Journal of Mathematical Psychology 60 (2014) 47–57

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

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 of qualitative independence and uncertainty included☆

P. Suppes

CSLI, Ventura Hall, 220 Panama Street, Stanford University, Stanford, CA 94305-4101, United States

HIGHLIGHTS

- Mutual non-definability of the three fundamental qualitative probability concepts.
- Representation theorem for comparative probability as an extensive quantity.
- Random-variable axioms of independence in terms of indicator functions.
- Qualitative uncertainty as a fundamental probability concept measured by entropy.

ARTICLE INFO

Article history:
Received 23 April 2013
Received in revised form
6 June 2014

Keywords:
Qualitative probability
Qualitative independence
Qualitative uncertainty
Entropy
Padoa's principle

ABSTRACT

First, Padoa's principle is used to prove the non-definability of the fundamental qualitative concepts of comparative probability, independence and comparative uncertainty in terms of each other. Second, the qualitative axioms of uncertainty leading to an entropy representation are new. Third, a qualitative random-variable axiomatization of these concepts is given, but the random variables are restricted to generalized indicator functions, their products and their iterates. A new axiom of independence in terms of such indicator functions is used in this axiomatization. Fourth, a standard extensive-quantity representation is then proved for comparative probability, and the new axiom of independence provides the basis for proving the desired absolute invariance theorem for the constructed probability measure.

1. Introduction

The first purpose of the present article is to show that, given purely qualitative axioms of comparative probability, independence and comparative uncertainty (Section 3), none of the three concepts can be defined in terms of the other two (Section 4). As far as I know, such a proof of the mutual non-definability of these qualitative probability concepts is not to be found anywhere else, although Padoa's principle for proving non-definability of a concept is well-known in the literature of logic and the foundations of mathematics. Yet, the standard response might well be, "Why all the fuss about this, just wait until you have a unique measure normed to one, then everything can be defined in terms of this measure", well, the answer is that the two assumptions that the measure is unique and norms to one do not hold for the purely qualitative conceptual axioms. To emphasize this point, having all the quantitative properties, including uniqueness of such standard probability measures available, introduces the possibility of defining concepts that are, in a purely intuitive and qualitative sense, clearly not so definable. Understanding this distinction is important for everyone, but especially students of probability, in order to appreciate independence and particularly uncertainty, as separate concepts in their own right.

It is perhaps more accurate to say that independence is usually recognized as a separate concept that states a property used to formulate a restriction needed in most structures to make them scientifically manageable. For example, if the present depends

☆ I am indebted to Theodore Alper and Tomohiro Hoshi for detailed criticisms and comments on an early draft, and to audiences at Stanford, University of California Irvine, University of Utrecht and Ludwig Maximilian University of Munich for many useful remarks. Claudio Carvalhaes helped compute Tables 1–4 and correct a number of oversights in a later draft. Marcos Perreau-Guimaraes has suggested some useful improvements in the "final" draft. The double quotes reflect the fact that the anonymous reviewers, selected by the editor, contributed even more to this version.

E-mail address: psuppes@stanford.edu.
URL: http://suppes-corpus.stanford.edu/.

http://dx.doi.org/10.1016/j.jmp.2014.06.003
0022-2496/

on the past, but in ways we cannot specify and limit, the resulting temporal processes – stochastic processes in the language of probability – would not be understandable in any detail. Fortunately, many natural processes closely approximate first-order Markov processes, which justifies the working assumption that in many cases the present is independent of all but the immediate past. The second purpose is to axiomatize qualitative uncertainty. In the older literature, the words "probability" and "uncertainty" were used as approximate synonyms. For example, in a book I much admire, Harold Jeffreys' *Theory of Probability*, the second edition (1948) states the *principle* of *uncertainty* this way (Jeffreys, 1948):

> "Hence a physical law is not an exact prediction, but a statement of the relative probabilities of variations of different amounts. It is only in this form that we can avoid rejecting casuality altogether as false, or as inapplicable..." (pp. 13–14)

Even as late as 1970, the two-volume treatise by Bruno de Finetti on the theory of probability (de Finetti, 1974, 1975, English translation) has no index reference to "uncertainty" or systematic discussion of the concept. In Volume 2, p. 150, there is casual mention of "the calculus of probability (the logic of uncertainty) is completely neutral with respect to the facts...". This absence of systematic separation of the concept of probability and that of uncertainty is especially notable, given deFinetti's sensitivity to the linguistic use of probability concepts in Italian, English, French and German. The one mention of *entropy* in Volume 2 is not developed.

The modern concept of uncertainty and its entropy measure did not exist, really, before the highly original work of Shannon (1948), which was given, not much later, a clear mathematical formulation (in Russian) by Khinchin (1953, 1956), and then published in English in Khinchin (1957) in a widely circulated Dover edition. The basic idea has become a centerpiece of modern information theory, an excellent exposition of which is given in Cover and Thomas (1991). At a more advanced level, the central role of the concept of entropy in ergodic theory is evident in, for example, the detailed analytic review of Ornstein and Weiss (1991). Cover and Thomas (1991) properly like to emphasize that entropy, as the measure of uncertainty, plays the role in the modern theory of information that measures of probability play in classical probability theory.

In the tradition of the theory of measurement, qualitative axioms of extensive measurement that are strong enough to prove the existence of a positive measure, unique up to a positive similarity transformation, are given in Section 5.2. It is important to be more explicit about these extensive-quantity axioms. The restricted set of random variables consisting of the indicator functions of events, their products and their iterates, e.g., $\mathbf{A}^i\mathbf{B}^i$, $\mathbf{A}^i + \mathbf{A}^i = 2\mathbf{A}^i$, etc. are added to the qualitative structure of events.

With this additional set of random variables and their qualitative relations available, it is then possible, as our third objective, to formulate a set of weak qualitative random-variable axioms to construct a strictly agreeing probability measure (Section 5). Stronger axioms using extended indicator functions (the set of indicator functions closed under addition) were given by Suppes and Zanotti (1976), from which can be proved the uniqueness of the probability measure up to a positive similarity transformation for any finite Ω, i.e., any finite set of possible outcomes.

Finally, a fourth purpose of this article is to show by a completely elementary argument using two qualitative axioms of independence (QIR1 and QIR6) probability must satisfy a stronger invariance principle than uniqueness up to a positive similarity transformation. It must be absolutely unique. The proof is given in Section 6. This proof makes explicit an informal point of Kolmogorov in his well-known book on the foundations of probability (Kolmogorov, 1933/1950), that independence is the concept that distinguishes probability as a positive measure. Historically, this was recognized later. In earlier work, such as that of de Moivre (1712) and Laplace (1812), the problem of the absolute uniqueness of the probability measure was solved by introducing a ratio definition, which obscures Kolmogorov's fundamental insight.

It is worth mentioning that problems also arise for Bayesians. For example, the well-known axioms of Savage (1954) prove the standard expected utility theorem, namely, that decision D_1 should be preferred to decision D_2 if and only if the expected utility of D_1 is greater than that of D_2. But this theorem assumes, as stated by Savage, a probability measure normed to one. Examination of the axioms and the basic theorem shows that multiplication of this measure by an arbitrary positive constant does not disturb anything, so that, in principle, the results are also applicable to probability as a standard extensive quantity. The necessary additional axioms and definitions are missing from Savage's formulation. The same remarks apply to Ramsey's proto-axioms of Ramsey (1931) or de Finetti's qualitative axioms (de Finetti, 1937/1964). Of course, de Finetti's qualitative axioms had to be strengthened to overcome the counterexample given by Kraft, Pratt, and Seidenberg (1959) that for these weak qualitative axioms, even the existence of a strictly agreeing measure cannot be proved. The problem of existence, but not uniqueness was solved positively by Scott (1964), who significantly added an axiom using the indicator functions of the events, instead of just the events themselves. When the expectation of acts or decisions is the main foundational focus, rather than the probability of events, as is the case in Ramsey (1931) and Savage (1954), probability can be defined as the dimensionless ratio of utility differences. This is not made explicit by either Ramsey or Savage. It is nearly so in the development of Ramsey's approach used in Davidson and Suppes (1956) and Suppes (1956).

2. Some preliminary formal definitions

First, let Ω be a nonempty set, which is the *domain* of possible outcomes. For any finite set A, let $|A|$ be the number of elements in A. Let $\mathcal{F}$ be a nonempty family of subsets of Ω, and the sets that are elements of $\mathcal{F}$ are the *events*, with the usual intuitive interpretation in probability. They are, by definition, sets of possible outcomes. If $\mathcal{F}$ is closed under union and complementation, it is an *algebra* of events. If it is also closed under countable unions, it is a σ-*algebra* of events. (In the older literature, such algebras were called fields.) In subsequent sections, I assume without remark the elementary set-theoretical properties of an algebra of events.

A *partition* of a nonempty set Ω is a nonempty family π of nonempty subsets of Ω, which are mutually exclusive and whose union equals Ω. These subsets are called *elements* of the partition π.

A partition π of Ω is an *experiment* if all of its elements are events, i.e., sets that are in $\mathcal{F}$. (Note that by this definition any experiment π is a subset of $\mathcal{F}$.) Random variables defined on Ω define partitions of Ω that are experiments. For example, let $\mathbf{X}$ be a random variable defined on Ω with integer values $1, \ldots, n$. Then the elements of the partition generated by $\mathbf{X}$ are defined by values of $\mathbf{X}$. For example, the sets A_i, $i = 1, \ldots, n$, where

$$A_i = \{\omega : \mathbf{X}(\omega) = i,\ \omega \in \Omega\},$$

are the elements of the experiment, which is a subset of $\mathcal{F}$, and so they are events. The *uncertainty* of a random variable depends only on the uncertainty of its generated partition, not on its values as such. An equivalent way of saying this is that the uncertainty or randomness of a random variable depends on its distribution, not its values. (What is here, and in the theory of information, called *uncertainty* is, in the terminology of ergodic theory, *randomness*. See, for example, the terminology used in Ornstein and Weiss (1991).) Hereafter, I shall use the term *experiment*, but also refer to random variables.

Set-theoretically, the product $*$ of two partitions π_1 and π_2 of Ω is defined as follows:

$$\pi_1 * \pi_2 = \{A_i \cap B_j : \text{for all } A_i \in \pi_1, B_j \in \pi_2\} - \{\emptyset\}.$$

$\pi_1 * \pi_2$ is a partition of Ω, and is the coarsest refinement of both π_1 and π_2. The product $*$ is associative, commutative, idempotent $(\pi * \pi = \pi)$ and for the trivial partition $\pi(\Omega)$, Ω is also the only element of this partition,

$$\pi * \pi(\Omega) = \pi.$$

In other words, $\pi\ (\Omega)$ is an identity element with respect to the product operation for partitions.

Another such set-theoretical concept, already mentioned informally, is also needed. The *indicator function* of a set is just the function that assigns the value 1 to elements of the set and the value 0 to all other elements of Ω. For simplicity of notation, if A is a set we shall denote by $\mathbf{A}^i$ its indicator function, which is, in a probability context, a random variable, if A is an event. (Note that random variables are ordinarily printed in boldface.) Thus,

$$\mathbf{A}^i(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

There is a subtle point about indicator functions that is critical in determining, independent of the kind of quantity being measured, the unit of measurement, which for standard extensive quantities such as length and weight, are not absolutely unique, but only up to a positive similarity transformation, which in the present case amount to multiplication by a positive real number. This standard result, discussed in many places, generates a contradiction with taking the positive value of an indicator function always to be 1, as in Eq. (1).

As an example of this problem, let Ω be a nonempty, finite set of weights, I want to measure in some standard unit such as grams or pounds. For example, assigning a uniform distribution to the elements of Ω, the set of all atomic events, will depend on the size of Ω. If Ω_1 has n events and $\Omega_2 n + 1$ events, the Ω_1-measure is $\frac{1}{n}$, for n atomic events, but $\frac{1}{n+1}$ for the Ω_2-measure of atomic events are being produced to match a 1 gram standard. This is not what we want, and not what is done in practice. An extensive measure like weight, with a qualitative random-variable axiomatization, should have a unit only unique up to a positive similarity transformation. Whereas, in terms of standard indicator functions

$$E(\mathbf{\Omega}_1^i) = P(\Omega_1) = E(\mathbf{\Omega}_2^i) = P(\Omega_2) = 1,$$

which is a unique measure, without any substantive random-variable axiom to support it.

So the subtle point is that the uniqueness up to a positive similarity transformation must be reflected in a *generalized* indicator function, with a parameter α identical to $E(\mathbf{\Omega}^i) = \alpha$, and subject to the standard positive similarity transformations, characteristic of extensive measurement. So (1) is generalized to:

$$\mathbf{A}^i(\omega) = \begin{cases} \alpha > 0 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

As is shown later, the property of qualitative statistical independence will require that $\alpha = 1$ which, of course, is not the case for the general theory of extensive quantities.

The concepts of set, partition and indicator function are purely set-theoretical concepts that do not in themselves depend in any way on probability theory. But when they are used in probability theory, they have a special status.

Sets that are elements of the algebra $\mathcal{F}$ of events are themselves *events*, and partitions that are subsets of $\mathcal{F}$ are formally *experiments*. I follow here the terminology of Kolmogorov (1933/1950). I do have some reservations about this terminology. In modern information theory, random variables have uncertainty, not experiments, and the entropy measure of uncertainty is usually defined for random variables, not experiments. In fact, the influential book on information theory by Cover and Thomas (1991) does not even index *experiment*. As for the third concept mentioned, indicator functions, when used in probability theory, are ordinarily random variables, and that is true here. (A *random variable* is a real-valued function $\mathbf{X}$ defined on Ω such that for each real number x, $\{\omega : \mathbf{X}(\omega) < x\}$ is in $\mathcal{F}$.)

A probabilistic concept needed in the formulation of the axioms is qualitative independence, $A \perp B$, to symbolize the qualitative independence of events in $\mathcal{F}$. The quantitative-probability definition for a finite algebra of events is that for events A and B in $\mathcal{F}$, they are independent if and only if $P(A \cap B) = P(A)P(B)$. Given the independence of pairs of events, two experiments $\pi_1 = \{A, \ldots, A_m\}$ *and* $\pi_2 = \{B, \ldots, B_n\}$, $\pi_1 \perp \pi_2$ *iff* $A_i \perp B_j, i = 1, \ldots, m$ *and* $j = 1, \ldots, n$.

If π is an experiment with n elements, with p_i the probability of element i, $i = 1, \ldots, n$, the entropy $H(\pi)$ of the experiment or random variable is defined as:

$$H(\pi) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{2}$$

(The choice of the base of the logarithm is optional, but 2 is common in information theory and computer science.)

It is easy to show that for finite sets of outcomes, the entropy measure H, as just defined in (2), has the following elementary properties for experiments or random variables:

(i) $H\ (\pi) \geqslant 0$,
(ii) $H\ (\pi\ (\Omega)) = 0$,
(iii) If $\pi_1 \perp \pi_2$, then $H\ (\pi_1 * \pi_2) = H\ (\pi_1) + H\ (\pi_2)$.

In stating the qualitative axioms of Section 3 and later, *iff* is sometimes used as an abbreviation for *if and only if*.

3. Purely qualitative axioms of comparative probability, independence and comparative uncertainty

The qualitative axioms introduced in this section fall into three groups. First, are the purely qualitative comparative probability axioms, probably first formulated clearly by de Finetti (1937/1964), although there are earlier predecessors in the literature, such as Bernstein (1917), mentioned by Kolmogorov. The single primitive relation of these axioms is the qualitative one $A \succcurlyeq B$, whose meaning is that event A is at least as probable as event B.

Qualitative Comparative Probability Axioms

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

For later purposes we define the probabilistic equivalence relation $\approx$ in the standard fashion. For events A and B

$A \approx B$ *iff* $A \succcurlyeq B$ *and* $B \succcurlyeq A$.

Second are the qualitative axioms on independence. They use just the binary relation $A \perp B$, i.e., A is independent of B. All of these axioms are stated for the independence of two events, but can be generalized to n events in a straightforward way, as is done later. The sources of these axioms are varied. The formal qualitative axioms, but not the standard quantitative definition in probability, have a recent history. A good summary of the earlier qualitative literature, with some significant additions, was given by

Zoltan Domotor (1969, 1970). A related study contributing some new axioms and an excellent discussion, can be found in Fine (1973), and also another qualitative axiom and some positive results in Luce and Narens (1978). A definition of $A \perp B$ in terms of indicator functions was given by Suppes and Alechina (1994). That definition is much simplified here.

Because independence is not a comparative qualitative order relation like comparative probability or comparative uncertainty, what qualitative axioms to assume is less clear. This point becomes obvious on reading the detailed and complex analysis given in Fine (1973). Moreover, the analysis given in Domotor (1969) is even longer and more complicated. Domotor proves a long list of qualitative properties of the binary relation $\perp$ of independence. However, more than half of these properties of $A \perp B$ use in their formulation the quadratic binary product

$$P(A \times B) = P(A)P(B),$$

and a separate axiomatization must be given of this nonstandard relation, which Domotor does.

Rather than include these "quadratic" properties of binary independence, I prefer later to axiomatize the desired properties by using the qualitative random variables that include the generalized indicator functions of events. This random-variable axiomatization is given in Section 5.

The qualitative axioms assumed in this section are the following five:

Qualitative Independence Axioms

QI1. $A \perp \Omega$.
QI2. *If* $A \perp A$, *then* $A \approx \Omega$ *or* $A \approx \emptyset$.
QI3. *If* $A \perp B$ *then* $B \perp A$.
QI4. *If* $A \perp B$ *then* $A \perp \neg B$.
QI5. *If* $A \perp B, A \perp C$ *and* $B \cap C = \emptyset$, *then* $A \perp B \cup C$.

Qualitative axioms of the third set are those of comparative uncertainty. These axioms have, as far as I know, not been previously stated in the qualitative form given here. The basic binary relation is that of experiment π_1 being at least as uncertain as π_2, written $\pi_1 \succcurlyeq_u \pi_2$, using, as you can see, a small subscript u for uncertainty. The measure H of uncertainty of an experiment, or a random variable is the standard entropy, as defined in Section 2, and discussed later in Section 7.2.

So the finite qualitative structures of this section are of the type, $(\Omega, \mathcal{F}, \succcurlyeq, \perp, \succcurlyeq_u)$, where experiments of Ω are required to be subsets of $\mathcal{F}$.

The five axioms of comparative uncertainty are formally very similar to the five axioms of comparative probability. In the critical qualitative axiom CU5, the condition of null intersection in CP5 is replaced by that of independence of experiments.

Qualitative Comparative Uncertainty Axioms for Experiments

CU1. *If* $\pi_1 \succcurlyeq_u \pi_2$ *and* $\pi_2 \succcurlyeq_u \pi_3$, *then* $\pi_1 \succcurlyeq_u \pi_3$.
CU2. $\pi_1 \succcurlyeq_u \pi_2$ *or* $\pi_2 \succcurlyeq_u \pi_1$.
CU3. $\pi \succcurlyeq_u \pi(\Omega)$.
CU4. *If* $\pi_1 \succ_u \pi_2$ *then* $\pi_1 \succ_u \pi(\Omega)$.
CU5. *If* $\pi_1 \perp \pi_3$ *and* $\pi_2 \perp \pi_3$, *then* $\pi_1 * \pi_3 \succcurlyeq_u \pi_2 * \pi_3$ *iff* $\pi_1 \succcurlyeq_u \pi_2$.

The first theorem, stated immediately after this comment, illustrates how a single non-necessary additional qualitative axiom, added to those already given in this section, can determine uniquely a quantitative model of the axioms. This model itself determines a unique joint probability distribution of all the probability atoms in $\mathcal{F}$, a quantitative relation of probabilistic independence, and a unique quantitative entropy as the measure of uncertainty.

Theorem 1. *Let* $\mathbf{\Omega} = (\Omega, \mathcal{F}, \succcurlyeq, \perp, \succcurlyeq_u)$ *be a finite qualitative probability structure satisfying the qualitative axioms* CP(1–5), QI(1–5) *and* CU(1–5). *In addition, let* $\mathbf{\Omega}$ *satisfy the uniformity condition: For any two events A and B in* $\mathcal{F}$, $A \approx B$ *iff* $|A| = |B|$, *where* $|A|$ *is the number of members of A, so, for example, if* $a \neq b$, $|\{a, b\}| = 2$. *Then there exist a unique probability measure P and a unique uncertainty measure H such that,*

(i) $P(\Omega) = 1 = |\Omega|/|\Omega|$,
(ii) $P(A) = |A|/|\Omega|$,
(iii) $A \succcurlyeq B$ *iff* $|A| \succcurlyeq |B|$ *iff* $P(A) \succcurlyeq P(B)$,
(iv) $A \perp B$ *iff* $\frac{|A \cap B|}{|\Omega|} = \frac{|A|}{|\Omega|} \cdot \frac{|B|}{|\Omega|}$ *iff* $P(A \cap B) = P(A)P(B)$,
(v) *If* $\pi = \{A_1, \ldots, A_n\}$, *then the entropy* $H(\pi) = -\sum_{i=1}^{n} \frac{|A_i|}{|\Omega|} \log \frac{|A_i|}{|\Omega|} = -\sum_{i=1}^{n} p_i \log p_i$, *where* $p_i = P(A_i)$,
(vi) *If the experiment* $\pi_1 \perp \pi_2$ *then* $H(\pi_1 * \pi_2) = H(\pi_1) + H(\pi_2)$.

The proof of this theorem is one of the direct calculations, and is omitted here. The uniform case is the simplest, but many other non-necessary unique probability distributions for finite domains can be characterized qualitatively in a similar way.

4. Proof of mutual non-definability of qualitative concepts using Padoa's principle

McKinsey (1935), Padoa (1902), Suppes (1957/1999), and Tarski (1935) give a thorough exposition and justification of the following method of proving when a concept of a theory cannot be defined in terms of the other concepts of the theory.

Find two models of the given axiomatic theory such that the concept C in question is extensionally different in the two models, but the other concepts of the theory are the same in the two models. It is clear that if C were definable, a contradiction would follow. (For detailed proofs, see the references given above.)

Here is a simple example taken from my logic textbook Suppes (1957/1999, p. 169) The theory of preference is based on the primitive relation P (for strict preference) and I (for indifference). The axioms of the theory are:

A1. *If* xPy *and* yPz, *then* xPz.
A2. *If* xIy *and* yIz, *then* xIz.
A3. *Exactly one of the following:* xPy, yPx *or* xIy.

We want to show that concept P is independent of concept I, that is, cannot be defined in terms of I. Let the domain of interpretation for both interpretations be the set $\{1, 2\}$. Let I be interpreted as identity in both cases. In one case let P be interpreted as $<$, and in the other case as $>$. In the first interpretation, we have:

$1\,P\,2$,

since

$1 < 2$,

and consequently by Axiom A3

not $2\,P\,1$.

But in the second interpretation, we have:

$2\,P\,1$,

since

$2 > 1$.

Now if P were definable in terms of I then P would have to be the same in both interpretations, since I is. However, P is not the same, and we conclude that P cannot be defined in terms of I.

Since standard probability models satisfy all the qualitative axioms, in the three lemmas of this section, we take two models for

Table 1
Comparative probabilities for Lemma 4.1.

Event	P	P'	Event	P	P'
$a+b+c+d$	1	1	c	0.480	0.480
$b+c+d$	0.910	0.910	$b+d$	0.430	0.430
$a+c+d$	0.840	0.839	$a+d$	0.360	0.359
$c+d$	0.750	0.749	d	0.270	0.269
$a+b+c$	0.730	0.731	$a+b$	0.250	0.251
$b+c$	0.640	0.641	b	0.160	0.161
$a+c$	0.570	0.570	a	0.090	0.090
$a+b+d$	0.520	0.520	$\emptyset$	0	0

Table 2
Experiment entropies for Lemma 4.1.

Experiment	$H(P)$	$H(P')$	Experiment	$H(P)$	$H(P')$
$\{a, b, c, d\}$	1.7540	1.7547	$\{c, a+b+d\}$	0.9988	0.9988
$\{a+b, c, d\}$	1.5183	1.5184	$\{a+c, b+d\}$	0.9858	0.9858
$\{a+d, b, c\}$	1.4619	1.4631	$\{a+d, b+c\}$	0.9427	0.9418
$\{a+c, b, d\}$	1.3953	1.3960	$\{d, a+b+c\}$	0.8415	0.8400
$\{a, c, b+d\}$	1.3445	1.3445	$\{a+b, c+d\}$	0.8113	0.8129
$\{a, d, b+c\}$	1.2347	1.2335	$\{b, a+c+d\}$	0.6343	0.6367
$\{a, b, c+d\}$	1.0469	1.0492	$\{a, b+c+d\}$	0.4365	0.4365
			$\{a+b+c+d\}$	0	0

one of the three concepts with two different probability measures P and P' such that, in the first case, the qualitative relations $\succcurlyeq$ and $\succcurlyeq'_u$ are the same in both models, i.e., $\succcurlyeq = \succcurlyeq'$ and $\succcurlyeq_u = \succcurlyeq'_u$, but in the case of independence $\perp \neq \perp'$. I emphasize that the use of explicit probability measures in the proofs is optional, not obligatory. It is particularly important to note that the two quantitative probability distributions P and P' cannot directly be used in applying Padoa's principle. For example, as can be seen in Table 1, P and P' differ quantitatively for 8 of the 16 events, but qualitatively they produce the same ordering $\succcurlyeq$.

In the three lemmas of this section, references to the three relations $\succcurlyeq$, $\perp$ and $\succcurlyeq_u$ of comparative probability, independence, and comparative uncertainty refer specifically to the qualitative properties stated in the axioms CP1–5, QI1–5, and CU1–5. If additional particular structural or existential axioms were assumed, the proofs given would often no longer hold. For example, if we also assumed in this section that the finite set of possible outcomes, i.e., the elements of Ω, were equally likely, as is done in Theorem 1, we could then use this property to define independence and uncertainty, as is done in Theorem 1.

The point of the proofs given in this section is to show the non-definability of these three basic notions of probability in terms of each other when only universal necessary qualitative properties are considered. Here, "necessary" means the property holds for any strictly agreeing standard probability measure.

Lemma 4.1. *Assuming only the qualitative axioms of Section 3, the independence relation $\perp$ cannot be defined in terms of the qualitative comparative probability relation $\succcurlyeq$ and the qualitative comparative uncertainty relation $\succcurlyeq_u$.*

Proof. Two models which are the same for both the comparative probability and uncertainty relations $\succcurlyeq$ and $\succcurlyeq_u$, but different for the independence relation $\perp$, are defined by two probability measures P and P' for the domain $\Omega = \{a, b, c, d\}$ and with $\mathcal{F}$ being simply the family of all subsets of Ω, so that if Ω has n elements, $\mathcal{F}$ has 2^n.

$$P : a = \frac{9}{100}, \qquad b = \frac{16}{100}, \qquad c = \frac{48}{100}, \qquad d = \frac{27}{100},$$

$$P' : a' = a, \qquad b' = b + \frac{0.1}{100}, \qquad c' = c, \qquad d' = d - \frac{0.1}{100}.$$

(Note that to simplify notation I am denoting the elements of Ω and their numerical values by the same symbols: a, b, c and d. Moreover, I use primes to distinguish the possibly different values of P', i.e., a', b', c' and d' from those of P. The models are simple enough not to require an elaborate explicit notation, e.g., $p(a) = \frac{9}{100}$ and $p'(a) = p(a)$ rather than $a = \frac{9}{100}$ and $a' = a$.)

In Table 1, I show the qualitative comparative probability relation $\succcurlyeq$ is qualitatively invariant for the slightly different 8 pairs of numerical probabilities assigned to the 16 pairs of events by P and P'.

Each event is denoted in the table by the sum of the probabilities of its elements. So here for example, $\Omega = a + b + c + d$.

It is evident from Table 1 that although the measures P and P' are slightly different, the qualitative comparative relation $\succcurlyeq$ is the same for both. The simple test is that for both P and P' the numerical values in both columns are strictly monotonically decreasing from top to bottom in both columns, one for P and one for P'. But the fact already mentioned that 8 of the pairs have different values, e.g., in line 4 of Table 1, $P(c + d) = 0.750$ and $P'(c + d) = 0.749$, shows that the probability distributions P and P' cannot themselves be used directly to prove non-definability. To do that they must be exactly the same for all pairs.

I now turn to the argument showing that the qualitative comparative uncertainty relation $\succcurlyeq_u$ is also qualitatively invariant with respect to P and P'. First, there are, purely from a set-theoretical viewpoint, 15 partitions of Ω, as shown in Table 2. The same abuse of notation is followed as in Table 1. Here the partitions should properly be shown as sets, but in fact the events that are members of the various partitions are shown, as in Table 1, by the sum of the probabilities of the elements belonging to the event in terms of the measures P and P'. So, for example, the experiment

$$\pi = \{\{a, b\}, \{c, d\}\},$$

is written

$$\pi = \{a + b, c + d\},$$

with the commas separating the description of the events that are the elements of the experiment in question, e.g., $a + c$, $b + d$. Note that it is experiments, not events, that are measured for uncertainty. As remarked earlier, this is the same concept of entropy used to measure the randomness of a random variable, and it is only the probability distribution of a random variable that is relevant, not its actual values.

The numerical computations of entropy for probability distributions P and P' are given in Table 2. Notice that two of the experiments, $\{c, a + b + d\}$ and $\{a + c, b + d\}$, differ under P and P' in entropy by 0.0030. But, Table 2 shows that $\succcurlyeq_u$ has the same strictly monotonically decreasing qualitative ordering of entropies, and this common qualitative ordering is not disturbed by this numerical difference being small, but positive.

Finally, we consider the independence relation $\perp$ under the measures P and P'. The independence equation

$$(a + b)(b + c) = b, \tag{3}$$

holds for measure P, i.e., $A \perp B$, where $A = \{a, b\}$, $B = \{b, c\}$, or, in standard notation for measure P:

$$P(A)P(B) = P(A \cap B).$$

In particular, here $P(A) = a + b = \frac{25}{100}$ and $P(B) = b + c = \frac{64}{100}$ and

$$\frac{25}{100} \cdot \frac{64}{100} = \frac{16}{100} = b,$$

as required for independence.

But, A is not independent of B under measure P', i.e.,

$$(a' + b')(b' + c') \neq b', \tag{4}$$

Table 3
Qualitative comparative probabilities for Lemma 4.2.

Event	P	P'	Event	P	P'
$a+b+c+d$	1	1	c	0.47	0.46
$b+c+d$	0.91	0.91	$b+d$	0.44	0.45
$a+c+d$	0.87	0.87	$a+d$	0.40	0.41
$c+d$	0.78	0.78	d	0.31	0.32
$a+b+c$	0.69	0.68	$a+b$	0.22	0.22
$b+c$	0.60	0.59	b	0.13	0.13
$a+c$	0.56	0.55	a	0.09	0.09
$a+b+d$	0.53	0.54	$\emptyset$	0	0

as is easily checked, or again in standard notation

$$P'(A)P'(B) \neq P'(A \cap B).$$

So (3) and (4) show the two models P and P' are qualitatively different for the independence relation $\perp$. Note that one case of qualitative difference is all that is required. Whether or not any other pairs of events are different under P or P' does not affect the argument. This completes the proof of Lemma 4.1.

Lemma 4.2. *Assuming only the qualitative axioms of Section 3, the qualitative comparative uncertainty relation $\succcurlyeq_u$ cannot be defined in terms of the qualitative comparative probability relation $\succcurlyeq$ and the independence relation $\perp$.*

Proof. Let $\Omega = \{a, b, c, d\}$ as in the previous proof. Let P and P' be as follows:

$$P : a = \frac{9}{100}, \quad b = \frac{13}{100}, \quad c = \frac{47}{100}, \quad d = \frac{31}{100}. \tag{5}$$

$$P' : a' = \frac{9}{100}, \quad b' = \frac{13}{100}, \quad c' = \frac{46}{100}, \quad d' = \frac{32}{100}. \tag{6}$$

The probabilities of the 16 events are shown in Table 3.

As is evident, the qualitative comparative probability relation $\succcurlyeq$ is the same for both P and P'. The strict monotonicity test holds for both P and P'.

I next turn to the proof that the independence relation $\perp$ is qualitatively invariant for P and P'. There are $16 \cdot 16 = 256$ pairs, but most of them have an a priori independence relation $\perp$ for both P and P'. For example, $\Omega \perp \Omega$. If we exclude the a priori cases involving either the empty set $\emptyset$ or Ω, then when $\emptyset \subset A$, $B \subset \Omega$, if $A \subseteq B$ or $A \cap B = \emptyset$, a priori we also know that A is not independent of B. So this leaves 30 cases that need to be computed for the given P and P' used in this proof of Lemma 4.2. The 30 cases are shown in Table 4. Note that in all these cases for, both P and P', the independence relation $\perp$ does not hold. Some are close to equality, e.g., see the last line, line 30. But independence is an exact relation.

Corresponding to Table 2, we have for this proof more entropy calculations, but now the results for the two models need to be different to prove the non-definability of comparative uncertainty. We show that $\succcurlyeq_u$ is not qualitatively invariant for P and P' as defined for Lemma 4.2. In particular, for P, $H(\{a, b, c+d\}) = 0.9749$ and $H(\{a+d, b+c\}) = 0.9710$, so that

$$\{a, b, c+d\} \succ_u \{a+d, b+c\},$$

but for P' this relation is reversed, since $H(\{a, b, c+d\}) = 0.9749$ and $H(\{a+d, b+c\}) = 0.9765$. We have now shown that for given P and P', we have two distinct models of the uncertainty relation $\succcurlyeq_u$, as required for the proof of Lemma 4.2.

Lemma 4.3. *Assuming only the qualitative axioms of Section 3, the qualitative comparative probability relation $\succcurlyeq$ cannot be defined in terms of the independence relation $\perp$ and the qualitative comparative uncertainty relation $\succcurlyeq_u$.*

Proof. Let $\Omega = \{a, b\}$, and

$$P : a = \frac{3}{4}, \qquad b = \frac{1}{4},$$

$$P' : a = \frac{1}{4}, \qquad b = \frac{3}{4},$$

$$P : a \succ b,$$

$$P' : b \succ a.$$

Table 4
Independence of events for Lemma 4.2.

	A	B	$A \cap B$	$P(A)P(B)$	$P(A \cap B)$	$P'(A)P'(B)$	$P'(A \cap B)$
1	bcd	acd	cd	0.7917	0.78	0.7917	0.78
2	bcd	abc	bc	0.6279	0.60	0.6188	0.59
3	bcd	ac	c	0.5096	0.47	0.5005	0.46
4	bcd	abd	bd	0.4823	0.44	0.4914	0.45
5	bcd	ad	d	0.364	0.31	0.3731	0.32
6	bcd	ab	b	0.2002	0.13	0.2002	0.13
7	acd	abc	ac	0.6003	0.56	0.5916	0.55
8	acd	bc	c	0.522	0.47	0.5133	0.46
9	acd	abd	ad	0.4611	0.40	0.4698	0.41
10	acd	bd	d	0.3828	0.31	0.3915	0.32
11	acd	ab	a	0.1914	0.09	0.1914	0.09
12	cd	abc	c	0.5382	0.47	0.5304	0.46
13	cd	bc	c	0.468	0.47	0.4602	0.46
14	cd	ac	c	0.4368	0.47	0.429	0.46
15	cd	abd	d	0.4134	0.31	0.4212	0.32
16	cd	bd	d	0.3432	0.31	0.351	0.32
17	cd	ad	d	0.312	0.31	0.3198	0.32
18	abc	abd	ab	0.3657	0.22	0.3672	0.22
19	abc	bd	b	0.3036	0.13	0.306	0.13
20	abc	ad	a	0.276	0.09	0.2788	0.09
21	bc	ac	c	0.336	0.47	0.3245	0.46
22	bc	abd	b	0.318	0.13	0.3186	0.13
23	bc	bd	b	0.264	0.13	0.2655	0.13
24	bc	ab	b	0.132	0.13	0.1298	0.13
25	ac	abd	a	0.2968	0.09	0.297	0.09
26	ac	ad	a	0.224	0.09	0.2255	0.09
27	ac	ab	a	0.1232	0.09	0.121	0.09
28	bd	ad	d	0.176	0.31	0.1845	0.32
29	bd	ab	b	0.0968	0.13	0.099	0.13
30	ad	ab	a	0.088	0.09	0.0902	0.09

Independence holds only for the trivial cases, such as $\Omega \perp \Omega$, for both P and P', similarly for the comparative uncertainty relation $\succcurlyeq_u$. There are just two experiments $\pi(\Omega)$ and $\pi_1 = \{\{a\}, \{b\}\}$. Their qualitative uncertainty relation is the same for both models P and P', i.e., in both cases $\pi_1 \succ_u \pi(\Omega)$, and thus $\succcurlyeq_u$ is the same for P and P'. But the qualitative comparative probability relation is different for the two models, $a \succ b$ for P and $b \succ a$ for P'. Moreover, it is easy to prove from CP1–3 that if $a \succ b$ then it is not the case $b \succ a$, and vice versa. So the relation $\succcurlyeq$ cannot be defined in terms of $\perp$ and $\succcurlyeq_u$, as required for the proof of Lemma 4.3.

This completes the demonstration that each of the three basic qualitative relations of probability theory cannot be defined in terms of the other two, as now summarized in Theorem 2.

Theorem 2. *None of the three qualitative concepts of probability, comparative probability, independence or comparative uncertainty, as axiomatized in Section 3 can be defined in terms of the other two.*

A final point, to be emphasized again, is that each of the axioms introduced in Section 3 is a universal necessary consequence of any standard probability measure.

Lemmas 4.1–4.3 show, in a mathematically decisive way, that the three fundamental qualitative concepts of the theory of probability, as axiomatized in Section 3, cannot, at this elementary qualitative level, be defined in terms of each other. Each makes a distinctive cognitive contribution to our understanding of probability.

5. Qualitative random-variable axiomatization and representation of comparative probability

The particular new feature of the axioms given in Section 5 is that they restrict the use of random variables to just the generalized indicator functions of events, their iterates and their products. The axioms are weaker than those of Suppes and Zanotti (1976) in not using sums of indicator functions. On the other hand, it is necessary to have iteration to prove, for *any* finite set of possible outcomes satisfying the qualitative axioms, a unique strictly agreeing probability measure exists. The need for having the iteration be unbounded is clear for the simple example of just two possible outcomes a and b such that the probability of a is $\sqrt{2} - 1$ and that of b, $2 - \sqrt{2}$.

As mentioned already, it will be useful to compare with the axioms given earlier, a purely qualitative axiomatization written in terms of the set of generalized indicator functions closed under iteration and product. The main reason for making this is well-known in the literature on the foundations of measurement. For structures that may be either finite or infinite, purely elementary axioms of the form given earlier (Section 3) cannot, by themselves, be used to formulate axioms sufficient to prove a necessary numerical representation theorem (for detailed theoretical discussion of this problem see Scott and Suppes (1958) and Luce, Krantz, Suppes, and Tversky (1990, Ch XXI)). This means that for such structures, an Archimedean axiom that is not elementary (in the sense of first-order logic) is required to complete the axiomatization. At the present stage of development of these ideas, a variety of forms have been given, which I will not attempt to review here. The only point to make is that when a random-variable axiomatization is given, the general form of the Archimedean axiom is homogeneous with the other axioms. An early example of such an axiomatization just for comparative probability was given in Suppes and Zanotti (1976). In fact, their axioms are used in formulating the random-variable comparative probability axioms CPR1–3 ('R' standing for a random variable).

Suppes and Zanotti (1976) used extended indicator functions, which are the functions obtained by closing the set of indicator functions under addition, e.g., $\mathbf{A}^i + \mathbf{B}^i$. Such functions form a natural class of random variables, but are less intuitive as a class than just the generalized indicator functions and their iterates, when, for example, thought of as outcomes to bet on. (But Suppes and Zanotti (1976) ignored the problem of needing *generalized* indicator functions, as discussed in Section 2.)

Something much simpler, as I have discovered rather late, will suffice to state the axioms, with only iteration and products. So for each positive integer n, $n\mathbf{A}^i$ is also a random variable. (This is a special case of the general axiom that if $\mathbf{A}$ is a random variable and c is any number not equal to 0, then $c\mathbf{A}$ is a random variable.)

From a formal axiomatic standpoint, we first define the random variables that are in $\mathcal{F}^*$. To begin with, for A any event of $\mathcal{F}$, and $\alpha > 0$, the α-generalized indicator function $\mathbf{A}^i = \{(\omega, \alpha) : \omega \in A\} \cup \{(\omega, 0) : \omega \in \neg A\}$ is in $\mathcal{F}^*$.

Definition 5.1. of $\mathcal{F}^*$.

(i) If A is in $\mathcal{F}$, $\mathbf{A}^i$ is in $\mathcal{F}^*$,
(ii) If $\mathbf{A}^i$ is in $\mathcal{F}^*$, then its nth iterate is also, i.e., $n\mathbf{A}^i \in \mathcal{F}^*$,
(iii) If $\mathbf{A}^i$ and $\mathbf{B}^i$ are in $\mathcal{F}^*$, then $\mathbf{A}^i\mathbf{B}^i$ is also,
(iv) No other random variables are in $\mathcal{F}^*$.

From an intuitive estimation or gambling standpoint, it is much easier to reflect on the subjective probability of $n\mathbf{A}^i$ than of $n\mathbf{A}^i + k\mathbf{B}^i$. For example, if $\mathbf{A}^i(\omega) = 1$ means "heads" in a toss of a coin with unknown bias, $5\mathbf{A}^i$ is just the estimate of 5 such tosses being "heads". The mixture of $\mathbf{A}^i + \mathbf{B}^i$ is less straightforward. It is this use only of iterates not of arbitrary sums of indicator functions that is a positive feature of this axiomatization. But do note that $\mathbf{A}^i + \mathbf{B}^i$ is meaningful when $A \cap B = \emptyset$, and so $\mathbf{A}^i + \mathbf{B}^i = \mathbf{C}^i$, where $A \cup B = C$.

5.1. Random-variable axioms of comparative probability

For random variables $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ in $\mathcal{F}^*$,

CPR1. *If* $\mathbf{A} \succcurlyeq \mathbf{B}$ *and* $\mathbf{B} \succcurlyeq \mathbf{C}$ *then* $\mathbf{A} \succcurlyeq \mathbf{C}$.
CPR2. $\mathbf{A} \succcurlyeq \mathbf{B}$ *or* $\mathbf{B} \succcurlyeq \mathbf{A}$.
CPR3. $\mathbf{\Omega}^i \succ \emptyset^i$.
CPR4. $\mathbf{A} \succcurlyeq \emptyset^i$.
CPR5. *If* $\mathbf{A} \cap \mathbf{C} = \emptyset$ *and* $\mathbf{B} \cap \mathbf{C} = \emptyset$, *then* $\mathbf{A} + \mathbf{C} \succcurlyeq \mathbf{B} + \mathbf{C}$ *iff* $\mathbf{A} \succcurlyeq \mathbf{B}$.
CPR6. (Archimedean Axiom) *For A and B in* $\mathcal{F}$, *if* $\mathbf{A}^i \succ \mathbf{B}^i$ *then there exists a k and an n such that*

$$n\mathbf{A}^i \succ k\mathbf{\Omega}^i \succ n\mathbf{B}^i.$$

The special form of this Archimedean axiom (CPR6) is required, because only iterations and products of generalized indicator functions, not their additions, are used.

5.2. Representation theorem for random-variable formulations of qualitative comparative probability as an extensive quantity

Theorem 3 is a standard theorem for comparative probability as an extensive quantity. Theorem 4 of Section 6.2 uses axioms of independence of random variables to extend Theorem 3 to prove a stronger uniqueness.

Theorem 3. *If a structure* $\mathbf{\Omega} = (\Omega, \mathcal{F}^*, \succcurlyeq)$, *with* Ω *a nonempty set, satisfies Axioms* CPR1–6, *there is an expectation function* E_S *defined on* $\mathcal{F}^*$ *such that for any generalized indicator functions* $\mathbf{A}^i$ *and* $\mathbf{B}^i$ *in* $\mathcal{F}^*$,

(i) $0 \le E_S(\mathbf{A}^i) \leqslant E_S(\mathbf{\Omega}^i)$,
(ii) $E_S(\mathbf{\Omega}^i) > 0$, *and* $E_S(\emptyset^i) = 0$,
(iii) $E_S(\mathbf{A}^i) \geqslant E_S(\mathbf{B}^i)$ *iff* $\mathbf{A}^i \succcurlyeq \mathbf{B}^i$,
(iv) *If* $A \cap B = \emptyset$ *then* $E_S(\mathbf{A}^i + \mathbf{B}^i) = E_S(\mathbf{A}^i) + E_S(\mathbf{B}^i)$.

(v) *The expectation E_S is unique up to multiplication by a positive constant, i.e., a positive similarity transformation S.*

The approach of the proof of Theorem 3 is straightforward. We shall assume without any significant loss of generality that every atom of Ω is an atomic event in $\mathcal{F}$, so $\mathcal{F}$ is the powerset of Ω. Notation needed in the proof and consideration of cases are simplified by this assumption. We first find the expectation, as a random variable, of each generalized indicator function of an event in $\mathcal{F}$. We begin by defining for every event $A \succ \emptyset$, the set L_A of positive rational fractions, using random variables that are iterates of generalized indicator functions, with $k \leq n$:

$$L_A = \left\{ \frac{k}{n} : k\mathbf{\Omega}^i \succcurlyeq n\mathbf{A}^i \right\}.$$

(To avoid an easy confusion I use here and in what follows 'k' as a numerical variable, and reserve 'm' to designate a measure m.)

The set L_A is nonempty, since it follows easily from the comparative probability axioms that $\mathbf{\Omega}^i \succcurlyeq \mathbf{A}^i$ and thus $\frac{1}{1} = 1$ is in L_A. L_A is bounded from below by 0, since all the fractions are positive. So, as a standard property of the real numbers, L_A has a greatest lower bound (glb). Similarly, we define

$$U_A = \left\{ \frac{k}{n} : k\mathbf{\Omega}^i \preccurlyeq n\mathbf{A}^i \right\},$$

which is also nonempty, since by Axiom CP6, there exists a k and an n such that $k\mathbf{\Omega}^i \preccurlyeq n\mathbf{A}^i$, and so $\frac{k}{n}$ is in U_A. Since $\mathbf{\Omega}^i \succcurlyeq \mathbf{A}^i$, U_A is bounded from above by 1, it has a least upper bound (lub). Use of these concepts is critical at several points in the argument that follows. So I state here, five elementary equivalences about inequalities for positive rational fractions $\frac{k}{n}$ with $k \leq n$:

$$\frac{k}{n} \in L_A \quad \textit{iff} \ \frac{k}{n} \geq \operatorname{glb} L_A, \tag{7}$$

$$\frac{k}{n} \in U_A \quad \textit{iff} \ \frac{k}{n} \leq \operatorname{lub} U_A, \tag{8}$$

plus the two negations

$$\frac{k}{n} \notin L_A \quad \textit{iff} \ \frac{k}{n} < \operatorname{glb} L_A, \tag{9}$$

$$\frac{k}{n} \notin U_A \quad \textit{iff} \ \frac{k}{n} > \operatorname{lub} U_A, \tag{10}$$

and finally,

$$\text{if } \frac{k}{n} \in L_A \text{ and } \frac{k'}{n'} \in U_A, \text{ then } \frac{k}{n} \geq \frac{k'}{n'}. \tag{11}$$

The next step is to prove that $\operatorname{lub} U_A = \operatorname{glb} L_A$. First, it follows at once from (7), (8) and (11) that

$$\operatorname{glb} L_A \geq \operatorname{lub} U_A.$$

Suppose now, that

$$\operatorname{glb} L_A > \operatorname{lub} U_A.$$

Then there is a k and an n such that

$$\operatorname{glb} L_A > \frac{k}{n} > \operatorname{lub} U_A.$$

Since $\frac{k}{n} > \operatorname{lub} U_A$, we have from (10) and the definition of U_A, it is not the case that $k\mathbf{\Omega}^i \preccurlyeq n\mathbf{A}^i$, and so

$$k\mathbf{\Omega}^i \succ n\mathbf{A}^i. \tag{12}$$

Since $\frac{k}{n} < \operatorname{glb} L_A$, we can also infer by (9) and the definition of L_A

$$k\mathbf{\Omega}^i \prec n\mathbf{A}^i. \tag{13}$$

So by transitivity (Axioms CPR1, CPR2, (12) and (13)) we infer both $k\mathbf{\Omega}^i \succ k\mathbf{\Omega}^i$ and $n\mathbf{A}^i \succ n\mathbf{A}^i$, either of which contradicts Axiom CPR2, which implies reflexivity for all events and random variables, in $\mathcal{F}^*$ e.g., $k\mathbf{\Omega}^i \succcurlyeq k\mathbf{\Omega}^i$ and $n\mathbf{A}^i \succcurlyeq n\mathbf{A}^i$. We conclude that

$$\alpha \operatorname{lub} U_A = \alpha \operatorname{glb} L_A = E_S(\mathbf{A}^i), \tag{14}$$

where α is the parameter of the generalized indicator functions. Using (14) and the fact that for all $\mathbf{A}^i$, $\operatorname{glb} L_A \geqslant 0$, condition (i) of Theorem 3 is established, and (ii) is obvious.

We next prove (iii) of the theorem. Assume first $A \succcurlyeq B$.

Case 1. $A \approx B$, i.e., $A \succcurlyeq B$ and $B \succcurlyeq A$.

We use again

$$L_A = \left\{ \frac{k}{n} : k\mathbf{\Omega}^i \succcurlyeq n\mathbf{A}^i \right\},$$

$$L_B = \left\{ \frac{k}{n} : k\mathbf{\Omega}^i \succcurlyeq n\mathbf{B}^i \right\}.$$

Since $\mathbf{A}^i \approx \mathbf{B}^i$, we infer at once that

$$n\mathbf{A}^i \approx n\mathbf{B}^i,$$

and so

$$\frac{k}{n} \in L_A \textit{ iff } \frac{k}{n} \in L_B.$$

Hence

$$L_A = L_B,$$

and so

$$E_S\left(\mathbf{A}^i\right) = E_S\left(\mathbf{B}^i\right),$$

as desired.

Case 2. $A \succ B$.

By virtue of Axiom CPR6, $\exists k, n$ such that $n\mathbf{A}^i \succ k\mathbf{\Omega}^i \succ n\mathbf{B}^i$, whence for this k and n, from (7) and (8)

$$\operatorname{glb} L_A > \frac{k}{n} > \operatorname{glb} L_B,$$

and thus

$$E_S\left(\mathbf{A}^i\right) > E_S\left(\mathbf{B}^i\right).$$

Now, going the other way, assume

$$E_S\left(\mathbf{A}^i\right) \geqslant E_S\left(\mathbf{B}^i\right).$$

Case 1.

$$E_S\left(\mathbf{A}^i\right) = E_S\left(\mathbf{B}^i\right).$$

This is obvious from the proof of Case 1 for $A \approx B$.

We next assume

$$E_S\left(\mathbf{A}^i\right) > E_S\left(\mathbf{B}^i\right),$$

which is equivalent to

$$\operatorname{lub} U_A > \operatorname{lub} U_B,$$

and thus $\exists k, n$ such that

$$\operatorname{lub} U_A > \frac{k}{n} > \operatorname{lub} U_B,$$

and so

$$\frac{k}{n} \in U_A \text{ and } \frac{k}{n} \notin U_B,$$

and thus

$$n\mathbf{A}^i \succcurlyeq k\mathbf{\Omega}^i,$$

but

not $n\mathbf{B}^i \succcurlyeq k\boldsymbol{\Omega}^i$,

and hence

$k\boldsymbol{\Omega}^i \succ n\mathbf{B}^i$,

so by strict transitivity

$n\mathbf{A}^i \succ n\mathbf{B}^i$,

and thus,

$A \succ B$,

which completes the proof of condition (iii) of Theorem 3.

We next need to prove condition (iv).

If $A \cap B = \emptyset$ then $E_S(A \cup B)^i = E_S\left(\mathbf{A}^i\right) + E_S\left(\mathbf{B}^i\right)$. It is obvious from prior arguments that

$E_S(A \cup B)^i = \alpha \operatorname{glb} L_{A\cup B} = \alpha \operatorname{lub} U_{A\cup B}$.

So we want to show that if $A \cap B = \emptyset$,

$\operatorname{glb} L_{A\cup B} = \operatorname{glb} L_A + \operatorname{glb} L_B$.

For any k, n, k' *and* n' we have:

If $\frac{k}{n} \in L_A$ and $\frac{k'}{n'} \in L_B$, then $k\boldsymbol{\Omega}^i \succcurlyeq n\mathbf{A}^i$ and $k'\boldsymbol{\Omega}^i \succcurlyeq n'\mathbf{B}^i$.

Multiplying the first inequality by n' and the second by n, followed by adding the two resulting inequalities, we have

$\left(n'k + nk'\right)\boldsymbol{\Omega}^i \succcurlyeq nn'\mathbf{A}^i + nn'\mathbf{B}^i$,

which since $A \cap B = \emptyset$, is, of course, equivalent to

$$\left(n'k + nk'\right)\boldsymbol{\Omega}^i \succcurlyeq nn'\,(A \cup B)^i\,, \tag{15}$$

and so

$\dfrac{n'k + nk'}{nn'} \in L_{A\cup B.}$

This argument shows that

$$\operatorname{glb} L_A + \operatorname{glb} L_B \le \operatorname{glb} L_{A\cup B.} \tag{16}$$

The corresponding argument using least upper bounds follows from

$k\boldsymbol{\Omega}^i \preccurlyeq n\mathbf{A}^i \quad \text{and} \quad k'\boldsymbol{\Omega}^i \preccurlyeq n'\mathbf{B}^i$,

and inequality (15) is reversed, that is,

$(n'k + nk')\boldsymbol{\Omega}^i \preccurlyeq nn'(A \cup B)^i$,

and so we now infer

$$\operatorname{lub} U_{A\cup B} \le \operatorname{lub} U_A + \operatorname{lub} U_B. \tag{17}$$

So, from (14)–(17) we have immediately if $A \cap B = \emptyset$,

$E_S(A \cup B)^i = E_S(\mathbf{A}^i) + E_S(\mathbf{B}^i)$.

Finally, because many proofs are available, we do not prove condition (v) of Theorem 3, which is the standard invariance result for extensive quantities, i.e., invariance under the group of positive similarity transformations S.

6. Random-variable axiomatization and representation of qualitative independence

The random-variable axioms of independence are still qualitative, like those of Section 3, but the new Axiom QIR6, not in Section 3, shows that indicator functions of events have more expressive power than the events themselves.

In Section 3, I mainly used the qualitative axioms of independence introduced by Domotor (1969). I mentioned that his many additional "quadratic" axioms, written $A \times B$, could not be formulated in the framework of Section 3. The situation is different for the random-variable formulation of qualitative axioms. Domotor's "quadratic" terms $A \times B$ are replaced here by the products of the generalized indicator functions, for example, $\mathbf{A}$ and $\mathbf{B}$, i.e., $\mathbf{A}^i\mathbf{B}^i$, which is well-defined as a random variable.

6.1. Random-variable axiomatization of qualitative independence

QIR1. $\mathbf{A}^i \perp \boldsymbol{\Omega}^i$.
QIR2. *If* $\mathbf{A}^i \perp \mathbf{A}^i$, *then* $\mathbf{A}^i \approx \boldsymbol{\Omega}^i$ *or* $\mathbf{A}^i \approx \emptyset^i$.
QIR3. *If* $\mathbf{A}^i \perp \mathbf{B}^i$ *then* $\mathbf{B}^i \perp \mathbf{A}^i$.
QIR4. *If* $\mathbf{A}^i \perp \mathbf{B}^i$ *then* $\mathbf{A}^i \perp \neg\mathbf{B}^i$.
QIR5. *If* $\mathbf{A}^i \perp \mathbf{B}^i$, $\mathbf{A}^i \perp \mathbf{C}^i$ *and* $B \cap C = \emptyset$, *then* $\mathbf{A}^i \perp (B \cup C)^i$.
QIR6. *If the n generalized indicator functions are mutually independent, written* $\perp(\mathbf{A}^i_1, \ldots, \mathbf{A}^i_n)$, *then*

(i) $(A_1 \cap A_2 \cap \cdots \cap A_n)^i \approx \left(\mathbf{A}^i_1\right)\cdots\left(\mathbf{A}^i_n\right)$,
(ii) *condition* (i) *holds for any subset of the n indicator functions having at least two members.*

The need for the subset condition (ii) in this definition can be made clear by a three-event counterexample that satisfies (i) but not (ii), I leave the construction to the reader. The generalization from 2 to n mutually independent generalized indicator functions is needed in many theoretical and applied developments of probability. The binary case of $n = 2$ fits the form of Axioms QIR 1–5 better, but is not sufficient in even the case of $n = 3$.

6.2. Theorem 4 on uniqueness of scale using independence

That Axiom QIR6 provides the basis for proving that the scale type is absolute is the subject of this subsection.

Theorem 4. *For any event A in* $\mathcal{F}$ *let the generalized indicator function* $\mathbf{A}^i$ *be defined, as in Section 2, as the function, for any* ω *in* Ω,

$$\mathbf{A}^i(\omega) = \begin{cases} \alpha > 0 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{18}$$

From the Axioms QIR1*and* QIR6, *it follows that* $\alpha = 1$. *In other words, the group of invariant transformation is reduced from the positive similarities to just the identity function, for an absolute scale.*

Proof. With the generalized definition of an indicator function, as given above, and putting '$\boldsymbol{\Omega}^i$' for '$\mathbf{A}^i$', in Axiom QIR1, we have $\boldsymbol{\Omega}^i \perp \boldsymbol{\Omega}^i$. Making the same substitution also for '$\mathbf{B}^i$' in Axiom QIR6, we infer from the two axioms

$$(\Omega \cap \Omega)^i \approx \boldsymbol{\Omega}^i \cdot \boldsymbol{\Omega}^i. \tag{19}$$

But since both sides of this equivalence just involve the constant numerical function $\boldsymbol{\Omega}^i = \alpha$, it follows that (19) implies

$\alpha = \alpha \cdot \alpha$,

since $(\Omega \cap \Omega)^i = \boldsymbol{\Omega}^i$. Moreover $\alpha > 0$, which is true only if

$\alpha = 1$,

which completes the proof of Theorem 4.

A remark on Theorem 4 is in order here. The statement of Axiom QIR6 can be strengthened to give it the form of a definition: $\perp(\mathbf{A}^i_1, \ldots, \mathbf{A}^i_n)$ *if and only if* (i) *and* (ii) *of Axiom QIR6 hold.*

Why not now claim mutual independence is really definable in terms of comparative probability? The answer is that in the theory of definition (for details, see (Suppes, 1957/1999, Ch 8)), it is not a *proper* definition, but a *creative* one. Being *creative* means that using the definition, a theorem can be proved only in the notation of the original theory (Eq. (19)), but cannot be proved in the original theory without the definition of independence. So this creative definition should really be treated as a new axiom, which is what has been done here, but with a slight weakening of the creative definition to the "if...then" form of Axiom QIR6.

7. Random-variable axiomatization and definition of comparative uncertainty

We first state the random-variable axiomatization of qualitative uncertainty in Section 7.1. In Section 7.2, we show that uncertainty is definable in terms of the unique probability distribution and quantitative independence relation given in Section 6. This definability of quantitative uncertainty stands in contrast to the non-definability of qualitative uncertainty proved in Section 4.

7.1. The random-variable axioms of uncertainty

To give a random-variable form of Axioms CU 1–5, if $\mathbf{A}$ is a random variable, then $\pi(\mathbf{A})$ is the experiment, i.e., the partition of Ω generated by $\mathbf{A}$. More formally, if $\mathbf{A}$ is a random variable with values $a_1, \ldots, a_n$, then

$$\pi(\mathbf{A}) = \{\mathbf{A}(\omega) = a_i, i = 1, \ldots, n, \omega \epsilon \Omega\}. \tag{20}$$

So only a minor change of notation is required to formulate Axioms CU1–5 as Axioms CUR1–5.

CUR1. *If* $\pi(\mathbf{A}) \succcurlyeq_u \pi(\mathbf{B})$ *and* $\pi(\mathbf{B}) \succcurlyeq_u \pi(\mathbf{C})$, *then* $\pi(\mathbf{A}) \succcurlyeq_u \pi(\mathbf{C})$.
CUR2. $\pi(\mathbf{A}) \succcurlyeq_u \pi(\mathbf{B})$ *or* $\pi(\mathbf{B}) \succcurlyeq_u \pi(\mathbf{A})$.
CUR3. $\pi(\mathbf{A}) \succcurlyeq_u \pi(\mathbf{\Omega})$.
CUR4. *If* $\pi(\mathbf{A}) \succ_u \pi(\mathbf{B})$ *then* $\pi(\mathbf{A}) \succ_u \pi(\mathbf{\Omega})$.
CUR5. *If* $\pi(\mathbf{A}) \perp \pi(\mathbf{B})$ *and* $\pi(\mathbf{B}) \perp \pi(\mathbf{C})$, *then* $\pi(\mathbf{A}) * \pi(\mathbf{C}) \succcurlyeq_u \pi(\mathbf{B}) * \pi(\mathbf{C})$ *iff* $\pi(\mathbf{A}) \succcurlyeq_u \pi(\mathbf{B})$.

7.2. Proving correctness of the definition of comparative uncertainty

Definition 7.1. $\pi_1 \succcurlyeq_u \pi_2$ *if and only if* $H(\pi_1) \geqslant H(\pi_2)$.

To justify the definition of comparative uncertainty, we need to prove the definition given is correct, in the sense of proving the qualitative axioms given in Section 3, which is equivalent to proving the corresponding random-variable axioms CUR1–5.

Proof of Axiom CU1. Transitivity of $\succcurlyeq_u$. Assume $\pi_1 \succcurlyeq_u \pi_2$ and $\pi_2 \succcurlyeq_u \pi_3$. Then by Definition 7.1 $H(\pi_1) \geqslant H(\pi_2)$ and $H(\pi_2) \geqslant H(\pi_3)$, from which we infer from the transitivity of numerical $\geqslant$, $H(\pi_1) \geqslant H(\pi_3)$, and so by Definition 7.1, $\pi_1 \succcurlyeq_u \pi_3$, as required.

Proof of Axiom CU2. It follows at once from the connectivity of numerical $\geqslant$ that

$$H(\pi_1) \geqslant H(\pi_2) \quad \text{or} \quad H(\pi_2) \geqslant H(\pi_1),$$

and so by Definition 7.1

$$\pi_1 \succcurlyeq_u \pi_2 \quad \text{or} \quad \pi_2 \succcurlyeq_u \pi_1.$$

Proof of Axiom CU3. Since $H(\pi(\Omega)) = 0$ and for any experiment π, $H(\pi) \geqslant 0$, using Definition 7.1, Axiom CU3 follows at once.

Proof of Axiom CU4. From the hypothesis that $\pi_1 \succ_u \pi_2$, we infer $H(\pi_1) > H(\pi_2)$, but we know for all experiments π, $\pi \succcurlyeq_u \pi(\Omega)$ and thus by Definition 7.1

$$H(\pi_2) \geqslant H(\pi(\Omega)),$$

and so by transitivity

$$H(\pi_1) > H(\pi(\Omega)),$$

which implies by definition that

$$\pi_1 \succ_u \pi(\Omega).$$

Proof of Axiom CU5. From the two hypotheses of CU5 that $\pi_1 \perp \pi_3$ and $\pi_2 \perp \pi_3$, we have

$$H(\pi_1 * \pi_3) = H(\pi_1) + H(\pi_3)$$

and

$$H(\pi_2 * \pi_3) = H(\pi_2) + H(\pi_3),$$

from which it follows at once that

$$H(\pi_1) + H(\pi_3) \geq H(\pi_2) + H(\pi_3) \quad \text{iff } H(\pi_1) \geq H(\pi_2).$$

From the first of these two inequalities we have:

$$\pi_1 * \pi_3 \succcurlyeq_u \pi_2 * \pi_3 \quad \textit{iff}$$
$$H(\pi_1 * \pi_3) \geq H(\pi_2 * \pi_3) \quad \textit{iff}$$
$$H(\pi_1) \geq H(\pi_2),$$

and so by Definition 7.1

$$\pi_1 \succcurlyeq_u \pi_2,$$

as required, to complete the derivation of Axiom CU5 from Definition 7.1.

This concludes the formal developments of this article. Some readers may feel that the characterization of the definition of qualitative independence or qualitative comparative uncertainty is not "complete" or "rigorous" enough. I offer no rigorous proof to the contrary, because only in the most special limited circumstance can such complete proofs be given. Adequacy of the characterization ordinarily used, and adopted here, of probabilistic independence, is a fine example. The extended discussion in different chapters of Fine (1973) shows how easy it is to disagree on what should be the properties of independence. Kolmogorov (1933/1950) stresses the importance of this problem of clarification (p. 9). But this article is not the place to consider in detail these contentious questions. Given the inevitable openness of several issues, some readers may want to offer further qualitative axioms of independence that go beyond the framework developed here.

For representation of the qualitative concepts of comparative probability, independence and comparative uncertainty, the standard measurement-theory approach to extensive quantities seems the most natural (see Krantz, Luce, Suppes, & Tversky, 1971). However, as we have seen, this approach takes us no farther than Kolmogorov's use of positive measures. The analysis I have given is very much in the spirit of Hilbert's foundational work in geometry, beginning as it does, with qualitative axioms and ending with quantitative representations such as $P(\Omega) = E(\mathbf{\Omega}^i) = 1$. Of course, as is well known, the 19*th*-century viewpoint toward foundations of geometry (Hilbert, 1897/1930; Pasch, 1882) has been the model for modern work in the foundations of measurement from the beginning of the 20th century until now.

References

Bernstein, S. (1917). On the axiomatic foundation of the theory of probability. *Mitteilungen der Mathematischen Gesellschaft in Charkov*, (in Russian).
Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons, Inc.
Davidson, D., & Suppes, P. (1956). A finitistic axiomatization of subjective probability and utility. *Econometrica*, *24*, 264–275.
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.
de Moivre, A. (1712). De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus. *Philosophical Transactions*, *27*, 213–264. comm. 1711.
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. L. (1973). *Theories of probability*. New York: Academic Press.
Hilbert, D. (1897/1930). *Grundlagen der geometrie* (7th ed.). Leipzig and Berlin: B. G. Teubner.
Jeffreys, H. (1948). *Theory of probability* (2nd ed.). Oxford: Clarendon 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. (1933/1950). *Foundations of the theory of probability*. New York: Chelsea Publishing Company, translated by N. Morrison. First published in German in 1933.
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.
Laplace, P. S. (1812). Théorie analytique des probabilités, Courcier, Paris.
Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement Vol. III: representation, axiomatization, and invariance*. San Diego, CA: Academic Press.
Luce, R. D., & Narens, L. (1978). Qualitative independence in probability theory. *Theory and Decision*, *9*, 225–239.
McKinsey, J. C. C. (1935). On the independence of undefined ideas. *American Mathematical Society. Bulletin*, *41*, 291–297.
Ornstein, D. S., & Weiss, B. (1991). Statistical properties of chaotic systems. *American Mathematical Society. Bulletin. New Series*, *24*, 11–116.
Padoa, A. (1902). C. r. deuxiéme congrés international des mathématiciens. In: *Un nouveau systéme irréductible de postulats pour l'algébre*, Paris.
Pasch, M. (1882). *Vorlesungen über neuere geometrie*. Leipzig: Verlag von Julius Springer.
Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). London: Kegan Paul, Trench, Trubner & Co.
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley, revised edition printed in 1972, Dover, New York.
Scott, D. (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology*, *1*, 233–247.
Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, *23*, 113–128.
Shannon, C. (1948). The mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423. 623–656.
Suppes, P. (1956). The role of subjective probability and utility in decision-making. In: *Proceedings of the third Berkeley Symposium on mathematical statistics and probability (1954–1955)*, *Vol. 5*, pp. 61–73.
Suppes, P. (1957/1999). *Introduction to logic*. New York: Van Nostrand, reprinted in 1999 by Dover, New York.
Suppes, P., & Alechina, N. (1994). The definability of the qualitative independence of events in terms of extended indicator functions. *Journal of Mathematical Psychology*, *38*, 366–376.
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.
Tarski, A. (1935). Einige methodologische untersuchungen über die definierbarkeit der begriffe. *Erkenntnis*, *5*, 80–100.