Chapter 9

TWO DOGMAS OF PROBABILISM

Carl G. Wagner*
The University of Tennessee

1. INTRODUCTION

Lehrer's epistemology, as articulated, for example, in *Rational Consensus in Science and Society* (Lehrer and Wagner 1981), has always emphasized that rational decision making must take account of the total available evidence. Yet dogmatic restrictions on the representation of uncertain judgment, or on the way in which such judgment may be revised, undermine the goal of faithfully representing the evidence. In this paper we discuss two such restrictions, *dogmatic Bayesianism* and the *dogma of precision*, and outline some ways in which probabilism has begun to be liberated from their grip.

Dogmatic Bayesianism, which asserts that the only acceptable method of revising a probability distribution is by conditionalizing on an event E that one has come to regard as certain, has already been substantially weakened by the discovery of principled ways of updating probabilities when conditionalization is simply inapplicable. Since descriptions of these alternative revision methods are accessible and clear, we shall simply mention

1) revising one's probability distribution by means of a *weighted average* of that distribution along with those of other informed individuals (Lehrer and Wagner 1981);

2) *probability kinematics* (Jeffrey 1965, 1983, 1988), a generalization of conditionalization in which new evidence alters the probabilities of a disjoint family of events; and

3) *reparation* (Jeffrey 1991, 1995; Wagner 1997, 1999), a revision method

E.J. Olsson (ed.), The Epistemology of Keith Lehrer, 143–152.

that raises the probability of hypothesis H when it is discovered that H implies previously known evidence E.[1]

Probabilism, which allows the representation of degrees of belief by subjective probabilities taking any values in the interval $[0, 1]$, itself offers a profound expansion of the classical dogmatic epistemological categories "accept," "reject," and "suspend judgment." Yet the expressive resources of probabilism need to be further enlarged. Recall that on the standard account (Ramsey 1990) subjective probability is a measure of one's degree of confidence in the truth of a proposition or the occurrence of an event, as reflected in one's willingness to take either side of certain bets. It is supposed that one is always capable of articulating the precise odds governing such bets. That this dogma of precision, as Walley (1991) has called it, is both unrealistic and unnecessary, is gaining acceptance among students of the foundations of probability. In what follows we outline the elementary parts of the theory of upper and lower probabilities, with the aim of giving these ideas wider currency among epistemologists.

2. SUBJECTIVE PROBABILITY AND THE DOGMA OF PRECISION

For the sake of simplicity it is assumed in what follows that your frame of discernment Ω regarding possible states of the world is finite.[2] It is also supposed that there is an infinitely divisible unit of utility. Suppose that you are able to assign to each event $A \subset \Omega$ a real number $p(A)$ such that

1° You are willing to pay anything less than $p(A)$ units of utility in exchange for receiving one such unit if A occurs, and nothing if A fails to occur; and

2° In exchange for receiving anything more than $p(A)$ units of utility, you are willing to obligate yourself to pay one such unit if A occurs, and nothing if A fails to occur.

The set function p is your *subjective probability* on events in Ω, with $p(A)$ being your *threshold price for* A.[3] Standard *Dutch book arguments* (see, e.g., Earman 1992, pp. 38–40) show that in order to avoid a sure loss, p must be *coherent*, i.e., satisfy the usual axioms for a probability measure,

$$p(A) \geqslant 0, \quad \text{for all } A \subset \Omega, \tag{2.1}$$

$$p(\Omega) = 1, \quad \text{and} \tag{2.2}$$

$$p(A \cup B) = p(A) + p(B), \text{ for all } A, B \subset \Omega \text{ such that } A \cap B = \varnothing. \tag{2.3}$$

In section 4 we show as a simple corollary of a much more general result that coherence is also sufficient to avoid a sure loss.

The demands placed on probability assessors by the dogma of precision are stringent, and unrealistic. How, for example, is one to assess the probability of getting a white ball in a random selection from an urn containing red and white balls in unknown proportion?[4] An additional such example, the case of incompletely specified contingency tables, is described in the next section. As we shall see, an elegant analysis of this case, due to Strassen (1964), leads naturally to a simple, intuitively appealing general account of upper and lower probabilities.

3. STRASSENIAN UPPER AND LOWER PROBABILITIES

Imagine a collection of objects, consisting of spheres, cylinders, cubes, and cones, each of which is colored red, white, or blue. Suppose that 50 % of the objects are red, 30 % are white, and 20 % are blue. There are no red cubes or red cones, no white cylinders or white cubes, and no blue spheres. An object is chosen at random from this collection. What is the probability that it is 1. a sphere, 2. a sphere or cylinder? This problem involves the incomplete contingency table

	sphere	cylinder	cube	cone	
red			0	0	.5
white		0	0		.3
blue	0				.2
					1.0

and furnishes another example in which information is insufficient to assess precise probabilities. On the other hand it is fairly easy to see, e.g., that no more than 80 % of the objects can be spheres, and that at least 50 % of the objects are spheres or cylinders.

Strassen (1964) furnished the following elegant analysis of the general problem of this type: Let Ω and Θ be frames of discernment regarding the state of the world, let p be a probability on events in Ω, and suppose that for each $\omega \in \Omega$ the set $T(\omega)$, comprising those outcomes $\theta \in \Theta$ compatible with the outcome ω, is nonempty. For all $A \subset \Theta$, let

$$A_* := \{\omega \in \Omega : T(\omega) \subset A\}, \qquad \text{and} \tag{3.1}$$

$$A^* := \{\omega \in \Omega : T(\omega) \cap A \neq \varnothing\}, \tag{3.2}$$

and let

$$\beta(A) := p(A_*), \qquad \text{and} \tag{3.3}$$

$$\alpha(A) := p(A^*). \tag{3.4}$$

The set functions β and α are, respectively, the *Strassenian lower and upper probabilities*[5] on events in Θ induced by p and the compatibility relation T. The reasons for this terminology will soon be made clear, but let us first note some basic properties of these set functions.

Theorem 3.1. *The set functions β and α defined by (3.3) and (3.4) have the following properties:*

(i) $0 \leqslant \beta(A) \leqslant \alpha(A) \leqslant 1$, *for all* $A \subset \Theta$.

(ii) $\beta(\varnothing) = \alpha(\varnothing) = 0$ *and* $\beta(\Theta) = \alpha(\Theta) = 1$.

(iii) β *and* α *are monotone, i.e., if* $A_1 \subset A_2$, *then* $\beta(A_1) \leqslant \beta(A_2)$ *and* $\alpha(A_1) \leqslant \alpha(A_2)$.

(iv) β *and* α *are conjugates, i.e.,* $\beta(A) + \alpha(\bar{A}) = 1$ *for all* $A \subset \Theta$.

(v) *For every positive integer* r, β *is* r*-monotone and* α *is* r*-alternating, i.e.*

$$\beta(A_1 \cup \cdots \cup A_r) \geqslant \sum_i \beta(A_i) - \sum_{i<j} \beta(A_i \cap A_j) + \cdots + (-1)^{r-1}\beta(A_1 \cap \cdots \cap A_r) \quad (3.5)$$

and

$$\alpha(A_1 \cap \cdots \cap A_r) \leqslant \sum_i \alpha(A_i) - \sum_{i<j} \alpha(A_i \cup A_j) + \cdots + (-1)^{r-1}\alpha(A_1 \cup \cdots \cup A_r), \quad (3.6)$$

from which it follows that

(vi) β *is superadditive and* α *is subadditive, i.e., if* $A_1 \cap A_2 = \varnothing$, *then* $\beta(A_1 \cup A_2) \geqslant \beta(A_1) + \beta(A_2)$ *and* $\alpha(A_1 \cup A_2) \leqslant \alpha(A_1) + \alpha(A_2)$.

Proof. The proofs of (i)–(iii) are straightforward. The proof of (iv) follows from the fact that A_* and $(\bar{A})^*$ are set-theoretic complements. To prove (3.5) note that $(A_1 \cup \cdots \cup A_r)_* \supset (A_1)_* \cup \cdots \cup (A_r)_*$ and that for every $I \subset \{1, \ldots, r\}$

$$\left(\bigcap_{i \in I} A_i\right)_* = \bigcap_{i \in I} (A_i)_*.$$

Then apply monotonicity of p and the principle of inclusion and exclusion for p. The inequality (3.6) follows from (3.5) and (iv). Assertion (vi) follows from the case $r = 2$ of (3.5) and (3.6). □

Theorem 3.2. *The following are equivalent:*

(i) *If $p(\omega) > 0$, then ω is compatible with exactly one outcome $\theta \in \Theta$.*

(ii) *β is a probability measure.*

(iii) *α is a probability measure.*

(iv) $\beta = \alpha$.

Proof. Straightforward. □

The above theorem simply confirms what one would expect, namely, that when T is in effect a Θ-valued random variable, then α and β coincide with the usual probability measure induced on events in Θ by p and T.

We now demonstrate the appropriateness of calling α and β upper and lower probabilities. In what follows, I_A denotes the indicator of the event A, i.e., $I_A(\theta) = 1$ if $\theta \in A$ and $I_A(\theta) = 0$ if $\theta \in \bar{A}$, and we routinely omit the phrase "units of utility." The betting commitments described in 1° and 2° of section 2 above will be tersely characterized, as a willingness to "buy I_A for $p(A) - \varepsilon$, for all $\varepsilon > 0$," and "sell I_A for $p(A) + \varepsilon$, for all $\varepsilon > 0$."

Suppose that p is your subjective probability on events in Ω, with T, α, and β as above. Except in the case described in Theorem 3.2, you have inadequate information to ground assessment of a subjective probability on events in Θ. There are, however, identifiable constraints on any such probability.

Theorem 3.3. *Any coherent probability q that might be a candidate for representing your threshold prices for events in Θ must satisfy*

$$\beta(A) \leqslant q(A) \leqslant \alpha(A), \qquad \text{for all } A \subset \Theta, \tag{3.7}$$

for if (3.7) *is violated, you will suffer a sure loss.*

Proof. Suppose that $q(A) < \beta(A)$ for some $A \subset \Theta$, with $\beta(A) - q(A) = \varepsilon > 0$. You'll sell I_A for $q(A) + \varepsilon/4$ and buy I_{A_*} for $p(A_*) - \varepsilon/4 = \beta(A) - \varepsilon/4$. Suppose that ω is the true Ω-state and that θ is the true Θ-state. If $\omega \in A_*$, then $\theta \in A$, and so your net gain is $(q(A)+\varepsilon/4-1)+(1-\beta(A)+\varepsilon/4) = -\delta/2 < 0$. If $\omega \notin A_*$, θ may or may not be an element of A. If $\theta \in A$, your net gain is $(q(A) + \varepsilon/4 - 1) + (0 - \beta(A) + \varepsilon/4) = -1 - \varepsilon/2 < 0$, and if $\theta \notin A$, your net gain is $(q(A) + \varepsilon/4 - 0) + (0 - \beta(A) + \varepsilon/4) = -\varepsilon/2 < 0$.

Suppose that $q(A) > \alpha(A)$. Then $q(\bar{A}) = 1 - q(A) < 1 - \alpha(A) = \beta(\bar{A})$, by Theorem 3.1 (iv), which leads, as above, to a sure loss. □

As noted above, you are perfectly justified in the above situation in refusing to announce threshold prices for events in Θ. But there are some additional bets that you ought to be willing to make:

Theorem 3.4. *If β and α are defined by* (3.3) *and* (3.4) *you ought to be willing, for each $A \subset \Theta$, and for all $\varepsilon > 0$, to*

1° *buy I_A for $\beta(A) - \varepsilon$, and*

2° *sell I_A for $\alpha(A) + \varepsilon$.*

Proof. The argument here is not that you will otherwise suffer a sure loss (no one can make a Dutch book against someone unwilling to bet), but, rather, that 1° and 2° are at least as good as bets you are willing to make. In the case of 1°, you'll buy I_{A_*} for $p(A_*) - \varepsilon = \beta(A) - \varepsilon$, so you ought to be willing to buy I_A for that price, since if the true Ω- and Θ-states are, respectively, ω and θ, and $\omega \in A_*$, then $\theta \in A$, so I_A pays off for you in every case that I_{A^*} does, and possibly other cases as well.

In the case of 2°, you'll sell I_{A^*} for $p(A^*) + \varepsilon = \alpha(A) + \varepsilon$, so you ought to be willing to sell I_A for that price. With ω and θ as above, if $\omega \notin A^*$, then $T(\omega) \cap A = \varnothing$, i.e., $T(\omega) \subset \bar{A}$. Since $\theta \in T(\omega)$, $\theta \notin A$. So in every case in which you avoid paying off on I_{A^*} (and perhaps in other cases as well) you will avoid paying off on I_A. □

4. UPPER AND LOWER SUBJECTIVE PROBABILITIES

Theorem 3.4 leads naturally to the following generalization of classical subjective probability: Suppose that for each event $A \subset \Omega$ you are able to assign real numbers $\lambda(A)$ and $\nu(A)$ such that, for all $\varepsilon > 0$, you are willing to

1° buy I_A for $\lambda(A) - \varepsilon$, and

2° sell I_A for $\nu(A) + \varepsilon$.

The set functions λ and ν are then, respectively, your *lower* and *upper subjective probabilities* on events in Ω. Whereas in section 2 above your threshold prices as bettor and bookie were required to be identical, here they may be distinct, with obvious gains in realism and expressive possibility. The following theorem gives necessary and sufficient conditions for avoiding a sure loss in the above situation:

Theorem 4.1. *Given the betting commitments* 1° *and* 2° *above, you will avoid a sure loss if and only if there exists a coherent probability q on events in Ω such that*

$$\lambda(A) \leqslant q(A) \leqslant \nu(A), \qquad \text{for all } A \subset \Omega. \tag{4.1}$$

Proof. See Walley 1981, p. 15. □

It follows immediately from Theorem 4.1 that coherence of subjective probabilities, as defined in section 2 above, is not only necessary, but also sufficient to avoid a sure loss (cf. Kemeny 1955 and Lehman 1955).

Note that (4.1) is a rather weak condition, which, for example, does not even imply monotonicity of λ and ν. Buehler (1976) has argued for much more stringent restrictions. Indeed, he claims *inter alia* that the lower probability λ must be additive! Buehler's argument is based on the following theorem.

Theorem 4.2. *Suppose that* $\lambda(A) \leqslant q(A)$ *for all* $A \subset \Omega$, *where* q *is a coherent probability, and that* λ *is self-conjugate, i.e.,* $\lambda(A) + \lambda(\bar{A}) = 1$ *for all* $A \subset \Omega$. *Then* $\lambda = q$.

Proof. Suppose that $\lambda(A) < q(A)$ for some A. Then, by self-conjugacy of λ, $\lambda(\bar{A}) = 1-\lambda(A) > 1-q(A) = q(\bar{A})$, contradicting the fact that λ is dominated by q. □

It follows from Theorems 4.1 and 4.2 that if λ avoids a sure loss, and is self-conjugate, then λ is in fact a coherent probability. Buehler's argument that λ must be self-conjugate goes as follows: Clearly, $\lambda(A) + \lambda(\bar{A}) \leqslant 1$; otherwise you will suffer a sure loss. Suppose that $\lambda(A) + \lambda(\bar{A}) < 1$. Let a and b be such that $\lambda(A) < a$, $\lambda(\bar{A}) < b$, and $a + b < 1$. Then you'll reject buying I_A for a and $I_{\bar{A}}$ for b, even though by accepting both bets you would be guaranteed of the net gain $1 - a - b > 0$. So you will miss out on a sure gain. Apart from the fact that missing a sure gain is considerably less serious than suffering a sure loss, this argument is further weakened by its dependence on your being offered I_A and $I_{\bar{A}}$ one at a time, with no knowledge that both will be offered. If you were offered these bets simultaneously, you would recognize immediately that you were being offered a certain payoff of 1, and would clearly agree to pay any price less than 1 in exchange.

Here is a simple way in which lower and upper probabilities satisfying (4.1) arise: Let $\mathcal{P}$ be a nonempty family of coherent probabilities on events in Ω and let

$$\lambda_{\mathcal{P}}(A) := \inf_{p\in\mathcal{P}}\{p(A)\}, \qquad \text{and} \tag{4.2}$$

$$\nu_{\mathcal{P}}(A) := \sup_{p\in\mathcal{P}}\{p(A)\}. \tag{4.3}$$

Then $\lambda_{\mathcal{P}}$ and $\nu_{\mathcal{P}}$ satisfy (4.1) as well as the conjugacy relation $\lambda_{\mathcal{P}}(A)+\nu_{\mathcal{P}}(\bar{A}) = 1$. The set functions $\lambda_{\mathcal{P}}$ and $\nu_{\mathcal{P}}$ are, respectively, the *lower* and *upper envelopes of* $\mathcal{P}$.

Examples of naturally occurring families $\mathcal{P}$ (called *probasitions* in Jeffrey 2001) include such things as 1. the family of all additive representations of some comparative probability relation (Roberts 1976) and 2. the family of all probabilities with respect to which a fixed random variable has a fixed expected value. The Strassenian lower and upper probabilities β and α are also envelopes, with $\mathcal{P}$ = the set of all marginalizations to Θ of all probabilities Q on $\Omega \times \Theta$ that are compatible with p and T in the sense that the marginalization of Q to Ω is p and $Q(\omega, \theta) = 0$ if $\theta \notin T(\omega)$ (Wagner 1992).

5. CONCLUSION

Dogmatic restrictions on the representation of uncertain judgment, or on the way in which such judgment is revised, undermine the goal of faithfully representing the evidence regarding the state of the world. While Bayesian dogmatism has begun to yield to other principled methods of probability revision, the dogma of precision is still dominant.

One source of resistance to working with non-additive upper and lower probabilities is the fear that such measures must necessarily be mathematically intractable. This greatly exaggerates the true state of affairs. While space does not permit a detailed account, we mention that there is a useful theory of upper and lower expectation (see Dempster 1967 and Walley 1981, 1991), as well as a generalization of probability kinematics in which new evidence places bounds on possible revisions of prior in the form of Strassenian upper and lower probabilities (Wagner 1992). Finally, to conclude this paper on the same note on which it began, we remark that there is a theory of consensus for upper and lower probabilities (Wagner 1989) which is remarkably similar to that in Lehrer and Wagner 1981.

ENDNOTES

* Research supported by the National Science Foundation (SES-9984005)

[1] Reparation thus provides a solution to the *old evidence problem*, first posed by Glymour (1980).

[2] The term "frame of discernment" is due to Shafer (1976). The elements of Ω are mutually exclusive and exhaustive, i.e., precisely one element of Ω, though typically unknown, represents the true state of the world. Multiple frames of discernment may, however, be brought to bear on a single problem, as, for example, when we classify the possible outcomes of selecting an object at random from a set of colored shapes by frames delineating 1) the possible colors and 2) the possible shapes. Mathematicians typically refer to Ω as a "sample space."

[3] It is implicit here that you are unwilling to pay more than $p(A)$ in 1° and unwilling to take less than $p(A)$ in 2°. In the usual treatment, you must be willing to pay $p(A)$ in 1° and to take $p(A)$ in 2°. We do *not* require this. A virtue of our treatment is that one can assign proper subsets of Ω (i.e., contingent events) probability one without being in the position of having no prospect of

positive gain, and the possibility of a loss (cf. Earman 1992, p. 41). Our treatment also allows for a natural segue to the account of upper and lower probabilities in section 4.

[4] Devotees of the principle of insufficient reason would adopt the uniform distribution here, thus employing the same distribution in the case of complete ignorance that they would given reliable information that exactly half the balls in the urn are white.

[5] In 1967 Dempster, unaware of Strassen's 1964 paper, published a similar analysis. Shafer (1976) offered a *sui generis* account of set functions having the monotonicity properties of the lower probability β, regarding such set functions, which he called *belief functions*, as directly assessable measures of degrees of belief.