, and let them correspond to atoms 1, 2, and 3 of $\mathbb{B}_3$. Now in $\mathbb{B}_3$, $1 \vee 2 = \{1, 2\} \neq 3$, but $p^1 + p^2 = p^3$ so we cannot claim that $\mathbb{B}_3$ is the Boolean algebra of statistical propositions since the map $\mathbb{B}_3 \to \mathcal{O}$ defining this condensations of the pure instrument of the upper polygonal boundary is not injective. on the other hand, this triple can be identified with the rat maze triple with $\lambda = 1/2$ and $-1 \leq \theta \leq 1$. Now $p^1 \vee p^2$ means the rat exited from 1 or 2, and p^3 means the rat exited from 3. Though $p^1 \vee p^2$ and p^3 are statistically equivalent, and thus define the same statistical proposition, they are distinct phenomenologically. We are thus faced with our old enemy of phenomenological distinctions being abolished by statistical identifications. It is only the phenomenological propositions that can be expected to form a lattice. Of course, since our theory formalizes only the statistical aspect of phenomena, there is no way of identifying these statistical coincidences; we can expect though to have a systematic way of suspecting them. The construction of the Boolean algebra $\mathbb{B}$ must therefore be viewed as the result of effecting a separation of all possible coincidences. It is with this idea in mind that we now proceed to study the general case.

Seeing that pure instruments correspond to those that cannot be interpreted as containing ad hoc interferences by the experimenter, we should expect its atoms to correspond to phenomenological propositions. Now a condensation of a pure instrument that corresponds to summing together some of the atoms, replaces the original propositions by logical disjunctions, and so must also correspond to phenomenological propositions. If a condensation is not pure though, it means that a logical construct starting from a measuring situation not interpretable as containing ad hoc interferences, is one that is so interpretable. This can now be viewed as a suspected statistical coincidence. We can try to remove it by introducing new hypothetical states in the following way. Let $\mathcal{A}$ be the set of $p \in \mathcal{O}$ which are atoms of pure instruments.

Definition 8.1 *An* instrument measure μ *of a statistical theory is a map* $\mu : \mathcal{A} \to [0, 1]$ *such that the following conditions are satisfied:*

1. $\mu(\mathbf{1}) = 1$,
2. *If* $J = (q_1, \ldots q_m)$ *is a condensation of* $I = (p_1, \ldots, p_n)$ *by means of a partial map* $\phi : \mathbf{n} \to \mathbf{m}$ *and both instruments are pure, then* $\mu(q_j) = \sum\{\mu(p_i) | \phi(i) = j\}$.

We note that every state $\sigma \in \mathcal{S}$defines an instrument measure μ_σ defined by $\mu_\sigma(p) = \langle\sigma, p\rangle$. Denoting by $\hat{\mathcal{S}}$ the set of instrument measures we have a natural inclusion $\mathcal{S} \subset \hat{\mathcal{S}}$. We want to think of $\hat{\mathcal{S}}$ as the set of states of a new statistical theory which incorporates new states but retains the same set of phenomenological propositions. Let us define now the new instrument set $\hat{\mathcal{I}}$. If $I = (p_1, \ldots, p_m) \in \mathcal{I}$ is a pure instrument we define the m-tuple $\hat{I} = (\hat{p}_1, \ldots, \hat{p}_m) \in (Conv(\hat{\mathcal{S}}, [0,1]))^m$ by setting $\hat{p}_i(\mu) = \mu(p_i)$. We define $\hat{\mathcal{I}}_n$ as the convex envelope of the set of n-atom condensations of m-tuples $\hat{I}$ as defined above. Thus $\hat{\mathcal{I}}$ is the instrument set obtained by maintaining the same set of phenomenological propositions arrayed in realizable instruments, but observing now the amplified set of states. In analogy with Theorem 8.1 we see that restricting $\hat{\mathcal{I}}$ to $\mathcal{S}$ we recover $\mathcal{I}$ at least in the the case when each $\mathcal{I}_n$ is finite dimensional. To introduce the new operation set $\hat{\mathcal{R}}$ is more problematic since it's not clear how a previously realizable operation should transform the new states, we thus introduce the largest possible set of operations that reduce to the old ones when applied to $\mathcal{S}$. First we define the sets $\hat{W}_n$ as being the set of all consistent n-tuples $(\hat{\theta}_1, \ldots, \hat{\theta}_n)$ of elements of $Conv(\hat{\mathcal{O}}, \hat{\mathcal{O}})$ satisfying Axiom 4.9$'$ and such that there is an n-exit operation $\theta \in \mathcal{R}$ with the property that given any n pure instruments $I_1, \ldots, I_n \in \mathcal{I}$ the restriction of $\hat{\theta}_1\hat{I}_1, \ldots, \hat{\theta}_n\hat{I}_n$ to $\mathcal{S}$ coincides with $\theta(I_1, \ldots, I_n)$. The set of operations $\hat{\mathcal{R}}$ is now defined as being generated from the $\hat{W}_n$ by means of the axioms of Chapter 4 as applied to the already existing sets $\hat{\mathcal{S}}$ and $\hat{\mathcal{I}}$. It is not clear whether to any $\theta \in \mathcal{R}_n$ a consistent n-tuple such as described above can be found. It's not hard to show that $\hat{T} = (\hat{\mathcal{S}}, \hat{\mathcal{I}}, \hat{\mathcal{R}})$ defines a statistical theory which we call the *state extension* of the theory $T = (\mathcal{S}, \mathcal{I}, \mathcal{R})$. Applied to a two dimensional triple $(\mathcal{S}, \mathcal{O}, \langle\cdot,\cdot\rangle)$ the state extension theory has as its triple the Boolean triple $(\mathcal{S}_{\mathbb{B}}, \mathcal{O}_{\mathbb{B}}, \langle\cdot,\cdot\rangle)$ constructed earlier.

Theorem 8.2 *For any statistical theory T we have $\hat{\hat{T}} = \hat{T}$.*

Proof: If $K = (k_1, \ldots, k_n)$ is a pure instrument of $\hat{T}$ then by definition of $\hat{I}$ it is either of the form $\hat{I}$ where $I \in \mathcal{I}$ is pure or it is a condensation of one of these. Now $\hat{I}$ itself is pure for if $\hat{I} = \sum \lambda_i \phi_{i*}\hat{I}_i$ where the I_i are pure and the ϕ_i define condensations, then this relation holds on $\mathcal{S} \subset \hat{\mathcal{S}}$, and since on these states, $\hat{I}$ and $\hat{I}_i$ coincide with I and I_i we would have $I = \sum \lambda_i \phi_{i*} I_i$. By purity of I we have $\phi_{i*}I_i = I$ and so $\phi_{i*}\hat{I}_i = \hat{I}$ and $\hat{I}$ is pure. Thus the pure instruments of $\hat{T}$ include $\hat{I}$, I pure in T. If now $\hat{\mu}$ is an instrument measure in $\hat{T}$, defining $\mu(p) = \hat{\mu}(\hat{p})$ for $p \in \mathcal{A}$ we see that μ is an instrument measure in T. Thus $\hat{\hat{\mathcal{S}}}$ can be identified with $\hat{S}$. Q.E.D

The state extension theory has a certain phenomenological content. Consider for example a theory that is adapted to describing the behavior of a certain type of organism within an enclosed environment. The organism along with its environment can be viewed as a copy of a physical state defined simply by a certain interval of energy. This new set of energetically defined states include now besides the original organism-environment combinations also crystals of iron, gamma radiation, beach sand, pea soup, molten sulphur, other types of organisms, etc. If we maintain the same set of phenomenological propositions as arrayed in a set of realizable instruments, and observe the new states, these provide us with instrument measures on the set of pure instruments. Thus in the reduced system corresponding to the enlarged set of physical states, the set of states is a subset of $\hat{S}$. The new states in $\hat{T}$ can be looked upon as resulting from removing any restrictions existing in the preparation procedures that would lead to the creation of a certain type of state of affairs. They correspond to any conceivable preparation procedures, and so $\hat{T}$ embodies the conceivable universe as seen by a set of propositions originally designed to study only a part of the real one. Theories for which $\hat{T} = T$ we call *state complete.* They do not contain any statistical coincidences resulting from merely restrictions on state production.

Theorem 8.3 *Any Boolean triple is a state complete theory.*

Proof: The pure instruments are of the form $(\chi_{A_1}, \ldots, \chi_{A_n})$ with A_i disjoint open-closed sets. Thus $\mathcal{A}$ can be identified with $\mathbb{B}$ itself. The condensation Axiom (2) of Definition 8.1 means that if μ is an instrument measure and $A = A_1 \cup \cdots \cup A_k$, $A_i \in \mathbb{B}$, A_i disjoint, then $\mu(A) = \sum \mu(A_i)$ which is to say that μ is a finitely additive measure in $\mathbb{B}$ and since **1** is identified to \$ in B we have $\mu(\$) = 1$ by Axiom (1) of Definition 8.1 and thus μ is in fact a state of the Boolean triple and $\hat{S} = S$. Q.E.D

Theorem 8.4 *If $\mathcal{H}$ is a Hilbert space with $3 \leq \dim \mathcal{H} = N < \infty$ then $T_c(\mathcal{H})$ is state complete.*

Proof: This is an immediate corollary of Theorem 7.12 and Gleason's theorem [11]. Q.E.D

State complete theories are a natural starting point for general statements about lattices of propositions, since these no longer exhibit certain types of coincidences. Other coincidences could conceivably still exist, leading still to impure condensations of pure instruments. This is a totally unexplored area.

Concerning the phenomenal world, we can now ask the following questions:

1. Does the phenomenal universe, as viewed by a realizable set of instruments, give rise to a state complete theory; in other words, is the real world statistically indistinguishable from the conceivable world?

2. In the state extension of any theory adapted to viewing the phenomenal universe, do there still exist pure instruments with impure condensations? If so this would mean that there are certain statistical laws of logic operating in the real world that up to now haven't been conceived.

Both Boolean triples and $T_c(\mathcal{H})$ come fairly close to being theories that view the phenomenal universe as a whole, each of course with its particular set of instruments. Both of these are state complete and with no impure condensations of pure instruments. Better candidates for real world theories would be $T_I(\mathcal{H})$ and even more so $T_Q(\mathcal{H})$. The determination of their state extensions would be quite illuminating.

Bibliography

[1] George W. Mackey "Mathematical Foundations of Quantum Mechanics" W. A. Benjamin Inc., New York, 1963.

[2] C. Piron "Foundations of Quantum Physics" W. A. Benjamin Inc., New York, 1967.

[3] G. Ludwig "Deutung des Begriffs 'Physicalische Theorie' und Axiomatische Grundlegung der Hilbertsraumstructur der Quantunmeckanik durch Hauptsaetze des Messens" Lecture Notes in Physics, Vol 4, Springer Verlag, Berlin, 1970.

[4] Carlos Malamut "Estruturas Algébricas de Teorias Estatísticas" Master's Thesis, Pontifícia Universidade Católica, Rio de Janeiro, 1970.

[5] R. Giles, "Foundation for Quantum Mechanics" J. Math. Phys., **11**, 2139 – 2160.

[6] Roman Sikorski "Boolean Algebras", Springer Verlag, Berlin, 1964.

[7] Harry H. Harmann "Modern Factor Analysis", 2^{nd} edition, University of Chicago Press, Chicago, 1970.

[8] Saunders MacLane "Categories for the Working Mathematician", Springer Verlag, Berlin, 1971.

[9] C. A. Hooker, editor, "The Logico-Algebraic Approach to Quantum Mechanics" Vol. I, D. Reidel Publiching Co., Boston, 1975.

[10] N. I. Akhiezer and I. M. Glazman "The Theory of Linear operators on Hilbert Space" Vol. 2, Ungar, New York, 1966.

[11] A. M. Gleason "Measures on the Closed Subspaces of a Hilbert Space" Journal of Mathematics and Mechanics, **6** (1957), 885-893. Appears also reprinted in [9].

Index