Name (to appear in publication): Haim Gaifman

Position (to appear in publication): Professor

Affiliation (to appear in publication): Columbia University

The 5 Questions

1. Why were you initially drawn to probability theory and/or statistics?

As an undergraduate student majoring in mathematics, philosophy and physics I was interested in foundational questions regarding probability and induction. Combining this interest with an interest in mathematical logic, I planned to work as a graduate student on a thesis devoted to probabilities in the context of formal languages. Abraham Robinson, who was then at the Hebrew University, agreed to sponsor this research. At that time Carnap was looking for a research assistant with a strong mathematical background, to work with him in his ongoing project in the foundations of probability. Upon the recommendation of Bar-Hillel, he offered me the position and I accepted. I continued to be a PhD student at the Hebrew University; it was understood that results obtained by me while working as Carnap's assistant could be incorporated into the thesis. After a year, having solved an open problem posed by Tarski about measures on Boolean algebras, I got an offer from him to come to Berkeley as his research assistant. This I did, enrolling at the same time in the mathematics PhD program. During my stay at Berkeley I was drawn to set theory and shortly afterwards to models of Peano arithmetic. Probability theory, in its foundational aspects, remained however a subject of great interest for me, to which I kept, and still keep, returning.

2. What is distinctive about your work in the foundations of probability or its applications?

The idea of studying probabilities in broad contexts — often determined by first-order or richer languages — underlies a considerable part of my work. The idea stems from my initial plan, mentioned above, prior to my work as Carnap's assistant.[1] Carnap's extremely ambitious project aimed at establishing a branch of logic that would do for empirical reasoning what deductive logic — in its modern development — does for mathematics. Coming years after Carnap gave up on the *Aufbau* (*The Logical Structure of the World*, 1928), it is informed nonetheless by a bottom-up approach, which resembles that of the *Aufbau.* An "*ur*-probability" (*the* original prior)

[1] An abstract of my 1964 paper "Concerning measures in first order calculi" was presented in 1960, at the first International Congress for Logic Philosophy and Methodology of Science and can be found in the abstracts section (pp. 77,78). The historical account given by Jeffrey in the collection *Studies in Inductive Logic and Probability I* p. 223, is inaccurate in this respect. My work with Carnap led directly to my "Applications of de-Finetti's theorem to inductive logic", which was published in that collection. It did not lead to the 1964 paper. I was however acquainted with Carnap's main system of the early 1950's, before I came to work with him.

is to be defined on quantifier free sentences ─ which are built from primitive predicates and individual constants ─ subject to strict symmetry constraints. The primitive predicates are supposed to reflect the basic epistemic organization that underlies human cognition.

My own approach differed from Carnap's in two crucial aspects. It was a top-down one which investigated probabilities defined on all sentences of a rich language, and there were no foundational presuppositions on the family of primitive predicates. That framework was supposed to illuminate the basic features of Bayesian methodology. Though I did not subscribe to Carnap's idea of logical probability, I was initially a Bayesian.[2] My position has shifted in time towards a more balanced view, as I became aware of the limitations of large-scale subjective probabilities.

I would like to include in this answer claims about probability, which will figure in answers to subsequent questions. The answer will be longer than what the question calls for, since I intend to use the occasion to suggest a possible research program.

One shortcoming of Carnap's approach was that the focusing on the "foundation" ─ the probabilities over the quantifier free sentences ─ usually leads to probabilities that, in the limit, assign 0 to various universal empirical hypotheses. To see how this works, assume that of s observed objects, s_1 fall under the (primitive) predicate P and $s-s_1$ do not; given this evidence, let $c(s_1, s)$ be the conditional probability that an additional object falls under P. Carnap argued quite plausibly that $c(s_1, s)$ should be a weighted mean of two numbers: a real number, w_P, strictly between 0 and 1, which constitutes the "*a priori* width" of P, and the relative frequency s_1/s, which constitutes the empirical evidence. The relative weight of the latter should grow with s. Since only the ratio of weights matters, we can choose s as the weight of the empirical evidence and let some number, $\lambda = \lambda(w_P, s, s_1)$ be the weight of w_P. In Carnap's λ-system λ is a fixed parameter that depends only on w_P. A larger λ means a smaller effect of the evidence, hence greater caution in jumping to conclusions. The decision to make λ fixed simplifies the system, but imposes a crucial limitation. It turns out that for any fixed positive λ, the *a priori* probability of the hypothesis that all objects fall under P approaches 0 as the number of objects increases; consequently, no matter what the evidence is, the conditional probability of the universal generalization is always 0. The same holds for hypotheses of the form "all Q's are P's", e.g., "all emeralds are green". This conflicts with the obvious intuition that such universal generalizations can be true and that we should be able to confirm them, even if we assume an unlimited number of objects.[3] To do justice to this intuition λ should decrease as function of s, at a sufficient rate.

[2] 'Byesianism' implies of course Bayesian conditionalization, the hallmark of this approach. Some philosophers who accepted subjective probabilities argued that this does not imply Bayesian conditionalization. A careful analysis will reveal that Bayesian conditionalization is indeed implied. One has to be careful though in applying it, because the new evidence, A, may be obtained in a way that gives us also additional relevant information that is not represented in A. In such cases conditionalization on A is not justified.

[3] There have been attempts to justify the awkward outcome by arguing that in science universal generalizations should be construed not literally, but as predictions concerning a limited number of future observations. Carnap himself held during a certain period such a view. There has been also an argument from betting situations, purporting to show that the probability of an empirical universal generalization involving an infinite number of cases

The moral is that if our prior probability is to accommodate wider possibilities of learning from experience ─ possibilities of confirming a greater number of hypotheses ─ we need to include this as a factor in choosing the prior probability. We cannot expect that it would be taken care of by considerations relating only to quantifier free sentences.

In general, the more hypotheses we want to include, as subject to possible learning, the more complicated the required prior is. We can take care of all the consistent hypotheses that are statable in a countable language, since their number is countable. But, as Putnam observed in 1963, for any given prior, one can define an empirical hypothesis for which the conditional probability of its holding in the next case, given its complete success in all previous cases, is always ≤ ½. [4] Therefore, any prior probability that takes care of all consistent hypotheses, which are statable in a given language, can be defined only in an essentially more expressive language. This is an *a priori* limitation on the applicability of Bayesian methods. The upshot is that prior probabilities are , in principle, determined by context. For example, assume that we are given a long binary sequence produced apparently by tossing a coin with unknown bias. The sequence passes the usual randomness tests and on the basis of the evidence we decide that it is a Bernoulli sequence, with a probability for 'heads' that lies in the interval [0.55, 0.6]. Then someone claims that this is a pseudo-random sequence, defined by such and such an algorithm; presumably the coin has been rigged by an implanted chip. 50 additional throws produce outcomes in complete agreement with the algorithm. At that point we switch abruptly and accept the algorithm hypothesis. Our original prior was a symmetric function (a mixture of Bernoulli probabilities) which did not take into account the possibility of pseudo-randomness. The abrupt change is not obtainable by Bayesian conditionalization.

Both the positive and the negative aspects of a Bayesian approach are investigated in my joint work with Snir, [1982], [5] where the language consists of a mathematical part (which includes arithmetic) and an empirical part consisting of empirical predicates and function symbols. Sentences in this language define sets of (possible) worlds, where a world is identified with an interpretation of the empirical part (the mathematical part is the same in all worlds). The probability of a sentence is the probability of the set of worlds in which the sentence is true.

On the positive side, there are convergence theorems that imply that, under quite general conceptions of "growing evidence", desired behaviors take place with probability 1 as the evidence grows. Namely: (i) the conditional probability of an hypothesis (i.e., sentence) tends to 1 if the hypothesis is true, to 0 ─ if it is false,[6] and (ii) the conditional probability of any hypothesis approaches the conditional probability that derives from any other prior, as long as the two priors assign the value 0 to the same hypotheses. "Probability 1" refers of course to the value of our given prior; hence these are coherence results that constitutes an inner justification.

must be 0. The argument is however fallacious. For a discussion of the issues see pp. 109 - 111 and 134 in my "Subjective probability natural predicates and Hempel's ravens", in *Erkenntnis*, 1979.

[4] "Degree of confirmation and inductive logic", in the Schilp volume *The philosophy of Rudolf Carnap.* 1963.

[5] "Probabilities over rich languages, testing and randomness", Journal of Symbolic Logic, 1982.

[6] In other words, the set of worlds in which the hypothesis is true and the limit is 1, or the hypothesis is false and the limit is 0, has probability 1.

It is when we come to justify the prior itself that the negative aspects emerge.

As already remarked, a probability function that assigns probability 0 to a consistent empirical hypotheses (a sentence that is true in some possible world) is dogmatic: no conditionalization on evidence that has positive probability, can move it from 0. Let us say that a prior is non-dogmatic, or open-minded for a class of sentences, Φ, if it assigns positive value to every consistent $\varphi \in \Phi$. A larger class Φ means that the probability is more open-minded. The results show that the more open-minded the probability, the more complex it is. In the paper, both "open-mindedness" and "complexity" are measured on a scale that corresponds to the arithmetic hierarchy.[7] It is, to be sure, ~~is~~ a crude scale; but the complexity-price for open-mindedness is a general phenomenon manifested in many ways.

Another tradeoff, due to limited computational resources is analyzed in a 2004 paper of mine,[8] dealing with the use of probabilistic algorithms that leads to assignments of high probabilities to unproven mathematical claims of finitary nature. E.g., one concludes that there is probability $\geq 1 - 2^{-40}$ that $2^{400} - 593$ is a prime number. Such statements can, in principle, be proven or refuted, but the checking requires unpractically long computations. The paper proposes a framework, based on a restricted notion of logical consequence, tailored to reflect limited deductive capacities, by which one can makes sense of the assignment of probabilities in a non-empirical context.

In Gaifman-Snir [1982] there is — as far as I know, for the first time in the literature — a general account of randomness, which treats a general empirical vocabulary, an arbitrary probability distribution, and arbitrary classes of sentences. The idea is simple: A world, w, is Φ-random, with respect to a probability Pr, where Φ is a class of sentences, just when w satisfies every $\varphi \in \Phi$ for which $Pr(\varphi) = 1$. Previous notions of randomness, due to Martin-Löf and to Schnorr come out as particular cases.[9] Given a probability Pr, a Φ– random world can be viewed as *typical* to degree Φ; the more extensive Φ, the more typical the world. In the paper randomness-degrees are arranged more or less along the lines of the arithmetic hierarchy (with certain adjustments, suggested by the previous randomness notions). A finer tuning would be required for examples that are based on specific statistical tests. Nowadays there is wide use of *pseudo-random* number generators — algorithms that generate sequences of digits that pass successfully various randomness tests (in fact, the decimal expansion of π scores reasonably well on the usual

[7] The complexity is that of the mathematically-defined function, which gives the probabilities of quantifier free sentences (the values for other sentences are determined by these); these sentences are treated via some Gödel numbering. A real-valued function $f(x)$ is defined by an arithmetical wff, if the wff defines a rational-valued function $g(x, y)$ such that $|f(x) - g(x, y)| \leq 1/y$ for all $y = 1,2,..$; rational numbers are treated, within Peano's arithmetic, in the obvious way.

[8] "Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements" *Synthese*, May 2004.

[9] They deal with probabilities ~~gendistributions, defined~~ over infinite binary sequences (i.e., the empirical vocabulary consists of a monadic predicate over the natural numbers), generated by Bernoulli trials in which the ~~parameter *p*~~ probability of "heads" is a recursive real number; the class of sentences, call them Φ^{ML} and Φ^{Sc}, are such that $\Sigma_1 \subset \Phi^{Sc} \subset \Phi^{ML} \subset \Sigma_2$. The choice of the class is motivated by tests used in statistics. These notions are generalized in the paper to any places in the arithmetical hierarchy: $\Sigma_n \subset \Phi_n^{Sc} \subset \Phi_n^{ML} \subset \Sigma_{n+1}$. In recent years, the topic of random reals and their connection to the complexity notions of Kolmogoroff, Solomonoff and Chaitin became a thriving, highly technical area of research.

tests). Since the sequences are mathematically determined, they cannot be considered "truly random"; presumably, they are not as random as a sequence derived from some physical phenomena (coin tosses, radioactive decay, etc.) But what is it that distinguishes the physical cases? After all, physical devices are not, as a rule, flawless; we expect them to produce good approximations to an idealized model. But how is this idealized model determined?

I suggest that, for a given *Pr*, "true randomness" is nothing but Φ-randomness for a "sufficiently large" Φ. In principle, we can consider a class consisting of all the sentences in a language that includes a powerful mathematical apparatus.[10]

The general notion of randomness underlies my view about a unified concept of probability, sketched in [1986*a*].[11] The idea is to start with a notion of creedal, or subjective probability, and to derive the notion of objective probability as a notion that has empirical content, a property that can be confirmed or disconfirmed by experiment. Objective probabilities are often defined in terms of relative frequencies. Now statements about relative frequencies have probabilities (they describe events that correspond to measurable sets in the σ-algebra). Large-number laws mean that certain statements of this form have probability 1. To say that the relative frequencies are such and such is simply to say that our world is Φ-random for a certain class of sentences, Φ. Such a probability *succeeds* (to degree Φ) in our world. The question of success arises for any subjective probability, *Pr*, which is defined over a large enough field of events. The more inclusive the class Φ, for which our world is Φ-random, the more successful *Pr* is. The notion of a *calibrated probability* — a probability assigned by an expert in a given area, which accords with observed frequencies — is a special case of this concept.

We cannot of course ignore the fact that the notion of randomness appeals to infinite models; in particular, relative frequencies are limits defined for infinite sequences. They cannot be verified or refuted by finite evidence. Yet we do have a methodology of accepting or rejecting such hypotheses on the basis of large enough finite evidence. The methodology is intricate; it is a rich subject by itself, subject to statistical research. But, in principle, it puts objective probabilities in the same category as theoretical notions with empirical content. The claim that a probability is objective is a factual claim about the world.

The notion of *propensity* used by philosophers to mark objective probabilities can be reduced to that of randomness. In a nutshell, propensity is a disposition of an experimental setup to produce Φ-random outcomes (usually in the form of sequences), for some presupposed class Φ.

Success is the first condition for the objectivity of a prior probability. There is another condition, *inner stability*. It means, roughly speaking, that prior probabilities do not undergo too much change upon conditionalization on information that is, in principle, accessible at the relevant time. The probability of H ("heads") on the first toss of a fair coin is 0.5. But suppose that a coin is randomly drawn from a pool of coins, half — with a 0.8/0.2 bias for H and half — with the same bias for T. The probability for H on the first toss is still 0.5; but conditionalization on the

[10] Say all sentences in any k^{th}-order logic that define (in the standard interpretation) *Pr*-measurable sets of worlds.
[11] "Towards a unified concept of probability", *Proceedings of the 1983 International Congress for Logic Methodology and Philosophy of Science*, R. Barcan Marcus, G.J.W. Dorn and P. Weingartner editors, North Holland 1986, pp. 319 -350 (the publication of the volume was delayed, the paper was completed in 1984). Unfortunately, this volume is not easily available and the paper is little known. It is my fault not to have followed up on it.

outcome of the lottery would significantly change it. In general, inner stability is relative to two Boolean algebras (or σ-algebras): the algebra over which the prior is defined and the subalgebra of events which we consider *knowable* in the given context. In [1986a] there is a proposal for measuring the *expected change* when we conditionalize on events from the second algebra. The smaller the expected change, the more stable the prior. Inner stability depends on what we consider "knowable events". These vary with time and with context. Even in a physically deterministic world it makes sense to assign objective probabilities to lottery outcomes, because the amount of information required to make better-than-random predictions is too large, hence it is not accessible knowledge. While inner stability is a feature of the probability distribution, success is a feature of the world. Both are a graded notions and the spectrum has more than one dimension.

My paper on higher-order probabilities [1986*b*] grew out of this conception.[12] It illustrates the interplay of objective and subjective probabilities in a single system. Essentially, the model is based on a subjective mixture of objective probabilities; the algebra of events contains also events of the form $a \leq Pr(A) \leq b$, to be read: "the objective probability of A is between a and b". Its main point is to present the theory in axiomatic form and to get suitable representation theorems.

3. How do you conceive of the relationship between probability theory and/or statistics and other disciplines?

I shall interpret "other disciplines" as referring to philosophical disciplines. The subject is otherwise too wide for the present format. Let me, nonetheless, point out that present day physics is unthinkable without probability theory, which enters into it at a foundational level. Probability and statistics are also ubiquitous in the experimental sciences (though not as foundational as in physics), as well as in the general culture. Probability theory is a major tool in various disciplines in CS, which deal with uncertain reasoning, AI and expert systems. These areas have contributed a great amount, if not the majority of publications that deal with applications of probability within the general framework of uncertain reasoning. I shall say more on this in subsequent answers.

As a purely mathematical theory, probability theory is part of the intricate web of mathematics. Tools from probability theory are applied on a broad scale and questions relating to probabilities are posed in diverse contexts. Probabilistic concepts have been generalized in abstract imaginative ways, based on structural affinities that can be appreciated only by experts. In mathematics the answer to your question is therefore descriptive; it has to do with the ever changing epistemic organization of mathematics as a whole. In philosophy, on the other hand, it is a systematic question that calls for systematic answers. Let me attempt one.

Probability theory is part of the logic of reasoning under conditions of uncertainty, where "logic" is taken in a broad sense: a system of general principles, or meta-rules of reasoning. (Statistics,

[12] "A theory of higher order probabilities" in *Theoretical Aspects of Reasoning About Knowledge*, ed. J. Halpern, 1986. An expanded version is in *Causation Chance and Credence*, ed. B. Skyrms and W. L. Harper, 1988.

Please send the completed questionnaire by **January 15, 2008** electronically to Vincent F. Hendricks (vincent@ruc.dk) or mail (fax) to Vincent F. Hendricks, Dept. of Philosophy and Science Studies, Roskilde University, DK4000 Roskilde, Denmark, Fax: +45 4674 3012

roughly speaking, deals with situations where large amounts of data, uniformly organized in standard forms, are available.) The uncertainty derives from epistemic shortcoming: lack of knowledge, usually due to the empirical nature of the subject matter. But it can also derive from limited deductive capacity, as in the case of unsettled mathematical conjectures, about which a mathematician may have various degrees of belief. It should not be surprising that some principles of uncertain reasoning may apply across the board; after all, these are strategies of the same mind, which is biologically underpinned by the human brain. A strategy that works in the empirical domain might work sometimes in mathematical investigations (see, e.g., Pólya's writing on the matter.[13]) Surely, the strategies are useful and one can speculate on their survival value in human evolution. The speculations cannot however provide philosophical justification, just as facts about brain mechanisms cannot provide answers to questions about mathematical validity and truth.

It follows that probability theory is part of the theory of rational thinking, with direct inputs into rational choice, the philosophy of science and epistemology in general. That characterization accommodates also objective probabilities, along the lines indicated in my previous answer. But the existence of such probabilities is a brute empirical fact, or, if you want, part of physics. (To the extent that physics gives us an ontological picture of the world, probability theory is part of ontology.) Let me elaborate more on that conception.

How basic are the norms of uncertain reasoning? We can gain insight into such questions by asking, what the giving up of the norms entails for our thinking and for the possibility of meaningful communication. Thus, an elementary part of sentential logic appears necessary for the very possibility of coherent discourse. Quine observed that even if truth can be viewed as a mere convention, elementary logic cannot. For we need elementary logic in order to derive the consequences of whatever convention we adopt. Conjunction and conditional are necessary to state the rules of any well defined game and some basic inference rules are necessary to derive the consequences. Here, it seems we have hit rock bottom. There is little or no hope for securing such a base for reasoning with uncertainties. Yet, particular judgments concerning uncertainties appear unavoidable. Suppose that guessing correctly the color of a ball to be drawn from a bag is extremely important (say it is a life or death question). Knowing that the bag contains four black balls and one red and knowing that the last 40 drawings have produced nine red balls randomly distributed, it would be irrational to guess 'red';[14] unless one has additional information (say, one knows that the next drawing is rigged, or one has noticed some regularities in the past pattern). Surely, there is no place for the following answer, given by someone, say John: "There is no way of telling what the next drawing will yield. The question is underdetermined; in this situation rationality mandates no particular choice and at this moment I feel like guessing 'red'."

[13] Pólya has written several books on uncertain reasoning and heuristic strategies in mathematics. The heuristics is based on finding the right analogies, which are usually specific to the mathematical structures and forms.. But general principles concerning qualitative degrees of belief apply in mathematics as they apply in empirical domains. As mentioned above, there are also cases in which numeric probabilities are assigned to mathematical statements.
[14] A scenario of this type was suggested by Peirce, 1878, "On the doctrine of chances with later reflections", in *Philosophical Writings of Peirce*, ed. J. Buchler, Dover 1955, p. 160.

Someone who applies John's type of reasoning does not have much chance of survival; but this observation presupposes already *our* type of reasoning. Essentially, we are dealing here with a version of Hume's problem of induction. I do not think that we can resolve it, but I think that someone who reasons like John has a different language from ours.[15] Our language is informed by a qualitative notion of *likeliness*, which enables us to judge that some event (proposition) is more likely than another event. Dispositional characterizations of the most common type ('safe', 'safer' 'dangerous', 'reliable', 'more reliable', etc. etc.) are grounded on that. A human language that does not accommodate some such notion is hardly conceivable. And it is part of this story that likeliness guides action. These considerations may help us in locating a hard core for the logic of uncertain reasoning. It will turn out, I think, weaker than qualitative probability as defined by de Finetti in 1937 and neatly presented in Savage's *The Foundations of Statistics.* The assumption of a total pre-order (i.e., that every two events are comparable) should be dropped. At this basic level, only a partial pre-order is justified. The presupposition of a linear ranking is often accepted uncritically, because linear rankings are all around us, from physics to beauty contests. We have linear rankings of likeliness in certain areas, where uniform organization obtains (say, we have sufficient statistics and judge other information irrelevant). But in many cases we find it difficult if not impossible to make comparisons — not because we are not smart enough, or introspective enough to analyze at great length what we know — but because no sufficient basis for ranking exists. If for practical reasons we must decide which of two events is (weakly) more likely, we will make a decision; but such a practical decision need not be grounded in an enduring rational framework.

I believe that the other axioms of qualitative probability should be retained, in particular, the axiom that states (in the Boolean algebraic framework) that if $A \cap B = A \cap C = \varnothing$, then: $A \cup B \lesssim A \cup C \;\Leftrightarrow\; B \lesssim C$ (where $\lesssim$ is the less-than-or-equally-likely relation). In a linguistic framework, in which $\lesssim$ is a relation between sentences describing the events, we appeal to the notion of logical implication (in the traditional sense of "logic"). Thus, using '$\vdash$' for logical implication, we can take as an axiom: $\alpha \vdash \beta \Rightarrow \alpha \lesssim \beta$. In a more refined system we would use $\vdash_k$ instead of $\vdash$, where '$\alpha \vdash_k \beta$' reads: α *is known to logically imply* β. Such a system is required for cases where the uncertainty derives also from limited deductive capacity. The Boolean algebraic framework is simpler, but is unable to represent this type of uncertainty. We might want, for reasons of simplicity, to ignore the complication and treat our agents as logically omniscient. But this is a methodological rather than a philosophical move.

From the hard core outlined above to numeric probabilities the way is long. There is a rich history of attempts to establish numeric probabilities on the basis of rationality principles. I

[15] This observation should not be taken as a solution of Hume's problem. It is conceivable that the empirical regularities will radically change, forcing us to adopt a new kind of language.

cannot go, within the scope of this answer, into details. I shall limit myself to the following points. The derivation of numeric probability from a qualitative one requires the axiom of a total pre-order and, in addition, axioms guaranteeing the existence of "fine partitions". For example, if for arbitrary large n's there are partitions of the universal event (the maximal set in the Boolean algebra) into n disjoint equally likely events, then unions of these events establish a numeric scale, 0, $1/n$, $2/n$,…, n/n; the probability of any event can be then determined up to an interval of length $1/n$, if we can compare the event to each of these unions. But the fine-division assumption can be justified only by appeal to a chance mechanism, such as repeatable, independent tosses of a fair coin. Probabilities can be derived, in less obvious ways, from other fine-partition assumptions, but all require an appeal to chance mechanisms.[16] Given the empirical fact of randomness phenomena, we are justified in applying, in such cases and for comparable events, objective probabilities, in the same way that we are justified in applying a successful scientific theory.

The problem of comparing arbitrary events with those that can get probabilities remains however. Let A, B, C be, respectively, the events that (i) the next US president will be a Democrat (ii) Iran will develop a nuclear bomb within the next 4 years, (iii) within 5 years the twin prime conjecture will be decided. Let $D = A \cap (B \cup C)$. Is my personal probability for D greater than 0.55? Appeal to bets I am willing to accept does not make the question easier. I find it impossible to give fair odds and prefer not to gamble. What if (assuming an hypothetical scenario) I am forced to offer odds that commit me, under threat of severe punishment? I shall give some odds, but my choice will be subject to mood and whim; no principle is available here for inferring "right odds" from what I know. Dutch book arguments fare no better. They require that the agent give odds for all events in a big collection, with a commitment to accept any system of bets that accord with the odds. Surely, if I must do so under threat of punishment, and if I think that the experimenter, who is out to get me, will choose a Dutch book against me, if s/he can, I shall take care to eliminate Dutch books: I shall assign odds that are based on some probability distribution. In this unlikely, artificial scenario, it is indeed recommended to use quantitative probabilities. But this is all that the argument proves.

Preferences among possible acts can sometimes be used to reveal one's beliefs concerning likeliness, or even to clarify to oneself one's own intuitions. But *being more likely* is a primitive of our conceptual apparatus. It can explain choice of action, but it is not definable in terms of the latter. I prefer the gamble on X to the same gamble on Y, *because* I think X more likely than Y; my belief does not reduce to the preference, it explains it. Proposals, such as Savage's, for deriving probabilities from preferences among acts, require a linear ranking of acts (complex entities of hypothetical nature), which is no easier, and often harder, than comparisons of likeliness. There are also complications regarding utilities, in particular, an adoption of a controversial sure-thing principle, which has counterexamples, such as Ellsberg's paradox.

The emerging picture is roughly this. A comparative notion of likeliness is a primitive of our conceptual apparatus, which informs our language. Some basic rationality rules can be traced to it. Given the empirical regularities of chance phenomena, we are rationally justified in adopting

[16] For example, the axiom: if A is strictly more likely than $\varnothing$, then, there is a partition of the universal event into disjoint events, each of which is strictly less likely than A.

the framework of numeric probabilities and in applying it in a great variety of cases. How wide is the field that admits justified use of numeric probabilities? This is an open question and a subject of dispute. The field is restricted by the fact (discussed in the previous answer) that any workable prior probability must be to some extent dogmatic: it must ignore many possibilities to which it assigns the value 0. In particular, the adoption of new surprising hypotheses, not to speak of switches to new scientific theories, is not achieved through Bayesian conditionalization,[17].

4. What do you consider the most neglected topics and/or contributions in probability theory and/or statistics?

Given the amount of papers that are being published in the area of uncertain reasoning, and that I am not a voracious reader, there is a chance that what I think "neglected" has been addressed somewhere. I shall not risk it. But I will comment on general trends.

We have today a great number of tools, theoretical, as well as operational, for representing uncertainty and for processing data. The development is largely driven by the knowledge engineering industry in computer science. Practical needs provide direct and indirect support for a broad spectrum of theoretical research. From the engineering point of view, tinkering is justified. One tries whatever might have some promise, one adopts whatever seems to work. What I am missing in all this is a philosophically motivated approach on a foundational level. I do not mean that we need "foundations" to make us feel philosophically secure. I mean a fine-grained analysis that illuminates basic features, explains the (usually implicit) presuppositions underlying various methods, how they stand to each other and why such and such methods work in specific areas. Usually, a proposed method will have some motivating intuition and a good handbook will make this clear. But most often the reader finds himself in the position of a shopper presented with recipes and recommendations (X is usually good for this, and Y for that; here are some arguments for X, and here are some criticisms; here are some arguments for Y and here are arguments against it). Besides, the motivation for some proposals seem to be mere formal analogies, or game-playing at constructing models.

I would like to mention in this context Putnam's "Degree of confirmation and inductive logic".[18] This short paper from 1963 pointed out the in-principle shortcomings of Carnap's program and, arguably, put an end to it. On this account it cannot be considered "neglected". But the paper indicates a positive line of research: the tradeoff between open-mindedness ─ the possibilities that the prior affords to learn from experience ─ and the prior's complexity. And this line seems to be neglected. There is some further development in [Gaifman-Snir], but nothing, to my knowledge, after that.[19] Since feasible computation can be of crucial importance, the subject has been addressed in the context of various proposals. But there has been no investigation of the computational complexity that is required for applying more open-minded priors. I shall say something more on this in the final section.

[17] Some philosophers, while willing to accept
[18] Published in the Schilpp volume for Carnap.
[19] I, more than anyone, am perhaps responsible for not following up this line, or for publishing related material in a little known collection.

5. What do you consider the most important open problems in probability theory and/or statistics and what are the prospects for progress?

To a large extent my answer is implied by the previous ones. If our goal is to understand the place of probabilistic reasoning within the enterprise of human knowledge, we need an analytic map of the various approaches, which is much more than a handbook-enumeration of tools. This is a daunting project, but there are good prospects for making progress at least in some directions.

As I see it, we can start by assuming that we are not going to solve Hume's problem. [20] That being said, we should avoid the other extreme of uncritical description of the various kinds of reasoning practiced by humans ("that is how we do it; end of story"). There is considerable place for rational analysis and criticism. Certain symmetries should be preserved; to take a trivial example, our reasoning regarding the tossing of a coin should not be affected by switching the names we use in referring to the two sides.[21] Yet our reasoning is tested in the actual world; rationality cannot be too much removed from truth and success. If the world rewards "irrational reasoning", i.e., if certain inferences judged irrational are a better guide to truth, then we should modify our criteria. A philosopher is called upon to toe a fine line between an impoverished rationality and an exaggerated one.

Here is a task where some progress is plausible: get better insights into the status of the maximal entropy principle (henceforth ME). Jaynes, who advocated ME as a way of choosing prior probabilities, got through skilful applications of it impressive results in physics. The fact that certain applications of ME have been very successful tells us something about the empirical world. What does it tell, beyond the fact itself? ME has been the focus of much research among physicists (research with which I am not acquainted). Perhaps we can get a good understanding of its strengths and weaknesses without getting into sophisticated physics. In this respect the work of Paris and Vencovská ─ who derive ME, in a discrete setup of finitely generated Boolean algebras, from a group of relatively simple formal principles (which ME satisfies) ─ is quite interesting. [22] Their principles were criticized on the familiar grounds that they are too strong and amount to inferring real knowledge from ignorance. They responded by pointing out that the counterexamples offered by their critics presuppose relevant knowledge that is not represented in the model. Thus, they argued, when all relevant knowledge is expressed in terms of the setup, ME is binding on *a priori* grounds of rationality. Is it?

As a prescription for choosing a prior probability, which will be then used in a Bayesian framework, ME can fail quite badly. The following very simple example will suffice for the present discussion. Consider tosses of a strange-looking coin, where the results are given in the form of a binary sequence (all the agent knows is that '0' and '1' refer to different sides of the coin). Symmetry dictates that, for a single toss, the probability of 1 (and of 0) is ½. For *n* tosses

[20] That is, solve it not by dismissing it in the name of some ordinary-language philosophy.
[21] It is conceivable that some naming ceremony might affect the behavior of a coin. We are justified in disregarding the possibility, unless the evidence forces it upon us. But even then, we can presuppose that the names we use in analyzing the problem have by themselves no effect. Otherwise the very possibility of a coherent analysis is undermined.
[22] "A note on the inevitability of maximum entropy", International. J. of Approx. Reason. 4(3), 183-224, 1990. " In Defense of the Maximum Entropy Inference Process", Ibid. 17(77), 77 – 103, 1997.

there are 2^n possibilities and ME mandates the uniform assignment that assigns probability $1/2^n$ to each. Thus, we get Bernoulli distribution, with $p = ½$. This prior is as dogmatic as a prior can be. Nothing is learnable from experience; given that all the first 100 tosses resulted in 1, the conditional probability for 1 on the next toss is still ½. Unless one fully believes that s/he knows that the coin is fair, one should never choose that prior. Note that the case falls within the purview of the Paris-Vencovská system. The example shows that the system is not capable of handling subjective probabilities in which the agent is uncertain about the objectivity of his or her evaluations.

Now the same situation can be modeled differently: the agent assumes that the tosses are Bernoulli trials, but has no inkling about the probability p (say, of 1). ME dictates a uniform probability density over [0,1]. The probability of getting a sequence of k 1's and $n-k$ 0's comes out as $\binom{n}{k}\int_0^1 p^k(1-p)^{n-k}dp$. It is the prior determined by $\lambda = 2$ in Carnap's continuum of inductive methods (mentioned in the first section). This is not an unreasonable prior. Yet, why should we prefer it to $\lambda = 1$ (which has a faster rate of learning from experience, though a higher risk of errors in the early stages)? As remarked in my first answer, all these probabilities are dogmatic, to the extent that they rules out the possibility that all tosses yield 1 — an hypothesis one will surely entertain, given say that the first 10.000 tosses have all resulted in 1.

This shows how crucial the right modeling of the situation is and that ME cannot be applied in a schematic general way. Jaynes, who was well aware of these points, relied considerably on his expert intuitions in physics. What is it then about the world that determines the cases where ME succeeds and which accounts for its success in those cases? Surely, it is more than the randomness phenomena and the regularities that account for numeric probabilities in general.

While the Paris-Vencovská system does not provide a rationality argument for ME, it yields nonetheless valuable insight into it and perhaps may provide justification for using it in cases where the agent is a true expert and some additional conditions hold (this would require detailed analysis).

Concerning the issue of open-mindedness versus feasible computability, we can look for results of the following form that establish mathematical connections between the two. Open-mindedness can be measured by the possibilities that are not ruled out *a priori* , that is, by a class of hypotheses whose prior probability is non-zero. One would do better to consider some limited natural class (rather than a high level one based on the arithmetical hierarchy); e.g., in the case of binary sequences, we may include, in addition to Bernoulli trials, also Bernoulli trials with varying probabilities, perhaps also certain Markov processes. We may also consider the speed of learning from experience, i.e., the rate of convergence as the evidence grows. Feasible computability can be measured on some well known complexity-scale from theoretical computer science. It would be very nice to have results of the form: a prior probability that is open-minded with respect to such and such hypotheses, requires, for its (approximate) computation at least such and such computational resources.