ABSTRACT. Popper (1959, Appendices *iv and *v) has given several axiom systems for probability that ensure, without further assumptions, that the domain of interpretation can be reduced to a Boolean algebra. This paper presents axiom systems for subtheories of probability theory that characterize in the same way lower semilattices (Section 1) and distributive lattices (Section 2). Section 1 gives a new (metamathematical) derivation of the laws of semilattices; and Section 2 one or two surprising theorems, previously derived only with the help of an axiom for complementation. The problem of the creativity of the axioms is explored in Section 3, enlarging on Popper (1963). In conclusion, Section 4 explains how these systems, and the full system of Popper (1959), provide generalizations of the relation of deducibility, contrasting our approach with the enterprise known as probabilistic semantics.

0. INTRODUCTION

Two of Patrick Suppes’s admirers, one 20 years older, one 20 years younger, salute him on his 70th birthday! We happily take this opportunity to report on joint work combining two of Suppes’s most abiding enthusiasms: the use of formal methods and the theory of probability. In this paper we present new axiom systems for subtheories of probability theory that characterize lower semilattices (Section 1) and distributive lattices (Section 2). Section 3 discusses the problem of the creativity of some of the axioms and definitions; results given here enlarge on the work of Popper (1963). To conclude, in Section 4 we explain briefly the way in which these systems, and the full system of Popper (1959, Appendices *iv and *v), provide generalizations of the relation of deducibility, and comment critically on the enterprise known as probabilistic semantics.

The treatment is austere. Each system is formulated in a language with infinitely many individual variables, a numerical function symbol p, and other operation symbols, named as they are introduced. There is of course also the usual logical apparatus, including the conditional ⇒ and biconditional ⇔, the conjunction sign &, the quantifiers ∀ and...
the identity sign $=,$ and the arithmetical symbols $\leq$ and $\geq$, 0 and $+$, and $\cdot$ (usually abbreviated to nothing at all). Full advantage is taken of routine arithmetical abbreviations. Distracting parentheses are omitted wherever it is possible to omit them. At no stage do we display expressions of any language being considered; they are named, not shown. If the variables free in a formula $\phi$ are $x_0, \ldots, x_{n-1}$ then we call any formula of the form $\forall x_0 \ldots \forall x_{n-1} \phi$, whatever the order of the variables, the closure of $\phi$. Displayed formulas with free variables are always understood to abbreviate their own closures.

Arithmetical axioms are taken for granted, and the logical and arithmetical notation is standardly interpreted. In contrast, the domain of interpretation for the individual variables is unspecified, and the operation symbols connecting them are equally malleable. Of course, there are intended interpretations here too. The (unextended) operation symbol of concatenation, for example, occurs in each system, and it is always interpretable as a meet operation. But it is an important feature of each system, missing from most axiomatizations of probability, that no assumption is made about the elements of the domain, or the algebraic operations that act on them, beyond what is stated in the probability axioms. The axiomatizations are autonomous (Popper, 1959, Appendix *iv). It is clear that our axiomatizations could all without difficulty be cast into the explicitly set-theoretical format preferred by Suppes.

Within each system of axioms we introduce the formal definitions

\[(0.0) \quad x \triangleq z = D_f \forall y [p(x, y) \leq p(z, y)] \]

\[(0.1) \quad x \sim z = D_f x \leq z \& z \leq x. \]

In (0.1) is defined the idea of probabilistic indistinguishability [in the first argument]. It is trivial that $\sim$ stands for an equivalence relation; and quite easy to establish also that if $Y^*$ results from $Y$ by replacing $X$ by $Z$ at any number of places then

\[(0.2) \quad X \sim Z \Rightarrow Y \sim Y^*. \]

(0.2) guarantees that probabilistically indistinguishable terms may be substituted for each other without damage in the first argument of $p$. To extend this result to the second argument, and thus ensure that $\sim$ is a congruence, will require an additional axiom. Once congruence is assured we may introduce an axiom that defines identity as indistinguishability:
Alternatively, each structure in which the system holds may be factored by the equivalence relation in the structure that interprets \( \sim \). By this artifice the models of the systems \( M^+, D^+, \) and \( B^+ \) below are reduced, respectively, to lower semilattices, distributive lattices, and Boolean algebras. The converse result, that on each non-trivial algebra in each variety we can define a function that satisfies the corresponding system of axioms, may also be established, provided the functions are permitted to take values in a [non-standard] extension field of the real line. (For the system \( M^+ \) it is not necessary to go beyond the reals.) But the pursuit of such results goes beyond the ambitions of the present paper.

1. AXIOMS FOR MEET

The language of our first system contains a single binary operation symbol [for the meet operation], represented by concatenation. (Note that concatenation represents also arithmetical product. This will cause no confusion.) Although we shall not repeat the performance for the other systems, on this occasion we give an explicit recursive definition of the algebraic terms (words) of the language. (There are numerical terms as well.) The letters 'x', 'y', 'z', 'w' are metalinguistic variables for variables; 'X', 'Y', 'Z' are metalinguistic variables for words. In what follows we refer to individual variables simply as variables. Words may be substituted for variables in the usual way.

\[
(1.0) \quad \begin{align*}
(i) & & \text{the individual variables are words} \\
(ii) & & \text{if } X \text{ and } Z \text{ are words, then } (XZ) \text{ is a word} \\
(iii) & & \text{these are all the words}
\end{align*}
\]

As advised above, unhelpful parentheses are omitted wherever possible. The axioms of each system are of two kinds: those that make no mention of the algebraic operations such as meet, and those that do. The letter \( A \) labels axioms of the first kind, whilst the others have titles suggestive of the operations they introduce. The nomenclature and notation of (Popper, 1959, Appendices *iv and *v) are accordingly amended. Our first system, \( M \), contains three \( A \)-axioms and two axioms for meet.
A01 \( 0 \leq p(x, z) \)

A02 \( p(x, z) \leq p(y, y) \)

A1 \( \exists x \exists z p(x, z) \neq 0 \)

M1 \( p(xz, y) \leq p(x, y) \)

M2 \( p(xz, y) = p(x, zy)p(z, y) \).

A1 is an extremely weak axiom, so weak that the system has models in which \( p \) identically equals 1. (We shall call such models \textit{degenerate}.) But it suffices for our purposes in this section and the next. M1 turns out to be a law of monotony for \( p \)'s first argument. M2 is the familiar multiplication law of probability.

We prove in this section that in M the meet operation is idempotent, commutative, and associative; that is, that the words \( xx \) and \( x \) are probabilistically indistinguishable, as are the pair \( xz \) and \( zx \), and the pair \( (xy)z \) and \( x(yz) \). Our first Theorem (unlike its Corollary) avoids the use of A01. (If A01 is used, there is an alternative proof that avoids M1. This is left as an exercise.) The same is true of much of Theorem 1: the first identity in each of (1.4) and (1.5), and the whole of the idempotence law (1.6), are proved without calling on A01.

THEOREM 0. For every \( y \)

\[(1.1) \quad p(y, y) = 1.\]

Proof. By A02, \( p(y, y) \) is independent of \( y \). Set \( p(y, y) = k \). Then \( p(yy, yy) = k = p(y(yy), y(yy)) \). From A02 and M1 follow both

\[

depth=2p(y, yy) \leq k = p(yy, yy) \leq p(y, yy) \\
p(y, y(yy)) \leq k = p(y(yy), y(yy)) \leq p(y, y(yy)).
\]

Thus each equals \( k \). By M2, \( k = p(yy, yy) = p(y, y(yy))p(y, yy) = k^2 \), and hence \( k = 0 \) or \( k = 1 \).

By the specification of \( k \), M2 again, M1 twice, and A02

\[
k \cdot p(x, z) = p(xz, xz)p(x, z) \\
= p((xz)x, z) \leq p(xz, z) \leq p(x, z) \leq k.
\]
If \( k = 0 \), therefore, \( p(x, z) = 0 \) for all \( x \) and \( z \), in contradiction to A1. Thus \( k = 1 \). (Note that \( p(xz, z) = p(x, z) \) whether or not \( k = 1 \).) ■

COROLLARY. For every \( x \) and \( z \)

\[ 0 \leq p(x, z) \leq 1. \tag{1.2} \]

*Proof.* Here we use A01. ■

**Theorem 1.** The following inequalities and equalities hold.

\[ p(xz, y) \leq p(z, y) \tag{1.3} \]

\[ p(x, xz) = 1 = p(x, zx) \tag{1.4} \]

\[ p(xz, z) = p(x, z) = p(zx, z) \tag{1.5} \]

\[ p(xz, z) = p(x, z) = p(zx, z) \tag{1.6} \]

*Proof.* Since by (1.2) the value of \( p(x, zy) \) lies in \([0, 1]\), we obtain (1.3) immediately from M2.

By (1.1), M1, and A02 we have

\[ 1 = p(xz, xz) \leq p(x, xz) \leq 1, \]

and likewise, using (1.3) in place of M1,

\[ 1 = p(zx, zx) \leq p(x, zx) \leq 1, \]

which together establish (1.4).

The first half of (1.5) was proved at the end of Theorem 0. The second half is proved from M2 and (1.4):

\[ p(xz, z) = p(z, xz)p(x, z) = p(x, z). \]

To prove (1.6), use M2 and (1.4) to obtain

\[ p(xz, z) = p(x, xz)p(x, z) = p(x, z), \]

then (1.5), M2, and (1.1) to obtain

\[ p(x, z) = p(xz, z) = p(x, zz)p(z, z) = p(x, zz). \]
This completes the proof. 

In (1.6) we have the law of idempotence (for both arguments). A direct, but intuitively dense, proof of the commutative and associative laws for the first argument, using only A01, A02, (1.1), M1, and M2, is given by Popper (1959), Appendix *v. Here we shall establish these laws, and more, in a simpler but longer way, by metamathematical induction on the length of words. We define the factors of a word $Y$ as follows.

\begin{align}
(1.7) \quad & (i) \quad Y \text{ is a factor of } Y \\
& (ii) \quad \text{if } Y = XZ \text{ then every factor of } X \text{ or factor of } Z \text{ is a factor of } Y \\
& (iii) \quad \text{these are all the factors of } Y.
\end{align}

It is obvious that a variable that is a factor of a word $Y$ must occur in $Y$. The converse is almost as obvious.

**Lemma 2.** If $y$ is a variable that occurs in $Y$ then $y$ is a factor of $Y$.

*Proof.* By induction on the length of $Y$, using (1.7). 

**Lemma 3.** If the variable $x$ occurs in $Z$ then $p(x, Z) = 1$.

*Proof.* Suppose that $x$ occurs in $Z$ and $p(x, Z) < 1$. We shall show by induction on the length of $X$ that, as a consequence, if $X$ is a factor of $Z$ and $x$ occurs in $X$ then $p(X, Z) < 1$. As $Z$ is a factor of $Z$, it follows that $p(Z, Z) < 1$, in contradiction to (1.1).

By supposition the result holds for any factor $X$ of $Z$ that contains $x$ and is of length 1. Suppose it holds for all factors that contain $x$ and are of length less than $n$. Let $X = YW$ be a factor that contains $x$ and is of length $n$. Then $p(X, Z) = p(YW, Z)$, which by M1 and (1.3) is no greater than either $p(Y, Z)$ or $p(W, Z)$. Since at least one of $Y$ and $W$ contains $x$ and is of length less than $n$, by the induction hypothesis at least one of $p(Y, Z)$ and $p(W, Z)$ is less than 1. In either case $p(X, Z) < 1$, which is what we strove to prove.

Interestingly enough, it does not seem to be possible to prove Lemma 3 by induction on the length of $Z$. It should be clear that in much the
same way we can show also that \( p(X, Z) = 1 \) whenever \( X \) is a factor of \( Z \). What we need, however, is a much stronger result.

**THEOREM 4.** Let \( X \) and \( Z \) be words such that every variable that occurs in \( X \) occurs also in \( Z \). Then \( p(X, Z) = 1 \).

**Proof.** The proof is by induction on the length of the word \( X \); that is, the number of variables, including repetitions, that occur in \( X \). The basis of the induction, at which \( X \) is supposed to have length 1, is incorporated in Lemma 3.

For the induction step, suppose that the result holds for any appropriate word of length less than \( n \). Let \( X = YW \) have length \( n \). Then

\[
p(X, Z) = p(YW, Z) = p(Y, WZ)p(W, Z)
\]

by M2. Since every variable of \( X \) occurs in \( Z \), certainly every variable of \( Y \) occurs in \( WZ \), and every variable of \( W \) occurs in \( Z \). Moreover each of \( Y \) and \( W \) has length less than \( n \). Thus the induction hypothesis applies to each, and

\[
p(Y, WZ) = 1 = p(W, Z).
\]

This completes the proof. \( \blacksquare \)

We now take note of the definition (0.0), and establish (in Theorem 7) that indistinguishable words are intersubstitutable in the first argument of \( p \).

**THEOREM 5.** The following equivalences hold in \( M \).

\[
(1.8) \quad z \leq x \iff \forall y \left[ p(x, zy) = 1 \right]
\]

\[
(1.9) \quad z \leq x \iff \forall y \left[ p(xz, y) = p(z, y) \right].
\]

**Proof.** Suppose first that \( z \leq x \). Then by instantiating \( y \) in (0.0) with \( zy \) we may infer that \( p(z, zy) \leq p(x, zy) \). By M1 and A02,

\[
p(zy, zy) \leq p(z, zy) \leq p(x, zy) \leq p(zy, zy),
\]

so by (1.1) \( p(x, zy) = 1 \).

Conversely, suppose that the latter identity holds for all \( y \). By M2 and M1 we have

\[
p(z, y) = p(x, zy)p(z, y) = p(xz, y) \leq p(x, y),
\]

which shows that \( z \leq x \). Clearly we have shown that the right-hand side of (1.9) also holds if \( z \leq x \).

For the converse of this, suppose that \( p(xz, y) \) always equals \( p(z, y) \). By M1, \( p(xz, y) \leq p(x, y) \), which proves what is wanted. \( \blacksquare \)
THEOREM 6. Each of the following identities holds in $M$.

(1.10) $p(xx, w) = p(x, w)$

(1.11) $p(xz, w) = p(zx, w)$

(1.12) $p(xyz, w) = p((xy)z, w)$

Proof. (1.10) is identical with part of (1.6). Now note that, according to Theorem 4, $p(xz, (zx)w) = 1$, so that by (1.8), $xz \leq zx$. By symmetry, $zx \leq xz$. (1.11) follows from (0.0) and (0.1). For (1.12), we need two applications of Theorem 4. Otherwise the proof is the same. \[\]

THEOREM 7. Suppose that $X$ is a factor of $Y$. Let $Y^*$ be any result of replacing or not replacing occurrences of $X$ in $Y$ by $Z$. Then

(1.13) $X \sim Z \Rightarrow Y \sim Y^*$.

Proof. By induction on the length of $Y$. \[\]

It is impossible to prove in $M$ that indistinguishable words may replace each other in the second argument of $p$ (Popper, 1959, pp. 339f.; Leblanc, 1981). To ensure that $\sim$ is a congruence, we therefore introduce a further axiom. (In the systems of van Fraassen, 1981, congruence is postulated outright by an axiom scheme called (2.1). On this and several other scores his probabilistic axiomatizations seem markedly less revealing than ours are.) Either of the following statements suffices:

$A2 \quad x \sim z \Rightarrow p(y, x) = p(y, z)$

$A2^+ \quad p(x, z) = p(y, y) = p(z, x) \Rightarrow p(y, x) = p(y, z)$.

$A2$ is clearly a consequence of $A2^+$ (in the presence of $A02$), and is the more immediate assumption. (It is shown in Section 3 that within $M$ it is properly weaker than $A2^+$.) But $A2^+$ has some formal advantages; in particular, it is a universal statement. The systems obtained by adding the axioms $A2$ and $A2^+$ to $M$ are called $M^+$ and $M^{++}$, respectively. It is easily shown that the new axioms do everything that is required of them.
THEOREM 8. Suppose that $X$ is a factor of $Y$. Let $Y^*$ be any result of replacing or not replacing occurrences of $X$ in $Y$ by $Z$. Then in $\mathbf{M}^+$ (and therefore in $\mathbf{M}^{++}$)

\[(1.14) \quad X \sim Z \Rightarrow p(w, Y) = p(w, Y^*).\]

Proof. Omitted. 

It follows from Theorems 6–8 that each structure in which $\mathbf{M}^+$ (or $\mathbf{M}^{++}$) holds may be reduced to a lower semilattice by the methods described at the end of Section 0. The converse in this case is easy.

THEOREM 9. If $\langle \mathcal{M}, \leq \rangle$ is a lower semilattice then the function $\mu$ defined on $\mathcal{M} \times \mathcal{M}$ by

\[\mu(a, c) = \begin{cases} 1 & \text{if } c \leq a \\ 0 & \text{otherwise} \end{cases}\]

satisfies the axioms of the system $\mathbf{M}^{++}$.

Proof. Omitted. 

We conclude this section with two striking equivalents of $A^2_+$ within $\mathbf{M}^+$. Popper (1963, p. 175; see also Popper, 1966/1984, pp. 278–280) has named (1.15) and (1.16) the principle of redundancy and the principle of stability respectively.

THEOREM 10. Within $\mathbf{M}$ the formulas

\[(1.15) \quad p(x, y) = 1 \Rightarrow p(zx, y) = p(z, y) = p(z, xy)\]
\[(1.16) \quad p(x, y) = 1 \Rightarrow [p(z, y) \neq 0 \Rightarrow p(x, zy) = 1]\]

are equivalent, and each is implied by $A^2_+$. Within $\mathbf{M}^+$ each implies $A^2_+$.

Proof. (1.16) follows from (1.15) by application of M2. For the converse, first assume that (1.16) holds, and that $p(x, y) = 1$. If $p(z, y) \neq 0$ then $p(x, zy) = 1$ by (1.16), and the consequent of (1.15) then follows by (1.11) and M2. If $p(z, y) = 0$, however, then by A01 and M1, $p(zx, y) = 0$. That $p(z, xy)$ also equals 0 follows from another use of M2. Thus (1.16) implies (1.15).
Suppose that \( p(x, y) = 1 \). By (1.4), \( p(y, xy) = 1 \); and by (1.5) and supposition, \( p(xy, y) = 1 \). By (1.1) and \( A2^+ \), \( p(z, y) = p(z, xy) \). That each of these is equal to \( p(zx, y) \) follows from \( M^2 \).

To prove \( A2^+ \) from (1.15), assume its antecedent: that \( p(x, z) = p(y, y) - p(z) \). By (1.1), each equals 1. By (1.15), used twice, we may conclude that \( p(y, z) = p(y, xz) \) and \( p(y, x) = p(y, zx) \). But by (1.11) and \( A2 \), \( p(y, xz) = p(y, zx) \). Hence \( p(y, x) = p(y, z) \). In other words, (1.15) implies \( A2^+ \) within \( M^2 \).

2. AXIOMS FOR JOIN

In this section and the next we investigate the effect of adding to the system \( M \) a single axiom for the join operation, namely the addition law

\[
J \quad p(xz, y) + p(x \lor z, y) = p(x, y) + p(z, y).
\]

The addition law, in some form or another, is often taken to epitomize probability theory, but in our opinion the multiplication law \( M^2 \) is much more fundamental than \( J \) is. Since probability theory is not a self-dual theory, it is hard to credit that its most distinctive law is self-dual. Where the addition law, in the form \( p(xz) + p(x \lor z) = p(x) + p(z) \), can well be regarded as primary is in the theory of absolute probability. But with relative probabilities, \( M^2 \) takes precedence.

\( D \) is obtained by adding \( J \) to \( M \). Our main result is that \( D \) does much more than impose a lattice structure on its interpretations. There is a well known connection (Birkhoff, 1973, p. 232) between the modular law and the existence in a lattice of a valuation function (a function \( p \) for which \( p(xz) + p(x \lor z) = p(x) + p(z) \)). It is remarkable that the simultaneous presence of the multiplication law \( M^2 \) enforces the much stronger condition of distributivity. There is accordingly no such thing as a (relative) probability measure on a non-distributive lattice. The significance of this fact (see Popper, 1968) for what is known as quantum logic is beyond the scope of this paper.

**THEOREM 11.** The following is a theorem of \( D \).

\[
(2.0) \quad (x \lor z)y \sim xy \lor zy.
\]
Proof. By M2, followed by J, followed by three further applications of M2,

\[ p((x \lor z)y, w) = p(x \lor z, yw)p(y, w) = p(x, yw)p(y, w) + p(z, yw)p(y, w) - p(xz, yw)p(y, w) = p(xy, w) + p(zy, w) - p((xz)y, w). \]

Now we know from Theorems 6 and 7 that \((xz)y \sim (xy)(zy)\), so we may rewrite the last term on the right, and apply axiom J once more to give

\[ p((x \lor z)y, w) = p(xy, w) + p(zy, w) - p((xy)(zy), w) = p(xy \lor zy, w), \]

which establishes the indistinguishability of \((x \lor z)y\) and \(xy \lor zy\). ■

THEOREM 12. Each of the following holds in the system D.

\begin{align*}
(2.1) & \quad x \lor x \sim x \\
(2.2) & \quad x \lor z \sim z \lor x \\
(2.3) & \quad x \lor (y \lor z) \sim (x \lor y) \lor z.
\end{align*}

Proof. Omitted. ■

The systems \(D^+\) and \(D^{++}\) are obtained in the same way from \(M^+\). It follows from (2.9) below that within D the Axioms A2 and A2+ are demonstrably equivalent. Hence the systems \(D^+\) and \(D^{++}\) are equivalent. But as sets of axioms they are different, and where it is necessary we continue to distinguish them. It is now straightforward within \(D^+\) (and thus in \(D^{++}\)) to establish anew that probabilistically indistinguishable words are intersubstitutable in all probabilistic contexts. A new proof is needed since the set of words, and also the definition of the factors of a word, change – in a way that is obvious – with the introduction of the operation symbol \(\lor\).
THEOREM 13. Suppose that $X$ is a factor of $Y$. Let the expression $Y^*$ stand for any result of replacing or not replacing occurrences of $X$ in $Y$ by $Z$. Then in $D^+$

$$\text{(2.4)} \quad X \sim Z \Rightarrow Y \sim Y^*.$$  

$$\text{(2.5)} \quad X \sim Z \Rightarrow p(w, Y) = p(w, Y^*).$$

Proof. Omitted.

THEOREM 14. The following are theorems of $D$.

$$\text{(2.6)} \quad x(x \lor z) \sim x$$  

$$\text{(2.7)} \quad x \sim x \lor xz.$$  

Proof. Omitted.

With the absorption laws (2.6) and (2.7) we have now proved all the defining laws of distributive lattices. It follows from Theorems 6 and 11–14 that each structure in which $D^+$ holds may be reduced to a distributive lattice by the methods described at the end of Section 0.

THEOREM 15. The following formulas, and also (1.15) and (1.16), are theorems of $D$.

$$\text{(2.8)} \quad p(x, y) + p(z, y) \leq p(xz, y) + 1$$

$$\text{(2.9)} \quad A2 \Rightarrow A2^+.$$  

Proof. The proof of (2.8) is immediate from J and (1.2). To prove (1.15), suppose that $p(x, y) = 1$. By (2.8), (1.3), and (1.11), $p(z, y) = p(zx, y)$, and the rest follows from M2. By Theorem 10, (1.16) is also a theorem of $D$, as is (2.9). (Hence $A2$ and $A2^+$ are equivalent in $D$.)

3. CREATIVE DEFINITIONS

A formula (or set of formulas) introducing a new symbol into a deductive system is called non-creative if all the new theorems derivable with
its assistance contain the new symbol; and it is called a \textit{definition} if, in addition, it allows the new symbol to be eliminated from all contexts (Suppes, 1957, p. 154). It is clear that within \( M \) the axioms M1 and M2 (which introduce the concatenation of two words) are creative. For with their assistance we can derive \((1.1)\), which involves no concatenation and is not derivable from A01–A02 and A1 alone. (For a counterexample, interpret \( p \) as \( 2\mu \), where \( \mu \) is any standard probability measure.) M1 and M2, which, at least in the form given here, offer no way of eliminating concatenation from all contexts, are thus properly regarded as axioms, not as a definition of a new symbol. The same is true in the system \( M^+ \).

The status of the formula J within the systems \( D, D^+, \) and \( D^{++} \) is somewhat different. It certainly provides a direct way of eliminating the symbol \( \lor \) wherever it occurs; and its main task seems to be to fix the interpretation of this symbol (or, more correctly, of the composite functor \( q \) given by \( q(x, y, z) = p(x \lor z, y) \)). In other words, at an intuitive level J is a \textit{bona fide} definition (Popper, 1963, Section IV). Nevertheless, by Suppes’s criterion of non-creativity (\textit{loc. cit.}), J too is creative. Indeed we may immediately derive from it the formulas

\begin{align*}
(3.0) & \forall x \forall z \exists w \forall y p(xz, y) + p(w, y) = p(x, y) + p(z, y), \\
(3.1) & \forall x \forall z \forall y \exists w p(xz, y) + p(w, y) = p(x, y) + p(z, y),
\end{align*}

in which \( \lor \) does not occur. Given the results of Sections 1 and 2, it will hardly be expected that either \((3.0)\) or \((3.1)\) is a theorem of \( M^{++} \). For an explicit counterexample, and for other counterexamples needed shortly, we make use of a family of models in which are satisfied not only all the axioms of \( M^{++} \) but also all the axioms of Boolean algebra. Each model is based on the four-element Boolean algebra \( \{0, b, -b, 1\} \), and concatenation is interpreted standardly by the Boolean meet. The symbol \( p \) is interpreted by a function \( \mu \) of the form given in Table I (the value of \( \mu(x, z) \) is presented in the row headed by \( x \) and the column headed by \( z \)). It is not hard to check that for all \( \beta \) in the interval \([0, 1]\) Table I yields a model of \( M^+ \). Moreover, \( A2^+ \) holds vacuously provided that \( \beta \neq 1 \). Thus Table I is a model of \( M^{++} \) whenever \( \beta \in [0, 1] \).

To show that \((3.0)\) is not a theorem of \( M^{++} \), give \( \beta \) any value except \( 1/2 \); and to show that \((3.1)\) is not a theorem, give \( \beta \) any value except \( 1/2 \) or \( 0 \). For example, if \( \beta = 1/4 \), the function \( \mu \) is the square of the uniform measure (it is not surprising that both \((3.0)\) and \((3.1)\) hold
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>( \mu )</th>
<th>0</th>
<th>(-b)</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( b )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( \beta )</td>
</tr>
<tr>
<td>(-b)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \beta )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

when \( \beta = 1/2 \), and rather obviously satisfies \( \mathbf{M}^{++} \); it is also easily checked that there is no element \( w \) of the algebra for which \( \mu(w, 1) = \mu(b \cdot -b, 1) + \mu(w, 1) = \mu(b, 1) + \mu(-b, 1) = 1/4 + 1/4 = 1/2 \). Hence (3.1) fails; and hence (3.0) fails. On the other hand, (3.1) holds if \( \beta = 0 \), though not uniformly for all \( y \). It follows that in \( \mathbf{M}^{++} \) the formula (3.0) is strictly stronger than is (3.1).

Thus within each of the systems \( \mathbf{D}, \mathbf{D}^+, \) and \( \mathbf{D}^{++} \) the formula \( \mathbf{J} \) is more than a mere abbreviating convention. Still, (3.0) and (3.1) are scarcely glamorous and controversial theorems; each is a (generalized) existential formula, and what in each case are asserted to exist are elements that are directly denoted by the functional notation introduced in \( \mathbf{J} \). A more interesting example of a formula derivable in \( \mathbf{D}^{++} \) but not in \( \mathbf{M}^{++} \), is (2.8), as may be seen by giving \( \beta \) any value in the open interval \( (1/2, 1) \). Central formulas derivable in \( \mathbf{D}^+ \) but not in \( \mathbf{M}^+ \) are \( \mathbf{A}^2^+, \) (1.15), and (1.16): each holds if \( \beta < 1 \), but not if \( \beta = 1 \), so none is a theorem of \( \mathbf{M}^+ \). The formulas (1.15) and (1.16), like (2.9), are even derivable in \( \mathbf{D} \), but not (of course) in \( \mathbf{M} \).

Formulas (1.15), (1.16), \( \mathbf{A}^2^+ \), (2.8), and (2.9), are all strictly universal; their prenex normal forms contain only universal quantifiers. Thus \( \mathbf{J} \), appended to any of the systems \( \mathbf{M}, \mathbf{M}^+, \) and \( \mathbf{M}^{++} \), is creative in a sense stronger than that embodied in Suppes’s criterion: it permits derivation of new universal theorems not involving the defined symbol \( \forall \). This idea is familiar to proof theorists, who would describe \( \mathbf{D}^{++} \) (for instance) as a non-conservative extension of \( \mathbf{M}^{++} \) for \( \Pi^0_1 \) formulas. (The idea of a non-conservative extension is wider than that of a creative definition.) The point may also be made model-theoretically (for the
model-theoretic terminology see for example Bridge (1977, pp. 13 and 23).

Let \( \mathbf{X} \) be a theory and \( \phi \) a formula introducing a new operation symbol \( \lambda \) into the language of \( \mathbf{X} \). Let \( \mathbf{Z} \) be the enlarged theory, and let \( \mathfrak{A} \) be a model of \( \mathbf{X} \). Then if \( \phi \) is a genuine abbreviatory definition of \( \lambda \), a model for \( \mathbf{Z} \) may be obtained merely by expansion of \( \mathfrak{A} \); no more need be done than to appoint some element or elements of the domain of \( \mathfrak{A} \) to serve as the interpretation of \( \lambda \). If, however, \( \phi \) is a creative definition, then the rule that Suppes, following Leśniewski, proposes for the definition of operation symbols (\textit{op. cit.}, p. 158) implies that it will sometimes be necessary first to recast the domain of the model \( \mathfrak{A} \), by adding new elements or purging existing ones; the extension (or submodel) \( \mathfrak{L} \) may then be expanded to a model \( \mathfrak{M} \) for \( \mathbf{Z} \). Such a process of extension-cum-expansion suffices if \( \mathbf{X} \) is the theory of dense order without endpoints, and \( \phi \) defines a least element. With the systems \( \mathbf{M}, \mathbf{M}^+, \) and \( \mathbf{M}^{++} \), and the definition \( \mathfrak{I} \), however, more drastic measures are in some cases needed in order to deliver a model of the enlarged theory. For an illustration, suppose that \( \mathfrak{A} \) is a model for \( \mathbf{M}^{++} \) in which (2.8) fails; that \( \mathfrak{L} \) is an extension of \( \mathfrak{A} \), and that \( \mathfrak{M} \) is a expansion of \( \mathfrak{L} \) in which \( \mathbf{D}^{++} \) holds. Then \( \mathfrak{M} \) is a model for (2.8), and so \( \mathfrak{L} \) is too. But a standard theorem of model theory (Bridge, 1977, p. 39) tells us that a strictly universal sentence that is true in a model is true in all its submodels. Thus (2.8) is true also in \( \mathfrak{A} \), contrary to specification.

If \( \beta \) is given any value in \((1/2, 1)\), say \( 3/4 \), the model shown in Table I is such a model \( \mathfrak{A} \). Since the meet of \( b \) and \( -b \) is the element \( 0 \), and 
\[
\mu(b, 1) + \mu(-b, 1) - \mu(b \cdot -b, 1) = 3/2,
\]
there is no way, if \( \mathfrak{I} \) is to hold, to extend the model to contain an element to play the role of the join of \( b \) and \( -b \). In contrast, it is easy to find a non-degenerate submodel of Table I that can be expanded to a model of \( \mathfrak{I} \): simply delete the elements \( b \) and \( -b \). Indeed, this procedure is available whichever non-degenerate model \( \mathfrak{A} \) of \( \mathbf{M}^{++} \) we start with. In effect we select two ordered elements of the domain of \( \mathfrak{A} \), and take the smaller of the two as their meet, the larger of the two as their join.

**THEOREM 16**. Let \( \mathfrak{A} = \langle \mathcal{K}, \cdot, \mu \rangle \) be a non-degenerate model of \( \mathbf{M}^{++} \). Then there is a non-degenerate submodel \( \mathfrak{L} \) of \( \mathfrak{A} \) that may be expanded to a model of \( \mathbf{D}^{++} \).

**Proof**. Choose elements \( a, c \) of \( \mathcal{K} \) such that \( \mu(a, c) \neq 1 \). Let 
\[
b = a \cdot c.
\]
Then \( \mu(b, c) \neq 1 \), by (1.5), and \( \mu(c, b) = 1 \), by (1.4). It
is easily established that if \( \cdot \) and \( \mu \) are now restricted to \( \{b, c\} \), then
the structure \( \mathcal{L} = \langle \{b, c\}, \cdot, \mu \rangle \) is a model of \( \mathbf{M}^{++} \). Now define
the operation \( + \) on \( \{b, c\} \) as follows: \( b + b = b \) and \( b + c = c + b = c + c = c \). It is trivial that \( \mathcal{M} = \langle \{b, c\}, \cdot, +, \mu \rangle \) is also a model of \( \mathcal{J} \). □

4. PROBABILITY AS A GENERALIZATION OF LOGIC

A system \( \mathbf{B}^+ \) of probability of which any model is reducible to a Boolean
algebra is obtained from \( \mathbf{D}^+ \) by adjoining this axiom for complement:

\[
\mathcal{C} \quad p(y, z) \neq p(z, z) \Rightarrow p(z, z) = p(x, z) + p(x', z).
\]

If \( \text{A1} \) is also strengthened slightly (ensuring that \( p \) takes two values)
\[
\text{A1}^+ \quad \exists x \exists z p(x, z) \neq p(z, z),
\]
we obtain a set of axioms equivalent to the system preferred by Popper
(1959, Appendices *iv and *v), though slightly diverging from it in detail. (Note that Axiom \( \mathcal{C} \) is given a variant formulation on p. 298 of Popper (1966/1984).) The principal differences are that in Popper’s system
the definition \( \mathcal{J} \), which is genuinely non-creative in \( \mathbf{B}^+ \), is dropped or replaced
by the explicit definition \( x \vee z = (x' z')' \); and that \( \text{A01} \) and \( \text{A02} \) are replaced by
the much weaker consequence \( \text{A3} \): \( p(x, x) \leq p(z, z) \)
(or the same formula with \( = \)). There is also a system \( \mathbf{B}^{++} \), deductively
equivalent to \( \mathbf{B}^+ \), in which \( \text{A2}^+ \) replaces \( \text{A2} \). Everything said below
about \( \mathbf{M}^+, \mathbf{D}^+, \) and \( \mathbf{B}^+ \), applies equally to \( \mathbf{M}^{++}, \mathbf{D}^{++}, \) and \( \mathbf{B}^{++} \). We
exclude degenerate models; models, that is, in which all probabilities
equal 1.

It is transparent that there is a close connection between \( \mathbf{B}^+ \) and
classical sentential logic (and between \( \mathbf{M}^+ \) and \( \mathbf{D}^+ \) and what might be
called conjunctive logic and distributive logic). But exactly how are
the standard logical relations to be defined in probabilistic terms? The
orthodox answer, in the enterprise known as probabilistic semantics
(for instance, Field (1977); Leblanc, (1979, 1983), and numerous later
papers; Paulos (1981); van Fraassen (1981); and Bendall (1982)), is
to start with a language \( L \) for sentential logic, and adopt the theory of
probability \( \mathbf{B}^+ \) as a metatheory for \( L \). The variables \( x, y, z \) therefore
range over the sentences of \( L \), and ‘\( p \)’ is a variable for real-valued
functions defined on pairs of sentences of \( L \). The free variable ‘\( p \)’ is
open to universal quantification; and in probabilistic semantics (though
this is not the most usual definition) \( z \) is said to imply \( x \) if for each \( p \) that
satisfies the appropriate axiom system, and for each \( y, p(z, y) \leq p(x, y) \); that is, if \( z \preceq x \). Then if \( p \) ranges over functions that satisfy \( \mathbf{B}^+ \), or \( \mathbf{M}^+ \), or \( \mathbf{D}^+ \), it is relatively straightforward to prove the soundness and completeness theorems for the standard formulations of classical, or conjunctive, or distributive logic (Leblanc, 1983, Section 5).

Our approach is quite different, and may rather be described as probabilistic syntax. Let \( X \) and \( Z \) be words of the language of \( \mathbf{B}^+ \), words in which concatenation and the symbols \( \lor \) and \( ' \) may occur. Interpreting these words in the customary way as sentences turns our original object language (the language of \( \mathbf{B}^+ \)) into a metalanguage, containing names, such as \( \lor ' \), for object-linguistic connectives; and the old metalanguage, in which the variables \( 'X' \) and \( 'Z' \) occur, becomes the metametalanguage. We may then define \( Z \vdash X \) in the corresponding system of logic to mean that the (closure of) the formula \( Z \preceq X \) holds in some model of \( \mathbf{M}^+ \), \( \mathbf{D}^+ \), or \( \mathbf{B}^+ \); for instance, \( Z \vdash X \) is valid in classical logic if and only if \( Z \preceq X \) holds for some non-degenerate probability function \( \mu \) satisfying \( \mathbf{B}^+ \). That this condition is necessary is undisputed. But it is also sufficient, since a closed formula true for one function \( \mu \) satisfying \( \mathbf{M}^+ \), \( \mathbf{D}^+ \), or \( \mathbf{B}^+ \) is true for any other. This follows from a result of Kalicki and Scott (1955) that semilattices, distributive lattices, and Boolean algebras are equationally complete; that is, if the closure of the formula \( Z \preceq X \) is true at all in such an algebra, it is a theorem. (No such result holds, however, for Heyting algebras. For this reason, and others, a satisfactory probabilistic definition of intuitionistic deducibility is much harder to achieve.)

It has been objected by Leblanc and van Fraassen (1979, p. 369; see also note 5), that this identification of \( Z \preceq X \) (or equivalently in \( \mathbf{B}^+ \), the identity \( p(X, ZX') = 1 \)) with \( Z \vdash X \) is spurious; more simply, that it is incorrect to interpret \( p(x, X') = 1 \) as meaning that \( X \) is logically true. Their argument is that there are models of \( \mathbf{B}^+ \) in which \( \mu(b, \neg b) = 1 \) even when \( b \) is not logically true. But this would be like arguing that, because it has models in which \( A \vdash C \) holds even when \( A \) does not imply \( C \) (for example, let \( \Gamma \vdash \Delta \) hold except when the elements of \( \Gamma \) are all true and those of \( \Delta \) all false), Gentzen’s sequent calculus does not properly capture in \( \vdash \) the relation of logical implication. The objection (which relies on reading ‘\( X \)’ and ‘\( Z \)’ as names for sentences, not as metametalinguistic variables) may therefore safely be rejected.

The great difference between probabilistic semantics and our own approach is shown by the fact that what Leblanc (op. cit, p. 264; we have
made inconsequential notational changes), calls "a soundness theorem, . . . roughly to the effect that if a Boolean identity $x = z \ldots$ is provable . . . then $p(x, y) = p(z, y)$ for any . . . $y$ and any binary probability function $p$" is from our viewpoint not a soundness theorem at all but a completeness theorem; in effect it establishes that the system $\mathbf{B}^+$ is able to deliver $X \sim Z$ whenever $X = Z$ is an identity of Boolean algebra. We shall say more elsewhere about this difference of approach.

There is, it must be emphasized, nothing technically incorrect about probabilistic semantics. For those who find the usual semantics of elementary logic more problematic than its syntax, it may even be a godsend. But it is not what was intended when these studies started.

Karl Popper
Kenley,
Surrey CR8 5HH,
U.K.

David Miller
Department of Philosophy,
University of Warwick,
Coventry CV4 7AL,
U.K.

REFERENCES


CONTRIBUTIONS TO THE FORMAL THEORY OF PROBABILITY


COMMENTS BY PATRICK SUPPES

New results on the formal theory of probability are presented by Popper and Miller in a clear and rigorous fashion. They complement and extend very nicely much earlier work of Popper. I found especially interesting the axioms that are intuitively just for multiplication of probabilities – this excludes the additivity axiom J. Such axiomatic systems have undoubtedly received too little study in the general literature. One intriguing question is how easy it is to use their axioms to obtain by obvious translation a qualitative theory of multiplicative probability, as I shall call it. A lot is known about such qualitative probability structures, but as far as I know they have not been studied in a context parallel to the axiom systems $M, M^+$ and $M^{++}$. (By a qualitative relation I mean one in their notation such that $x, y \preceq z, w$ iff $p(x, y) \leq p(z, w)$). The more problematic question is how to formulate qualitatively the multiplicative axiom M2. In Suppes and Zanotti (1982) the following axiom is the critical one for multiplication:

$$\text{If } x \rightarrow y \text{ then } z, x \preceq w, x \text{ iff } zx, y \preceq wx, y$$
I have translated it into Popper and Miller’s notation, but it is unclear to me how it relates in detail to M2. Of course, it is obvious that to get a numerical representation from such a qualitative relation it will be necessary to add some stronger axioms. However, I would prefer to raise another question in a different direction. What about using such qualitative theories – which can of course be formulated strictly in first-order logic – as the first step in generalizations of the relation of derivability. My conjecture is that almost all of what we want as a qualitative generalization of the relation of derivability could be obtained from the purely qualitative axioms that model as closely as possible the current axioms of Popper and Miller.

Popper and Miller already provide hints in this direction in their remarks about Leblanc and van Fraassen where they identify logical implication with a probability 1 statement. But a more elaborate development is required to check the conjecture I have just made. I like their system and I do hope they will encourage further work along the line indicated, which would establish some interesting close connections with the extensive literature on qualitative probability and the theory of measurement, much discussed in various volumes of *Foundations of Measurement*.

Popper and Miller remark at the beginning of Section 2 that their multiplication axiom M2 is more fundamental than the axiom J of addition. This runs counter to the popular view of probability as an additive set measure with norm 1, and I am sorry they did not say more about this point. In general terms I think of the fundamental concepts of probability as being the following:

1. Additivity for exclusive events;
2. Independence of some events from others;
3. Conditional dependence of some events on others;
4. Randomness of occurrence of events.

On the surface, this list would seem to go against Popper and Miller’s view, but in fact their M2 expresses a fundamental property of conditional probability. Moreover, this property holds when additivity is weakened to the subadditivity of upper probability measures or the superadditivity of lower probability measures. I give the simple proof for upper probability measures. First, define conditional upper probability in the usual way: if $P^*(B) > 0$ then

$$P^*(A \mid B) = P^*(AB)/P^*(B).$$
Then if $P^*(BC) > 0$

$$P^*(AB | C) = \frac{P^*(ABC)}{P^*(C)} = \frac{P^*(A | BC)P^*(B | C)P^*(C)}{P^*(C)} = \frac{P^*(A | BC)}{P^*(B | C)}$$

The point of the proof is to show that additivity is not required, only the definition of conditional probability and the Boolean properties of events. (Note, of course, that Popper and Miller do not require my condition that $P^*(BC) > 0$, a computational advantage of their system in dealing with conditional probability.)

**Final Note.** I marvel at Karl Popper’s energy and intellectual vigor at 90. Two decades from now I hope to be able to salute David Miller on his 70th birthday with something comparable.

**REFERENCE**