WHAT KIND OF UNCERTAINTY IS THAT? USING PERSONAL PROBABILITY FOR EXPRESSING ONE'S THINKING ABOUT LOGICAL AND MATHEMATICAL PROPOSITIONS*

> What is essential for the future development of probability considerations, as for the development of science in general, is that trained minds play upon its problems freely and that those engaged in discussing them illustrate in their own procedure the characteristic temper of scientific inquiry—to claim no infallibility and to exempt no proposed solution of a problem from intense criticism. Such a policy has borne precious fruit in the past, and it is reasonable to expect that it will continue to do so.
>
> —Ernest Nagel, *Principles of the Theory of Probability*, Concluding Remarks[1]

Try to use probability to formalize your uncertainty about logical or mathematical assertions. What is the challenge?

Concerning the normative theory of personal probability, in a frank presentation titled *Difficulties in the theory of personal probability*,[2] L. J. Savage writes,

> The analysis should be careful not to prove too much; for some departures from theory are inevitable, and some even laudable. For example, a person required to risk money on a remote digit of π would, in order to comply fully with the theory, have to compute that digit, though this would really be wasteful if the cost of computation were more than the prize involved. For the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know. Is it possible to improve the theory in this respect, making allowance within it for the cost of thinking, or would that

* We thank Jessi Cisewski and Rafael Stern for their helpful comments with prior drafts of this paper.

[1] Ernest Nagel, *Principles of the Theory of Probability* (Chicago: University Press, 1939), pp. 76–77.

[2] This text is taken from a draft of Leonard J. Savage's manuscript, *Difficulties in the*

entail paradox, as I am inclined to believe but unable to demonstrate? If the remedy is not in changing the theory but rather in the way in which we are to attempt to use it, clarification is still to be desired.

But why does Savage assert that "a person required to risk money on a remote digit of p

section ii, strategy (2) for responding to Savage's challenge is to relax the conditions that $\mathcal{B}$ is as large as a field of sets. That creates some elbow room for having uncertainty about what is otherwise incorporated as part of the mathematical background assumptions of a measure space.

P is a (countably additive) probability over $\mathcal{B}$ used to represent YOUR uncertainty. We express Savage's challenge to YOU in representing your uncertainty about logical/mathematical constants as follows. In addition to the events that constitute the elements of $\mathcal{B}$, the received theory of mathematical probability introduces a class χ of (possibly bounded) random variables X as ($\mathcal{B}$-measurable) real-valued functions from Ω to $\mathbb{R}$. Denote by $E_P[X]$ the P-expected value of the random variable X. Let I_G be an indicator function for an event G. That is,

$$I_G(\omega) = 1 \text{ if } \omega \in G \text{ and } I_G(\omega) = 0 \text{ if } \omega \in G^c.$$

Then $E_P[I_G] = P(G)$. Thus, in the received theory, probability is an values Consider a problem in probability that relies on three familiar bits of knowledge from high-school geometry.

The area of a circle with radius r equals pr^2.
The area of a square is the square of the length of its side.
The Pythagorean Theorem: Given a right triangle, with side lengths a and b and hypotenuse length c, then $a^2 + b^2 = c^2$.

Let W be the set of points interior to a circle C with radius r. A point from coherent probability assessment.[9] With strategy (1), next we illustrate how to convert this “bug” into a “feature” that opens the door to using commonplace numerical methods as a response to Savage’s challenge.

i. strategy (1)

We extend Example 1 to illustrate strategy (1): Loosen the grip of the Total Evidence Principle Use a Statistician's Stooge to replace the original uncertain quantity Xp_6 with a different one, q, that the Stooge knows (but YOU do not know) is coextensive with Xp_6

YOU about the region S is that it belongs to the algebra $\mathcal{B}$. Then the Y_i form an iid sequence of Bernoulli(q) variables, where q is the area(S)/2p. As it happens, q = 2/p. But this identity is suppressed in the following analysis, with which both YOU and the Stooge concur.

YOU and the Stooge know that $\sum_{i=1}^{n} Y_i$ is Binomial(n,q). Let $\overline{Y}_n = \sum_{i=1}^{n} Y_i / n$ denote the sample average of the Y_i. $\overline{Y}_n$ is a sufficient statistic for q, that is, a summary of the n draws X_i that preserves all the relevant evidence in a coherent inference about q based on the data of the n-many iid Bernoulli(q) draws.

The Stooge samples with n = 10^{16}, obtains $\overline{Y}_n$ = 0.63661977236, and carries out ordinary Bayesian reasoning with YOU about the Binomial parameter q using YOUR “prior” for q. According to what the Stooge tells YOU, q is an uncertain Bernoulli quantity of no special origins. YOU tell the Stooge your “prior” opinion about q. For convenience, suppose that YOU use a uniform conjugate Beta(1, 1) “prior” distribution for q, denoted here as P(q). So, the Stooge reports, given these data, YOUR “posterior” probability is greater than .999, that 0.63661971 £ q £ 0.63661990. Then, since the Stooge knows that q = 2/p, the Stooge reports for YOU that the probability is at least .999 that the sixth digit of p is 2. Of course, in order for YOU to reach this conclusion you have to suppress the information that S is an inscribed square within C, rather than some arbitrary geometric region within the algebra of ruler-and-compass constructions. The Stooge needs this particular information, of course, in order to determine the value of each Y_i.àExample

This technique, strategy (1), generalizes to include the use of many familiar numerical methods as a response to Savage’s question: How do YOU express uncertainty about a mathematic93.6rl8mathema2.7(ov

For finitely many contracts YOUR outcome is the sum of the separate contracts.

$$\sum_{i=1}^{n} b_i [X_i (w) - P(X_i)].$$

In de Finetti's theory, the state space $W = \{w\}$ is formed by taking all the mathematical combinations of those random variables $c = \{X\}$ that YOU have assessed with YOUR previsions. We illustrate this technique in Example 2, below.

Definition. YOUR Previsions are collectively incoherent provided that there is a finite combination of acceptable contracts with uniformly negative outcome—if there exists a finite set $\{b_i\}$ $(i = 1, \ldots, n)$ and $e > 0$ such that, for each w Î W,

$$\sum_{i=1}^{n} b_i [X_i (w) - P(X_i)] < -e.$$

With this choice of $\{b_i\}$ the Gambler has created a sure loss for YOU—a Dutch Book. Otherwise, if no such combination $\{b_i\}$ exists, YOUR previsions are coherent.

Let $c = \{X_j$: j Î J$\}$ be an arbitrary set of variables, defined on W. What are the requirements that coherence imposes on YOU for giving coherent previsions to each random quantity in the set c? That is, suppose YOU provide previsions for each of the variables X in a set c where each variable X is defined with respect to W, that is, the function X: W→Â is well defined for each X. When are these a coherent set of previsions?

De Finetti's Theorem of Coherent Previsions[10]

YOUR Previsions are coherent if and only if there is a (finitely additive) probability $P(\cdot)$ on W with YOUR Previsions equal to their P-expected values.

$$P(X) = E_P[X].$$

This theorem yields the familiar result that, when all the variables in c are indicator functions—when all of the initial gambles are simple bets on events—YOUR previsions are immune to the Gambler having a strategy for making a Book against you if and only if your previsions are a (finitely additive) probability.

[10] de Finetti, Probabilismo: Saggio critico sulla teoria della probabilità e sul valore della scienza

De Finetti’

the intersection of two events each of which you have assessed with well-defined previsions. In Example 2, YOU give determinate previsions $P(\{3,6\})$ and $P(\{1,2,3\})$, but are not required by coherence to assess $P(\{3\})$. Alas, however, this approach through de Finetti's Fundamental Theorem does not solve YOUR question of how to depict uncertainty about mathematical/logical constants.à Example

Example 3: For convenience, label the four events in c: $F_1 = \{1\}$, $F_2 = \{3,6\}$, $F_3 = \{1,2,3\}$, and $F_4 = \{1,2,4\}$. Consider the following specific sentential proposition, H, about which we presume YOU are unsure of its validity—until, that is, you calculate truth tables.

$$H: \quad [(F_2 \;\; [F_1 \;\; (F_4 \;\; F_3)]) \rightarrow [(F_2 \;\; F_1) \;\; (F_2 \;\; F_4)]]$$

Analogous to the variable Xp_6, the sixth decimal digit in p, the indicator variable I_H is a constant: it takes the value 1 for each state in W. 1 = I_W £ I_H. So, by the Fundamental Theorem, in order to be coherent YOUR prevision must satisfy $P(I_H) = 1$. Assume that, prior to a truth-table calculation, YOU are unsure about H. Alas, de Finetti's theory of coherent previsions leaves YOU no room to express this uncertainty. The closure of coherent previsions required by the linear span of the random variables that YOU have coherently assessed does not match the psychological closure of your reasoning process.

Here is the same problem viewed from another perspective.

Example 3 (continued): Garber (1983) suggests YOU consider the sentential form of the problematic hypothesis as a way of relaxing the structural requirements of logical omniscience.

$$H: \quad [(F_2 \;\; [F_1 \;\; (F_4 \;\; F_3)]) \rightarrow [(F_2 \;\; F_1) \;\; (F_2 \;\; F_4)]]$$

This produces the schema:

$$\mathcal{H}': \quad [\mathcal{P} \;\; (\mathcal{Q} \;\; (\mathcal{R} \;\; \mathcal{S}))] \rightarrow [(\mathcal{P} \;\; \mathcal{Q}) \;\; (\mathcal{P} \;\; \mathcal{R})]$$

Evidently $\mathcal{H}'$ is neither a tautology nor a contradiction. So, each value 0 £ $P(\mathcal{H})$ £ 1 is a coherent prevision, provided that we have

The same problem recurs when, instead of imposing the norms merely of a sentential logic, as in Garber's suggestion, we follow Gaifman's (2004) intriguing proposal for reasoning with limited resources. Gaifman offers YOU a (possibly finite) collection P of *sentences* over which you express your degrees of belief. As Gaifman indicates, in his approach sentences are the formal stand-ins for Fregean *thoughts*—"senses of sentences," as he puts it (2004, p. 102). This allows YOU to hold different degrees of uncertainty about two *thoughts* provided that they have different *senses*. In Gaifman's program, YOUR opinions about sentences in P are governed by a restricted logic. He allows for a *local algebra* of sentences that are provably equivalent in a restricted logic. Then YOUR assessments for the elements of P might not respect logical equivalence, as needed in order to escape the clutches of logical omniscience. Just as with de Finetti's rule of closure under the linear span of assessed events, also in Gaifman's system of a *local algebra* YOU are not required to assess arbitrary well-formed subformulas of those in P.

We are unsure just how Gaifman's approach responds to Savage's challenge. First, as a practical matter, we do not understand what YOUR previsions for such sentences entail when previsions are used as betting rates. When YOU bet on a sentence ***s*** (in a *local algebra*), what are the payoffs associated with such a bet? That is, how does a *local algebra* fix the payoffs when YOU bet on ***s*** with prevision P(***s***)? It cannot be that the truth conditions for ***s*** determine the payoffs for the bet. That way requires YOU to be logically omniscient if you are coherent, of course.

Second, and more to the point of Savage' mathematical propositions. Nor does YOUR performance match the norms of a sentential logic, as per Garber's proposal. Nor does YOUR performance match the norms of a local algebra, as per Gaifman's proposal. What reason makes plausible the view that YOUR thinking about a mathematical proposition, your actual performance when judging the value of Xp_6

the three elements of W. And let the following be three incoherent prevision functions over c.

$P_1(w_i) = P_1(I_i) = <0.5, 0.5, 0.5>$, for $i = 1, 2, 3$.
$P_2(w_i) = P_2(I_i) = <0.6, 0.7, 0.2>$, for $i =$

reported in section 6 of our (2003), explains how to calculate a prevision for a new variable without increasing YOUR existing rate of incoherence.

Assume YOU assess previsions for each element of a (finite) partition $p = \{h_1, \ldots, h_m\}$, with values $P(h_i) = p_i$, $i = 1, \ldots, m$. YOU are asked for YOUR prevision $P(Y)$ for a (p-measurable) variable Y, with $Y(h_i) = c_i$.

- Calculate a pseudo-expectation using *YOUR* possibly incoherent previsions over p: $P(Y) = \sum_i p_i c_i$
- Then you will not increase *YOUR Rate of Incoherence* extending your previsions to include the new one for Y, $P(Y) = \sum_i p_i c_i$

When YOU are coherent, YOUR rate of incoherence is 0. Then pseudo-expectations are expectations, and the only way to extend YOUR previsions for a new variable, while preserving YOUR current 0-rate of incoherence, is to use the pseudo-expectation algorithm. However, when YOU are incoherent, there are other options for assessing $P(Y)$ without increasing YOUR rate of incoherence. But, without knowing how incoherent YOU are, still YOU can safely use the pseudo-expectation algorithm and be assured that your rate of incoherence does not increase. The pseudo-expectation algorithm is *robust*.

One intriguing case of this result arises when Y is the variable corresponding to a *called-off* (conditional) gamble.[14] Then using a pseudo-expectation with respect to YOUR (possibly) incoherent previsions for Y suggests how to extend the principle of *confirmational conditionalization*[15] to include incoherent conditional previsions. When YOU hypothesize expanding your *corpus of knowledge* to include the new evidence $(X = x)$, YOUR possibly incoherent previsions $P(\cdot)$ become $P(\cdot \mid X = x)$, as calculated according to the Bayes algorithm for pseudo-expectations.

This leads to the following Corollary, which is an elementary generalization of familiar results about the asymptotic behavior of a coherent posterior probability function given a sequence of identically, independently distributed (*iid*) variables.[16]

[14] We discuss this in section 6 of Schervish, Seidenfeld, and Kadane, "Two Measures of Incoherence," Technical Report #660, Department of Statistics, Carnegie Mellon University (1997).

[15] See Levi, *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance* (Cambridge: MIT, 1980).

[16] See Savage, *Foundations of Statistics*, p. 141, Theorem 1, for the special case of a finite parameter space, and Doob's theorem, as reported by Schervish, *Theory of Statistics* (New York, Springer-Verlag, 1995), T.7.78, p. 429, for the general version, as used here.

Corollary. Let

previsions about X_c. But how to match q against what we are thinking about X_c? What exactly is our Stooge reporting to us about q?

(2) Relax the structure of a measure space in order to accommodate a more psychologically congenial closure condition on the set of variables to be assessed (Hacking, 1967). What fits the bill? De Finetti's use of the linear span in place of an algebra of events