Minds & Machines
DOI 10.1007/s11023-015-9385-7

Qualitative Axioms of Uncertainty as a Foundation for Probability and Decision-Making

Patrick Suppes[1] · **Arthur Paul Pedersen**[†]

Received: 10 March 2014 / Accepted: 9 October 2015

Abstract Although the concept of uncertainty is as old as Epicurus's writings, and an excellent quantitative theory, with entropy as the measure of uncertainty having been developed in recent times, there has been little exploration of the qualitative theory. The purpose of the present paper is to give a qualitative axiomatization of uncertainty, in the spirit of the many studies of qualitative comparative probability. The qualitative axioms are fundamentally about the uncertainty of a partition of the probability space of events. Of course, it is common to speak of the uncertainty, or randomness, of a random variable, but only the partition defined by the values of the random variable enter into the definition of uncertainty, not the actual values. It is straightforward to add axioms for decision making following the general line of Savage from the 1950s. Indeed, in the spirit of Epicurus, it is really our intuitive feeling about the uncertainty of the future that motivates much of our thinking about decisions. Here, the distinction between the concepts of probability and uncertainty can be made by citing many familiar examples. Without spelling out the technical details, the axiomatization of qualitative probability with uncertainty as the most important primitive concept, it is possible to raise a different kind of question about bounded rationality. This new question is whether or not one should bound the uncertainty in thinking and investigating any detailed framework of decision making. Discussion of this point is certainly different from the question of bounding rationality by not maximizing expected utility. In practice, we naturally bound

[†] This paper was submitted for the present special issue shortly before Pat Suppes passed away. The manuscript required substantial corrections and revisions, not only for routine copy-editing, but also for mathematical accuracy and completeness. With Pat's permission, this work was carried out by one of the guest editors, Arthur Paul Pedersen, whose remarks in the following manuscript are denoted by 'APP'.—*Gregory Wheeler, Editor-in-Chief.*

✉ Patrick Suppes

✉ Arthur Paul Pedersen
pedersen@mpib-berlin.mpg.de

[1] Stanford, CA, USA

Published online: 22 October 2015

uncertainty in our analysis of decision-making problems. As in the case of formulating an alternative for maximizing expected utility, so is the case of rational alternatives to maximizing uncertainty. There are several issues to consider. In the spirit of my other work in qualitative probability, I explore alternatives rather than attempt to give a definitive argument for one single solution.

Keywords Bounded rationality · Bounded uncertainty · Qualitative probability · Qualitative independence · Qualitative uncertainty · Entropy

Introduction

The central concern of this article is to develop the concept of bounded uncertainty, which has been neglected, but should have a significant role in detailed discussions of bounded rationality. I hope to succeed in this article in explaining what the concept of bounded uncertainty is and why it is important to consider it in any general theory of bounded rationality.

Before starting a formal development of the concepts to be considered, it seems desirable to give some examples whose content is easily understood and thus can serve in expressing the intuitions behind the concept of uncertainty—or, its equivalent, the concept of information—about which more is said later.

Bill and Mary find a coin in circumstances that suggest the coin might be biased. They look at each other and say, "Let's toss the coin a few times to see if it is biased or fair". (The coin is *fair* if about half the tosses turn up heads). Mary says, "O.K. Let's flip it 5 times". Bill responds, "Why 5?" After a short discussion, Bill persuades Mary to go a little higher and toss the coin 7 times. Mary looks at him quizzically and says, "Well, smarty, why not eleven or three or ten tosses?"

At a more complicated level of experimentation, this kind of discussion over "How many trials?" goes on in almost all parts of experimental science, and in other situations, especially ones of making bets with friends on the outcome of sporting events. Almost always in these informal settings no systematic calculation of outcomes is made, by experimenter or friend.

In this discussion I have mentioned two kinds of events: experimental ones, like coin flipping, which can easily be repeated, and sporting events, which cannot. Here my interest is in *experiments*, and I shall restrict myself to them in what follows.

An experiment—should it be a coin-tossing experiment or a more elaborate one—is designed to confirm a theoretical prediction or just to learn the probability of a given kind of event, like a coin toss resulting in heads. The experimental procedures themselves are presumed to have little effect on the probability of outcomes, although quantum mechanics and several other areas of science provide exceptions to this rule.

Return to Bill and Mary. "Well," Mary says, "I don't know how we measure the information coming from each trial of an experiment such as coin tossing or throwing dice, but since the trials are meant to be independent and identical in structure, the information must add up. So n trials gives us n units of information." (Here it is more natural to talk about information than uncertainty.) Bill says, "That sounds right. So the more trials we have in an experiment, the more information we get. So why not just run lots of trials and not argue about how many?" Mary replies, "That will work for us, but not for expensive scientific trials. For them, we need some kind of theory, because they cost too much and take too long."

The rest of this article is devoted to developing the foundations for such a theory. To begin, I state qualitative axioms of probability in "Qualitative Axioms for Probability" section. Then, in "Qualitative Axioms of Independence and Uncertainty" section, I state qualitative axioms of independence as well as uncertainty. In "Qualitative Random-Variable Axioms" section, I state slightly stronger qualitative random variable axioms using indicator functions of events rather than events themselves. As part of the development of concepts in these two sections, it is also necessary to introduce the concept of the qualitative independence of experiments. Much of the material in the first few sections is meant to be a rapid review of qualitative axioms about probability concepts taken from the scientific literature. These sections take material from Suppes (2014), but they also present some new material.

In "Standard Representation Theorem for Uncertainty" section, I state a quantitative representation theorem for uncertainty. This theorem is new. The final section, "Bounded Uncertainty" section, focuses on bounded uncertainty.

Of equal importance is the development of the concept of entropy in "Qualitative Random-Variable Axioms" section as a measurement of uncertainty, now widely accepted in the modern theory of information.

Qualitative Axioms for Probability

I begin with standard comparative probability axioms for events. The axioms are completely qualitative in character. They have a fairly long history in the modern theory of qualitative probability. Certainly, they were given prominence by de Finetti (1937/1964, 1974, 1975). Presented below are the five axioms, essentially those of his original formulation. The relation of qualitative comparative probability is denoted by the wavy inequality symbol '$\succcurlyeq$,' read 'at least as probable as.' The strict binary relation $\succ$ is defined in the usual way, so $A \succ B$ iff (i) $A \succcurlyeq B$ and (ii) it is not the case that $B \succcurlyeq A$. The symbol '$\varnothing$' denotes the empty set, i.e., the impossible event, and '$\cap$' and '$\cup$' are the usual symbols used for intersection and union of sets. The capital Greek omega, 'Ω,' in Axiom **CP4** denotes the set (or universe) of all possible outcomes.

CP1. If $A \succcurlyeq B$ and $B \succcurlyeq C$, then $A \succcurlyeq C$.
CP2. Either $A \succcurlyeq B$ or $B \succcurlyeq A$.
CP3. $A \succcurlyeq \varnothing$.
CP4. $\Omega \succ \varnothing$.
CP5. If $A \cap C = \varnothing$ and $B \cap C = \varnothing$, then $A \cup C \succcurlyeq B \cup C$ iff $A \succcurlyeq B$.

The only probabilistic concept used in these axioms is the qualitative relation of event A being at least as probable as event B, i.e., $A \succcurlyeq B$.

The structures that satisfy these axioms can be simple probability distributions for finite sets of events. These axioms alone, however, are too weak to prove a quantitative or numerical probability representation theorem. A counterexample that decisively shows this is so was given many years ago (Kraft et al. 1959). Later I will discuss axioms that must be added to these original five in order to prove a numerical representation theorem. But the task to be focused on here is not the standard probability representation, but rather the representations of independence and uncertainty. I turn to this task in the next section.

Qualitative Axioms of Independence and Uncertainty

Before we can consider qualitative axioms of uncertainty, it is necessary to spend some time on the qualitative axioms of independence. On the one hand, the basic idea behind these axioms is straightforward. The independence of events in the theory of information plays the role that the null intersection of events plays in probability. On the other hand, detailed discussion about qualitative axioms of independence is not familiar to many people who know well the general theory of probability.

Presented below is a standard list of qualitative axioms of independence. This list is by no means unique and is derived from several sources, especially Domotor (1969, 1970). But it is detailed enough to give a feel for such qualitative axioms. The notation '$\mathrm{A}\perp\mathrm{B}$' is read 'Event A is independent of event B.' If A is an event, then $\neg\mathrm{A}$ denotes the complementary event, so $\mathrm{A}\cup\neg\mathrm{A}=\Omega$ (quantitatively, we expect $\mathbf{P}(\mathrm{A})+\mathbf{P}(\neg\mathrm{A})=\mathbf{P}(\Omega)=1$).

QI1. $\mathrm{A}\perp\Omega$.
QI2. If $\mathrm{A}\perp\mathrm{B}$, then $\mathrm{B}\perp\mathrm{A}$.
QI3. If $\mathrm{A}\perp\mathrm{B}$, then $\mathrm{A}\perp\neg\mathrm{B}$.
QI4. If $\mathrm{A}\perp\mathrm{B}$, $\mathrm{A}\perp\mathrm{C}$, $\mathrm{B}\cap\mathrm{C}=\varnothing$, $\mathrm{A}\perp\mathrm{B}\cup\mathrm{C}$.
QI5. If $\mathrm{A}\perp\mathrm{B}$, $\mathrm{B}\perp\mathrm{C}$, and $\mathrm{A}\perp\mathrm{B}\cap\mathrm{C}$, then $\mathrm{C}\perp\mathrm{A}\cap\mathrm{B}$.

Qualitative axioms for uncertainty have only a recent history in the theory of probability. Fundamentally, the search for such axioms was initiated by the surprising original work of Claude Shannon on solving problems of coding and sending messages over channels of limited capacity. Not very long after his first publication in 1948, the theory of information suggested by his work was given a clear mathematical representation (in Russian) by Khinchin (1953, 1956). When his work was translated into English (1957), it was widely read and had a big influence in the development of the theory of information in the 1950s and 1960s. (Khinchin treats as equivalent information and uncertainty, and this is now a common view.)

Presented below is a set of five qualitative axioms of uncertainty that match rather closely the classical qualitative axioms of de Finetti for comparative probability. Before presenting the qualitative axioms of uncertainty, I need to introduce a few additional concepts.

Let Ω be a nonempty set. A set π is called a (*finite*) *partition* of Ω if π is a nonempty finite family of sets $\mathrm{P}_1,\ldots,\mathrm{P}_n$ such that:

(i) P_i is a subset of Ω, for each $i = 1, \ldots, n$;
(ii) $P_i \neq \emptyset$ for each $i = 1, \ldots, n$;
(iii) $P_i \cap P_j = \emptyset$ for every i, j with $1 \leq i < j \leq n$; and
(iv) The union of all sets P_i in π is equal to Ω, i.e., $\bigcup_{i=1}^{n} P_i = \Omega$.

Adopting language from Kolmogorov (1933/1950), a finite partition π of Ω is said to be an *experiment* if every element of π, i.e., every P_i in π, is also an event. We define an *event* to be a set that is an element of a given *algebra* over Ω.

Formally, an *algebra of events* over Ω is a collection $\mathscr{A}$ of subsets of Ω satisfying the following properties:

(i) Ω is an element of $\mathscr{A}$;
(ii) If A is an element of $\mathscr{A}$, then so is $\neg$A; and
(iii) If A and B are elements of $\mathscr{A}$, then so is $A \cup B$.

Now to be more precise, an *experiment* over a given algebra of events $\mathscr{A}$ is a subcollection of $\mathscr{A}$ that is a finite partition of Ω. Let $\mathcal{E}(\mathscr{A})$ denote the set of all experiments over $\mathscr{A}$.

We also need to define the concept of the *product* of experiments over an common algebra of events. Set-theoretically, the product $\odot$ of partitions π and ξ of Ω, $\pi \odot \xi$, is defined as follows:

$$\pi \odot \xi = \left\{ C \subseteq \Omega : \ C \neq \emptyset \ \text{ and } \ C = A \cap B \ \text{ for some } A \in \pi \text{ and } B \in \xi \right\}.$$

Observe that for a given an algebra of events $\mathscr{A}$ over Ω, if π and ξ are experiments in $\mathcal{E}(\mathscr{A})$, then so is their product $\pi \odot \xi$. It is readily verified that the product $\odot$ is associative, commutative, and idempotent over $\mathcal{E}(\mathscr{A})$.

Finally, we need to define the concept of *independence* for experiments over a given algebra of events $\mathscr{A}$. Two experiments π and ξ in $\mathcal{E}(\mathscr{A})$ are said to be *independent* if every event A from π is independent of any event B from ξ—that is, $A \perp B$ for all $A \in \pi$ and $B \in \xi$. The same symbol $\perp$ is used for independence of experiments as for events, so $\pi \perp \xi$ just in case π is independent of ξ.

The following qualitative axioms of uncertainty slightly simplify the set of axioms first introduced in Suppes (2014). I use '$\succcurlyeq_u$' as the symbol for the uncertainty relation—read 'at least as uncertain as'—where the relationship between *experiments* over a given algebra of events, not events belonging a given algebra of eventS. As you can see, the subscript '*u*' is used for the uncertainty relation between experiments. In Axiom **CU3**, as well as hereafter, $\pi(\dot{\Omega})$ denotes the partition of Ω whose only element is the set Ω itself, i.e., $\pi(\dot{\Omega}) = \{\Omega\}$ (the reason for the dot above Ω will become clear in the next section).

CU1. If $\pi \succcurlyeq_u \xi$ and $\xi \succcurlyeq_u \vartheta$, then $\pi \succcurlyeq_u \vartheta$.
CU2. Either $\pi \succcurlyeq_u \xi$ or $\xi \succcurlyeq_u \pi$.

CU3. $\pi \succcurlyeq_u \pi(\dot{\Omega})$.
CU4. If $\pi \perp \vartheta$ and $\xi \perp \vartheta$, then $\pi \odot \vartheta \succcurlyeq_u \xi \odot \vartheta$ if and only if $\pi \succcurlyeq_u \xi$.

Unfortunately, these three sets of purely qualitative axioms are not strong enough to prove a numerical representation in terms of a standard probability distribution over events and (to be discussed later) a standard entropy measure over experiments.

There are many ways to add further axioms, especially of an Archimedean character, that are strong enough to prove a standard representation theorem. Personally, rather than mixing the logical character of the axioms—for example, formulated in first-order logic or not—I prefer to state all the axioms in terms of restricted qualitative random variables which naturally arise from the algebra of interest. The task of establishing a representation theorem for such a system of axioms is undertaken in the next section.

Qualitative Random-Variable Axioms

Before taking on the main task of this section, we must begin with some formal remarks about notation and terminology. Given a subset E of a nonempty set Ω, let E be the real-valued function on Ω such that for all $\omega \in \Omega$:

$$\dot{E}(\omega) \quad = \quad \begin{cases} 1 & \text{if } \omega \in \mathrm{E}; \\ 0 & \text{otherwise.} \end{cases}$$

The function E is called the *indicator function* for E relative to the set Ω. More generally, an extended indicator function is a sum of indicator functions for subsets of Ω. That is, an *extended indicator function* relative to Ω is a real-valued function X on Ω such that for some positive integer n, subsets $\mathrm{E}_1, \ldots, \mathrm{E}_n$ of Ω, and non-negative integers $k_1, \ldots, k_n$:

$$X = \sum_{i=1}^{n} k_i \cdot \dot{E}_i. \qquad (*)$$

The possible values of an extended indicator function are non-negative integers, and the set of possible values is finite. Addition and multiplication of extended indicator functions are defined pointwise as usual.

For each non-negative integer m, let $X^{-1}(m)$ be the level set of m under X:

$$X^{-1}(m) = \quad \left\{ \omega \in \Omega \ : \ X(\omega) = m \right\}.$$

Given an algebra $\mathscr{A}$ of events over a nonempty set Ω, an extended indicator function X on Ω is said to be *$\mathscr{A}$-measurable* if $X^{-1}(m) \in \mathscr{A}$ for each nonnegative integer m. A real-valued function X on Ω is a member of $\mathscr{A}$-measurable just in case it has the form of $(*)$ for some (unique) positive integer n, distinct non-negative integers $k_1, \ldots, k_n$, and nonempty sets $\mathrm{A}_1, \ldots, \mathrm{A}_n \in \mathscr{A}$ such that $\mathrm{A}_i = X^{-1}(k_i)$ for each $i = 1, \ldots, n$ and the collection $\{A_1, \ldots, A_n\}$ forms a partition of Ω. The unique partition that X generates is therefore an experiment and shall be denoted by $\pi(X)$.

The set of $\mathscr{A}$-measurable extended indicator functions on Ω, denoted by $\mathscr{A}^*$, is called the *algebra of extended indicator functions* generated by $\mathscr{A}$. Observe that $\mathscr{A}^*$ is the smallest semigroup under functional pointwise addition that contains the indicator functions for all events in $\mathscr{A}$ (cf. Suppes and Zanotti 1976).

Random-Variable Axiomatization of Comparative Probability

In the random-variable axioms for comparative probability, the strict binary relation $\succ$ is defined from $\succcurlyeq$ in the usual way as before. Here, however, $\succcurlyeq$ is a binary relation over a given algebra of extended indicator functions $\mathscr{A}^*$.

CPR1. If $\mathrm{X} \succcurlyeq \mathrm{Y}$ and $\mathrm{Y} \succcurlyeq \mathrm{Z}$, then $\mathrm{X} \succcurlyeq \mathrm{Z}$.
CPR2. Either $\mathrm{X} \succcurlyeq \mathrm{Y}$ or $\mathrm{Y} \succcurlyeq \mathrm{X}$.
CPR3. $\mathrm{X} \succcurlyeq \dot{\varnothing}$.
CPR4. $\dot{\Omega} \succ \dot{\varnothing}$.
CPR5. $\mathrm{X} \succcurlyeq \mathrm{Y}$ if and only if $\mathrm{X} + \mathrm{Z} \succcurlyeq \mathrm{Y} + \mathrm{Z}$.
CPR6. **(Archimedean Axiom)** If $\mathrm{X} \succ \mathrm{Y}$, then there exist positive integers k and n such that:

$$n\mathrm{X} \succ k\dot{\Omega} \succ n\mathrm{Y}.$$

Using axioms **CPR1** to **CPR6**, the following representation theorem can be proved (cf. Theorem 3, Suppes 2014).[1]

Theorem 1 *A structure* $(\Omega, \mathscr{A}^*, \succcurlyeq)$ *satisfies axioms* **CPR1** to **CPR6** *if and only if there is real-valued function* $\mathbb{E}$ on $\mathscr{A}^*$ *such that for all* $\mathrm{X}, \mathrm{Y} \in \mathscr{A}^*$:

(i) $\mathbb{E}(\mathrm{X}) \geq 0$;
(ii) $\mathbb{E}(\dot{\varnothing}) = 0$ *and* $\mathbb{E}(\dot{\Omega}) > 0$;
(iii) $\mathbb{E}(\mathrm{X} + \mathrm{Y}) = \mathbb{E}(\mathrm{X}) + \mathbb{E}(\mathrm{Y})$;
(iv) $\mathrm{X} \succcurlyeq \mathrm{Y}$ *if and only if* $\mathbb{E}(\mathrm{X}) \geq \mathbb{E}(\mathrm{Y})$.

Moreover, if $\mathbb{E}$ *and* $\mathbb{E}'$ *are real-valued functions on* $\mathscr{A}^*$ *satisfying properties* (i) *to* (iv) *, then there is a positive real number* λ *such that* $\mathbb{E} = \lambda\, \mathbb{E}'$; *that is, any function* $\mathbb{E}$ *satisfying properties* (i) *to* (iv) *is unique up to a positive similarity transformation.*

While Suppes and Zanotti (1976) advance a similar axiomatization and theorem for comparative probability, the axiomatization advanced here uses a simpler Archimedean axiom (**CPR6**).

[1] *APP's Note*: Theorem 3 of (Suppes 2014) does not hold without suitably modifying the axioms and structural assumptions stated therein. In an effort to charitably reflect Suppes' intentions, Theorem 3 of (Suppes 2014) has been recast in the present paper with such modifications. The proof of the corrected result proceeds very much in the spirit of the argumentation offered for Theorem 3 of (Suppes 2014).

Random-Variable Axiomatization of Qualitative Independence

To obtain a system of axioms for qualitative independence, the random-variable axioms for comparative probability (**CPR1** to **CPR6**) are supplemented with the following axioms:

QIR1. $X \perp \dot{\Omega}$.
QIR2. $X \perp Y$ implies $X \perp nY$ for every positive integer n.
QIR3. $V \perp W$ if and only if for all $X, X', Y, Y', Z, Z' \in \mathscr{A}^*$ and integers m, n:

(1) $m < |\pi(X)|$ and $n < |\pi(Y)|$
(2) $X \perp X'$ and $Y \perp Y'$ and $Z \perp Z'$,
(3) $V^m \perp Y$ and $W^n \perp Y'$ and $V^m \cdot W^n \perp Y \cdot Y'$, and
(4) $Z \succcurlyeq V^m \cdot Y \succcurlyeq X$ and $Z' \succcurlyeq W^n \cdot Y' \succcurlyeq X'$
imply
(5) $Z \cdot Z' \succcurlyeq (V^m \cdot Y) \cdot (W^n \cdot Y') \succcurlyeq X \cdot X'$.

The notation in which the axioms have been formulated is standard. For example, $|\pi(X)|$ denotes the cardinality of $\pi(X)$, while the numerical superscripts adorning the random variables refer to integer powers of the respective random variables. Thus:

$$V^m = \underbrace{V \cdot V \cdot \ldots \cdot V}_{m-\text{times}}$$

An alternative axiomatization underscores the reduction of independence from random variables to events:

QIR1.′ $\dot{A} \perp \dot{\Omega}$.
QIR2.′ $X \perp Y$ if and only if $\dot{A} \perp \dot{B}$ for every $A \in \pi(X)$ and $B \in \pi(Y)$.
QIR3.′ $\dot{A} \perp \dot{B}$ if and only if for all $X, X', Y, Y', Z, Z' \in \mathscr{A}^*$:

(1) $X \perp X'$ and $Y \perp Y'$ and $Z \perp Z'$,
(2) $\dot{A} \perp Y$ and $\dot{B} \perp Y'$ and $\dot{A} \cdot \dot{B} \perp Y \cdot Y'$, and
(3) $Z \succcurlyeq \dot{A} \cdot Y \succcurlyeq X$ and $Z' \succcurlyeq \dot{B} \cdot Y' \succcurlyeq X'$
imply
(4) $Z \cdot Z' \succcurlyeq (\dot{A} \cdot Y) \cdot (\dot{B} \cdot Y') \succcurlyeq X \cdot X'$.

Building upon the insights of Suppes and Alechina (1994), the following representation theorem can be proved.[2]

[2] *APP's Note*: The axiomatization advanced in the present paper differs from the axiomatization appearing in the submitted manuscript. The original axiomatization, based on that presented in §6.1 of (Suppes 2014), is incomplete in formulation and execution. In an effort to capture Suppes's original intentions, an alternative axiomatization that draws upon Suppes's earlier work (Suppes and Alechina 1994) has been formulated in the present paper. The remarks and technical discussion following Theorem 2 and concluding §4.2 are my own.

Theorem 2 *A structure* $(\Omega, \mathscr{A}^*, \succcurlyeq, \perp)$ *satisfies axioms* **CPR1** *to* **CPR6** *and* **QIR1** *to* **QIR3** *if and only if there is a unique real-valued function* $\mathbb{E}$ *on* $\mathscr{A}^*$ *such that for all* $X, Y \in \mathscr{A}^*$:

(i) $\mathbb{E}(X) \geq 0$;
(ii) $\mathbb{E}(\dot{\varnothing}) = 0$ *and* $\mathbb{E}(\dot{\Omega}) > 0$;
(iii) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
(iv) $X \succcurlyeq Y$ *if and only if* $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
(v) $X \perp Y$ *if and only if* $\mathbb{E}(\dot{A} \cdot \dot{B}) = \mathbb{E}(\dot{A}) \cdot \mathbb{E}(\dot{B})$ *for all* $A \in \pi(X)$ *and* $B \in \pi(Y)$.

In particular, if $\mathbb{E}$ *is a real-valued function satisfying properties* (i) *to* (v), *then* $\mathbb{E}(\dot{\Omega}) = 1$.

Of course, a real-valued function $\mathbb{E}$ satisfying properties (i) to (v) of Theorem 2 is an ordinary finitely additive expectation function on $\mathscr{A}^*$ which, when restricted to $\mathscr{A}$, is equivalent to a finitely additive probability function $\mathbb{P}$. The proof of 2 is omitted to due to space constraints. A key step in the proof invokes a result using Vandermonde determinates that Renyi (1970a, p. 120; 1970b, pp. 170–171) credits to L.V. Kantorvich.

Suppes and Alechina (1994) define pairwise independence for events within the formal context of a random-variable axiomatization of comparative probability due to Suppes and Zanotti (1976). Specifically, expressed in terms of the notation of the present paper, Suppes and Alechina (1994) define an event A to be independent of an event B just in case the following two conditions are satisfied:

(1) For all positive integers m, m', n, n':

$$n\dot{A} \succcurlyeq m\dot{\Omega} \text{ and } n'\dot{B} \succcurlyeq m'\dot{\Omega} \text{ imply } nn'(A \dot{\cap} B) \succcurlyeq mm'\dot{\Omega}.$$

(2) For all positive integers m, m', n, n':

$$n\dot{\Omega} \succcurlyeq m\dot{A} \text{ and } n'\dot{\Omega} \succcurlyeq m'\dot{B} \text{ imply } nn'\dot{\Omega} \succcurlyeq mm'(A \dot{\cap} B).$$

Suppes and Alechina (1994) show that if a structure $(\Omega, \mathscr{A}^*, \succcurlyeq)$ satisfies Suppes and Zanotti's random-variable axioms for comparative probability, then there is a unique real-valued function $\mathbb{E}$ satisfying properties (i) to (iv) of Theorem 1 above as well as the property that for all events $A, B \in \mathscr{A}$, $\mathbb{E}(\dot{A} \cdot \dot{B}) = \mathbb{E}(\dot{A})\,\mathbb{E}(\dot{B})$ just in case A is independent of B in accordance with conditions (1) and (2).

The present account of qualitative independence, by contrast, rests upon axioms formulated more generally for extended indicator functions. While axiom **QIR2** expresses a familiar property of pairwise independence that applies to extended indicator functions and not just to events, variants of this axiom appearing in well-known axiomatic presentations of qualitative independence have been formulated only to apply to events (e.g., Krantz et al. 1971, §5.8; Fine 1973, IIF). Of course, the random-variable axioms underlying the present account entail that properties (1) and (2) hold of $\dot{A}$ and $\dot{B}$ just in case $\dot{A} \perp \dot{B}$.

The foregoing developments only treat the concept of *pairwise independence*. The related concept of *mutual independence* can be explicated qualitatively in terms of the notion of pairwise independence (cf. Domotor 1970, §3.1; Fine 1972, IIF). To this end, define by recursion on positive integers n the n-ary relation $\perp\!\!\!\perp_n$ for each family $(X_i : i \in I) \in (\mathscr{A}^*)^I$ of extended indicator functions whose index set has cardinality $|I| = n$. For $n = 1$, a case included for the sake of convenience, define $\perp\!\!\!\perp_n$ by stipulating that $\perp\!\!\!\perp_n(X_i : i \in I)$ holds for every family $(X_i : i \in I)$ of extended indicator functions for which I is a singleton. For $n = m + 1$, assuming $\perp\!\!\!\perp_m$ has been defined for every family whose index set has finite cardinality m, define $\perp\!\!\!\perp_n$ by setting for every family $(X_i : i \in I) \in (\mathscr{A}^*)^I$ with $|I| = n$:

$$\perp\!\!\!\perp_n\Big(X_i : i \in I\Big) \quad \text{if and only if} \quad \text{for every } \big(\dot{A}_i : i \in I\big) \in \mathop{\times}_{i \in I} \pi(X_i) \text{ and } j \in I:$$

$$(1) \quad \perp\!\!\!\perp_m\Big(\dot{A}_i : i \in I \backslash \{j\}\Big) \quad \text{and}$$

$$(2) \quad \dot{A}_j \perp \prod_{i \in I \backslash \{j\}} \dot{A}_i.$$

An arbitrary family $(X_i : i \in I) \in (\mathscr{A}^*)^I$ is thereby said to be *mutually independent*, written $\perp\!\!\!\perp(X_i : i \in I)$,, if $\perp\!\!\!\perp_{|I_0|}(X_i : i \in I_0)$ for each nonempty finite subset $I_0 \subseteq I$. Observe that if I is finite, then $\perp\!\!\!\perp(X_i : i \in I)$, if and only if $\perp\!\!\!\perp_{|I|}(X_i : i \in I)$. It is readily verified that $\perp\!\!\!\perp$ agrees with the familiar numerical concept of mutual independence. That is, if $\mathbb{E}$ is a real-valued function satisfying properties (i) to (v) of Theorem 2, then $\perp\!\!\!\perp(X_i : i \in I)$, just in case for each nonempty finite subset $I_0 \subseteq I$ and family $\big(\dot{A}_i : i \in I_0\big) \in \times_{i \in I_0} \pi(X_i)$:

$$\mathbb{E}\left(\prod_{i \in I_0} \dot{A}_i\right) = \prod_{i \in I_0} \mathbb{E}\big(\dot{A}_i\big).$$

Of course, in anticipation of what's next, the notation for pairwise independent experiments may be extended to cover mutually independent experiments: A family of experiments $(\pi_i : i \in I)$ is said to be *mutually independent*, also written $\perp\!\!\!\perp(\pi_i : i \in I)$, if there is a family $(X_i : i \in I) \in (\mathscr{A}^*)^I$ such that $\pi_i = \pi(X_i)$ for every $i \in I$.

Random-Variable Axioms of Comparative Uncertainty

To ensure that the axioms for comparative uncertainty admit a numerical representation, axioms **CU1** to **CU1** must supplemented with additional axioms, including an Archimedean axiom. To the axioms for qualitative independence (and hence qualitative probability), we add the following axioms for comparative uncertainty.

CUR1. $\pi \succsim_u \xi$ and $\xi \succsim_u \vartheta$ imply $\pi \succsim_u \vartheta$.
CUR2. Either $\pi \succsim_u \xi$ or $\xi \succsim_u \pi$.

CUR3. $\pi \succcurlyeq_u \pi(\dot{\Omega})$.

CUR4. If $\pi_1 \perp \xi_1$, $\pi_1 \succcurlyeq_u \pi_2$, and $\xi_1 \succcurlyeq_u \xi_2$, then:

(1) $\pi_2 \perp \xi_2$ implies $\pi_1 \odot \xi_1 \succcurlyeq_u \pi_2 \odot \xi_2$;
(2) $\pi_2 \odot \xi_2 \succcurlyeq_u \pi_1 \odot \xi_1$ implies $\pi_2 \perp \xi_2$.

CUR5. $\pi \succcurlyeq_u \pi \odot \xi$ implies $\pi \odot \xi \approxeq \pi$.

CUR6. $\pi \approxeq \xi$ implies $\pi \approx_u \xi$.

The main problem we now face is to formulate a proper Archimedean axiom for uncertainty using the product operation for experiments. The change required is not drastic, but conceptually important. We use a concept familiar in the study of quantitative probability and statistics, but not in the study of qualitative probability. This is the concept of a sequence, finite or infinite, of mutually independent, identically distributed (IID) random variables.

Like the notion of mutual independence, the concept of *distributional equivalence* can be treated qualitatively within the present account. As before, we introduce the notion of distributional equivalence for extended indicator functions as well as experiments. Extended indicator functions X and Y from $\mathscr{A}^*$ are said to be *identically distributed*, written $X \approxeq Y$, if $X^{-1}(m) \approx Y^{-1}(m)$ for each nonnegative integer m. Similarly, experiments ρ and θ are said to be *identically distributed*, also written $\rho \approxeq \theta$, if there are extended indicator functions X and Y for which $\rho = \pi(X)$ and $\theta = \pi(Y)$ and $X^{-1}(m) \approx Y^{-1}(m)$ for each nonnegative integer m. More generally, given an extended indicator function $X \in \mathscr{A}^*$ and a family of extended indicator functions $(X_i : i \in I) \in (\mathscr{A}^*)^I$, if $X \approxeq X_i$ for every $i \in I$, then the family $(X_i : i \in I)$ is said to be *identically distributed* in accordance with X, written $X \approxeq (X_i : i \in I)$. Likewise, if a given experiment $\pi \approxeq \pi_i$ for every $i \in I$ of a given family of experiments $(\pi_i : i \in I)$, then $(\pi_i : i \in I)$ is said to be *identically distributed* in accordance with π, written $\pi \approxeq (\pi_i : i \in I)$.

In what follows, let $\bigodot_{i=1}^n \pi_i$ denote the n-fold product of a given family of experiments $(\pi_i)_{i=1}^n \in (\mathscr{A})^n$:

$$\bigodot_{i=1}^{n} \pi_i = \left\{ C \subseteq \Omega : \ C \neq \varnothing \ \text{ and } \ C = \bigcap_{i=1}^{n} A_i \ \text{ for some } \ (A_i)_{i=1}^n \in (\pi_i)_{i=1}^n \right\}.$$

Of course, the case where $n = 2$ agrees with the binary product $\cdot$ defined in "Qualitative Axioms for Probability" section.

Using these ideas, we can now formulate an appropriate Archimedean axiom (**CUR7**).

(i) $\perp\!\!\!\perp (\pi_i)_{i=1}^m$ and $\perp\!\!\!\perp (\xi_j)_{j=1}^n$ and $\perp\!\!\!\perp (\vartheta_k)_{k=1}^n$;

(ii) $\pi \approxeq (\pi_i)_{i=1}^m$ and $\xi \approxeq (\xi_j)_{j=1}^n$ and $\vartheta \approxeq (\vartheta_k)_{k=1}^n$;

$$\bigodot_{j=1}^{n} \pi_j \succ_u \bigodot_{i=1}^{m} \vartheta_i \succ \bigodot_{k=1}^{n} \xi_k. \tag{iii}$$

We next need to explain the quantitative measurement $\mathbb{H}$ of the entropy of an experiment. An excellent, clear argument is given by Khinchin (1957, pp. 2–9). The argument assumes elementary quantitative probability distributions (for finite domains). The concept of uncertainty is used informally and intuitively. I take a somewhat different path but reach the same result. I do so by initially restricting cases just to the maximum entropy φ for n atomic events, i.e., the uniform distribution for each n, the number of atomic events.

So by the basic principle already stated:

$$\varphi(1,\ 0,\ldots,0) = 0 \qquad (\textit{nouncertainty})$$
$$\varphi\left(\frac{1}{2},\frac{1}{2}\right) \prec \varphi\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right),$$

and so on. Simple additivity will not work, since the result would be $\varphi\left(\frac{1}{n},\ldots,\frac{1}{n}\right) = 1$ for n—and so, for example, $\varphi\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{2}+\frac{1}{2} = 1$ and $\varphi\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) = \frac{1}{3}+\frac{1}{3}+\frac{1}{3} = 1$.

Neither will simple multiplication, since the inequality is the opposite of what is desired:

$$\varphi\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} > \varphi\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3} = \frac{1}{27}.$$

So we try the logarithm of the product as a simple function. To keep $\varphi\left(\frac{1}{2},\frac{1}{2}\right)$ non-negative, since φ negative would have a peculiar uncertainty value if φ were below 0 for certainty, we need to take the negative of the logs of proper fractions to obtain positive uncertainty:

$$\varphi\left(\frac{1}{2},\frac{1}{2}\right) = -\log\left(\frac{1}{2}\cdot\frac{1}{2}\right) = -\left(\log\frac{1}{2}+\log\frac{1}{2}\right) = -2\log\frac{1}{2} = 2 \text{ (for log base 2)}$$

If we now weight each log by $\frac{1}{n}$, for each positive integer n, we have the following simple result for uniform distributions:

$$\varphi\left(\frac{1}{2},\frac{1}{2}\right) = \log 2\ ,$$
$$\varphi\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) = \log 3,$$
$$\varphi\left(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right) = \log n.$$

Moreover, it is easy to show that using this for each positive integer n, $\varphi(p_1,\ldots,p_n) \le \log n$. So we baptize φ as the entropy function $\mathbb{H}$, with the properties we considered important to match our intuitive idea of uncertainty:

$$\mathbb{H}\left(p_1, \ldots, p_n\right) = \varphi\left(p_1, \ \ldots, p_n\right) = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

Standard Representation Theorem for Uncertainty

Given the axioms presented in "Qualitative Random-Variable Axioms" section, the following theorem can be proved.

Theorem 3 *A structure* $(\Omega, \mathscr{A}^*, \succsim, \perp, (\mathscr{A}), \succsim_u)$ *satisfies axioms* **CPR1** *to* **CPR6**, **QIR1** *to* **QIR3**, *and* **CUR1** *to* **CUR7** *if and only if there is a unique real-valued function* $\mathbb{E}$ *on* $\mathscr{A}^*$ *satisfying properties* (i) *to* (v) *of Theorem 2 as well as an entropy function defined on the set of experiments over* Ω*—that is, a real-valued function* $\mathbb{H}$ *on* $(\mathscr{A})$ *such that for all* $\pi, \xi \in (\mathscr{A})$:

(1) $\mathbb{H}(\pi(\dot{\Omega})) = 0$
(2) $\pi \approxeq \xi$ *implies* $\mathbb{H}(\pi) = \mathbb{H}(\xi)$
(3) $\mathbb{H}(\pi) + \mathbb{H}(\xi) \geq \mathbb{H}(\pi \odot \xi)$, *with equality if and only if* $\pi \perp \xi$;
(4) $\mathbb{H}(\pi) \geq \mathbb{H}(\pi \odot \xi)$ *implies* $\pi \approxeq \pi \odot \xi$;
(5) $\pi \succsim_u \xi$ *if and only if* $\mathbb{H}(\pi) \geq \mathbb{H}(\xi)$;
(6) *For each* $X \in \mathscr{A}^*$ *and positive integer* n, *there exist* n *mutually independent random variables* $X_1, \ldots, X_n$ *in* $\mathscr{A}^*$, *each with the same distribution as* X.

Moreover, if $\mathbb{H}$ *and* $\mathbb{H}'$ *are real-valued functions on* $(\mathscr{A})$ *satisfying properties* (1) *to* (6) , *then there is a positive real number* λ *such that* $\mathbb{H} = \lambda\,\mathbb{H}'$; *that is, any function* $\mathbb{H}$ *satisfying properties* (1) *to* (6) *is unique up to a positive similarity transformation.*

The proof of Theorem 1 applies techniques similar to those in the theory of measurement (see, e.g., Krantz et al. (1971)). The proof has been omitted to due to space limitations.

Observe that the probability function $\mathbb{P}$ on $\mathscr{A}$ corresponding to the expectation function guaranteed by Theorem 3 is *strongly continuous* in the sense that for every real number $\epsilon > 0$, there is a partition $\{E_1, \ldots, E_n\} \subseteq \mathscr{A}$ of Ω (i.e., an experiment) such that $\mathbb{P}(E_i) < \epsilon$ for every $i = 1, \ldots, n$. It follows that if $\mathscr{A}$ is σ-field, then $\mathbb{P}$ is in fact *strongly nonatomic* in the sense that for every $A \in \mathscr{A}$ and real number $\epsilon \geq 0$, if $\epsilon \leq \mathbb{P}(A)$, then there exists $B \in \mathscr{A}$ such that $B \subseteq A$ and $\mathbb{P}(B) = \epsilon$ (Rao and Rao 1983, p. 142); this would also follow if $\mathscr{A}$ satisfied a slightly weaker condition called the *Seever property* Rao and Rao 1983, p. 210). If $\mathbb{P}$ is strongly nonatomic, it follows from a result due to (Aczél et al. 1974, p. 135) that the real-valued function $\mathbb{H}$ guaranteed by Theorem 1 is in fact a mixture of the Shannon entropy measure and the so-called *Hartley* entropy measure. The Hartley measure ψ of an experiment

π is $\log N(\pi)$, where $N(\pi)$ is the number of elements of π with nonzero probability. Thus, Theorem 1 implies the following corollary.

Corollary 1 *If the field of events $\mathscr{A}$ in Theorem 3 is a σ -field (or satisfies the Seever property), then the axioms of Theorem are necessary and sufficient for there to exist a unique real-valued function $\mathbb{E}$ on $\mathscr{A}^*$ satisfying the stated conditions as well as an entropy measure $\mathbb{H}$ that is a mixture of the Shannon entropy measure φ and the Hartley entropy measure ψ—that is, there are nonnegative real numbers α and β such that for every $\pi \in (\mathscr{A})$:*

$$\mathbb{H}(\pi) \quad = \quad \alpha\varphi(\pi) \; + \; \beta\psi(\pi).$$

Bounded Uncertainty

I begin the analysis of bounded uncertainty by considering the classical definition of bounded rationality given by Herbert Simon (1957, p. 198):

> *The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world — or even for a reasonable approximation to such objective rationality.*

Of course, Simon spends many pages expanding on this thesis and claiming that it is a great mistake of the behavioral sciences not to recognize the limited capacities humans have for solving the complicated problems they are presented with. He especially objects to a simple classical notion of the "economic man." It is not my task here to criticize what he has to say, or even to state what I agree with. Nonetheless, the general thesis about the capacities of humans, or even of organizations of humans, to solve complex problems being clearly limited is a truism that in today's environment can hardly be rejected by any informed person.

My limited task is just to explain, and give at least an example or two, of the importance of thinking about bounded uncertainty in the design of experiments, and even, sometimes, in the design of the analysis of data. Fortunately the ideas of independence and of probability distributions have already been introduced in earlier sections of this article. These ideas can be used to state in an efficient and clear way the problem of asking for too much in the design of experiments.

For the more detailed analysis, let us consider one of the simplest possible experiments, which is that of tossing a coin of unknown bias repeatedly in a way that produces independent trials, in order to estimate the bias, or to put it another way, the probability of getting heads in such repeated trials. In formal statistical terms, this is a problem of studying the behavior of independent random variables that are identically distributed, and the problem I have selected is among the simplest possible that can fit this description. So for this simplest of all problems, how many tosses, or repeated trials, if you wish, should be planned for? (I leave out the Bayesian aspects of this discussion just to simplify and emphasize the main points I want to make.)

Let $\mathbf{A}_k$ be the indicator function representing the kth trial of tossing the coin. This random variable can have only two values. Here let 1 be the value of $\mathbf{A}_k$ for the outcome of the toss resulting in "heads," and let 0 be the value for the outcome of the toss resulting in "tails." We can easily compute the entropy of the k^{th} trial as a function of its unknown probability p_k, which is in explicit terms the probability $p(\mathbf{A}_k = 1)$. Using notation and ideas from the earlier sections, we can write down the equation for the entropy $\mathbb{H}(\mathbf{A}_k)$, which is expressed in the following familiar equation:

$$\mathbb{H}\left(\mathbf{A}_k\right) = -\left(p_k \log p_k + (1 - p_k)\log(1 - p_k) \right).$$

Usually the entropy is computed in terms of log base 2. Given this equation and the additivity of entropy for independent random variables, we can then easily compute the total entropy of an experiment with n trials, which is the following expression:

$$\mathbb{H}\left(\sum_{k=1}^{n} \mathbf{A}_k\right) = \sum_{k=1}^{n} \mathbb{H}(\mathbf{A}_k) = n\,\mathbb{H}(\mathbf{A}_k),$$

Since the entropy of a single trial is positive, the entropy of n IID trials will be positive and monotonically increasing in uncertainty.

If there were no constraints on cost or time, the ideal experimental investigation of any sort would benefit from having unbounded uncertainty of outcomes in a sequence of independent and identically distributed random variables. Put in purely statistical terms, the more IID trials the better the test of the null hypothesis, whatever it may be. But the slogan of "the more the better" is naïve if cost and time are not considered. To put it simply, ideal investigations are idle to think about if no concerns for cost and time are considered. Nonetheless, many aspects of mathematic statistics reflect a concern for stopping rules in the design of experiments; the relevant literature is now very large, and entropy measurements of various kinds are considered.

The limited purposes of this article are, first, to show that at a qualitative foundational level, uncertainty, as measured by entropy, can be introduced as a fundamental concept on a par with comparative probability; and second, cannot be optimized in any pure sense of considering only uncertainty, as well represented statistically by large collections of independent identically distributed random variables, but must be bounded by external constraints of cost and time in any real-world regime of experimentation.

There is one important problem that might lead to confusion about the role of uncertainty in determining the number of trials to run in a given investigation. This is the very different role of the study of independent and identically distributed random variables that have been studied so intensely from a mathematical standpoint in the limit theorems that describe their asymptotic behavior. This is one of the oldest serious mathematical topics in statistics, beginning as it does, with the early work of Bernoulli, de Moivre and Laplace in the 18^{th} century. But the objective of those studies, important as they are, is quite different from the concrete design-of-experiment question of how many IID trials should be run to provide a practical test of significance of the given hypothesis.

This general topic goes under the title of “stopping rules” in modern statistics. As the name suggests this is a study of how many replications are needed, but still practical, to give confidence about the outcome. Such studies are of particular importance when they involve, for example, the prescription of a new drug for a disease or illness, but they are of importance in all kinds of other topics as well, such as the design of airplane engines or agricultural regimes for rotating crops.

It is not possible here to go into the mathematical and technical statistical literature on stopping rules, but it will be worthwhile trying to give some idea of how stopping rules are in themselves related quite directly to the concept of uncertainty as developed in this article.

In the extensive literature on stopping rules, computations of entropy are uncommon. I believe it would be desirable to carefully study additional computations, or rather additional use of such computations, in the formulation of stopping rules. I sketch a simple case.

The example I consider, is testing the simple theoretical hypothesis that a coin is unbiased. I write this hypothesis as $\mathbb{H}_T\left(\frac{1}{2},\frac{1}{2}\right)$, where T stands for ‘theoretical’ and $\left(\frac{1}{2},\frac{1}{2}\right)$ corresponds to the probability $\frac{1}{2}$ of “heads” and $\frac{1}{2}$ of “tails” on a single toss of a given coin. So the entropy of this hypothesis on a single toss is

$$\begin{aligned}\mathbb{H}_T\left(\frac{1}{2},\frac{1}{2}\right) &= -\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2}\\ &= -\log\frac{1}{2}\\ &= -\log 1 + \log 2\\ &= 1.\end{aligned}$$

So, in theory, the entropy of n tosses is simply

$$n\mathbb{H}_T\left(\frac{1}{2},\frac{1}{2}\right) = n.$$

Our experiment π^n of n trials to test this hypothesis is a sequence of n IID random variables as already described. Since the probability p_k of “heads” in the k^{th} trial of the experiment is, in terms of entropy,

$$\mathbb{H}_T\Big(p_k,(1-p_k)\Big) = -p_k\log p_k - (1-p_k)\log(1-p_k),$$

we must estimate p_k from the n IID experimental tosses of the coin from the actual sequence of n outcomes of trials, the main empirical result of the experiment π^n. We can do so by estimating p_k from the empirical data, with some using the mean number of “heads” in the empirical sequence of outcomes, and, no doubt, some also using a Bayesian prior in estimating p_k. Another possibility, which I will not explore here, is to use the theory of Kolmogorov complexity (1965, 1968) to estimate more directly the entropy of the actual sequence of "heads" and "tails" in the experiment. However the estimate is made, let $\mathbb{H}_E(\pi^n)$ be the experimental entropy.

The proposal is then to replace the standard statistical computations for accepting or rejecting the theoretical hypothesis $\mathbb{H}_T\left(\frac{1}{2}, \frac{1}{2}\right)$ by analyzing the absolute difference of uncertainty, as measured by entropy:

$$\triangle\mathbb{H} = \left| n\mathbb{H}_T\left(\frac{1}{2}, \frac{1}{2}\right) - \mathbb{H}_E(\pi^n) \right|$$

It is the quantity $\triangle\mathbb{H}$ of entropy difference that should be an object of statistical analysis. Moreover, in any real experiment $\triangle\mathbb{H}$ is finite and non-negative. Indeed, not just finite, but bounded before the actual outcome is realized by the design of the experiment. This last point is the critical one in the present context. The necessarily bounded uncertainty of the experimental design of any actual experiment implies one important aspect of bounded rationality.

References

Aczél, J., Forte, B., & Ng, C. T. (1974). Why the Shannon and Hartley entropies are 'natural'. *Advances in Applied Probability*, *6*, 131–146.

de Finetti, B. (1937/1964). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, *7*, 1–68, Translation in Kyburg and Smokler (1964) [93–158].

de Finetti, B. (1974). *Theory of probability* (Vol. 1). New York: Wiley. Translated by A. Machi and A. Smith.

de Finetti, B. (1975). *Theory of probability* (Vol. 2). New York: Wiley. Translated by A. Machi and A. Smith.

Domotor, Z. (1969). *Probabilistic relational structures and their applications*. Tech. Rep. 144, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA.

Domotor, Z. (1970). Qualitative information and entropy structures. In J. Hintikka & P. Suppes (Eds.), *Information and inference* (pp. 148–194). Holland: Reidel.

Fine, T. (1973). *Theories of probability: An examination of foundations*. London: Academic Press.

Khinchin, A. (1953). The entropy concept in probability theory (in Russian). *Uspekhi Matematicheskikh*, *8*(3), 3–20. nauk.

Khinchin, A. (1956). On the fundamental theorems of information theory (in Russian). *Uspekhi Matematicheskikh*, *11*(1), 17–75. nauk.

Khinchin, A. (1957). *Mathematical foundations of information theory*. New York: Dover Publications Inc.

Kolmogorov, A. N. (1933/1950). *Foundations of the theory of probability*. New York: Chelsea Publishing Company. Translated by N. Morrison. First published in German in 1933.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, *1*, 4–7.

Kolmogorov, A. N. (1968). Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, *IT–14*, 662–664.

Kraft, C. H., Pratt, J., & Seidenberg, A. (1959). Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, *30*, 408–419.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement Vol. I: Additive and polynomial representations*. New York: Academic Press.

Rao, K. P. S., & Rao, M. (1983). *Theory of charges: A study of finitely additive measures*. London: Academic Press.

Rényi, A. (1970a). *Probability theory*. Amsterdam: Elsevier.

Rényi, A. (1970b). *Foundations of probability*. San Francisco, CA: Holden-Day.

Shannon, C. (1948). The mathematical theory of communication. *Bell System Technical Journal*, *27*(379–423), 623–656.

Simon, H. A. (1957). *Models of man: Social and rational; mathematical essays on rational human behavior in a social setting*. New York: Wiley.

Suppes, P. (2014). Using Padoa's Principle to prove the non-definability, in terms of each other, of the three fundamental qualitative concepts of comparative probability, independence and comparative uncertainty, with some new axioms and definitions of qualitative uncertainty included. *Journal of Mathematical Psychology*, *60*, 47–57.

Suppes, P., & Zanotti, M. (1976). Necessary and sufficient conditions for existence of a unique measure strictly agreeing with a qualitative probability ordering. *Journal of Philosophical Logic*, *5*, 431–438.

Suppes, P., & Alechina, N. (1994). The definability of the qualitative independence of events in terms of extended indicator functions. *Journal of Mathematical Psychology*, *38*(3), 366–376.