QUALITATIVE PROBABILITIES REVISITED

ABSTRACT. Granted a de Finetti style qualitative comparative probability relation on a Boolean algebra, necessary and sufficient conditions are given for the existence of an agreeing probability measure on the algebra in finite, countable and arbitrary situations. Partially ordered linear spaces and order-preserving linear functionals are used in proving the results and in explaining why the axiomatization of qualitative probability relations is bound to be complex. The inherent technical difficulties can be overcome by relying on nonstandard representations that are also provided. Extensive work done by Suppes in this area is also discussed, in conjunction with the problem of uniqueness and simplicity. The central aim of this work is to provide a more holistic setting for the axiomatization of comparative probability and its associated representational methodology.

1. INTRODUCTION

The principal object of this paper is to provide axioms for a de Finetti style qualitative probability relation, necessary and sufficient for the existence of an agreeing probability measure in arbitrary Boolean algebras. I take this contribution to be a continuation of the works of Scott (1964), Suppes (1969), and Suppes and Zanotti (1976), mainly in bringing development of the theory of comparative belief more nearly abreast with that of representational measurement theory. In the course of expounding the necessary probabilistic apparatus, I comment amply on what I take to be Suppes’s pivotal ideas and results in comparison with related literature.

In 1964 Dana Scott published his widely known paper on finite measurement models, in which he applies vector space separation theorems to great effect – to obtain several measurement representation results in general, and to characterize finite qualitative probability structures in particular.

My starting point is the familiar ‘Scott condition’ for qualitative probabilities in finite Boolean algebras, which was quickly included also in Suppes (1967) for its remarkable power and simplicity. But first let me state briefly how I became aware of the issues surrounding the ‘measurement’ of subjective probability.
I owe a primary dept to my teacher, Patrick Suppes, who above all brought me to Stanford in the mid 1960s and taught me how to speak and think in ‘measurementese’. I recall most vividly his persistent and gentle prodding in the course of my laughable attempts to find what he called “simple and powerful” measurement theoretic axioms that result in “important representation theorems”. During my student years, Suppes’s dictum had become so important to many of us that my budget sharing house mate – a well known philosopher and frugal cook, who attended Suppes’s intellectually contagious seminars around that time – refused to offer any food on days that had passed without my proving a representation theorem. Those around us, at the time, knew well why I ate so little.

I recall, too, that it was Patrick Suppes who drew my attention to Scott’s work, whose significance he had quickly appreciated. Some years earlier, they wrote a joint paper on the nature of measurement theories, and of course Suppes was interested in measurement theory at least as early as 1951, as evidenced by his lesser known paper on axioms for extensive quantities, reprinted in Suppes (1969).

I have had the great fortune of being a student of these inspiring teachers. As a student and later as a frequent visitor at Stanford, I almost always found the door to Suppes’s Ventura Hall office open. In spite of the waiting pile of work each day – “today I’m snowed under”, he used to say apologetically – Suppes gave me generously of his time. I am tremendously grateful to him.

As many of us are aware, Patrick Suppes has contributed very significantly to a number of areas in measurement theory in general, and measuring subjective probability in particular. I find his papers and books always striking, not only for their importance, but also for the elegance and clarity with which they are written. As far as I can make out, in his work in measurement theory he is driven to find the simplest axioms and the right definitions that enable him to prove almost effortlessly various results under very few assumptions.

In this essay, I look at a family of those representation theorems for qualitative probabilities that share the Scott condition and can be derived from various convex set separation theorems or linear functional existence theorems. I begin with the simplest, i.e., Suppes’s preferred finite situation, and then gradually work my way to countable and arbitrary (infinite) qualitative probability spaces. At the end of the paper I introduce an ordered linear space framework that in my view explains why
representation of qualitative probabilities in infinite domains is bound to be complex. As we shall see, one reason why the decisive axioms for general qualitative probability relations tend to be messy is that they provide existence criteria for order-preserving linear functionals (and hence indirectly also for finitely additive measures) in a very roundabout way – the direct way being normed topological axioms.

It appears that no satisfying account has yet been given of the representation of comparative probabilities in infinite Boolean algebras. In particular, I take it to remain an unfinished task to provide simple axioms for a qualitative probability relation in an infinite algebra of events, necessary and sufficient for the existence of an agreeing finitely additive probability measure. Since the situation does not seem to be getting any better in this respect, and since there is no point in having a comparative belief that nobody seems to know how to measure simply, I take this opportunity to present a list of necessary and sufficient conditions that I managed to extract so far, and let others judge whether I was able to live up to Suppes's celebrated dictum of simplicity.

Also, I will briefly comment on Scott's (old) unpublished results, summarized in Domotor (1969). It is always a pleasure to acknowledge the originality of his ideas. Along the way, I state a couple of simple facts about the uniqueness of probabilistic representation.

2. MEASURING BELIEF: SOME DEFINITIONS AND PRELIMINARY OBSERVATIONS

The measurement-theoretic viewpoint of 'subjective' probability has become standard after the mid fifties. It was pioneered by Savage (1954) and popularized by Suppes (1956) who, anticipating an important trend in the development of the foundations of probability (and utility), was instrumental in bringing measurement and probability theories even closer together. Indeed, measurement of rational belief was looked at by Suppes as a special case of extensive measurement.

Piecing together incompatible events was more like adding physical masses, and comparing the probabilities of occurrence of these events looked very much like comparing objects with respect to their masses. This sort of physical analogy is successfully exploited in Suppes (1969, introductory pages), where uniqueness is the principal bonus. For a systematic codification of the flurry of subsequent research activity in this area, see Krantz, Luce, Suppes and Tversky (1971).
Measurement theory of rational belief can be seen as part of a larger theory of observation. Here we shall confine ourselves to granting that an observer reasons about the objects of his or her chancy world and parts thereof in terms of

- possible, mutually exclusive, and collectively exhaustive outcomes of an experiment, intended to be performed on the object of interest;
- observable events, associated with the experiment, whose occurrence depends exclusively on the experimental outcomes;
- comparative judgment of the probability of occurrence of events.

In accord with the standard trick of modern science and relative to the extant stage of our empirical knowledge, we proceed to attach to these carefully distinguished epistemic items suitable mathematical objects. In particular, the totality of possible outcomes is usually parametrized by a nonempty set \( \Omega \), called the sample space, that the observer regularly revises in face of new information.

Relative to the idea of occurrence, events are forced to band together so as to form a logic-like algebra. In the case of classical experiments it has proven adequate to parametrize the realm of observable events by a Boolean algebra \( \mathcal{A} \) of subsets of the sample space \( \Omega \), called the event algebra. So certain subsets of \( \Omega \) are singled out as parametrizers of particular events, and we make sure that all empirically meaningful combinations of events fit the Boolean laws of subsets of \( \Omega \). Should the nature of complementary events be called into question, as is the case in constructivist situations, then the event algebra \( \mathcal{A} \) will have to be revised into something considerably more abstract and general, e.g., an intuitionistic Heyting algebra (of open sets of \( \Omega \)) – the ‘logic’ of finite observations. Now if the distributivity law fails in an instance of conjunctively combined events, as is the case in quantum experiments, then the event algebra \( \mathcal{A} \) is revised again. For example, Suppes (1966) suggests using Dynkin algebras instead of Boolean algebras. These are nonempty families of subsets of the sample space \( \Omega \), closed under complementation and disjoint unions of events.

In the world of classical systems and the one to be considered in this paper, the collection of observable events forms a Boolean algebra. Along traditional lines, the attached parametrizing space-algebra pair \( \langle \Omega, \mathcal{A} \rangle \) is called a Boolean space. We say that a Boolean space is finite just in case its algebra of events is finite. Likewise, a Boolean space is called countable iff its Boolean algebra is countable. Since
not every Boolean algebra can carry a strictly positive and finitely additive probability measure, and since some Boolean algebras are more complete than others, these distinctions will be most instrumental in settling the strength of axioms for qualitative probabilities.

Our notation for set-theoretic and probabilistic concepts is standard. We use \( \emptyset \) to refer to the impossible event (zero element), and we write \( \Omega \) to designate the sure event (unit element) in the event algebra \( \mathfrak{A} \). Finally, the event calculus comes with its customary Boolean operations, including complementation \( \overline{A} \), union \( A \cup B \), disjoint union \( A + B \), intersection \( A \cap B \), and the Boolean subevent relation \( A \subseteq B \) for all events \( A \) and \( B \) in \( \mathfrak{A} \).

With regard to the observer, the assumption is that the judgments of probability are representable in terms of a single binary relation \( \leq \) between events in \( \mathfrak{A} \), called the observer's comparative or qualitative probability. So for a pair of events \( A \) and \( B \) in \( \mathfrak{A} \), we write

\[ A \leq B \]

to mean that event \( A \) is judged by the observer as at least as probable as the event \( B \).

The list of derived notions we shall need includes

1. **equiprobability:**

\[ A \sim B =_{df} A \leq B \land B \leq A, \]

intended to mean that events \( A \) and \( B \) are judged to be equally probable: \( A \) is as probable as \( B \).

2. **strict probability:**

\[ A < B =_{df} A \leq B \land \text{not } A \sim B, \]

intended to mean that \( A \) is judged to be (strictly) less probable than \( B \).

By way of a familiar mathematical representation trick, we have now arrived at a parametrizing qualitative probability structure \( \langle \Omega, \mathfrak{A}, \leq \rangle \) that has a mathematical life of its own. In particular, we can begin to apply the rich resources of representational measurement theory to it, to look for conditions constraining \( \leq \) that guarantee an agreeing Kolmogorovian probabilistic representation, and so forth.

Moving in the reverse direction, we can search for new empirical or epistemological applications of the freshly isolated mathematical object \( \langle \Omega, \mathfrak{A}, \leq \rangle \).
Here I wish to proceed in the forward direction and return to the measurement theoretic problem of representation of qualitative probability structures.

I begin, as a way of entering the core of our subject, by specifying a minimal list of conditions that every classical comparative probability relation ought to satisfy. The list is well known and is usually attributed to Bruno de Finetti, although there are traces of it also in Bernstein (1917, pp. 10–25). So we state it as a definition, using de Finetti’s name as part of our terminology.

**DEFINITION 1.** A mathematical object \( \langle \Omega, \mathcal{A}, \preceq \rangle \) is called a *de Finetti space* if and only if

1. \( \langle \Omega, \mathcal{A} \rangle \) is a Boolean space;
2. \( \preceq \) is a binary comparative probability relation on \( \mathcal{A} \), satisfying the following axioms for all \( A, B, C, \) and \( D \) in \( \mathcal{A} \):
   - (i) \( \emptyset \preceq \Omega \);
   - (ii) \( \emptyset \preceq A \);
   - (iii) \( [A \preceq B \land B \preceq C] \implies A \preceq C \);
   - (iv) \( A \preceq B \lor B \preceq A \);
   - (v) \( [A \preceq B \land C \preceq D] \implies A + C \preceq B + D \), given that the event pairs \((A, C)\) and \((B, D)\) are disjoint;
   - (vi) \( A \preceq B \iff A + C \preceq B + C \), where the pairs \((A, C)\) and \((B, C)\) are disjoint.

On the question of terminology, in this paper I will use the term ‘de Finetti space’ throughout. De Finetti spaces encapsulate the essence of rationality, exhibited in traditional judgments of comparative belief. Leaving the problem of justification of de Finetti’s criteria for rational belief aside, first we must ask whether his comparative belief can be measured. That is, we wish to know whether for every de Finetti space \( \langle \Omega, \mathcal{A}, \preceq \rangle \) there exists an agreeing finitely additive probability measure \( P \) on \( \langle \Omega, \mathcal{A} \rangle \) that ‘measures’ or realizes the given comparative belief \( \preceq \). Here ‘agreeing’ refers to the following representation condition:

\[
\forall A, B \in \mathcal{A}[A \preceq B \iff P(A) \leq P(B)].
\]

We know from Savage (1954) that infinite de Finetti spaces do not come with guaranteed agreeing measures in general. Furthermore, we know from the work of Kraft, Pratt, and Seidenberg (1959) that finite de Finetti
spaces are equally deficient. Simply, in general, de Finetti’s rationality constraints governing $\preceq$ are not sufficient for guaranteeing the existence of an agreeing finitely additive probability measure on $\langle \Omega, \mathcal{A} \rangle$. This shortcoming of de Finetti spaces preempts all the other questions that we might ask otherwise (e.g., is transitivity of the equiprobability relation $\sim$ justified?).

Lack of measurement seems to make the very notion of a de Finetti space useless. Nevertheless, we shall continue to use it for convenience of expression: it will serve well as the shorthand for the familiar and longer list of necessary conditions (nontriviality, nonnegativity, transitivity, comparability, weak and strong monotonicity) that we plan to impose on all qualitative probability relations $\preceq$, studied in this paper.

It is now clear that the measurement program for qualitative probability with which we began remains unfulfilled, if we have no guarantees for the measurement of $\preceq$ with an agreeing probability. We must therefore seek additional empirically justifiable constraints, to be imposed on $\langle \Omega, \mathcal{A}, \preceq \rangle$, sufficient for the existence of an agreeing measure on $\langle \Omega, \mathcal{A} \rangle$. To get much further with the theory of qualitative probability spaces, we must introduce more axioms about the ways in which qualitative probability can be manipulated.

Conceptually there are three fundamentally different ways in which this can be carried out and, as we shall soon see, all of them have been used in the literature in one way or another:

- Constraints formulated in a higher-order language of Boolean spaces $\langle \Omega, \mathcal{A} \rangle$ (including finiteness and sigma-additivity);
- Axioms stated in a second-order language of comparative probability $\preceq$ (including Archimedeaness and continuity);
- Conditions expressed in a suitably enlarged language of random variables, ordered by a comparative expectation relation that extends $\preceq$.

Having set the stage, we now turn to presenting various representation theorems, first in finite, then in countable and finally in arbitrary Boolean spaces.

3. QUALITATIVE PROBABILITY IN FINITE BOOLEAN SPACES

The present section is devoted to a detailed discussion of the nature and basic consequences of the familiar Scott condition:
SCOTT'S CONDITION. For any pair of finite sequences $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ in $\mathfrak{A}$ we have

$$\forall i < n [A_i \leq B_i] \implies B_n \leq A_n,$$

given that

$$\bigcup_{1 \leq i_1 < \cdots < i_k \leq n} A_{i_1} \cap \cdots \cap A_{i_k} = \bigcup_{1 \leq i_1 < \cdots < i_k \leq n} B_{i_1} \cap \cdots \cap B_{i_k}$$

holds for all $k$ with $1 \leq k \leq n$.

Notation. For any de Finetti space $\langle \Omega, \mathfrak{A}, \leq \rangle$ we write

$$\$(\Omega, \mathfrak{A}, \leq)$$

as an abbreviation for the foregoing Scott condition, interpreted in and applied to the de Finetti space $\langle \Omega, \mathfrak{A}, \leq \rangle$. Thus all sample space and event algebra variables are presumed to take their values in the de Finetti space $\langle \Omega, \mathfrak{A}, \leq \rangle$.

Readers familiar with Scott's paper will quickly note that the condition stated above is different from that in Scott (1964). Dana Scott uses an equational constraint

$$\sum_{1 \leq i \leq n} \hat{A}_i = \sum_{1 \leq i \leq n} \hat{B}_i,$$

involving characteristic functions $\hat{A}_i$ of events $A_i$. We know, of course, that every event $A \in \mathfrak{A}$ comes with its unique two-valued characteristic function $\hat{A} : \Omega \rightarrow \mathbb{R}$, where $\hat{A}(\omega) = \text{df} 1$ if $\omega \in \Omega$ and 0 otherwise.

Using mathematical induction, one quickly verifies that the condition I used above is indeed equivalent to that of Scott. All one has to realize is that for any $\omega \in \Omega$ we have $\sum_{1 \leq i \leq n} \hat{A}_i(\omega) \geq k$ exactly when $\omega$ belongs to the union

$$\bigcup_{i \leq i_1 < \cdots < i_k \leq n} A_{i_1} \cap \cdots \cap A_{i_k}.$$

This formula is easily seen to hold -- e.g., by induction. And there is more. In the same way as above, we quickly verify that the following
more general correspondence also holds:

**LEMMA 1.** For any pair of finite sequences of events we have

\[
\sum_{1 \leq i \leq n} \hat{A}_i \leq \sum_{1 \leq j \leq m} \hat{B}_j \iff \bigcup_{1 \leq i_1 < \cdots < i_k \leq n} A_{i_1} \cap \cdots \cap A_{i_k} \subseteq \bigcup_{1 \leq j_1 < \cdots < j_k \leq m} B_{j_1} \cap \cdots \cap B_{j_k}.
\]

Hence all finitary inequalities on all finitary combinations of characteristic functions of events are noncreative in that they are always reducible to suitable Boolean equations between events.

While the equation involving sums of characteristic functions is notationally and conceptually convenient, proofs are often easier to state and follow in terms of equations between unions of events.

Fine (1973) and others complain about Scott’s condition, because, for one, it is not stated solely within the language of de Finetti spaces. My formulation takes care of this objection. More importantly, the Boolean variant above explains why Scott’s *equational* constraint is necessary.

Using Kolmogorov’s axioms for a finitely additive probability measure \( P \) on \( (\Omega, \mathcal{A}) \), it is routine to check that for any finite string \( A_1, \ldots, A_n \) of events in \( \mathcal{A} \) we have

\[
\sum_{1 \leq i \leq n} P(A_i) = \sum_{1 \leq k \leq n} P \left( \bigcup_{1 \leq i_1 < \cdots < i_k \leq n} A_{i_1} \cap \cdots \cap A_{i_k} \right),
\]

a smooth generalization of the familiar \( P(A) + P(B) = P(A \cup B) + P(A \cap B) \).

It is equally straightforward to see that we have the equality

\[
\sum_{1 \leq i \leq n} P(A_i) = \sum_{1 \leq i \leq n} P(B_i)
\]

whenever \( \sum_{1 \leq i \leq n} \hat{A}_i = \sum_{1 \leq i \leq n} \hat{B}_i \) holds for all events belonging to the event algebra \( \mathcal{A} \). The claim remains valid even if the occurrences of equality in the premise and conclusion are replaced with occurrences of weak inequality.

Now in proving the necessity of Scott’s condition in any de Finetti space, e.g., in the instance of a probability measure \( P \), we tacitly presume
that the vectorial equality \( \sum_{1 \leq i \leq n} \hat{A}_i = \sum_{1 \leq i \leq n} \hat{B}_i \) implies the numerical equality \( \sum_{1 \leq i \leq n} P(A_i) = \sum_{1 \leq i \leq n} P(B_i) \). The implicit assumption works precisely because the equality involving characteristic functions is necessary for being able to prove a fundamental correspondence between finitely additive probability measures on a Boolean space and linear functionals (expectation integrals) on the associated linear space of characteristic functions. We of course need linear space constructs for formulating existence criteria for linear functionals.

It is fair to say that for finite de Finetti spaces, the simplest and most complete result is that of Scott (and independently of Ernest Adams):

**THEOREM 1.** For any finite de Finetti space \( \langle \Omega, \mathcal{A}, \preceq \rangle \) there exists an agreeing probability measure \( P \) on \( \mathcal{A} \), i.e., the following representation condition

\[
\forall A, B \in \mathcal{A} [A \preceq B \iff P(A) \leq P(B)]
\]

holds if and only if the Scott condition \( $\langle \Omega, \mathcal{A}, \preceq \rangle $ \) is satisfied.

The existence proof is in Scott (1964).

**Geometric Meaning of Suppes's Uniqueness**

Suppes (1969, pp. 5–8) provides a deceptively simple list of axioms for a finite de Finetti space that is sufficient for the existence of a unique agreeing probability measure. The pivotal axiom \( A \preceq B \Rightarrow \exists C[B \sim A \cup C] \) is used also in Suppes (1951) and Behrend (1956), albeit for different purposes and in different disguises.

Since the converse of Suppes’s axiom is a theorem, the constraint amounts to reducing the comparative probability relation to an equiprobability. What is not noted is that Suppes’s uniqueness is a consequence of his constraining a comparative probability so that geometrically, its agreeing measure is bound to be a distinguished interior point of the simplex \( P(\Omega, \mathcal{A}) \) of all probability measures, namely its barycentre – a probabilistic counterpart of the center of gravity. That is, the centered measure whose barycentric coordinates are equal to \( 1/n \), where \( n \) is the number of atoms in the event algebra, not equivalent to zero. In general, there are many barycentres in a simplex of probability measures, and Suppes’s axiom picks just one of them – notably, the simplest. Other
barycentres may be obtained by limiting the combination of equivalence constraints $A \sim B$ to proper subsets of atoms.

Before stating my position on uniqueness, I need one more piece of notation and terminology.

Routine calculation quickly reveals that whenever two probability measures $P$ and $Q$ on $\langle \Omega, \mathcal{A} \rangle$ satisfy the representation condition for $\preceq$ in the theorem, then so does their convex mixture $P^+ P \preceq Q$ for all reals $0 \leq a \leq 1$.

Recall that a convex mixture of two probability measures $P$ and $Q$ in proportions $a : (1 - a)$ is again a probability measure, defined argumentwise as follows:

$$ (P^+ P \preceq Q)(B) = aP(B) + (1 - a)Q(B) $$

for all events $B \in \mathcal{A}$. Convex mixing satisfies the familiar von Neumann axioms for mixture spaces and it generalizes smoothly to any finite or countable number of mixands. We shall use these mixing operations repeatedly also within the context of random variables. A set of probability measures is convex iff it is closed under the above defined convex mixture operation.

So an agreeing probability measure for $\preceq$ is unique, modulo a convex set of measures. These convex sets are characterized by the barycentric subdivision of order 1 of the convex space $\mathbf{P}(\Omega, \mathcal{A})$ of all probability measures. We know that in a finite situation space $\mathbf{P}(\Omega, \mathcal{A})$ is a simplex, i.e., a common generalization of a line segment, equilateral triangle, and regular tetrahedron. Its barycentres are given by the totality of all uniform convex mixture probability measures of the form $1/nP_1 + \cdots + 1/nP_n$, where $P_1, \ldots, P_n$ are strings of 0–1-valued Dirac probability measures—vertices of a unique face (i.e., subsimplex) in $\mathbf{P}(\Omega, \mathcal{A})$. Geometrically, the barycentre of a line segment is its midpoint, and that of a triangle is its center point.

Thus in the space of all probability measures of an algebra, there are as many barycentres as there are faces or subsimplexes, including the zero-dimensional ones. Now if we view the barycentres of all faces of $\mathbf{P}(\Omega, \mathcal{A})$ as new vertices, we obtain a refined family of subsimplexes, called the barycentric subdivision of space $\mathbf{P}(\Omega, \mathcal{A})$. Some of these newly defined faces of probability measures in the subdivision are singletons, comprised by barycentres. The set of all newly formed faces is in one-to-one correspondence with the set of representable qualitative probability relations.
It is easy to verify that the set of faces arising from the foregoing subdivision, induced by barycentres, is also isomorphic to the set \( P(\Omega, \mathcal{A})/\equiv \) of equivalence classes of measures, where the abstracting equivalence relation \( \equiv \) is defined by

\[
P \equiv Q \iff \forall A, B \in \mathcal{A}[P(A) \leq P(B) \iff Q(A) \leq Q(B)].
\]

Upon confining ourselves to atomic events, we see at once that the equivalence relation \( \equiv \), applied to the representation provided by Suppes, gives a singleton.

In my opinion – and I presume Suppes agrees – lack of representational uniqueness is a virtue of qualitative probabilities, rather than a defect. In our daily comparative judgments of occurrence of events (earthquakes, weather, stock market decline and such), if asked, we do not assume we are able to routinely offer our distinguished agreeing personal probability measure. What we can give are bets and estimates that realistically do not amount to the uniqueness of such a measure.

In addition to the Laplace-style uniqueness of Suppes that geometrically corresponds to barycentric measures, there is uniqueness that can be obtained by a barycentric subdivision of order 2. Simply, each convex set of measures that represents a particular qualitative probability relation has its own unique barycentre or local center of gravity, calling for the next round of facial subdivision. Although these second-order barycentric probabilities are not definable in terms of comparative probability relations, they can be singled out by various maximum entropy principles.

Returning to qualitative probability relations, a Scott-style condition applies also to the case of equiprobability \( \sim \) as the only qualitative probabilistic primitive. (Recall that for any pair of events \( A \) and \( B \), the intended interpretation of \( A \sim B \) is: events \( A \) and \( B \) are judged as equally probable.) Equiprobability relations are closely related to Suppes's barycentric comparative probabilities. The following theorem characterizes their representability:

**THEOREM 2.** Let \( \langle \Omega, \mathcal{A} \rangle \) be a finite Boolean space, impressed with a binary qualitative equiprobability relation \( \sim \). Then for all event sequences \( A, A_1, \ldots, A_n \) and \( B, B_1, \ldots, B_n \) in \( \mathcal{A} \) the conditions:

(i) \( \not \emptyset \sim \Omega \);
(ii) \( A \sim A \);
(iii) \( A \sim B \implies B \sim A; \)

(iv) \( \forall i \leq n[A_i \sim B_i] \implies (A \sim B \iff \forall j \leq m[C_j \sim 0]), \)

where \( k(\hat{A} - \hat{B}) = \sum_{i<j \leq m} \hat{C}_j + \sum_{1 \leq i \leq n} (\hat{A}_i - \hat{B}_i) \) with some \( k \geq 1, \)

are necessary and sufficient for the existence of a probability measure \( P \) on \((\Omega, \mathfrak{A})\) such that

\[
\forall A, B \in \mathfrak{A}[A \sim B \iff P(A) = P(B)].
\]

**Proof.** The proof is a special case of Theorem 5 in the next section, formulated for countable Boolean spaces. An alternative proof can be got from Scott (1964).

Axiom (iv) is a finitary version of the norm condition

\[
\| \pm (\hat{A} - \hat{B}) + C(\leq) + C(\sim) \| = 0 \implies A \sim B
\]

of Theorem 10, Section 6.

If \((\Omega, \mathfrak{A})\) is finite, then the premise in the norm condition of Theorem 10 says that

\[
\hat{A} - \hat{B} = \sum_{i \leq j \leq m} b_j \hat{C}_j + \sum_{1 \leq i \leq n} a_i(\hat{A}_i - \hat{B}_i)
\]

with some \( A_i \sim B_i, B_j \sim O \) and \( a_i \geq O \) for all \( i = 1, \ldots, n \) and \( j = 1, \ldots, m. \)

Upon holding the events fixed for a moment and replacing the vectors with their numerical values for all \( \omega \in \Omega, \) and looking at the coefficients \( a_i \) as unknowns, we obtain a list of first order linear algebraic equations with integer parameters. Since the equations have a solution in the field of rationals, we may as well assume that all coefficients \( a_i \) are rational. By clearing fractions and allowing repetitions in the sum, we arrive at the desired equation in (iv). To enhance intuitiveness, I stated the conclusion in the form of an equivalence. The necessity is obvious.

If we enrich the list of conditions in Theorem 2 with an extra constraint, stating that most atoms are equiprobable, then Suppes's uniqueness result follows, and the axioms - save the constraint - are both necessary and sufficient.

Note that the usual monotonicity and transitivity properties of \( \sim \) trivially follow from the Scott-style condition. In particular, we have the following corollary.
COROLLARY 1. The following clauses hold for all events $A$, $B$, $C$, $D$, $A_i$ and $B_i$ in a finite Boolean space with an equiprobability relation $\sim$ that satisfies the axioms of Theorem 2:

(i) $\forall i \leq n [A_i \sim B_i] \implies A \sim B$,

where $k(\hat{A} - \hat{B}) = \sum_{1 \leq i \leq n} (\hat{A}_i - \hat{B}_i)$;

(ii) $\forall i \leq n [A_i \sim B_i] \implies \forall j \leq m [C_j \sim \emptyset]$,

where $\sum_{1 \leq j \leq m} (\hat{C}_j + \sum_{1 \leq i \leq n} (\hat{A}_i - \hat{B}_i) = 0$;

(iii) $[A \sim A \wedge B] \implies A \sim B$;

(iv) $[A_1 \sim A_2 \sim A_3 \wedge B_1 \sim B_2 \sim B_3] \implies A_3 \sim B_1$,

where $A_1 + A_2 + A_3 = \Omega = B_1 + B_2 + B_3$ are partitions of the sample space;

(v) $A \sim B \implies \hat{A} \sim \hat{B}$;

(vi) $A \sim \hat{A} \implies \text{not } A \sim \emptyset$;

(vii) $A \sim B \iff A + C \sim B + C$,

where the pairs $(A, C)$ and $(B, C)$ are disjoint;

(viii) $[A \sim B \wedge C \sim D] \implies A + C \sim B + C$,

where the pairs $(A, C)$ and $(B, D)$ are disjoint;

(ix) $A \sim \emptyset \implies A \cup B \sim B$.

Proof. The first two clauses are easily seen to be special cases of the Scott-style condition.

The third conditional is based on the equation

$$2(\hat{A} - \hat{B}) = (\hat{A} - \hat{A}) + (\hat{B} - \hat{B})$$

and the given pair of premises.

Similarly, the conclusion of the fourth item follows immediately from the equation

$$3(\hat{B}_1 - \hat{A}_3) = (\hat{A}_1 - \hat{A}_3) + (\hat{A}_2 - \hat{A}_3)$$

$$+ (\hat{B}_1 - \hat{B}_2) + (\hat{B}_1 - \hat{B}_3)$$

and the assumptions.

The fifth clause is based on the equation $\hat{A} + \hat{A} = \hat{B} + \hat{B}$. The remaining clauses also follow on the nose from the Scott-style condition. In particular, the last clause relies on the equation

$$(A \cap B) + (A \cap \overline{B}) + (\emptyset - \hat{A}) = 0$$
that gives $A \cap B \sim \emptyset$ and $A \cap \overline{B} \sim \emptyset$ from $A \sim \emptyset$. Thus we get $A \sim \emptyset \implies A \sim A \cap B$. Now, since $\hat{A} + \hat{B} = (A \cup B) + (A \cap B)$, we have at once $A \cup B \sim B$, as wanted.

**Nonstandard Representation**

Although the Scott condition is not sufficient for the existence of an agreeing probability measure in infinite de Finetti spaces (e.g., it does not imply Archimedeaness), it is both necessary and sufficient for a nonstandard representation. In any de Finetti space, the Scott condition guarantees the existence of an agreeing probability measure that takes its values in a nonstandard model of the reals. This is a simple observation and it ought to be folklore knowledge, but I have not yet seen any references to it in print.

Mimicking a closely related idea in Domotor (1978), I will now present the nonstandard variant of a probabilistic representation. We begin with a simple lemma which plays an important role in what follows.

**Lemma 2.** In any de Finetti space the Scott condition $$(\Omega, \mathcal{A}, \preceq)$$ holds if and only if Scott's condition $$(\Omega, \mathcal{B}, \preceq)$$ is satisfied for every finite Boolean subalgebra $\mathcal{B} \subseteq \mathcal{A}$.

For suppose – going from left to right – that the Scott condition $$(\Omega, \mathcal{A}, \preceq)$$ holds and let $\mathcal{B}$ be a finite Boolean subalgebra of $\mathcal{A}$. Now pick any paired lists $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ of events in $\mathcal{B}$ such that $A_i \preceq B_i$ for all $1 \leq i \leq n$ and $\sum_{1 \leq i \leq n} \hat{A}_i = \sum_{1 \leq i \leq n} \hat{B}_i$. Since the equation between the sums of characteristic functions is equivalent to an equation between unions of events that continues to hold in $\mathcal{A}$, we can conclude that $B_n \preceq A_n$.

Moving in the other direction, let the Scott condition be true in all finite de Finetti spaces $(\Omega, \mathcal{B}, \preceq)$ with $\mathcal{B} \subseteq \mathcal{A}$. Then the premise $A_i \preceq B_i$ with $A_i, B_i$ in $\mathcal{A}$ for all $1 \leq i \leq n$ and $\sum_{1 \leq i \leq n} \hat{A}_i = \sum_{1 \leq i \leq n} \hat{B}_i$, can be confined to a finite Boolean subalgebra, in which the equation between sums of characteristic functions remains equivalent to a corresponding equation between unions of events. But the equation in the finite Boolean subalgebra holds, because the algebra is closed.
under arbitrary finite unions and intersections. So we can conclude that $B_n \leq A_n$ also holds, as wanted.

Next we establish a representation theorem for the Scott condition, using nonstandard reals.

**THEOREM 3.** For any de Finetti space $\langle \Omega, \mathcal{A}, \preceq \rangle$, there exists a finitely additive nonstandard-real valued probability measure $P : \mathcal{A} \to \mathbb{R}^*$ satisfying

$$A \preceq B \iff P(A) \leq P(B)$$

if and only if the Scott condition $\$\langle \Omega, \mathcal{A}, \preceq \rangle$ holds.

**Proof.** The necessity part is easy checking. As for sufficiency, we begin with constructing a nonstandard field $\mathbb{R}^* = d_f \mathbb{R}^\Delta / \equiv$, defined by the usual ultra-power construction on reals, involving the set $\Delta$ of all finite nonempty sets $\{\mathcal{B}_1, \ldots, \mathcal{B}_m\}$ of finite Boolean subalgebras of $\mathcal{A}$. For any finite Boolean subalgebra $\mathcal{B} \subseteq \mathcal{A}$ we set

$$\mathfrak{B} = df \{X \in \Delta \mid \mathcal{B} \in X\}.$$ 

In particular, $\{\mathcal{B}_1, \ldots, \mathcal{B}_m\} \in \mathfrak{B}_i$ for all $1 \leq i \leq m$.

Now we see that the set $\Phi = df \{\mathfrak{B} \mid \mathcal{B} \subseteq \mathcal{A} \text{ and } \mathcal{A} \text{ is finite}\}$ satisfies the finite intersection property. That is, we have

$$\forall i \leq m[\mathfrak{B}_i \in \Phi] \implies \bigcap_{1 \leq i \leq m} \mathfrak{B}_i \neq \emptyset.$$ 

Therefore there exists an ultrafilter $\Xi$ on the set of finite families of finite Boolean subalgebras $\Delta$, containing $\Phi$, so that we can complete the field construction by setting

$$f \equiv g \iff \{X \mid f(X) = g(X)\} \in \Xi$$

for all real-valued maps $f, g : \Delta \to \mathbb{R}$. The resulting equivalence classes $[f]$ of maps $f : \Delta \to \mathbb{R}$, modulo $\equiv$, will serve as our nonstandard reals. In particular, the usual definitions

$$[f] + [g] = [h] \iff \{X \mid f(X) + g(X) = h(X)\} \in \Xi,$$

$$[f] \cdot [g] = [h] \iff \{X \mid f(X) \cdot g(X) = h(X)\} \in \Xi,$$

$$[f] \leq [g] \iff \{X \mid f(X) \leq g(X)\} \in \Xi,$$
convert $\mathbb{R}^*$, defined above, into a totally ordered field that contains a copy of the set $\mathbb{R}$ of standard reals. We know that if $\mathbb{R} \neq \mathbb{R}^*$, then $\mathbb{R}^*$ is non-Archimedean — it contains an ideal of infinitesimals.

The final step is to define a probability measure $P : \mathcal{A} \to \mathbb{R}^*$ with the desired properties. Recall that in view of Lemma 2 above, the Scott condition holds in every finite de Finetti space $\langle \Omega, \mathcal{B}, \preceq \rangle$ with finite $\mathcal{B} \subseteq \mathcal{A}$. Therefore we have an agreeing probability measure $Q_\mathcal{B}$ on every finite subalgebra $\mathcal{B}$. Now let us set $P(A) = \text{df} [q_A]$, where $q_A : \Delta \to \mathbb{R}$ is defined by $q_A(\mathcal{X}) = Q_\mathcal{B}(A)$ if $A$ belongs to the finite Boolean subalgebra $\mathcal{B} = \text{df} \, \text{Bool}[\cup \mathcal{X}]$ — the smallest Boolean algebra, generated by the finite union of Boolean subalgebras comprising $\mathcal{X}$, and 0 otherwise.

Evidently, the function $P$ is finitely additive on $\mathcal{A}$, i.e., we have $\{ \mathcal{X} | q_A(\mathcal{X}) + q_B(\mathcal{X}) \} \in \Xi$ for all pairs of disjoint events $A$ and $B$. The reason being that the condition above is equivalent to the following valid additivity condition $\{ \mathcal{X} | Q_\mathcal{B}(A) + Q_\mathcal{B}(B) = Q_\mathcal{B}(A + B) \} \in \Xi$, where $\mathcal{B}$ is the smallest finite Boolean subalgebra of $\mathcal{A}$, containing all subalgebras comprising $\mathcal{X}$. Along similar lines one quickly checks that $P$ is nonnegative and normalized, and thus a probability measure.

To see that it is an agreeing probability measure, suppose we have $A \preceq B$ for a pair of events $A, B$ in $\mathcal{A}$. Now consider a finite Boolean subalgebra $\mathcal{C}$ that contains $A$ and $B$, and let $\mathcal{X} \in \mathcal{C}$, i.e., $\mathcal{C} \subseteq \mathcal{X}$. By definition we have $\mathcal{C} \subseteq \mathcal{B}$, where $\mathcal{B}$ is the smallest finite Boolean subalgebra, generated by the finite union of all subalgebras in $\mathcal{X}$. Since the Scott condition $\langle \Omega, \mathcal{B}, \preceq \rangle$ holds and we have $A \preceq B$ in $\mathcal{B}$, there exists a probability measure $Q_\mathcal{B}$ such that $A \preceq B \Rightarrow Q_\mathcal{B}(A) \leq Q_\mathcal{B}(B)$. But in light of the definition above this is equivalent to $A \preceq B \Rightarrow q_A(\mathcal{X}) \leq q_B(\mathcal{X})$. Hence we have $\{ \mathcal{X} | q_A(\mathcal{X}) \leq q_B(\mathcal{X}) \} \in \Xi$ and therefore $[q_A] \leq [q_B]$, i.e., $P(A) \leq P(B)$. The argument works also in the other direction. So, the map $P$ is an agreeing probability measure with values in a non-Archimedean field $\mathbb{R}$.

**Remark**

Since nonstandard numbers have standard parts and the part function is additive, we can set $P_{st}(A) = \text{df}$ standard part of $P(A)$, and obtain a real-valued probability measure $P_{st}$ on $\mathcal{A}$, satisfying the considerably weaker representation

$$ A \preceq B \Rightarrow P_{st}(A) \leq P_{st}(B). $$
Being of a formalist conviction, the leading exponent of nonstandard methods, Abraham Robinson, was quite happy with the metaphysics of nonstandard reals. In measurement theory, however, the problem of numerically representing properties or relations of real world objects presupposes some sort of empirical meaningfulness. For example, the two-dimensional geometric probability of picking a point from the main diagonal of a unit square, given that the point is incident with one of the two diagonals of the square, equals 1/2 under nonstandard infinitesimal calculations and equals 0/0 under standard calculations (since a two-dimensional area measure of a line segment is zero). So geometric sample spaces might possess nontrivial qualitative probability relations that admit meaningful nonstandard representation.

It is elementary that in general the Scott condition does not imply the Archimedean axiom. One of the advantages of the ‘nonstandard method’ is that all constructions are carried out in more or less the same way. In particular, a nonstandard representation of equiprobability and semiordered qualitative probability relations engages the pertaining Scott conditions in virtually identical ways.

Now suppose that we want a standard representation of comparative probability. Since in general the existence of linear functionals (hence measures) depends also on suitable topological or norm requirements, the Archimedean requirement alone may not suffice.

To make further progress, we shall rely on a geometric representation of the problem of measuring belief. Among the benefits I want to mention is an automatic possession of the Archimedean property.

For readers who want to verify the proofs of theorems in the next two sections, I recommend reading Section 6 in parallel.

4. QUALITATIVE PROBABILITIES IN COUNTABLE BOOLEAN SPACES

Countable de Finetti spaces are next to the finite ones on the scale of conceptual simplicity – and it shows. In particular, Scott’s theorem from the previous section generalizes to the countable case in the following way:

**THEOREM 4.** For any countable de Finetti space \(<\Omega, \mathcal{A}, \preceq>\) there exists a finitely additive probability measure \(P\) on \(<\Omega, \mathcal{A}>\) such that

\[ A \preceq B \iff P(A) \leq P(B) \]
for all $A, B \in \mathcal{A}$ if and only if for any pair of finite sequences of events $A, A_1^n, \ldots, A_{k_n}^n$ and $B, B_1^n, \ldots, B_{k_n}^n$ in $\mathcal{A}$ we have

$$\forall i \leq k_n [B_i^n \preceq A_i^n] \iff A \preceq B,$$

whenever

$$\lim_{n \to \infty} \sum_{1 \leq i \leq k_n} (\hat{A}_i^n - \hat{B}_i^n) = \hat{B} - \hat{A}.$$ 

The proof of this theorem relies on another, more general theorem, derived in Section 6.

It is important to note that here it is the sequence of sums of characteristic functions that is assumed to go to infinity and not the sum of such sequences! Consequently, as in Scott’s condition from the preceding section, for each $n$, the equation with sums can be rewritten into an equation with Boolean unions of finitary intersections. In light of Theorem 9, the equation with sums of characteristic functions can be reformulated so that it involves only events and their qualitative ordering.

**Proof.** Necessity: Suppose we have an agreeing probability measure $P$ on $\mathcal{A}$. Then the continuous linear functional $\mathcal{P}$ that extends $P$ to the linear space $\langle \Omega, \mathcal{A}, \preceq \rangle$ satisfies

$$\lim_{n \to \infty} \sum_{1 \leq i \leq k_n} [\mathcal{P}(\hat{A}_i^n) - \mathcal{P}(\hat{B}_i^n)] = \mathcal{P}(\hat{B}) - \mathcal{P}(\hat{A}).$$

Now since $\mathcal{P}$ is order-preserving, the assumption implies $\mathcal{P}(\hat{B}_i^n) \leq \mathcal{P}(\hat{A}_i^n)$, and thus the desired conclusion follows.

Sufficiency: Since $\mathcal{A}$ is assumed to be countable, we can use the class of vectors of the form $\hat{A} - \hat{B}$ as a basis for the convex cone $\mathcal{C}(\preceq)$ of $L(\Omega, \mathcal{A}, \preceq)$, associated with $\langle \Omega, \mathcal{A}, \preceq \rangle$. In particular, the norm condition of Theorem 9 now simplifies to

$$0 < \|\hat{A} - \hat{B} + \mathcal{C}(\preceq)\| \quad \text{for all} \quad B \prec A.$$ 

Upon contraposing the condition above and substituting the definition of distance between a vector and an arbitrary set of vectors, we get

$$\inf_{Y \in \mathcal{C}(\preceq)} \|\hat{A} - \hat{B} + Y\| = 0 \implies A \preceq B$$

for all $A, B \in \mathcal{A}$. 
Now after replacing each vector $Y$ with its defining form

$$\sum_{1 \leq i \leq n} a_i (\hat{A}_i - \hat{B}_i)$$

and upon eliminating the infimum from the norm condition, we arrive at the following conditional:

*If for any $n$ there exist sequences $\hat{A}_i^n$ and $\hat{B}_i^n$ such that*

$$\| \hat{A} - \hat{B} + \sum_{1 \leq i \leq k_n} a_i^n (\hat{A}_i^n - \hat{B}_i^n) \| \leq 1/n,$$

*then $\hat{B}_i^n \leq \hat{A}_i^n$ implies $A \leq B$.  

In view of the definition of our supremum norm, we can rewrite the foregoing inequality into a cascade of numerical inequalities

$$\hat{A}(\omega) - \hat{B}(\omega) - 1/n \leq \sum_{1 \leq i \leq k_n} a_i^n [\hat{A}_i^n(\omega) - \hat{B}_i^n(\omega)]$$

$$\leq \hat{A}(\omega) - \hat{B}(\omega) + 1/n$$

for all $\omega \in \Omega$.

Looking at the nonnegative coefficients $a_i^n$ as unknowns in our system of inequalities – having $\pm 1 \pm 1/n$ as its rational parameters, we notice that the unknowns $a_i^n$ have rational solutions. So we may as well assume that the coefficients are rational. By clearing the corresponding fractions and allowing repetitions, we can rewrite the sum in the inequalities again, this time with $a_i^n = 1$. Thus we get a limit

$$\lim_{n \to \infty} \left\{ \hat{A}(\omega) - \hat{B}(\omega) + \sum_{1 \leq i \leq k_n} [\hat{A}_i^n(\omega) - \hat{B}_i^n(\omega)] \right\} = 0$$

for all $\omega$, in which $\hat{A}(\omega) - \hat{B}(\omega)$ can be transferred to the right-hand side of the equation.

Using Theorem 10 in a similar way as above, it is easy to show a corresponding result for probabilistic indifference.

**Theorem 5.** Let $\langle \Omega, \mathcal{A}, \sim \rangle$ be a countable Boolean space together with an equiprobability relation $\sim$. Then for there to exist a probability measure $P$ on $\mathcal{A}$ such that

$$A \sim B \iff P(A) = P(B) \quad \text{for all} \quad A, B \in \mathcal{A},$$
it is necessary and sufficient that the following conditions hold for all
finite sequences $A, A_1^n, \ldots, A_k^n, B, B_1^n, \ldots, B_k^n$ of events in $\mathfrak{A}$:

(i) $\emptyset \sim \Omega$;
(ii) $A \sim A$;
(iii) $A \sim B \Rightarrow B \sim A$;
(iv) $\forall i \leq q_n[A_i^n \sim B_i^n] \Rightarrow (A \sim B \iff \forall j \leq p_n[C_j^n \sim \emptyset])$,

whenever

$$\lim_{n \to \infty} \left\{ \sum_{1 \leq j \leq p_n} \hat{C}_j^n + \sum_{1 \leq i \leq q_n} (\hat{A}_i^n - \hat{B}_i^n) \right\} = k(\hat{A} - \hat{B})$$

for some $k \geq 1$.

Here we start with the simplification of the norm condition

$$\| \pm (\hat{A} - \hat{B}) + C(\leq) + C(\sim) \| = 0$$

that allows us to separate all pairs of equiprobable events from all other pairs.

We reason in the same way as in the proof of Theorem 4. The
necessity follows from the equation

$$\lim_{n \to \infty} \left\{ \sum_{1 \leq j \leq p_n} \mathcal{P}(C_j^n) + \sum_{1 \leq i \leq q_n} [\mathcal{P}(A_i^n) - \mathcal{P}(B_i^n)] \right\} = k(\mathcal{P}(A) - \mathcal{P}(B)),$$

where $\mathcal{P}(A_i^n) = \mathcal{P}(B_i^n)$ and $\mathcal{P}(C_j^n) = 0$ hold, because $\mathcal{P}$ is order-preserving. As for sufficiency, we first translate the norm condition
into a limit condition, and then with the help of the usual continuity argument, we proceed to eliminate the real coefficients. What results is
a Scott style condition, displayed in the theorem.

Since every countable Boolean algebra carries a strictly positive
probability measure, we do not have to be concerned with the algebraic structure of $\mathfrak{A}$ itself. As we shall see momentarily, the situation is much
more complex in the uncountable case.

5. QUALITATIVE PROBABILITIES IN ARBITRARY BOOLEAN SPACES

The present section is devoted to investigating additional properties of
qualitative probability relations that are needed for their representation
in arbitrary infinite de Finetti spaces.
Using the resources of extensive measurement, Suppes and Zanotti (1976) offer a simple list of necessary and sufficient conditions that indeed guarantees the existence of an agreeing probability measure.

The authors trade Suppes's preferred simplicity of axioms for complications arising from the choice of objects for qualitative comparison. The objects in question are not events but integer-valued random variables. Consequently, the uniqueness in their result concerns the representing order-preserving linear functional (expectation functional) and not the probability measure, to which the functional is normalized.

One might object that the authors have conveniently redefined their original representation task, which was to come up with axioms for a measurable qualitative probability relation in a Boolean algebra. My complaint comes from the opposite direction. The class of random variables employed by Suppes and Zanotti is very limited. Even the simplest random variables, such as temperature in a room or position of a military target, fall outside their domain.

Archimedeaness should not be an excuse, since it can be formulated in any partially ordered linear space. In what follows, I will present a representation theorem that has this more general nature.

**THEOREM 6.** For any de Finetti space \( \langle \Omega, \mathfrak{A}, \preceq \rangle \) there exists a finitely additive probability measure satisfying

\[
A \preceq B \implies P(A) \leq P(B) \quad \text{for all} \quad A, B \in \mathfrak{A}
\]

if and only if

(i) \( A \prec B \iff \exists n [1/n\hat{\Omega} + \hat{A} \preceq \hat{B}] \);

(ii) \( 1/n \leq \|1/m \sum_{1 \leq i \leq m} (\hat{A}_i - \hat{B}_i) + C(\preceq)\| \)

for all \( n, m \) and \( 1/n\hat{\Omega} + \hat{B}_i \preceq \hat{A}_i \).

**Proof.** The necessity follows from having

\[
P(A) \geq P(B) + 1/n \iff 1/n\hat{\Omega} + \hat{B} \preceq \hat{A}.
\]

The sufficiency part relies on substituting \( n^{-1}\hat{\Omega} + C(\preceq) \) for \( C_n \) in Theorem 12, Section 6. In particular, we begin with the construction of a partially ordered linear space \( L(\Omega, \mathfrak{A}, \preceq) \) and its convex cone \( C(\preceq) \). Its strictly positive cone is given by

\[
C(\prec) = \bigcup_{a > 0} a \bigcup A \prec B (\hat{B} - \hat{A} + C(\preceq))
\]
and the convex subset $C_n$ is generated by
\[ \bigcup_{1/n\hat{\Omega} + \hat{A} \leq \hat{B}} [\hat{B} - \hat{A} + C(\preceq)]. \]
The rest of the work consists in applying Theorem 12 to the cones defined above.

Observe that in view of $0 \preceq n^{-1}\hat{\Omega}$, we have a convex set
\[ C_n = \{ X \mid n^{-1}\hat{\Omega} \preceq X \} = n^{-1}\hat{\Omega} + C(\preceq). \]

Now the assumption that this set covers the strictly positive cone, i.e.,
\[ C(\prec) = \bigcup_n [n^{-1}\hat{\Omega} + C(\preceq)], \]
is clearly equivalent to the Archimedean axiom
\[ 0 \prec X \iff \exists n [\hat{\Omega} \leq nX]. \]
The general infinite case of qualitative probabilities with representation in an Archimedean field is complicated by the problem of existence of strictly positive measures. These measures are needed for the axiom $0 \prec A$ for all $A \neq \emptyset$ that can be added any time to an already available compatible set of axioms.

We know from the work of Kelley (1959) that a necessary and sufficient condition for the existence of a strictly positive measure on a Boolean algebra is a strict positivity of the norm $\| \sum_{1 \leq i \leq n} \hat{A}_i \|$ of sums of all possible finite sequences of characteristic functions of events $A_i$, belonging to some infinite sequence $\mathcal{E}_1, \mathcal{E}_2, \ldots$ of families of events that covers $\mathcal{A}$ (i.e., the union of the sequence is equal to the algebra). Kelley’s norm condition on characteristic functions will not disappear without further conceptual concessions, including a passage to Boolean sigma algebras.

In what follows, we shall explore the role of Archimedeaness in qualitative probability theory in more detail.

**DEFINITION 2.** A weakly increasing sequence
\[ A_1 \leq A_2 \leq A_3 \leq \cdots \]
of events is called an *Archimedean chain* for a pair $(A, B)$ iff there exist disjoint event pairs $(B_1, C_1), (B_2, C_2), (B_3, C_3), \ldots$ such that
Suppose $P$ is an agreeing probability measure, representing $\leq$. For any strict pair $(A, B)$ of events – satisfying $A \prec B$, direct calculation gives the estimate

$$0 < P(B) - P(A) \leq 1/n \cdot [P(A_{n+1}) - P(A_1)]$$

that certainly fails for sufficiently large $n$. This suggests that Archimedean chains must be finite for all strict events pairs! That is, all weakly increasing sequences of the sort above have a finite length.

We write $\lg(A, B) = n$, whenever the length of any Archimedean chain for the pair $(A, B)$ is at most $n$ and there exists at least one Archimedean chain for $(A, B)$ that has length $n$.

Upon introducing the graded comparative relation

$$A \prec_n B = \text{df} \{ A \prec B \land \lg(A, B) = n \}$$

we can order the totality of strict pairs $A \prec B$ of events into a countable sequence of sets, gauged by the length of Archimedean chains.

In particular, we take $C_n$ to be the smallest convex set in the linear space of random variables $L(\Omega, \mathfrak{A}, \prec)$, generated by the union

$$\bigcup_{A \prec_n B} [\hat{B} - \hat{A} + C(\prec)].$$

The countably infinite sequence $C_1, C_2, \ldots$ is readily seen to cover all of $C(\prec)$. That is, we have

$$C(\prec) = \bigcup_{a > 0} \bigcup_n C_n.$$  

Using Theorem 12 from Section 6 and the definition of a norm, we obtain in a straightforward manner the following representation theorem for arbitrary de Finetti spaces.

**THEOREM 7.** For any de Finetti space $\langle \Omega, \mathfrak{A}, \leq \rangle$, there exists a finitely additive probability measure on $\langle \Omega, \mathfrak{A} \rangle$ satisfying

$$A \preceq B \iff P(A) \leq P(B) \quad \text{for all} \quad A, B \in \mathfrak{A}$$
if and only if the following conditions hold for all events $A$, $B$, $A_i$, $B_i$, $C_j$, $D_j$, in $\mathfrak{A}$:

1. All Archimedean chains for $A < B$ are finite;
2. The Scott-style condition

$$\forall i \leq p [A_i \leq B_i] \implies \text{not } \forall j \leq q [C_j \prec_n D_j]$$

holds, whenever the inequality

$$\sup_{\omega} \left| \sum_{1 \leq i \leq p} a_i [\hat{B}_i - \hat{A}_i](\omega) + \frac{1}{q} \sum_{1 \leq j \leq q} [\hat{D}_j - \hat{C}_j](\omega) \right| < \frac{1}{n}$$

involving values of characteristic functions of events is satisfied for all $n$, $p$, $q$ and $a_j \geq 0$.

Condition 2 has several obvious alternative formulations, but none of them is significantly simpler formally or conceptually than the one displayed above.

When compared with the Suppes–Zanotti axiomatization, this theorem has the virtue of being ontologically parsimonious. By and large, the axiomatic conceptual apparatus is that of de Finetti space – save the unavoidable technical presence of characteristic functions that is prompted by the inherent link between finitely additive probability measures and continuous linear functionals.

Since in general Boolean algebras tend to be more subtle mathematical objects than semigroups, I find the functional analytic approach to measurement representation more adequate. The next and last section is devoted to establishing important connections between de Finetti spaces and ordered linear spaces, with special emphasis on existence criteria for order-preserving linear functionals and measures.

6. EXISTENCE OF ORDER-PRESERVING LINEAR FUNCTIONALS

We picked de Finetti spaces to reason about the states of the world and about a rational agent’s states of belief about the world. Now the question is how to reason about de Finetti spaces themselves.

In this section I discuss what I take to be the simplest and at the same time most effective way in which qualitative probability spaces and ordered linear spaces can be related. Why ordered linear spaces? Because these spaces provide powerful criteria for the existence of
order-preserving linear functionals that often enjoy empirically meaningful formulations in terms of qualitative probabilistic order relations. Furthermore, the usual measure problems in Boolean algebras reappear in analogous ways also in the category of linear spaces. For example, just like there are Boolean algebras with no strictly positive real-valued measure, there are (metrizable) linear spaces that carry no nontrivial (i.e., not everywhere zero) real-valued continuous linear functionals!

Banach and those around him of course knew that linear spaces furnished with a norm are generally blessed with many continuous linear functionals, including those that respect order.

Hence our technical strategy will be a translation of all pertinent information about de Finetti spaces into the language of ordered linear spaces that carry a norm. Based on this information and the rich resources of linear analysis, we proceed to find suitable order-preserving linear functionals that we eventually retranslate into the framework of de Finetti spaces as agreeing probability measures.

The reader uninteresteded in formal technicalities can omit this section and take it for granted that the representation theorems presented in the preceding two sections are correct. What is important here is the emerging method of finding necessary and sufficient conditions for the existence order-preserving additive functions, intrinsic to the program of extensive measurement theory.

Constructing Linear Spaces from Boolean Spaces

With every Boolean space $(\Omega, \mathcal{A})$ we shall associate a linear space $L(\Omega, \mathcal{A})$ of simple random variables (step functions) over the field of reals, generated by the usual characteristic functions of events in $\mathcal{A}$. Thus $L(\Omega, \mathcal{A})$ consists of all real-valued $\mathcal{A}$-measurable functions $X : \Omega \rightarrow \mathbb{R}$ of the form

$$X = \sum_{1 \leq i \leq n} a_i \hat{A}_i,$$

where $A_i \in \mathcal{A}$ and $a_i \in \mathbb{R}$ is a string of real numbers for all $1 \leq i \leq n$. Of course, variable $n$ is assumed to take its values in the domain of positive natural numbers.

Random variables or vectors $X$ have many useful formulations (including the Stone space topological variants). It is good to keep in mind, however, that in the present approach every nonzero vector $X$ has
one standard partition representation, in which all scalar coefficients $a_i$ are nonzero and mutually different, and all events $A_i$ are nonempty and pairwise disjoint.

The usual vector addition $X + Y$ and scalar multiplication $aX$ operations, and the zero $0$ and unit $1$ elements, are defined in a pointwise fashion:

$$
(X + Y)(\omega) = \text{df} \ X(\omega) + Y(\omega);
$$
$$
(aX)(\omega) = \text{df} \ aX(\omega);
$$
$$
0(\omega) = \text{df} \ 0;
$$
$$
1(\omega) = \text{df} \ 1
$$
for all $\omega \in \Omega$.

It is trivial to check that $L(\Omega, \mathcal{A})$ is indeed a linear vector space with respect to the operations defined above. In particular, we have $\hat{\emptyset} = 0$, $\hat{\Omega} = 1$, $(A \cup B) = \hat{A} + \hat{B}$ if $A \cap B = \emptyset$, $1\hat{A} + 0\hat{B} = \hat{A} + 0 = \hat{A}$, and $\hat{A} + \hat{A} = 2\hat{A}$.

These linear space operations work equally well also on sets of vectors. So for any pair of subsets $X, Y$ in $L(\Omega, \mathcal{A})$, a vector $X$ and a real number $a$, we set

$$
X + Y = \text{df} \ \{X + Y \mid X \in X \land Y \in Y\};
$$
$$
aX = \text{df} \ \{aX \mid X \in X\};
$$
$$
X - Y = \text{df} \ X + (-1)Y;
$$
$$
X + Y = \text{df} \ \{X\} + Y;
$$
$$
X \setminus Y = \text{df} \ \{X \mid X \in X \land X \notin Y\}.
$$

Notice that $X + \emptyset = \emptyset \neq X - X$. A subset $X$ of vectors is called a convex cone iff $X + X \subseteq X$ and $aX \subseteq X$ for all $a \geq 0$.

Space $L(\Omega, \mathcal{A})$ carries a canonical natural partial ordering $X \leq Y$ of its vectors, defined once again pointwise:

$$
X \leq Y = \text{df} \ \forall \omega \in \Omega[X(\omega) \leq Y(\omega)].
$$

We see at once that under the foregoing order relation the space $L(\Omega, \mathcal{A})$ is actually a partially ordered linear space. That is, the usual compatibility (i.e., monotonicity and positive homogeneity) conditions hold. This is equivalent to claiming that the set of nonnegative random variables $C(\leq)$ is a convex cone.
Note that relative to the standard partition representation we have

\[ 0 < X \iff 0 < a_i \quad \text{for all} \quad 1 \leq i \leq n. \]

Furthermore, \( A \subseteq B \iff \hat{A} \leq \hat{B} \).

So we have an order structure on random variables that extends the Boolean inclusion ordering of events. Although we shall have little use for it here, perhaps I should mention that with respect to this natural order relation the space \( L(\Omega, \mathfrak{A}) \) is actually a Riesz space – a lattice ordered linear vector space. The familiar minimum and maximum lattice operations are defined pointwise.

The associated vector space not only encodes every possible bit of information about the Boolean space \( \langle \Omega, \mathfrak{A} \rangle \), it also carries important facts about the relationships between Boolean spaces. Simply, the association is functorial.

Linear spaces are important for us, since every norm bounded linear functional (expectation integral) on \( L(\Omega, \mathfrak{A}) \) determines a probability measure on \( \mathfrak{A} \) and conversely, every probability measure on \( \mathfrak{A} \) can be ‘extended’ to a unique linear functional on \( L(\Omega, \mathfrak{A}) \). For further details, see, e.g., Kappos (1969).

The foregoing conceptual tree will begin to bear fruit as soon as the vector space construction is appropriately refined with order structures induced by comparative probability relations. Observe that \( L(\Omega, \mathfrak{A}) \) can be ordered in another (and for us far more important but weaker) way, by arranging a correspondence

\[ a\hat{A} + \hat{C} \preceq a\hat{B} + \hat{C} \iff A \preceq B \]

for all \( A, B \) and \( C \) in \( \mathfrak{A} \), and \( a > 0 \).

Abusing the notation along traditional lines, we shall use the same symbol \( \preceq \) for all ‘empirical’ partial order relations, irrespective of whether they are defined on events, random variables or sets thereof.

Our next task is the analytic algebraization of comparative probability.

**Associating Convex Cones with Qualitative Probability Relations**

We know that order in a space can be specified by identifying its distinguished set of *positive* elements. Accordingly, we associate with every qualitative probability space \( \langle \Omega, \mathfrak{A}, \preceq \rangle \) a preordered linear space \( L(\Omega, \mathfrak{A}, \preceq) \), consisting of
the linear space $L(\Omega, \mathfrak{A})$ attached to $\langle \Omega, \mathfrak{A} \rangle$ as described above;

- a positive convex cone $C(\preceq)$, defined by the totality of vectors $X$ in $L(\Omega, \mathfrak{A})$ of the form
  \[ X = \sum_{1 \leq i \leq n} a_i (\hat{B}_i - \hat{A}_i) \]
  with $A_i \preceq B_i$ and $0 \leq a_i$ for all $1 \leq i \leq n$. The cone $C(\preceq)$ comes automatically with its negative counterpart $-C(\preceq)$, determined by the $(-1)$-multiples of vectors in the positive cone. The linear subspace $C(\preceq) \cup -C(\preceq)$ consists of all qualitative order comparable vectors in $L(\Omega, \mathfrak{A}, \preceq)$, and it should not be confused with the algebraic difference
  \[ C(\preceq) - C(\preceq) = L(\Omega, \mathfrak{A}, \preceq). \]

- a strictly positive convex cone $C(\prec) = df C(\preceq) \setminus C(\preceq)$, defined by the set of strictly positive vectors in $C(\preceq)$. We have $C(\prec) \subset C(\preceq)$ and $C(\preceq) + C(\prec) = C(\prec)$. The cone comes with its diametral strictly negative counterpart $-C(\prec)$, determined by the $(-1)$-multiples of vectors in the strictly positive cone.

- a linear subspace
  \[ C(\sim) = df C(\preceq) \cap -C(\preceq) \]
  in $L(\Omega, \mathfrak{A})$, given by all vectors that are simultaneously nonnegative and nonpositive. We have the symmetry property $C(\sim) = -C(\sim)$ and strict vs. linear part decomposition $C(\preceq) = C(\prec) + C(\sim)$.

The positive cone $C(\preceq)$ is obviously convex, i.e., it is closed under convex mixing of its vectors, and is proper (the cone is not a subspace). Last but not least, it satisfies the definitional characterization of convex cones:

(i) $[X \in C(\preceq) \land Y \in C(\preceq)] \implies X + Y \in C(\preceq)$;
(ii) $[X \in C(\preceq) \land 0 < a] \implies aX \in C(\preceq)$.

There is a natural one-to-one correspondence between convex cones and pre-orderings, obtained by

\[ X \preceq Y = df Y - X \in C(\preceq). \]

This comes to the same as requiring

\[ X \preceq Y \iff Y \in [X + C(\preceq)] \]
or

\[ X \preceq Y \iff X \in [Y - C(\succeq)]. \]

In the case of de Finetti spaces, the relation \( \preceq \) is reflexive and transitive but not antisymmetric in general. In addition, it extends the qualitative ordering of events in expected ways. Specifically, we have \( \hat{A} \preceq \hat{B} \iff A \preceq B \). Indeed, the converse direction holds by definition. Now if we had \( \hat{A} \preceq \hat{B} \) and \( B \prec A \), then \( \hat{A} - \hat{B} \in C(\prec) \) and \( \hat{B} - \hat{A} \in C(\prec) \) would give \( 0 \in C(\prec) \), i.e., \( 0 \prec 0 \). Here we are tacitly assuming that the cone \( C(\prec) \) is not empty. Since we have \( \emptyset \prec \Omega \), the vector \( \hat{\Omega} - \emptyset = 1 - 0 = 1 \) belongs to \( C(\prec) \). Because \( \emptyset \prec \hat{A} \), the positive cone \( C(\succ) \) is not empty either.

It is also evident that \( \preceq \) is monotonic in \( L(\Omega, \mathcal{A}, \preceq) \). In fact, it is the smallest relation of its kind that extends the qualitative probability relation \( \preceq \) to a comparative expectation relation on simple random variables.

Observe that in some de Finetti spaces we might have \( A \sim \emptyset \) for \( A \neq \emptyset \). In such situations both \( \hat{A} \) and \( -\hat{A} \) must belong to \( C(\preceq) \), even if \( A \) is in fact a negative random variable on \( \Omega \). So in this case we have \( \hat{A} \leq 0 \) while \( 0 \preceq \hat{A} \). This shows that the ordering of \( L(\Omega, \mathcal{A}) \), induced by a qualitative probability relation, can in general be rather different from the natural, set-theoretic inclusion based ordering, discussed in the previous subsection.

As customary, we define \( X \sim Y =_{df} X \preceq Y \wedge Y \preceq X \) and \( X \prec Y =_{df} X \preceq Y \wedge \neg Y \preceq X \).

These definitions are consistent with the cone characterizations, displayed earlier. We conclude this subsection with a list of basic properties of the extended qualitative order relation:

**Lemma 3.** Given a de Finetti space \( (\Omega, \mathcal{A}, \preceq) \), the following clauses are valid for all random variables \( X, Y, Z \) and \( W \) in \( L(\Omega, \mathcal{A}, \preceq) \) and events \( A \in \mathcal{A} \):

1. \( 0 \prec a\hat{\Omega} \) for all \( a > 0 \);
2. \( 0 \preceq a\hat{A} \) for all \( a \geq 0 \);
3. \( [X \preceq Y \wedge 0 \leq a] \implies aX \preceq aY \);
4. \( [X \preceq Y \wedge Z \preceq W] \implies X + Z \preceq Y + W \).
5. \( X \preceq Y \iff X + Z \preceq Y + Z \);
(6) \( X \leq Y \iff -Y \leq -X \).  

The verification is easy.

We say that the positive cone \( C(\leq) \) is Archimedean iff

\[
\forall n[(-nX + C(\leq)) \cap C(\leq) \neq 0] \implies X \in C(\leq).
\]

The definition is seen to be equivalent to the more familiar condition:

\[
\forall X \{ \exists Y [0 \leq Y \land \forall n(nX \leq Y)] \implies X \leq 0 \}.
\]

Although Archimedean cones are desirable objects for representation purposes, their structure is not readily definable in terms of Archimedeanness of the underlying qualitative probability. We shall return to this important problem in the final subsection.

**Norm-topological Characterization of Qualitative Probability Relations**

In addition to the algebraic and order structures introduced above, the mathematical workspace \( L(\Omega, \mathcal{A}) \) is also blessed with a natural geometric norm \( \| \| : L(\Omega, \mathcal{A}) \to \mathbb{R} \), a nonnegative real number valued function, defined by the supremum

\[
\| X \| = \sup_{\omega \in \Omega} |X(\omega)|,
\]

that satisfies the standard norm conditions (nullity, triangle inequality, and absolute homogeneity). Clearly, for any pair of distinct events \( A, B \in \mathcal{A} \) we have \( \| \hat{A} - \hat{B} \| = 1 \). On the other hand, \( \| \hat{A} - \hat{B} \| = 0 \iff A = B \). Hence \( L(\Omega, \mathcal{A}, \leq) \) is, among other things, also a normed linear space and therefore a metric space, and by default a (locally convex) topological space. The ensuing species of structures – linear algebraic, order, and metric topological – can be brought together in a number of different ways. Here we adhere to the simplest alternatives.

If vector \( X \) is in its standard partition form, then its norm is equal to the absolute value of the largest coefficient in its presentation: \( \| X \| = \sup_{1 \leq i \leq n} |a_i| \). For any pair of vectors \( X \) and \( Y \), their modulus \( \| X - Y \| \) provides a metric that can be interpreted as the distance between vectors \( X \) and \( Y \). So the workspace \( L(\Omega, \mathcal{A}) \) comes with its own analytic apparatus and built-in geometry that we will exploit heavily later.
Along traditional lines, the topological closure $\text{Cl} \mathbf{C}$ of any set of vectors $\mathbf{C}$ is given by

$$
\text{Cl} \mathbf{C} = \{ Y \mid 0 = \| \mathbf{C} - Y \| \}.
$$

It is the totality of vectors $Y$ whose distance

$$
\| \mathbf{C} - Y \| = \inf_{X \in \mathbf{C}} \| X - Y \|
$$

from the given set $\mathbf{C}$ is zero.

One can check that the vector addition and scalar multiplication are continuous operations with respect to the norm topology defined above. In particular, the class of open unit balls $\{ Y \mid 1/n > \| X - Y \| \}$ determines a convex neighborhood basis for vector $X$.

This completes the construction of the desired partially ordered and normed linear space $L(\Omega, \mathcal{A}, \preceq)$, determined by a de Finetti space $(\Omega, \mathcal{A}, \preceq)$.

We are now ready to present various theorems on the existence and positivity of continuous linear functionals on a normed linear space.

Existence of Separating Linear Functionals

The following result is fundamental in our treatment of the existence and uniqueness of order-preserving linear functionals.

**THEOREM 8 (Mazur-Orlicz).** Given a Boolean space $(\Omega, \mathcal{A})$, for any family $\mathbf{C} \subseteq L(\Omega, \mathcal{A})$ of random variables and a real-valued map $\Lambda : \mathbf{C} \rightarrow \mathbb{R}$ there exists a (continuous) linear functional $P : L(\Omega, \mathcal{A}) \rightarrow \mathbb{R}$ such that

1. $P(X) \leq \| X \|$ for all $X$ in $L(\Omega, \mathcal{A})$;
2. $\Lambda(X) \leq P(X)$ for all $X$ in $\mathbf{C}$
   if and only if

$$
\sum_{1 \leq i \leq n} a_i \Lambda(X_i) \leq \sum_{1 \leq i \leq n} a_i \| X_i \|
$$

holds for all $X_1, \ldots, X_n \in \mathbf{C}$ and arbitrary nonnegative reals $a_1, \ldots, a_n$.

A simple proof of a more general variant of this theorem is available in Peressini (1967) and we do not need to reproduce it here.
The main use of this result is, of course, in facilitating the construction of and finding axioms for order-preserving linear functionals. Typically, the procedure consists in defining \( \Lambda \) as a function of \( ||C|| \), where \( C \) is a suitable convex cone, induced by a given qualitative probability relation.

**COROLLARY 2.** Given a Boolean space \( (\Omega, \mathcal{A}) \), for any map \( \Lambda : L(\Omega, \mathcal{A}) \rightarrow [-\infty, +\infty) \) there exists a linear functional \( \mathcal{P} : L(\Omega, \mathcal{A}) \rightarrow \mathbb{R} \) satisfying

\[
\Lambda(X) \leq \mathcal{P}(X) \leq ||X|| \quad \text{for all } X \in L(\Omega, \mathcal{A})
\]

if and only if

\[
\sum_{1 \leq i \leq n} \Lambda(X_i) \leq \| \sum_{1 \leq i \leq n} X_i \|
\]

for all \( X_1, \ldots, X_n \) in \( L(\Omega, \mathcal{A}) \).

This friendly modification of the Mazur-Orlicz theorem was suggested by Scott, and its usefulness is discussed in Domotor (1969). The point of the corollary is that in view of the continuity of the vector addition and scalar multiplication with respect to the topology given by the norm \( || \cdot || \), we can safely eliminate all nonnegative coefficients present in the sum condition of the Mazur-Orlicz theorem. If needed, the coefficients can be put back by means of the corollary by generating arbitrary finitary repetitions in any of the sequences \( X_1, \ldots, X_n \) that are then collapsed into integer multiples and possibly divided by arbitrary positive naturals. Since the inequalities remain the same, in view of the norm continuity of linear space operations, all sequences of rational coefficients will converge to real coefficients without changing the inequalities involved. The obvious proof by induction on the length of sequences \( X_1, \ldots, X_n \) is omitted.

To obtain an order-preserving linear functional that we can use, set \( \Lambda \) equal to a value that changes smoothly with the ordering \( \preceq \).

As a first application of the Mazur-Orlicz theorem, I present a simple necessary and sufficient condition for the existence of an order-preserving linear functional in a countable situation.

Stated geometrically, the condition says that if the distance between each vector from the strictly negative cone \( -C(\prec) \) and the nonnegative cone \( C(\preceq) \) is always nonzero, then these convex cones can be *separated*
from each other by a linear functional \( \mathcal{P} \) in such a way that the supremum of all possible values of \( \mathcal{P} \) on the first cone will strictly smaller than the infimum of all values of \( \mathcal{P} \) achieved on the second cone. In addition, the separating value is zero.

**Theorem 9.** For any qualitative probability space \( (\Omega, \mathcal{A}, \preceq) \) with its normed, preordered linear space \( L(\Omega, \mathcal{A}, \preceq) \) and order cone \( \mathbf{C}(\preceq) \) having a countable basis, there exists a (continuous) linear functional \( \mathcal{P} : L(\Omega, \mathcal{A}, \preceq) \rightarrow \mathbb{R} \) satisfying

\[
X \preceq Y \iff \mathcal{P}(X) \leq \mathcal{P}(Y) \quad \text{for all } X, Y \in L(\Omega, \mathcal{A}, \preceq)
\]

if and only if

\[
0 \leq \|X + \mathbf{C}(\preceq)\| \quad \text{for all } X \in \mathbf{C}(\prec).
\]

**Proof.** (Necessity) Suppose we have an order-preserving linear functional \( \mathcal{P} \). Hoping for a contradiction, let us assume in addition that \( \|X + \mathbf{C}(\preceq)\| = 0 \) for some \( X \in \mathbf{C}(\prec) \). Then from the definition of \( \|\| \) it follows that for every \( n \) there exists a vector \( Y \in \mathbf{C}(\preceq) \) such that \( \|X + Y\| < 1/n \). In particular, since the functional is order-preserving, from \( X \in \mathbf{C}(\prec) \) we have \( \mathcal{P}(X) > 0 \) and for large enough \( n \) we get the strict inequality \( \|X + Y\| > 1/n \leq \mathcal{P}(X) \).

Now, because \( \mathcal{P}(Y) \geq 0 \), we also have \( \mathcal{P}(X) \leq \mathcal{P}(X + Y) \leq \|X + Y\| \), which is impossible. So the additional assumption is false and therefore the necessity holds.

(Sufficiency) Here I will implement Scott's idea. Since the cone \( \mathbf{C}(\preceq) \) is presumed to have a countable basis, the strict cone \( \mathbf{C}(\prec) \) will have one, too, and we may as well assume that it is given by the sequence \( X_1, X_2, \ldots \).

Now, in order to be able to use the corollary above, we define \( \delta \) for each \( n \) separately as follows:

\[
\Lambda_n(Y) = \begin{cases} 
\|X_n + \mathbf{C}(\preceq)\| & \text{if } Y \in X_n + \mathbf{C}(\preceq), \\
0 & \text{if } Y \in \mathbf{C}(\preceq) \setminus [X_n + \mathbf{C}(\preceq)], \\
-\infty & \text{otherwise}.
\end{cases}
\]

We see at once that under this definition of \( \Lambda_n \) the Mazur–Orlicz inequality is satisfied. Indeed, we have the inequality

\[
k \cdot \|X_n + \mathbf{C}(\preceq)\| + 0 + (-\infty) \leq \|Y_1 + \cdots + Y_k + Z\|,
\]
where \( Z \in C(\preceq) \) and \( Y_1, \ldots, Y_k \in [X_n + C(\preceq)] \). Note that we were allowed to throw away all those vectors that are giving \(-\infty\) on the left-hand side of the inequality, because they make it true trivially.

At this stage Corollary 2 gives us a linear functional \( \mathcal{P}_n \) such that \( \Lambda_n(Y) \leq \mathcal{P}_n(Y) \leq \|Y\| \) for all \( Y \in L(\Omega, \mathcal{A}, \preceq) \). Multiply the members of these inequalities by \( 1/2^n \) and sum them up. The resulting functional \( \mathcal{P}(Y) = \sum_{1 \leq n \leq \infty} 1/2^n \mathcal{P}_n(Y) \) is again additive and it is bounded by the norm of its arguments. Furthermore, in view of the definition of \( \Lambda_n \), the functional \( \mathcal{P} \) is strictly positive on all elements of

\[
C(\preceq) = \bigcup_{a > 0} a \bigcup_n [X_n + C(\preceq)]
\]

and nonnegative on \( C(\preceq) \). Finally, since \( X \preceq Y \) is the same thing as \( Y - X \in C(\preceq) \), the functional \( \mathcal{P} \) is order-preserving. Now the equality for \( C(\preceq) \) with the double union holds, because the cone is countably generated, so that its vectors have the form \( Y = \sum_{1 \leq i < m} a_i X_{k_i} \), where \( X_{k_i} \) are vectors from the sequence \( X_1, X_2, \ldots \) and \( a_i > 0 \). We must have \( X_{k_i} \in C(\preceq) \) at least for one \( i \), otherwise all of them would be in \( -C(\preceq) \) and their linear combination \( Y \) would also be in \( -C(\preceq) \), which is wrong. Therefore \( Y \) has the form \( Y = a_i [X_{k_i} + \sum_{1 \leq j \leq m, j \neq i} a_j/a_i X_j] \), so that \( Y \in a_i [X_{k_i} + C(\preceq)] \). The converse inclusion is clear.

Note that if \( C(\preceq) \) is closed, i.e., \( C(\preceq) = \{ X \mid 0 = \|C(\preceq) - X\| \} \), then \( \|X + C(\preceq)\| > 0 \) for all \( 0 \prec X \). Hence topological closedness of all order segments \( \{ Y \mid X \preceq Y \} \) with \( 0 \prec X \) is already sufficient for the existence of an order-preserving functional.

The following theorem is proved in a similar way as above:

**THEOREM 10.** For any countable probabilistic indifference space \( \langle \Omega, \mathcal{A}, \sim \rangle \), determining a normed linear space \( L = L(\Omega, \mathcal{A}, \sim) \) and a linear subspace \( C(\sim) \), there exists a (continuous) linear functional \( \mathcal{P} : L \to \mathbb{R} \) satisfying

\[
X \sim Y \iff \mathcal{P}(X) = \mathcal{P}(Y) \quad \text{for all} \quad X, Y \in L
\]

if and only if

\[
\| \pm X + C(\preceq) + C(\sim) \| = 0 \implies X \in C(\sim)
\]

for all \( X \in L \).
Proof. We begin with clarifying the notation. The linear subspace \( C(\sim) \) of \( L \), induced by equiprobable events, is defined by

\[
C(\sim) = \{ \sum_{1 \leq i \leq n} a_i (\hat{A}_i - \hat{B}_i) \mid A_i \sim B_i \text{ and } a_i \geq 0 \text{ for all } 1 \leq i \leq n \}.
\]

To guarantee that the linear functional remains positive on the characteristic functions of events, we need the positive cone

\[
C(\leq) = \{ \sum_{i \leq i \leq n} a_i \hat{A}_i \mid a_i \geq 0 \text{ and not } A_i \sim \emptyset \text{ for all } 1 \leq i \leq n \}.
\]

Finally, we set \( C(\lt) = C(\leq) + C(\sim) \) as an abbreviation for the convex cone of all nonnegative random variables. As before, we automatically have the strictly positive part \( C(\lt) = C(\lt) \setminus C(\leq) \), on which the representing strictly positive part \( C(\lt) = C(\lt) \setminus C(\leq) \), on which the representing linear functional will be strictly positive.

Since \( L \) is countably generated, we may take \( X_1, X_2, \ldots \) to be the strictly positive base vectors that satisfy \( 0 < \|X_n + C(\lt)\| \). Now defining \( \Lambda \) for each \( n \) in the same way as in Theorem 9, the Mazur–Orlicz inequality holds. In particular, we have

\[
\Lambda_n(Y) = \begin{cases} 
\|X_n + C(\lt)\| & \text{if } Y \in X_n + C(\lt), \\
0 & \text{if } Y \in C(\lt) \setminus [X_n + C(\lt)], \\
-\infty & \text{otherwise.}
\end{cases}
\]

Now the inequality

\[
k \cdot \|X_n + C(\lt)\| + 0 + (-\infty) \leq \|Y_1 + \cdots + Y_k + Z\|
\]

with \( Z \in C(\leq) \) and \( Y_i \in [X_n + C(\lt)] \) for all \( 1 \leq i \leq n \) is clearly satisfied, since the pertinent sets are convex. From here on, in close analogy with Theorem 9, we proceed with the application of Corollary 2, in which the inequality is used in its contrapositive form. In particular, we get a linear functional \( \mathcal{P} \) that satisfies
for all $X$ in $L$.

The last condition is justified by $X \in C(\sim) \Rightarrow 0 \leq \mathcal{P}(X)$ and $0 \neq \mathcal{P}(X)$ gives two alternatives. If $\mathcal{P}(X) < 0$, then from (i) above we get $X \notin C(\leq)$ and hence $X \notin C(\sim)$. If $\mathcal{P}(X) > 0$, then $\mathcal{P}(-X) < 0$ and therefore, as before, $-X \notin C(\sim)$. Since $C(\sim)$ is symmetric, we also have $X \notin C(\sim)$.

Upon assuming that $\sim$ is an equivalence relation and not $\emptyset \sim \Omega$, we get $\mathcal{P}(\hat{\Omega}) > 0$, and therefore a representing probability measure $P$ for $\sim$ may be obtained by setting

$$P(A) = \mathcal{P}(\hat{A})/\mathcal{P}(\hat{\Omega})$$

for all $A \in \mathfrak{A}$.

In the case of arbitrary infinite de Finetti spaces, we have the following theorem.

**Theorem 11.** For any qualitative probability space $\langle \Omega, \mathfrak{A}, \preceq \rangle$, determining a normed preordered linear space $L = L(\Omega, \mathfrak{A}, \preceq)$ with its positive cone $C(\leq)$, there exists a (continuous) linear functional $\mathcal{P} : L \to \mathbb{R}$ satisfying

$$X \preceq Y \iff \mathcal{P}(X) \leq \mathcal{P}(Y) \text{ for all } X, Y \in L,$$

whenever

$$0 < \|C(\prec)\|.$$

**Proof.** (Sufficiency) We follow the same strategy as above, except that this time we use directly the Mazur–Orlicz theorem. We set $C = C(\preceq)$ and define $\Lambda$ on $C$ simply as follows:

$$\Lambda(X) = \begin{cases} \|C(\prec)\| & \text{if } X \in C(\preceq), \\ 0 & \text{otherwise.} \end{cases}$$

In view of $C(\prec) = C(\prec) + C(\leq)$, the Mazur–Orlicz inequality

$$n \cdot \|C(\prec)\| \leq \| \sum_{1 \leq i \leq n} X_i + \sum_{1 \leq j \leq m} Y_j \|,$$
with the first sum from $C(\prec)$ and the second sum from $C(\preceq)$, is clearly satisfied. Hence there exists a linear functional $\mathcal{P} : L \to \mathbb{R}$ that is strictly positive on $C(\prec)$ and nonnegative on $C(\preceq)$.

If the strict cone $C(\prec) = \bigcup_{a>0} a \cup_n C_n$ can be pasted together from a sequence of convex sets $C_n$ such that their distance

$$\|C_n + C(\preceq)\| > 0$$

remains strictly positive, then the following much stronger result can be shown to hold:

**THEOREM 12.** For any normed, preordered linear space $L = L(\Omega, \mathcal{A}, \preceq)$, induced by a qualitative probability space $(\Omega, \mathcal{A}, \preceq)$, there exists a (continuous) linear functional $\mathcal{P} : L \to \mathbb{R}$ satisfying

$$X \preceq Y \iff \mathcal{P}(X) \leq \mathcal{P}(Y) \quad \text{for all} \quad X, Y \in L$$

if and only if

$$0 < \|C_n + C(\preceq)\| \quad \text{for all} \quad n,$$

where $C_n$ are convex subsets such that $C(\prec) = \bigcup_{a>0} a \cup_n C_n$.

In response to some suggestions in Kelley (1959), this theorem was originally formulated and shown by Dana Scott.

*Proof.* (Necessity) Granted an order-preserving functional $\mathcal{P}$, we set

$$C_n =_{df} \{ X \mid 1/n \leq \mathcal{P}(X) \}.$$  

We see that $C_n$ is convex and $1/n \leq \|C_n + C(\preceq)\|$ for all $n$. Hence Scott's condition holds.

(Sufficiency) We define function $\Lambda_n$ for each $n$ as follows:

$$\Lambda_n(X) =_{df} \begin{cases} \|C_n + C(\preceq)\| & \text{if } X \in C_n, \\ 0 & \text{if } X \in C_n \setminus C(\preceq), \\ -\infty & \text{otherwise.} \end{cases}$$

With this definition of $\Lambda$, the Mazur–Orlicz condition has the form

$$n \cdot \|C_n + C(\preceq)\| + 0 + (-\infty) \leq \| \sum_{1 \leq i \leq n} X_i + Y \|,$$
where the respective summands on the right-hand side belong to $C_n$ and $C(\subseteq)$. We threw away the vectors that give $-\infty$, because they make the inequality trivial. The inequality is seen to follow from the definition of distance and the convexity of $C_n$.

Hence we have a sequence $\mathcal{P}_n : L \to \mathbb{R}$ of linear functionals such that $\Lambda_n(X) \leq \mathcal{P}_n(X) \leq \|X\|$ for all $X \in L$. We now divide the inequalities by $2^n$ and sum up its members. The inequalities still hold, but this time with the desired linear functional $\mathcal{P}(X) = \sum_{1 \leq n \leq \infty} 2^{-n} \mathcal{P}_n(X)$, which clearly has the desired properties.

The foregoing representation theorems and their geometric equivalents are essential to explaining the conceptual role of Scott style axioms in representing qualitative probability relations.

Department of Philosophy,
University of Pennsylvania,
Philadelphia, PA 19104–3325, USA

REFERENCES

Bernstein, S. N.: 1917, Axiomatic Foundations of Probability Theory (Collected Works IV), Academy of Sciences, USSR.


**COMMENTS BY PATRICK SUPPES**

Zoltan Domotor provides a very thorough and useful analysis of the significant uses that may be made in Boolean algebras that are not finite of Scott’s condition. As he brings out in his paper, Scott’s condition is the most well-known formulation to use as the principal axiom in showing that a qualitative comparative probability relation on a finite Boolean algebra satisfies necessary and sufficient conditions for the existence of a strictly agreeing probability measure – strictly agreeing, that is, with the qualitative comparative relation. Even as a graduate student, Zoltan always had a natural taste for generalization, and that can be seen working very well in the present paper. Given his tendency also to be very thorough in referring to earlier literature, the only point on which I was surprised is his not remarking that the extensive use of linear spaces for the treatment of foundational problems in probability was already begun extensively in de Finetti’s theory of random quantities as developed in de Finetti (1937). Another point of interest here is that de Finetti did not want to be restricted to a sample space. This runs contrary to certain of his deeply held Bayesian views, and so random quantities form a linear space but are not built up from a Boolean algebra of events and a fixed space $\Omega$ of possible outcomes. Of course with $\Omega$ assumed, random quantities become standard random variables. I also want to mention two other points of particular importance in relation to previous literature. Zoltan’s formulation of Scott’s condition without the use of indicator functions is a desirable simplification, and his formulation for the case of countable Boolean algebras (Theorem 4)
is elegant and natural. The latter part of the paper, which emphasizes
the use of linear spaces constructed from Boolean spaces, is valuable
because of its thorough discussion of the many rich results about linear
spaces that can be used. This idea, too, is found already in de Finetti and
in a number of papers about probability theory, some foundational in
character, in the ’50s, but Domotor’s discussion has the virtue of being
thorough and up to date.

The one point on which we do not really agree is on the treat-
ment of uniqueness. In the extensive discussion following Theorem 1,
Zoltan expresses a view about how we obtain uniqueness of the measure
imposed on a qualitative probability structure with a viewpoint that is
certainly different from mine. He suggests choosing the measure that
is the barycenter of the convex set of possible measures. But this is
simply a geometric argument as far as I’m concerned, not at all related
to qualitative views about the probabilities of events, and it has no spe-
cial status. This approach bypasses the problem of giving qualitative
conditions that lead to a unique measure, in favor of using some simple
quantitative condition that itself in my judgment does not have concep-
tual justification. It covers up, for example, the fact that, though one
can give natural necessary and sufficient conditions for the existence
of a probability measure in the case of a finite Boolean algebra with a
qualitative ordering relation, the conditions, as is evident from the cur-
rent literature, are not at all simple for the case of uniqueness. In fact,
no really reasonable, necessary and sufficient conditions for uniqueness
are known. This is exactly what I like about the quite simple axioms
of Suppes and Zanotti (1976). By considering simple random variables
as well as events, very reasonable necessary and sufficient conditions
for uniqueness of the measure, when extended to the expectation of the
random variables, can be given.

Let me make one other remark about uniqueness. Why is it important
and not simply straightforward to use any of the existing measures, as,
for example, those guaranteed by Domotor’s Theorem 1? The answer is
that conditional probabilities vary so widely for the different probabili-
ties and independence is not invariant for the set of existing measures.
In other words, the two principal ways in which we deal with uncertain
events, namely by conditioning or by declaring them independent, do
not, for the qualitative ideas, have an adequate quantitative representa-
tion when the measure is not unique.