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 theory of personal probability*, dated April 1, 1967. Savage gave one of us (JBK) this draft while both were members of the Statistics faculty at Yale University. This text agrees with the quotation on p. 311 of Ian Hacking, "Slightly More Realistic Personal Probability," *Philosophy of Science*, xxxiv, 4 (December 1967): 311–25. In the published version, Savage, "Difficulties in the Theory of Personal Probability," *Philosophy of Science*, xxxiv, 4 (December 1967): 305–10, this text appears (p. 308) with printing errors, which are duplicated also in the version appearing in *The Writings of Leonard Jimmie Savage: A Memorial Selection* (Washington, DC: American Statistical Association and the Institute of Mathematical Statistics, 1981), p. 511.

0022-362X/12/0908-09/516–33 © 2012 The Journal of Philosophy, Inc.

> 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 π would, in order to comply fully with the theory, have to compute that digit"? His short answer is that the postulates of the theory of personal probability "imply that you should behave in accordance with the logical implication of all that you know."

In this essay we discuss three strategies for addressing Savage's challenge:

(1) Adapt I. J. Good's[3] idea to use a *Statistican's Stooge* in order to change the object of uncertainty for the agent. The *Stooge* replaces the problematic constant π by a nonproblematic random variable θ that the *Stooge* knows is co-extensive with π. From the perspective of the *Statistician*, the theory of personal probability affords nonproblematic probability judgments about θ. Viewed from the perspective of the *Stooge*, the *Statistician*'s nonproblematic uncertainty about θ expresses her/his problematic uncertainty about π. But how does the *Statistician* understand the random variable θ so that, without violating the *Total Evidence* requirement, her/his uncertainty about π is related to her/his uncertainty about θ? *Total Evidence* obliges the rational agent to formulate personal probabilities relative to a space of possibility consistent with *all* her/his evidence.

(2) Adapt the requirements for "what you know" to a less than logically omniscient agent. One way to do this is to change the *closure* conditions for what probabilistic assessments rationality demands of a coherent agent. Hacking[4] signals this idea; Garber[5] and Gaifman[6] provide variants of this strategy, as does de Finetti[7] with his theory of *coherence*, which we illustrate below in section II. Then, consonant with Savage's challenge, the agent's uncertainty about the digits of π is no different in kind than the agent's uncertainty about the

[3] I. J. Good, "Twenty-seven Principles of Rationality (#679)" (1971), in *Good Thinking: The Foundations of Probability and Its Applications* (Minneapolis: Minnesota UP, 1983), pp. 15–19.

[4] Hacking, *op. cit.*

[5] Daniel Garber, "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory," in John Earman, ed., *Testing Scientific Theories* (Minneapolis: Minnesota UP, 1983), pp. 99–131.

[6] Haim Gaifman, "Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements," *Synthese*, CXL, 1/2 (May 2004): 97–119.

[7] Bruno de Finetti, *Theory of Probability*, vol. 1 (Chichester, UK: Wiley, 1974).

digits of any other mathematical constant. But how to formalize the concept of possibility for such a boundedly rational agent? What is the normative theory of probability for an agent with bounded rationality?

(3) Modify de Finetti's criterion of coherence, which is a dichotomous distinction between coherent and incoherent judgments of personal probability, to accommodate *degrees of incoherence*.[8] Thus, as Savage's comment suggests, the agent's judgment about the digits of π is represented by an incoherent probability assessment. But the modified theory allows for reasoning with incoherent judgments and provides the agent with guidance how to use, for example, ordinary calculations to reduce her/his degree of incoherent uncertainty about the digits of π.

Contemporary probability theory, in particular the mathematical theory of personal probability, relies on a mathematical device, a *measure space* $<\Omega, \mathcal{B}, \mathbf{P}>$, which embeds mathematical and logical structural assumptions. We begin our discussion of these three strategies for addressing Savage's challenge by relating them to the three components of a measure space. Following de Finetti's convention, hereafter, we refer to the reasonable person whose uncertainty about mathematical and logical propositions is the subject of Savage's challenge with the pronoun, "YOU."

The first component of a measure space, $\Omega = \{\omega_i: i \in \mathrm{I}\}$ is a partition of YOUR space of serious possibilities, indexed by a set I. The ω_i are called *states*. This attribution as so-called "states" does not require special metaphysical features for the elements ω_i of the partition. These states need not be *atomic* in an absolute sense. Upon further reflection of YOUR opinions, YOU might refine the space, for example, by using a finer partition $\Omega' = \{\omega'_j: j \in \mathrm{J}\}$ where each $\omega_i \subseteq \Omega'$. YOU might need to refine Ω when considering, for example, a new random quantity that is not defined with respect to Ω. With respect to Savage's challenge, the problems for YOU in formulating Ω include, for example, that you are unsure whether YOU have succeeded identifying a partition: YOU are unsure whether different elements of Ω are disjoint and whether their union exhausts all the possibilities YOU judge are serious.

$\mathcal{B}$ is a Boolean (sigma) field of subsets of Ω. The elements of $\mathcal{B}$ are the abstract *events* over which YOUR uncertainty is to be represented with a probability function. As we illustrate, below in

[8] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane, "Measures of Incoherence: How Not to Gamble If You Must, with Discussion," In J. M. Bernardo et al., eds., *Bayesian Statistics* 7 (New York: Oxford, 2003), pp. 385–401.

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 $\Re$. Denote by $E_{\mathbf{P}}[X]$ the **P**-expected value of the random variable X. Let $\mathbf{I}_G$ be an indicator function for an *event* G. That is,

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

Then $E_{\mathbf{P}}[\mathbf{I}_G] = \mathbf{P}(G)$. Thus, in the received theory, probability is an instance of mathematical expectation. But in the received theory of personal probability, π is a constant variable. It takes the same value in each *state*: $\pi(\omega) = \pi$. So, $E_{\mathbf{P}}[\pi] = \pi$. YOU are required to know π. However, under strategy (3) (as explained in section III), in response to Savage's challenge YOU use an *incoherent* expectation function in order to model YOUR uncertainty about mathematical propositions.

Reflect on Savage's challenge in some detail. Let $X\pi_6$ be the variable whose value is the sixth decimal digit of π. Here, we emphasize the point that YOUR uncertainty about the decimal representation of π may occur without having to consider a "remote" digit. In an ordinary measure space $X\pi_6$ is the constant 2, independent of ω, because π is a constant whose value is independent of the elements of Ω. In an ordinary measure space, with probability 1 the event "$X\pi_6 = 2$" obtains, since as a mathematical result, it obtains in each state ω. Thus, in any ordinary measure space, there is no elbow room for a nonextreme probability about $X\pi_6$ or an expectation other than 2 for its value. Savage's admonition applies:

> For the postulates of the theory imply that you should behave in accordance with the logical implication of all that you know.

The construction of an ordinary measure space requires that you know what constant π is. That fact is part of the mathematical knowledge taken as background also in order to formulate probability *values* in a measure space, as we illustrate, next.

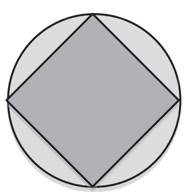
Example 1. Here is an illustration of the use of the mathematical background knowledge for a measure space for giving probability *values*. Consider a problem in probability that relies on three familiar bits of knowledge from high-school geometry.

The area of a circle with radius r equals πr^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 Ω be the set of points interior to a circle **C** with radius r. A point from Ω is chosen at random, with a uniform probability: equal probability for congruent subsets of **C**. Let $\mathcal{B}$ be the algebra of geometric subsets of **C** generated by ruler-and-compass constructions. That is, YOUR personal probability **P** is uniform over these geometric subsets Ω: congruent regions that belong to $\mathcal{B}$ have equal probability. YOU understand that YOUR probability that the random point is contained in a region **S** (for a region **S** that is an element of $\mathcal{B}$) is the ratio of the *area*(**S**) to the *area*(**C**). YOU are aware that YOUR probability of the event "The random point is in **S**" is the fraction *area*(**S**)$/\pi r^2$.

Let **S** be a square inscribed inside the circle **C**. (See Figure 1.) Then by the Pythagorean Theorem and the rule for the area of a square, area(**S**) $= 2r^2$. So, YOU are aware that YOUR probability that the random point is in the square **S** is $2/\pi$. Suppose YOU are aware that the first five decimal digits in the expansion of π are 3.14159. But YOU cannot identify the sixth decimal digit of π. Using the familiar long-division algorithm, then you are unable to calculate precisely YOUR personal probability ($2/\pi$) beyond the first four digits (0.6366) that the random point is in **S**. YOU know that the fifth digit is either 1 or 2. But, for instance, then YOU are unable to answer whether a bet on the random point is in **S** at odds of .63662:.36338 is favorable, fair, or unfavorable for YOU.$\Diamond_{\text{Example}}$

Figure 1



Thus, the challenge Savage poses affects both *the numerical values* that YOU can identify for YOUR (coherent) probability assessments, as well as the *random quantities* to which YOU can assign a 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 $X\pi_6$ with a different one, θ, that the *Stooge* knows (but YOU do not know) is coextensive with $X\pi_6$. Then YOU may hold nonextreme but coherent probabilities about the substitute variable θ. In this way, familiar numerical methods, including Monte Carlo methods, permit YOU to learn about $X\pi_6$ by shifting the failure of the Total Evidence principle to the *Stooge*.

Example 1 (continued). As an instance of I. J. Good's *Statistician's Stooge*, YOUR assistant, the *Stooge*, creates an elementary statistical estimation problem for the quantity $2/\pi$ using *iid* repeated draws from the uniform distribution on a circle **C**. The *Stooge* chooses **C** to be the circle with center at the origin (0, 0) and radius $r = \sqrt{2}$. Then the inscribed square **S** has corners with coordinates $(\pm 1, \pm 1)$. Let $X_i = (X_{i_1}, X_{i_2})$ $(i = 1, \ldots, n)$ be n random points drawn by the *Stooge* using the uniform distribution on **C**. After each draw the Stooge determines whether or not $X_i \in \mathbf{S}$, that is, whether or not both inequalities obtain: $-1 \leq X_{i_j} \leq +1$ $(j = 1, 2)$, which involves examining only the first significant digit of X_{i_j}.

Now, the *Stooge* tells YOU whether event Y occurs on the i^{th} trial, $Y_i = 1$, if and only if $X_i \in \mathbf{S}$ for a region **S**. But all the *Stooge* tells

[9] Example 1 opens the door also to a discussion of *higher-order* probabilities. YOU might try to assign a second-order personal probability distribution $\mathbf{P}^*$ to the quantity $2/\pi$ in order to represent the added higher-order uncertainty you have in YOUR first-order uncertainty **P** that the random point is in the region **S**. Higher-order probability is a topic beyond the focus of this essay. Here, we express our agreement with the Savage-Woodbury rejoinder—Savage, *The Foundations of Statistics*, 2nd ed. (New York: Dover, 1954/1972), p. 58. That rejoinder questions whether such a higher-order personal probability has operational content. The Savage-Woodbury response establishes that $\mathbf{P}^*$ provides YOU with a resolution of your first-order uncertainty: Use $\mathbf{P}^*$ to create an expected value for $2/\pi$, just as you would use personal probability to determine an expected value for any random quantity. Then this expected value is your first-order expected value for $2/\pi$, and YOU have no added uncertainty about your (first-order) probability that the random point is in **S**. Then there is no residual higher-order uncertainty.

Example 1 also opens the door to upper and lower previsions that govern one-sided gambles. We consider this in connection with de Finetti's *Fundamental Theorem of Previsions*, which we discuss in section II in connection with Example 2. We indicate why upper and lower previsions do not resolve Savage's challenge, either.

YOU about the region **S** is that it belongs to the algebra $\mathcal{B}$. Then the Y_i form an *iid* sequence of Bernoulli(θ) variables, where θ is the area(**S**)/2π. As it happens, $\theta = 2/\pi$. 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,θ). 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 θ, that is, a summary of the n draws X_i that preserves all the relevant evidence in a coherent inference about θ based on the data of the n-many *iid* Bernoulli(θ) 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 θ using YOUR "prior" for θ. According to what the *Stooge* tells YOU, θ is an uncertain Bernoulli quantity of no special origins. YOU tell the *Stooge* your "prior" opinion about θ. For convenience, suppose that YOU use a uniform conjugate Beta(1, 1) "prior" distribution for θ, denoted here as $\mathbf{P}(\theta)$. So, the *Stooge* reports, given these data, YOUR "posterior" probability is greater than .999, that $0.63661971 \leq \theta \leq 0.63661990$. Then, since the *Stooge* knows that $\theta = 2/\pi$, the *Stooge* reports for YOU that the probability is at least .999 that the sixth digit of π 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.$\Diamond_{\text{Example}}$

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 mathematical term τ? The numerical method provides evidence in the form of a random variable, Y, whose value $Y = y$ is determined by an experiment with a well-defined likelihood function, $\mathbf{P}(Y{=}y \mid \theta)$, that depends upon a parameter θ, known to the *Stooge* but not to YOU to equal the problematic quantity τ. YOU express a coherent "prior" probability for θ, $\mathbf{P}(\theta)$. By Bayes's Theorem, YOUR "posterior" probability, $\mathbf{P}(\theta \mid Y{=}y)$ is proportional to the product of this likelihood and "prior":

$$\mathbf{P}(\theta \mid Y{=}y) \propto \mathbf{P}(Y{=}y \mid \theta)\, \mathbf{P}(\theta).$$

As Good notes, playing fast and loose with the Total Evidence principle—in the example, by permitting the *Stooge* to suppress the problematic information that $\theta = \tau$—allows YOU, a coherent Bayesian statistician, to duplicate some otherwise non-Bayesian, Classical statistical inferences. For instance, a Classical α-level Confidence Interval for a quantity θ based on a random variable X, $CI^{\alpha}_{lower}(X) \leq \theta \leq CI^{\alpha}_{upper}(X)$ becomes a Bayesian posterior probability α for the same interval estimate of θ given X, by suppressing the observed value of the random variable, $X = x$, and leaving to the *Stooge* the responsibility of filling in that detail.

What troubles us about this approach as a response to Savage's challenge is that YOUR coherent uncertainty about the substitute parameter θ may reflect very little of what YOU know about the problematic quantity τ. Employing the *Stooge*, as above, allows YOU to express a coherent prior probability for θ. In Example 1, θ is the ratio: area of some arbitrary rule-and-compass region chosen (by the *Stooge*) from $\mathcal{B}$ divided by 2π, $\tau = 2/\pi$, and, unknown to YOU, $\theta = \tau$. But YOUR prior for θ, when that quantity is identified to YOU as just some region chosen by the *Stooge*, may have very little in common with YOUR uncertainty about τ, which depends upon the problematic information that **S** is the inscribed square and which the *Stooge* conveniently suppresses for YOU.

II. STRATEGY (2)

In this section we examine an instance of strategy (2)—modify the closure conditions on the space of uncertain events in order to avoid requiring YOU are logically/mathematically omniscient. Hacking (1967) responds to Savage's challenge this way. Here, we review de Finetti's (1974) theory of *coherent Previsions*: $P(\cdot)$ as an instance of this strategy.

In de Finetti's theory, YOU are required to offer a *fair price*, a *prevision* $P(X)$, for buying and selling the random variable X. X is defined for/by YOU with respect to a partition Ω. That is, for each state $\omega \in \Omega$, $X(\omega)$ is a well-defined real number. That is, the function $X{:}\Omega \rightarrow \Re$ is known to YOU. (In connection with the *Dutch Book* argument, de Finetti often refers to YOU as the *Bookie*.) To say that $P(X)$ is YOUR fair price for the random quantity X means that YOU are willing to accept all contracts of the form $\beta_{X,P(X)}[X - P(X)]$, where an opponent (called the *Gambler*) chooses a real-value, $\beta_{X,P(X)}$. This term, $\beta_{X,P(X)}$, is constrained in magnitude in order to conform to YOUR wealth, but allowed to depend on both the variable X and YOUR prevision for X. With $|\beta| > 0$ small enough to fit YOUR budget, YOU are willing to engage in the following contracts:

when $\beta > 0$ YOU agree to pay $\beta P(X)$ in order *to buy* (that is, to receive) βX in return;
when $\beta < 0$ YOU agree to accept $\beta P(X)$ in order *to sell* (that is, to pay) βX in return.

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

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

In de Finetti's theory, the state space $\Omega = \{\omega\}$ is formed by taking all the mathematical combinations of those random variables $\chi = \{X\}$ 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 $\{\beta_i\}$ ($i = 1, \ldots, n$) and $\varepsilon > 0$ such that, for each $\omega \in \Omega$,

$$\sum_{i=1}^{n} \beta_i [X_i(\omega) - P(X_i)] < -\varepsilon.$$

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

Let $\chi = \{X_j : j \in J\}$ be an arbitrary set of variables, defined on Ω. What are the requirements that coherence imposes on YOU for giving coherent previsions to each random quantity in the set χ? That is, suppose YOU provide previsions for each of the variables X in a set χ where each variable X is defined with respect to Ω, that is, the function $X: \Omega \rightarrow \Re$ 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 $\mathbf{P}(\cdot)$ on Ω with YOUR *Previsions* equal to their $\mathbf{P}$-expected values.

$$P(X) = \mathrm{E}_{\mathbf{P}}[X].$$

This theorem yields the familiar result that, when all the variables in χ 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* (Naples, Italy: Perrella, 1931), translated as "Probabilism: A Critical Essay on the Theory of Probability and on the Value of Science," *Erkenntnis*, XXI, 2/3 (September 1989): 169–223; and de Finetti, "La prévision: ses lois logiques, ses sources subjectives," *Annales de L'Institut Henri Poincaré*, VII (1937): 1–68, translated as (and with new notes by the author) "Foresight: Its Logical Laws, Its Subjective Sources," in Henry E. Kyburg, Jr., and Howard E. Smokler, eds., *Studies in Subjective Probability*, 2nd ed. (Huntington, NY: Krieger, 1980).

De Finetti's theory of coherent previsions commits YOU to having precise previsions for all variables in the linear span—for all linear combinations—of those variables $X \in \chi$ that you have already assessed with previsions. As we explain, below, this is a different closure condition than requiring YOU to have previsions determined even for all events in the smallest logic/algebra generated by Ω.

Again, suppose YOU provide coherent previsions for all variables X in the set χ. Let Y be another variable defined with respect to Ω but not necessarily in χ.

Let: $\underline{A} = \{X\colon X(\omega) \leq Y(\omega)$ and X is in the *linear span* of $\chi\}$
$\overline{A} = \{X\colon X(\omega) \geq Y(\omega)$ and X is in the *linear span* of $\chi\}$

Let: $\underline{P}(Y) = sup_{X \in \underline{A}}\, P(X)$ and $\overline{P}(Y) = inf_{X \in \overline{A}}\, P(X)$.

De Finetti's *Fundamental Theorem of Previsions*:

Extending YOUR previsions P to P^* in order to give a coherent prevision for Y, $P^*(Y)$, allows it to be any (finite) number from $\underline{P}(Y)$ to $\overline{P}(Y)$.

Outside this interval, the extension P^* is incoherent.

Next we illustrate these two results of de Finetti and explain their relevance to Savage's challenge.

Example 2. Consider a roll of a six-sided die with faces numbered in the usual way, 1, 2, 3, 4, 5, 6, and with opposite sides always summing to 7. Suppose YOU think about the following four events (which define the set χ) and identify YOUR *Previsions* in accord with the assessment that the die is *fair*:

$$P(\{1\}) = 1/6;\ P(\{3,6\}) = 1/3;\ P(\{1,2,3\}) = P(\{1,2,4\}) = ½.$$

The set of events for which YOUR coherent prevision is already determined by the previsions for these four events is given by the *Fundamental Theorem.* That set does *not* form an algebra. Only 24 of 64 events (only 12 pairs of complementary events) have determinate previsions.

For instance, by the *Fundamental Theorem*:

$\underline{P}(\{6\}) = 0 < \overline{P}(\{6\}) = 1/3;$

likewise $\underline{P}(\{4\}) = 0 < \overline{P}(\{4\}) = 1/3;$

however, $P(\{4,6\}) = 1/3.$

The smallest algebra for the four events in χ is the power set of all 64 subsets of Ω. Thus, de Finetti's theory of *coherence* does *not* require that YOUR previsions are well defined for all the propositions in the elementary logic formed from YOUR beliefs about the constituents. YOU do not have to close the set of YOUR previsions under even sentential logical operations. For instance, YOU are not required to provide a well-defined prevision for an event that is 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.$\Diamond_{\text{Example}}$

Example 3. For convenience, label the four events in χ: $F_1 = \{1\}$, $F_2 = \{3,6\}$, $F_3 = \{1,2,3\}$, 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.

$$H: \qquad [(F_2 \vee [F_1 \wedge (F_4 \vee F_3)]) \rightarrow [(F_2 \vee F_1) \wedge (F_2 \vee F_4)]]$$

Analogous to the variable $X\pi_6$, the sixth decimal digit in π, the indicator variable $\mathbf{I}_H$ is a constant: it takes the value 1 for each state in Ω. $1 = \mathbf{I}_\Omega \leq \mathbf{I}_H$. So, by the *Fundamental Theorem,* in order to be coherent YOUR prevision must satisfy $P(\mathbf{I}_H) = 1$. 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.

$$H: \qquad [(F_2 \vee [F_1 \wedge (F_4 \vee F_3)]) \rightarrow [(F_2 \vee F_1) \wedge (F_2 \vee F_4)]]$$

This produces the schema:

$$\mathcal{H}': \qquad [\mathcal{P} \vee (\mathcal{Q} \wedge (\mathcal{R} \vee \mathcal{S}))] \rightarrow [(\mathcal{P} \vee \mathcal{Q}) \wedge (\mathcal{P} \vee \mathcal{R})]$$

Evidently $\mathcal{H}'$ is neither a tautology nor a contradiction. So, each value $0 \leq P(\mathcal{H}) \leq 1$ is a coherent prevision, provided that we have the full set of truth-value interpretations for the sentential variables $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{S}$.$\Diamond_{\text{Example}}$

The replacement of H by $\mathcal{H}$ ignores the underlying mathematical relations among the variables in H. Suppose that YOU assess YOUR prevision for $\mathcal{H}$, $P(\mathcal{H}) = .6$. Does YOUR psychological state of uncertainty about H match the requirements that *coherence* places on a prevision, $P(\mathcal{H}) = .6$? Do *YOU* identify $\mathcal{H}'$ as the correct variable for what you are thinking about when you are reflecting on YOUR uncertainty about H, before you do the calculations that reveal H is a logical constant? We think the answer is "No."

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 $\boldsymbol{s}$ (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 $\boldsymbol{s}$ with prevision $P(\boldsymbol{s})$? It cannot be that the truth conditions for $\boldsymbol{s}$ 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's challenge, we do not see why YOUR uncertainty about mathematical propositions should match the normative constraints of an algebra closed under some finite number of iterations of a given rule of inference. Why should YOUR uncertainty over mathematical propositions match what is provable in a restricted *local algebra* of the kind sketched by Gaifman? Might it not be that YOU recall the seventh digit of π but not the sixth? Then, mimicking YOUR uncertainty about the digits of π with a restricted deductive system that generates the digits of π in a proof, according to a computation of π, will not capture YOUR uncertainty about π.

In Levi's terms,[11] YOUR *commitments* to having coherent (precise) previsions according to de Finetti's norms of coherence do not match YOUR *performance* when assessing YOUR uncertainty about

[11] Isaac Levi, *The Fixation of Belief and Its Undoing: Changing Beliefs Through Inquiry* (New York: Cambridge, 1991).

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 $X\pi_6$, matches the commitments of any such normative theory? We are doubtful of strategy (2)!

III. STRATEGY (3)

One feature common to strategies (1) and (2) is the goal of showing that YOU are coherent when you hold nonextreme personal probabilities for mathematical propositions.

Strategy (1) allows YOU to replace a problematic mathematical variable, for example, $X\pi_6$, one that is constant across the space of all possibilities (Ω), with another random variable θ that is not problematic in the same way. Numerical methods for computing the problematic variable, $X\pi_6$, then can be modeled as ordinary statistical experiments generating data Y about θ. There is no incoherence when YOU use nonextreme personal probabilities for θ, given Y. But, as we saw, in addition to failure to adhere to the Total Evidence principle, the effectiveness of strategy (1) depends upon YOUR willingness to use your prior probability for θ to express your thinking about $X\pi_6$.

Strategy (2) allows YOU to replace the familiar algebra $\mathcal{B}$ of a measure space with some other mathematical structure that can support a different set of coherent personal probability assessments—a set that is less demanding on YOUR logical reasoning abilities. For de Finetti, that other mathematical structure is the linear span formed by those previsions you are willing to make. For Garber it is the structure of a sentential logic. For Gaifman it is a *local algebra.* Though each of these might capture some aspect of YOUR thinking about a mathematical proposition, why should YOU think according to the norms of any one of these alternative mathematical structures? None of them is intended as a realistic psychological theory of how YOUR mind reasons.

We propose, instead, strategy (3): Concede that, regarding uncertainty about mathematical and logical propositions, despite the phenomenological similarities with uncertainty about nonconstant "empirical" variables, nonextreme *previsions* for mathematical propositions are incoherent. This is exactly what Savage points out is the problem with the theory of Personal Probability.

> *That is, we concede that YOUR commitment to being coherent is not discharged by what is an entirely predictable shortfall in YOUR performance.*

However, the problem is exacerbated by the fact that de Finetti's distinction between *coherent* and *incoherent* previsions is dichotomous. Perhaps a more nuanced theory of incoherence can guide incoherent thinkers on how to reason without abandoning their commitment to coherence? That is the core idea for strategy (3).

In several of our papers we develop a theory of *degrees of incoherence*.[12] When the *Gambler* can make a Book against the *Bookie*'s incoherent previsions, then many Books can be made. The different Books may be compared by *scaling* the (minimum) *sure gain* to the Gambler—or equivalently scaling the minimum *sure loss* to the Bookie, or adopting a Neutral index, which incorporates both perspectives.

Here are three indices that may be used to scale the sure gains/losses in a Book. For simplicity, in the following discussion we scale the finite set of gambles in a Book using the sum of the individually scaled gambles. This is a special case of our general theory.

> *Rate of Loss (for the Bookie)*: Scale the minimum sure loss to the *Bookie* by the total amount the *Bookie* is compelled to wager from the *Gambler*'s strategy.

What proportion of the *Bookie*'s budget can the *Gambler* win for sure?

> *Rate of Profit (for the Gambler)*: Scale the sure gain to the *Gambler* by the total amount used in the *Gambler*'s strategy.

What proportion of the *Gambler*'s stake does the *Gambler* have to escrow to win one unit for sure from the *Bookie*?

> *A Neutral Rate*: Scale the sure loss to the *Bookie* by the combined amounts (the total stake) wagered by both players according the *Gambler*'s strategy.

As we explain in our (2003) paper, this index is better designed than either of the first two for assessing incoherent previsions for constants. With the Neutral Rate, if $X_c(\omega) = c$ is a constant variable and $P(X_c)$ is a prevision for X_c, then the degree of incoherence for this one prevision is $|c - P(X_c)|$.

Relative to each of these indices, the *rate of incoherence* for an incoherent *Bookie*'s previsions is the greatest (scaled) loss/gain that the *Gambler* can achieve across the different strategies for making a Book.

Example 4: Illustrating Differences among These Three Rates of Incoherence. Consider a 3-element state space, $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Let $\chi = \{\mathbf{I}_i: i = 1, 2, 3\}$ be the set of the three indicator functions for

[12] For an overview, see Schervish, Seidenfeld, and Kadane, *op. cit.*

the three elements of Ω. And let the following be three incoherent prevision functions over χ.

$$P_1(\omega_i) = P_1(\mathbf{I}_i) = <0.5, 0.5, 0.5>, \text{ for } i = 1, 2, 3.$$
$$P_2(\omega_i) = P_2(\mathbf{I}_i) = <0.6, 0.7, 0.2>, \text{ for } i = 1, 2, 3.$$
and $$P_3(\omega_i) = P_3(\mathbf{I}_i) = <0.6, 0.8, 0.1>, \text{ for } i = 1, 2, 3.$$

With prevision $P_j(\cdot)$, $j = 1, 2$, or 3, each contract of the form $\alpha_i(\mathbf{I}_i - P_j(\omega_i))$ is judged fair. In all three cases ($j = 1, 2, 3$), $P_j(\omega_1) + P_j(\omega_2) + P_j(\omega_3) = 1.5$, and not $= 1$. Clearly, these are incoherent previsions. For the first incoherent prevision, $P_1(\cdot)$, all three rates of incoherence lead the *Gambler* to the same strategy, bet against the incoherent *Bookie*$_1$ with equal stakes on all three states, $\alpha_i = (1, 1, 1)$. For the second incoherent prevision, $P_2(\cdot)$, both the Rate of Loss and the Neutral Rate are maximized with the *Gambler* using the equal-stakes strategy, $\alpha_i = (1, 1, 1)$. But the Rate of Profit against incoherent *Bookie*$_2$ is maximized with the strategy $\alpha_i = (1, 1, 0)$, that is, gamble only on the first two states, and with equal stakes. For the third incoherent prevision, $P_3(\cdot)$, the Rate of Loss is maximized against *Bookie*$_3$ with the *Gambler*'s strategy of equal stakes on all three states, $\alpha_i = (1, 1, 1)$, whereas the other two rates of incoherence are maximized with the other strategy $\alpha_i = (1, 1, 0)$.$\Diamond_{\text{Example}}$

Thus, the three rates lead the *Gambler* to three different combinations of strategies. These are different ways to index a *rate of incoherence*. Of course, each coherent prevision has a 0-rate of incoherence with each index.

We first developed our ideas about rates of incoherence in order to engage familiar debates about Bayesian versus Classical Statistical procedures. Bayesians argue that where a particular Classical procedure is incoherent, therefore it is unacceptable. But this is a coarse-level analysis. We inquire, instead, how incoherent is the Classical procedure. Since Classical statistical procedures are often simple to calculate, they can be warranted in the special case when the rate of incoherence is small and a rival, full Bayesian analysis is computationally infeasible. We provide an illustration of this analysis in our (2000), where we investigate the rate of incoherence of fixed α-level hypothesis tests regardless of sample size.[13]

How do we propose to use our ideas about rates of incoherence to address Savage's challenge of how to use probabilities to formalize uncertainty about mathematical propositions? In the spirit of de Finetti's *Fundamental Theorem*, the following result,

[13] Schervish, Seidenfeld, and Kadane, "A Rate of Incoherence Applied to Fixed-Level Testing," *Philosophy of Science*, LXIX, S3 (September 2000): S248–64.

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 $\pi = \{h_1, \ldots, h_m\}$, with values $P(h_i) = p_i$, $i = 1, \ldots, m$. YOU are asked for YOUR prevision $P(Y)$ for a (π-measurable) variable Y, with $Y(h_i) = c_i$.

- Calculate a pseudo-expectation using *YOUR* possibly incoherent previsions over π: $P(Y) = \sum_i p_i c_i$
- 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 Θ be a finite-dimensional parameter space. Consider a nonextreme, *pseudo-prior density function* $p(\theta) > 0$, which may be incoherent. Suppose, however, that a *pseudo-likelihood density function* $p(X = x \mid \theta)$ has a 0-rate of incoherence; that is, these conditional probabilities are coherent. Suppose, also, they are different conditional probability functions for different values of θ. Let $X_i (i = 1, \ldots)$ form a sequence of conditionally *iid* variables, given θ, according to $p(X = x \mid \theta)$. Use the *pseudo-Bayes-algorithm* to create a sequence of *pseudo-posterior functions* $p_n(\theta \mid X_1, \ldots, X_n)$, $n = 1, \ldots$.

Then, almost surely with respect to the true state, $\theta^* \in \Theta$, the Neutral rate of incoherence for the *pseudo-posterior* converges to 0, and that *pseudo-posterior* concentrates on θ^*.

Example 1 (concluded). Reconsider the version of Example 1 involving *iid* repeated sampling of the bivariate variable X, a point randomly chosen from a circle **C**. **S** is a particular inscribed square. Let $Y_i = 1$, if $X_i \in$ **S**, and $Y_i = 0$, if $X_i \notin$ **S**. Let $\theta = 2/\pi = P(Y{=}1 \mid \theta)$. Suppose YOU assign a smooth but incoherent pseudo-prior to θ, for example, use a Beta(1, 1) pseudo-prior. Then, given the sequence $Y_n (n = 1, \ldots)$, by the Corollary, the sequence of YOUR pseudo-posteriors, $P_n(\Theta \mid Y_1, \ldots, Y_n)$ converges (even uniformly) to $2/\pi$. With the Neutral Rate, if $X_c(\omega) = c$ is a constant variable and $P(X_c)$ is a prevision for X_c, then the degree of incoherence for this one prevision is $|c - P(X_c)|$. Therefore, almost surely, also the Neutral Rate of incoherence in YOUR pseudo-posterior converges to $0.\Diamond_{\text{Example}}$

Thus, we see how to use data from familiar numerical methods, methods that have well-defined, coherent likelihood functions—as in the growing family of MCMC algorithms—to improve the rate of incoherence in our previsions for mathematical propositions.

IV. SUMMARY

We have reviewed three strategies for addressing the question whether probability theory can be used to formalize personal uncertainty about ordinary mathematical propositions. We posed the problem in the following form. Variable X_c is a mathematical/logical constant that YOU are unable to identify. So, according to the theory of Personal Probability YOUR nonextreme prevision for X_c is incoherent.

(1) Relax the Total Evidence requirement, for example, use I. J. Good's *Statistician's Stooge*, in order to substitute a related variable θ, about which ordinary statistical inference is coherent, for the problematic variable X_c. With the *Stooge*'s help in censuring some empirical information (for example, $X_c = \theta$), you can reason coherently about θ. It is the *Stooge* who converts those conclusions into incoherent

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

(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 does not work. Nor does either Garber's proposal to use sentential logic, or Gaifman's *local algebra*. We do not see how to match YOUR coherent assessments, where you are aware of these, with a domain of propositions defined by mathematical operations. The mathematical operations used for "closing" the domain of propositions form a Procrustean bed against the domain of YOUR coherent assessments.

(3) Concede that nonextreme probabilities for mathematical propositions are incoherent. Then provide normative criteria for reasoning with incoherent previsions in order to show how to reduce YOUR rate of incoherence. The dichotomy between coherent/incoherent assessments appears too coarse to explain how we use, for example, numerical methods to improve our thinking about mathematical quantities. With our approach to Savage's challenge, using the machinery of rates of incoherence, we expand an old Pragmatist idea—one that runs from Peirce through Dewey. We illustrate how to make the operation of a numerical calculation into an experiment whose outcome may be analyzed using familiar principles of statistical inference. Here, we have taken a few, tentative steps in this direction.

We do not know, however, how far our approach goes in addressing the scope of Savage's challenge. For example, a commonplace decision for a mathematician unsure about a specific mathematical conjecture is how to apportion her/his efforts between searching for a proof of the conjecture and searching for a counterexample to the same conjecture. In sections 14.14–14.15 of his (1970), dealing with search problems, DeGroot establishes Bayesian algorithms for optimizing sequential search.[17] Can these algorithms be adapted to the mathematician's decision problem by allowing for some incoherence in her/his assessments about the conjecture? Are the algorithms DeGroot proves optimal also with pseudo-expectations? We take this as a worthy conjecture.

TEDDY SEIDENFELD
MARK J. SCHERVISH
JOSEPH B. KADANE

Carnegie Mellon University

[17] Morris H. DeGroot, *Optimal Statistical Decisions* (New York: McGraw-Hill, 1970).