Ten great ideas about chance

P Diaconis and B Skyrm (Princeton UP 2018).

1. Measurement Photo: G Cardano

1

The first great idea is simply that chance can be measured. It emerged during the sixteenth and seventeenth centuries, and it is something of a mystery why it took so long.

2 Length: discussed.
3 Probability as a ratio, using equiprobable cases: birthday coincidence example.
4 Nothing provides us better candidates for equiprobable cases than vigorous throws of symmetric dice or draws from a well-shuffled deck of cards. This is where the measurement of probability began.
The idea was clearly there in the sixteenth century work on gambling by Gerolamo Cardano.
5 In constructing equiprobable cases, both Galileo and Cardano appear to make implicit use of independence.

The first substantial work in the mathematics of probability appears to be the correspondence between Pascal and Fermat, which began in 1654. (the Chevalier de Méré problem).
7 Problem of point. The ideas in the Pascal-Fermat correspondence were taken up and devel-
9 oped by Christiaan Huygens after he heard about the correspondence on a visit to Paris. He then worked them out by himself and wrote the first book on the subject in 1656. It was translated into english by John Arbuthnot in 1692 as Of the Laws of Chance. 12 Huygens begins his book with a fundamental principle:

10
> That any one Chance or expectation to win any thing is worth just such a Sum, as would procure in the same Chance and expectation at a fair Lay.

He then goes on to justify this measure by a fairness argument.

In the preface to the translation of Huygens, Arbuthnot, who was a follower of Newton, makes the following noteworthy remark: “It is impossible for a Die, with such determind force and direction, not to fall on such determind side, only I dont know the force and direction which makes it fall on such determind side, and therefore I call it Chance, which is nothing but the want of art.

Arbuthnot thus introduces the fundamental question of the proper conception of chance in a deterministic setting. His answer is that chance is an artifact of our ignorance.
[**C**] Notice that this argument is based on unfounded metaphysics (just as Huygens' premise, to be fair). Determinism is sometimes a good approximation! That is all, empirically.

Coin tossing is physics, not random! To demonstrate this, we had the physics department build us a coin-tossing machine. The coin starts out on a spring, the spring is released, the

11 coin spins upward and lands in a cup. Because the forces are controlled, the coin always lands with the same side up.

How then is the probabilistic treatment of coin flips so widespread and so successful? The basic answer is due to Poincaré. If the coin is flipped vigorously, with sufficient vertical and angular velocity, there is sensitive dependence on initial conditions.

Bernoulli's *Art Conjectandi* 1713. He defined the conditional probability
12 On the basis of these definitions, he shows that the probabilities of mutually exclusive events add and that probabilities satisfy the multiplicative law, $P(A \wedge B) = P(A)P(B \mid A)$. These simple rules form the heart of all calculations of probability. But Bernoullis major contribution was to establish a rigorous connection between probability and frequency that had heretofore only been conjectured. He called this his golden theorem: the weak law of large numbers.

SUMMING UP: Probability, like length, can be measured by dividing things into equally likely cases, counting the number of successful cases and dividing by the total number of cases. This definition satisfies the following:

1. Probability is a number between 0 and 1.
2. If A never occurs, $P(A) = 0$. If A occurs in all cases, $P(A) = 1$.
3. If A and B never occur in the same case, then $P(A \vee B) = P(A) + P(B)$.
4. Conditional probability for B given A is defined by.... Then, $P(A \wedge B) = P(A)P(B|A)$.

The law of large numbers, to which we will return in chapter 4 (and again in chapter
18 6), proves that chances can be approximated (with high probability) by frequencies in repeated independent trials.

Coin-tossing initial condition subsets illustrated.

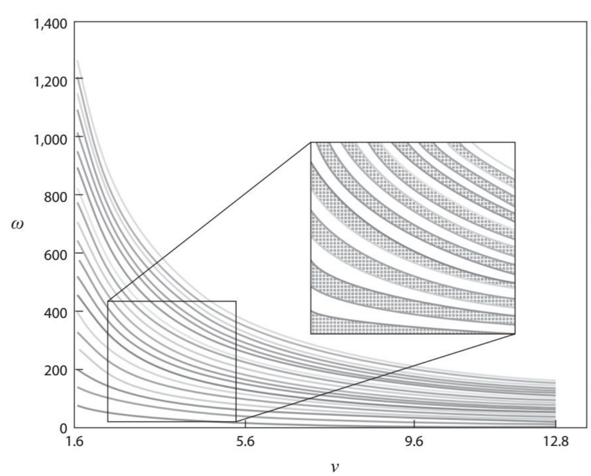

Figure 0.0.1: The hyperbolas separating heads from tails in part of phase space. Initial conditions leading to heads are hatched, tails are left white, and ω is measured in s^{-1}.

21 Proposition: With N people and C possibilities, N and C large, the chance of no match is

$$P(C,N) \sim e^{-N(N-1)/2C}. \tag{0.0.1}$$

Diaconis, Persi. Ten Great Ideas about Chance (p. 21). Princeton University Press. Kindle Edition.

2. Judgement Photo: F P Ramsey

22 Our second great idea is that judgments can be measured and that coherent judgments are probabilities.

In this chapter we will see how degrees of belief implicit in judgments about all sorts of cases can also be measured. When they are so measured, coherent judgments turn out to have the same mathematical structure as that discovered by Cardano and Galileo by counting equiprobable outcomes in gambling.

23 Often we do not have a nice set of intuitively equiprobable cases that allowed us to calculate by counting with supposedly perfectly fair dice. ... That does not mean that they cannot be measured.

Below, we will be talking about assessing probabilities by betting.... It is very natural to take the current market price as the markets probability. The expected value of a bet that pays off \$100 if C; \$0 otherwise is \$57, if the probability of C is 0.57. If the market prices do not obey the mathematics of probability, then the market can be arbitraged.[1]

24 **Part I Gambling and judgmental probabilities**

We measure the expected value imputed to an event by measuring the price that an individual will pay for a wager on that event.

Your judgmental probabilities are then the quantities which, when used in a weighted average, give that expected value. In particular,

The probability of A is just the expected value of a wager that pays off 1 if A and 0 if not. ... The balance point, where you are indifferent between buying the wager or not, measures your judgmental probability for A.[2]

COHERENT JUDGMENTS

25 Do judgmental probabilities, in general, have the mathematical structure gotten in chapter 1 by counting? Bruno de Finetti showed that if an individuals betting behavior is coherent, her judgmental probability, so defined, does indeed have the mathematical structure of a probability.

The basic argument can be given very simply.[3] In short, the participants do not have neither sure-winning or losing strategies. Such strategies are impossible. This is the coherent condition. In the following, a sure-losing strategy is devised for all the cases.

26 De Finetti showed that coherence is equivalent to ones judgments having the mathematical structure of probability.

COHERENT JUDGMENTS ARE PROBABILITIES

To say ones judgments have this mathematical structure is just to say that they behave as proportions: If there are 20% red beans and 35% white beans in a bag then there are 55%

[1] arbitrage = the simultaneous buying and selling of securities, currency, or commodities in different markets or in derivative forms in order to take advantage of differing prices for the same asset.

[2] It is farfetched idealization to assume that people can effortlessly and reliably make such fine discriminations. But taking the first steps of the approximation is perhaps all we need to do for many decisions.

[3] At this point the argument proceeds as if we could just use money as a measure of value. This assumption will be lifted in the second half of this chapter.

beans that are red or white.

I. Coherence implies probability.

1. Minimum of zero.

Demo As a bookie you bet $p > 0$[4] when you must pay 1; if you win you pay p and get -1. If you lose, you get nothing must pay p;, so always you lose.

2. Tautology (always true or sure-fire case) gets probability of 1.

Demo Suppose that you give a tautological proposition a probability p. It is either greater than or less than 1.

∗ If it is greater than 1, you would pay more than 1 for a bet that pays off 1 if p, nothing otherwise. When bets are settled you would win only 1, for a net loss.

∗ If you were to give the tautology probability less than 1, you would sell a bet that pays off 1 if p, nothing otherwise for less than 1. When bets are settled, your buyer would collect 1 and you would be left with a net loss.

> Sure-strategy approach is easier:
>
> 1. Minimum of zerop. If one to pay $P(p) < 0$ for \$1: if he wins, he get $1 - P$, if he loses, he gets $-P$. Sure winning.
>
> 2. Sure-case; if one pays $P(p)$, if he wins he gets $1 - P(p)$; he never loses, so $1 - P(p)$ is the
> 27 gain, whose sign is definite unless $P(p) = 1$.

3. Mutually exclusive parts add. Now consider additivity. Suppose p, q, are mutually exclusive. (Their conjunction is inconsistent.) But notice that an arrangement that pays off 1 if p or q; nothing otherwise can be made in two different ways. It can be made directly, as a bet on p or q, or it can be achieved indirectly by simultaneous bets on p and on q.

∗ If not $P(p \vee q) = P(p) + P(q)$, an arbitrage is possible.

II. Probability implies coherence.

To show this, we show sure strategy is impossible: If judgments are probabilities, then expected values of bets add: $E(b_1 + b_2 + \cdots) = E(b_1) + E(b_2) + \cdots$. But a Dutch book (sure-losing) requires that $E(b_1 + b_2 + \cdots)$ be negative, while $E(b_i)$ are all nonnegative. A
28 Dutch book is impossible. Judgments that are mathematical probabilities are coherent.

Conditional bets:

De Finetti had an additional important idea about coherence, one that relates conditional probabilities to conditional bets. A conditional bet is one that is called off provided that the condition isnt met

\$a if q and p,
\$b if not q and p,
0 if not p

The imputed conditional probability, $P(q|p)$, of a conditional bet judged as fair is $b/(a + b)$.

[4] This p is not he p in tthe book.

29 Other choices make sure-strategy possible. (illustration only up tp p29) A strategy to prove this: first make a sure losing strategy under p. Let the loss be x. To distribute this loss to the non p case as well.

Is there a sense of coherent belief change that applies? ... Conditioning on the evidence, Bayesian updating, is the unique coherent rule for updating probabilities in such a situation.
30 Any other rule leaves one open to a Dutch book against the rule—a Dutch book across time, a diachronic Dutch book. Here are a model and a precise version of the argument.

COHERENT BELIEF CHANGE IMPLIES CONDITIONING ON THE EVIDENCE

The epistemologist (scientist, statistician) acts as bookie. She has a principled way of updating on the evidence, but we dont presuppose what it is. Today she posts her probabilities and does business. Tomorrow she makes an observation (with a finite number of possible outcomes, each with positive prior probability.) She updates her probabilities according to her updating rule, a function that maps possible observational outcomes to revised probabilities. Her updating rule is public knowledge. The day after tomorrowafter her observation she posts revised fair prices and does business.

A bettors strategy consists of

(1) a finite number of transactions today that the epistemologist considers fair according to her probabilities, and

(2) a function taking possible observations to sets of finite transactions the day after tomorrow at the prices the epistemologist then considers fair according to her updating rule.

Let $P(A|e)$ be $P(A \wedge e)/P(e)$ and $P_e(A)$ be the probability that the bookies nonstandard updating rule gives A if e is observed. Suppose $P(A|e) > P_e(A)$ and let the discrepancy $\delta = P(A|e) - P_e(A)$. Here is the bettors strategy that makes a Dutch book.

Today: Offer to sell the bookie at her fair price:

1. \$1 if A and e, 0 otherwise.
2. \$$P(A|e)$ if not e, 0 otherwise.
3. \$$\delta$ if e, 0 otherwise.

Tomorrow: If e was observed (i.e., today), offer to buy \$1 if A, 0 otherwise, from the bookie at its current fair price: $P_e(A) = P(A|e) - \delta$. Then, in every possible situation, the bookie loses \$$\delta P(e)$.

Demo Today's fair prices are 1: $P(A \wedge e)$, 2: $(1 - P(e)P(A|e)$, 3: $P(e)\delta$. Tomorrow: $-(P(AS|e) - \delta)$ if e happens. Suppose e does not happen today: the bettor wins only 2: $-P(A \wedge e) + (1 - P(e)P(A|e) = P(A|e)$ The bookie loses \$$\delta P(e)$. If e happens, the loss is $P(A|e)$ and gain is $P_e(A) = P(A|e) - \delta$, so totally the bettor gains (the bookie loses) $\delta P(e)$. This is sure-wining. (p227)

34 **PART II: UTILITY AND JUDGMENTAL PROBABILITIES**

Suppose that according to your preferences, some outcome, call it GOOD, is the best and another, call it BAD, is the worst. We choose utility of GOOD to be 1 and utility of BAD to be 0.

Now, for the utility of any other outcome, O, we find a gamble GOOD with chance p, BAD otherwise, so that you are indifferent between it and O for sure. Then you take the utility of O to be equal to p. That is to say that the utility of O is equal to the expected utility of the gamble that you judge equally good.

38 Simplifying assumptions:

(First t)i) The preferences must totally order the gambles.

(ii) Continuity: If p preferred p preferred p, then there is a probability, a, such that ap + (1 a)p indifferent p.

(iii) Independence: p preferred p if and only if ap + (1 a)p preferred ap + (1 a)p, for every a and every p.

Numerical utility is just the probability of GOOD.

40 Ramsey:

Ramseys idea of an ethically neutral proposition. This is a proposition, p, whose truth or falsity, in and of itself, makes no difference to an agents preferences. That is to say, for any collection of outcomes, B, the agent is indifferent between B with p true and B with p false. Intuitively, the nice thing about gambles on ethically neutral propositions is that the expected utility of gambles on them depends only on their probability and the utility of the outcomes. Their own utility is not a complicating factor. Now we can identify an ethically neutral proposition, h, with probability 1/2 as follows. Consider two outcomes, A, B, such that you prefer the first to the second. Then the ethically neutral proposition, h, has probability 1/2 for you if you are indifferent between [A if h; B otherwise] and [B if h; A otherwise]. Now we have identified a subjective surrogate for a fair coin toss. This is the key idea.

44 We have seen the development of the view that probability theory is logic—the logic of coherent degrees of belief.

48 **3. Psychology** Photo A Tversky (+ D Kahneman)

Our third great idea is that the psychology of chance and the logic of chance are quite different subjects. The ideal normative assumptions of chapter 2 are often violated in practice. Normative and descriptive aspects of judgmental probability and decision theory come apart, leaving a substantial gulf.

49 The case that the gap is unbridgeable was made with great force by Daniel Kahneman and Amos Tversky in a whole series of studies.

Here we give an introduction to the contrast between the psychology of decision and the logic of decision.

50 Savage sees independence (or sure-thing reasoning) as a fundamental principle of rational decision. But it was immediately challenged by the French economist Maurice Allais.

52 Uncertainty and risk are distinct.
There are two urns. The first has 100 balls, some red and some blackyou dont know the proportion. You know that the second has 50 red and 50 black balls.

54 Some people just get an unpleasant feeling making decisions under uncertainty that they dont get making decisions under risk. Who is to say that they shouldnt? (Others might get a thrill making decisions under uncertainty that is bigger than the thrill making decisions under risk. Who is to say that they shouldnt?) But if such psychological payoffs enter into Ellsberg cases, then they too should be counted as part of the consequences. And when they are counted, we no longer have a counterexample to the Savage axioms.

In 1974, Amos Tversky and Daniel Kahneman published Judgment under Uncertainly: Heuristics and Biases. ... The article describes biases arising from three heuristics: representativeness, availability, and adjustment and anchoring.

55 In 1984, Kahneman and Tversky published Choices, Values and Frames13 in American Psychologist. Here we concentrate on framing

62 **4. Frequency** Photo Jacob Bernoulli
Early probabilists were well aware of the limitations of relying on intuitively equiprobable cases.

Leibniz and Bernoulli did not themselves identify probability with frequency. For them, probability was a form of rational degree of belief. What, then, was the formal connection between frequency and probability? Jacob Bernoulli succeeded in answering part of the question, with a version of our fourth great idea, the law of large numbers. It establishes one of the most important connections between chance and frequency. The initial form, the weak law of large numbers, was proved by Bernoulli in his Ars Conjectandi. It was subsequently strengthened to the Strong Law of Large Numbers by Borel and Cantelli. This great idea has a disreputable twin. It is the idea that chance is frequency.
64 As we saw, Bernoulli was interested in using the bound in empirical applications. But his bound was not very good and conjures up very large numbers of trials.

Bernoullis motivation for his golden theorem was the determination of chance from empirical data.
65 What does it mean to determine chances a posteriori from frequencies? The question is, given the datathe number of trials and the relative frequencies of success in those trialswhat is the probability that the chances fall within a certain interval. It is evident that this is not the problem that Bernoulli solved. He solved an inference from chances to frequencies,

not the inverse problem from frequencies to chances. The inverse problem had to wait for Thomas Bayes. Yet Bernoulli somehow convinced himself that he had solved the inverse inference problem. How did he do so? It was by a vague argument using the concept of moral certainty. Bernoulli uses the term moral certainty to refer to a probability so close to 1 that for all intents and purposes, one may treat it as a certainty.

Bernoulli argued that he had shown that with a large enough number of trials, it would be morally certain that relative frequency would be (approximately) equal to chance. But if frequency equals chance, then chance equals frequency. So, the argument goes, we have solved the problem of the inference from frequency to chance. This is Bernoullis swindle.

66 What Bernouilli shoed: They are probabilities that frequencies fall within a specified interval of frequencies given an exact value of chances, rather than probabilities that chances fall within a specified interval given an exact report of frequencies.
Bernoulli gives the chance that frequencies fall within an interval given the chance on a single trial, rather than rational degree of belief that the chances fall within an interval given the frequencies.
The argument that the law of large numbers solves the inverse problem is a fallacy. But it is a rather slippery fallacy,* especially when stated informally. And it is a remarkably persistent fallacy, easy to swallow in the absence of rigorous thinking.
We find it in the French mathematician and philosopher Cournot (1843),4 who holds that small-probability events should be taken to be physically impossible. He also held that this principle (Fréchet named it Cournots principle) is the one that connects probabilistic theories to the real world. It is taken, as in Bernoulli, as showing that we should identify probability with relative frequency in a large number of (independent? identically distributed?) trials. This mantra was repeated in the twentieth century by very distinguished probability theorists, including Émile Borel, Paul Lévy, Andrey Markov, and Andrey Kolmogorov. We cannot help but wonder whether this was to some extent a strategy for brushing off philosophical interpretational problems, rather than a serious attempt to confront them.
Cournots principle, taken literally, is absurd. Throw a dart at a target. The chance that it hits any specific point is very small. Then are we supposed to conclude that for any point,
67 it is physically impossible that the dart hits that point? Later statements tend to try to get around this by modifying the principle as saying that an event of very small probability, singled out in advance, is physically impossible. So you have to pick out a point in advance. Why should picking it out in advance make it physically impossible?

By now, no theorist would be fooled by Bernoullis swindle. But it seems to have an afterlife in popular thinking: Drug testing. a drug company runs randomized trials on a new drug. .... To those who do not understand statistics, this is an invitation to Bernoullis swindle. It is morally impossible to get this value if the drug is ineffective. ... The untutored think that they are getting the probability of effectiveness given the data, while they are being given conditional probabilities going in the opposite direction.

68 Hard core frequentism

Venn: The Victorian frequentistsJohn Stuart Mill, Leslie ellis, John Vennclaimed more than this. They claimed that relative frequency was the primary sense of probability and sometimes maintained that alternative senses were not important at all. ... This was, to some extent, the english empiricists against the Continental Rationalists.

70 What number (of trials) is enough? evidently only an infinite number of tosses will do, and this is the conclusion that Venn settles on.

The move to limiting relative frequency raises some mathematical issues: existence of limits,

The next question is what sort of events go into the sequence. Suppose that there are lots of dice, some symmetric and some oddly shaped, and lots of ways of throwing them. Should we just take any sequence of throws and use this to define probabilities? evidently not. ...

So, Venn finds himself pushed to a hypothetical relative frequency view. The probability
71 of this die coming up a 6 is the limiting relative frequency that it would have had if it were
thrown an infinite number of times with no change in die, the tossing, or any of the relevant circumstances. The material view of logic has been forced to define probability as a counterfactual! And the concept of relevant circumstances still appears to be a little vague.

In LLN There is the probability on a single trial and the assumption that it doesnt change from trial to trial; there is the statement of independence of trials; there is the probability that relative frequency is equal to the single trial probability. None of these probabilities are legitimate, according to Venns frequentism. He cannot state the law of large numbers!

72 Venns limiting relative frequencies are not of this world either. They live in a hypothetical world of infinite sequences of similar trials. What connection, then, does this hypothetical world have to the actual world?

There is another question that the frequency view leaves hanging. That is how probability can beas Cicero saida guide to life. ... His answer is roughly that as degrees of belief in a single event, we should take the corresponding relative frequency in a series of like events: ... How can we define like events without circularity?

73 Venn leaves us with many questions for the view that probability is frequency. Some are about the mathematical structure of the theory. Some are about the metaphysics of the theory, which is ultimately a theory of counterfactual or fictional series. Many are about the connection of the theory with real frequencies and real applications to decision.

von Mises: Hilbert's Problem 6 was to give a mathematical treatment of the axioms of
74 physics. Special emphasis was given to the role of probability in statistical physics: .... Von
Mises was not responding to Venn but to Hilbert. mind. Von Mises interpreted probability as relative frequency in a certain kind of infinite sequence, an idealization of the kind found in statistical physics. This sort of a sequencecalled a Kollektivwas to exhibit the intuitive ideas of local disorder and global order that had been emphasized by Venn.

Von Mises did not simply assume, like Venn, that limiting relative frequencies would exist. It needed to be postulated as a defining attribute of a Kollektiv. Local disorder was captured by a requirement of randomness.

What is randomness? Any reasonable place-selecting function cannot change frequen-
75 cies..... Von Mises really had in mind a more restrictive, intensional notion of a function, but he did not have a means of making it precise.

Then came Church.

76 That would seem to be the end of the story, except for a complication noted by Jean Ville in 1939. Invariance of relative frequencies under computable place selection is satisfied by some sequences that do not have all the properties that we would want of a truly random sequence. In particular, Ville showed that relative to any countable class of place selection functions, there is a Kollektiv in which relative frequency of H is 1/2, but for all but a finite number of initial segments, the relative frequency is not less than 1/2. The limiting relative frequency is approached from above. This leaves the very interesting question of a satisfactory definition of a random sequence still open. We will return to this later, in chapter 8.

77 Idealization: The relation of mathematical scientific theories to the world is a deep question, and it is not obvious that it has a univocal answer. It might have different answers in different applications. .... In any case, is there a methodology for connecting the mathematics to reality, or must the discussion remain at the informal level? This is a question that alternative views also need to address. ....
Frequency evidence could still be brought to bear on the validity of a chance model. This is what is now called a propensity view of probabilities. This seems to be the view of Frchet, who views chances as physical quantities with objective existence in the world. The proper mathematical framework within which to formulate chance models would then be the measure-theoretic framework that is the topic of chapter 5 rather than the apparatus of Kollektivs of von Mises.

rmp78Frequentism radically restricts the range of probability theory, and it must be judged inadequate as a general account of probability. But a deep question has been uncovered in the development of the frequentist view. What is the nature of a random sequence?

5. Mathematics Photo Kolmogorov

79 Our fifth great idea is the formulation of probability theory as part of mathematics within the modern theory of measure and integral. It was completed by Andrei Kolmogorov in a monograph published in 1933.

The connections between mathematics and probability (or indeed any area of applied science) is a lively, controversial subject even today.

83 Assigning length to such esoteric sets was possible because of the work of Borel, Lebesgue, and others. It turns out that not every set can be assigned a length (see our appendix.) At the end, Borel and his coworkers had a rigorous mathematical model with real similarities to coin tossing, with which they could use new mathematics to compute answers to interesting questions. There was more to do.

Consider the following recollection by a great probabilist, Mark Kac:
When I came to Lwow as a student in 1931, I had never heard of probability theory. ... the subject did not exist.... It must have been in 1933 or 1934 that I came across A. A. Markovs Wahrscheinlichkeitsrechnung. This book, a 1912 translation of the original Russian, made
84 a tremendous impression on me.... the technicalities of the moment a problem and related analytic subtleties difficult to absorb. ... What I could not make out was what were the random quantities .... He records his amazement on reading Kolmogorovs unification. Kac wasnt alone in wondering how to mathematize probability.

HILBERTS SIXTH PROBLEM
The great mathematician David Hilbert set out a list of 23 problems at the start of the twentieth century. His sixth problem asks (David Hilbert, Mathematical Problems, Bulletin of the American Mathematical Society 8, no. 10 (1902): 43779): The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part;

85 His 1933 monograph, Grundbegriffe der Wahrscheinlichkeitsrechnung (Foundations of the Theory of Probability), does three main things:

1. He puts the theory of probability on a clear mathematical foundation.
2. He properly formalizes conditional probability.
3. He proves the Kolmogorov extension theorem.

In addition, the monograph is peppered with smaller beautiful gems (Kolmogorovs 01 law) and presents a unified treatment of his earlier work on limit laws in probability for general random variables. .... It is remarkable that Kolmogorovs formulation is still the absolute standard for all flavors of probability: subjective, objective, and everything in between.
86 The treatment is abstract and axiomatic. This is not a book about interpretation; it is about the mathematical structure of probability theory.*
87 In this framework, it is easy to give a precise definition of a real-valued random variable, the notion that Kac found mysterious in early expositions of probability theory. It is a measurable function from the basic set to the real numbers.

CONDITIONAL PROBABILITY AS A RANDOM VARIABLE

Conditional probability is now generalized to conditional probability as a special kind of random variable. P(heart ‖ sex), as the random variable that takes the first value for men and the second value for women.

88 Conditional expectation as a random variable is defined in the same way.

89 Kolmogorov extension theorem allows us to define infinite dimensional stochastic processes.

Kolmogorov used his generalized concept of conditional probability to give a rigorous definition of Markov processes. And he proved the definitive form of the strong law of large numbers for general random variables: the average of a sequence of independent random variables converges if and only if it has finite mean values.

92 Let us say that A has a density if this ratio tends towards the limit. Using density as a careful version of probability in these settings runs into problems:
- Not all sets have a density.
- Density is not countably additive.

93 A: Set Without Density Integers beginning with 1. This frequency oscillates forever. B This can be avoided with clever weighting p94.

96 This philosophical attitude toward the infinite is evident both in the Grundbegrigfe and in Kolmogorov's later writings. .... He writes The notion of an elementary event is an artificial superstructure imposed on the concrete notion of an event. In reality, events are not composed of elementary events, but elementary events originate in the dismemberment of composite events.
... The modern theory of measure and integral is viewed pragmatically, not as metaphysics but rather as an idealization that allows the infinite to approximate the finite.

100 **6. Inverse inference** photo T Bayes

101 Bayes states his own great idea thus:
Given the number of times in which an unknown event has happened and failed:
Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.
This is how he begins his 1763 Essay towards Solving a Problem in the Doctrine of Chances. His predecessors, up to and including Bernoulli and de Moivre/ had reasoned from chances to frequencies. Bayes gave a mathematical foundation for statistical inference-inference from frequencies to chances.

This is a program to find the predictive probability, the probability that something will happen next time, from the past statistics.

Bayes vs Hume
Bayes appears to have been motivated not by practical concerns of law or medicine, but
102 rather by questions of mathematical philosophy. BAYES VERSUS HUME Price emphasizes
the philosophical magnitude of the step-its importance for inductive reasoning:

> Every judicious person will be sensible that the problem now mentioned is by no means a curious speculation in the doctrine of chances, but necessary to be solved in order to a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter.

Put this way, the project can be seen as an answer to Hume. In the Enquiry Concerning Human Understanding (1748), Hume writes

> Though there be no such thing as Chance in the world; our ignorance of the real cause of any event has the same influence on the understanding, and begets a like species of belief or opinion.
>
> . . . Being determined by custom to transfer the past to the future, in all our inferences; where the past has been entirely regular and uniform, we expect the event with the greatest assurance, and leave no room for any contrary supposition.
>
> . . . Let any one try to account for this operation of the mind upon any of the
> 103 received systems of philosophy, and he will be sensible of the difficulty. For my part ,
> I shall think it sufficient, if the present hints excite the curiosity of philosophers , and make them sensible how defective all common theories are in treating of such curious and such sublime subjects. (Section VI)

There is a very good case made in S. L. Zabell ("Laplace's Rule of Succession," Erkenntnis 31 (1989): 283-321.) that Bayes achieved his main results shortly after Hume issued this challenge.... It is natural for Bayes to be seen as an answer to Hume.

In Price's appendix

> Let us imagine to ourselves the case of a person just brought forth to life into this world and left to collect from his observation of the order and course of events what powers and causes take place in it. The Sun would , probably, be the first object that would engage his attention; but after losing it the first night he would be entirely ignorant whether he should ever see it again.

Price goes on to show that after a million observations, the chance of the sun's rising lies in a small interval close to 1 with high probability. It is clear from this (and from the preceding passage that explicitly mentions uniformity of nature) that Price takes Bayes' essay to be the answer to Hume.

105 But Hume, although a great philosopher, was not a good mathematician, and it is not likely
that he understood Bayes' contribution.

BAYES ON PROBABILITY
The essay begins with a development of probability, with some remarkable anticipations of the modern coherence views that we met in chapter 2:

> The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening.

to p107 outlining Bayes introduction.

108 Bayes already had the idea of coherence to characterize probability and conditional probability.

110 Bayes' work was not immediately taken up in England, but in France by Laplace.

111 Laplace calculated the probability of a success on the next trial given m successes in n trials:

$$\frac{\int_0^1 dx \binom{n}{m} x^{m+1}(1-x)^{n-m}}{\int_0^1 dx \binom{n}{m} x^{m}(1-x)^{n-m}} = \frac{m+1}{n+2}. \tag{0.0.2}$$

This is Laplace's rule of succession. Notice that for large numbers of trials an application of Laplace's rule is very close to simply taking the relative frequency of heads as one's probability for heads the next time. In this setting, with a lot of data, naive frequentism does not go far wrong.

GENERALIZED LAPLACE Suppose a flat prior is not appropriate, say the following beta distribution

$$\propto x^{\alpha-1}(1-x)^{\beta-1} \tag{0.0.3}$$

In this case, the resulting rule of succession reads

$$\frac{m+\alpha}{n+\alpha+\beta}. \tag{0.0.4}$$

114 Arguably, you can get anything you might reasonably want to represent your prior mixture of knowledge and ignorance. ... But if you were going to risk a lot on the next few trials, it would be prudent for you to devote some thought to putting whatever you know into your prior.

CONSISTENCY Suppose you adopt the point of view that chances are physical properties in the world-there is a true bias of the coin-and that you start with prior beliefs about the chances and feed in frequency evidence to update your degrees of belief. Will you always (or with high probability) learn the true chances? The answer can be no if you start with a pigheaded prior. Suppose you put prior probability one on the coin being biased toward heads. (The integral over the possible biases towards tails is zero.) Then you can never learn that the coin is really biased toward tails.

Let us say that your prior is consistent if no matter what the true single-case chance is, you will learn it with chance 1.

115 Basically, all the possibility is in the support of the density, it is consistent.

Why most published research findings are false

Something like this was predicted by John Ioannidis in a paper entitled, "Why Most Pub-
116 lished Research Findings are False." ....Medical research articles should not be interpreted

only on p-values.

117 There is now a movement to go beyond mechanical use of p-values, convenient as they may be. The proper guidance is Bayesian.

BAYES, BERNOULLI, AND FREQUENCY From Bayes' point of view, the conclusion of Bernoulli's swindle doesn't look too bad. Given a reasonable prior and lots of data from independent identically distributed trials, it is not unreasonable to infer that the chance is close to the frequency. But a reasonable conclusion doesn't make an argument valid.

BAYES CHANGED THE WORLD Bayes had a philosophical idea that changed the world. We have a prior probability over the chances. We get data and update using Bayes theorem to get a posterior. We put everything we know before getting the data into the prior over what we don't know. Then we feed in data and update.

118 The German naval enigma codes were deciphered by a group led by Alan Turing,. ... The evolution of western civilization might have taken a different course were it not for this work. Bayes' idea really did change the world.

121 There is a useful literature on Bayesian robustness.

122 **7. Unification** Photo de Finetti

But what, after all, is chance? We know how to deal with it computationally , but we don't seem to know what it is.

Bruno de Finetti tells us not to worry: Chance does not exist.

Yet, it is perfectly legitimate to carry on just as if it did! How is this possible? This is de Finetti's great idea, and the story of this chapter.

That probability as a physical quantity-physical propensity, objective chance-does not exist is a position that has been taken by many philosophers, most notably David Hume. But de Finetti did more than just take this philosophical stance. He showed that, in a certain
123 precise sense, that if we dispense with objective chance nothing is lost.

The mathematics of inductive reasoning remains exactly the same.

To say that order doesn't matter or to say that the only thing that does matter is frequency are two ways of saying exactly the same thing. This property is called exchangeability by de Finetti.

De Finetti proved the converse. Suppose your degrees of belief are exchangeable. Call

124 an infinite sequence of trials exchangeable if all of its finite initial segments are. De Finetti proved that every such exchangeable sequence can be gotten in just this way. It is just as if you had independence in the chances and uncertainty about the bias. It is just as if you were Thomas Bayes.

What the prior over the bias would be in Bayes is determined in the representation. Call this the imputed prior probability over chances the de Finetti prior. If your degrees of belief about outcome sequences have a particular symmetry, exchangeability, they behave just as if they are gotten from a chance model of coin flipping with an unknown bias and with the de Finetti prior over the bias. So it is perfectly legitimate to use Bayes' mathematics even if we believe that chance does not exist, as long as our degrees of belief are exchangeable.

De Finetti's theorem helps dispel the mystery of where the prior belief over the chances comes from. From exchangeable degrees of belief, de Finetti recovers both the chance statistical model of coin flipping and the Bayesian prior probability over the chances.

De Finetti has replaced them with a symmetry condition on degrees of belief. This is, we think you will agree, a philosophically sensational result.

READING DE FINETTI

Suppose you are interested enough to want to go to the source. We recommend chapter 9 of de Finetti's Probability, Induction, and Statistics as the best stand-alone account of his ideas and his program.

127 The appendix to this chapter gives a careful statement (with an error term) for this finite version of the theorem.

What does this show? If your judgment of symmetry—that probabilities don't depend on order of trials—doesn't itself depend on the number of trials, then you essentially have de Finetti's theorem.

131 MARKOV CHAINS

So far, so good, but what if your degrees of belief are not exchangeable?

Already in 1938, de Finetti suggested that exchangeability needed to be extended to a more general notion of partial exchangeability. The idea was that where full exchangeability fails, we might still have some version of conditional exchangeability: Markov exchangeability.

We are considering a loosening of the assumption of patternlessness in the data stream to one where the simplest types of patterns can occur.

132 A stochastic process is Markov exchangeable if sequences of the same length having the same transition counts and the same initial state are equiprobable. David Freedman showed that any stationary Markov exchangeable process is representable as a mixture of stationary Markov chains.

For another form of de Finetti's theorem for Markov chains, we need the notion of a recurrent stochastic process. A state of a stochastic process is called recurrent if the probability that it is visited an infinite number of times is 1. The process is recurrent if all its states are recurrent.

Diaconis and Freedman show that a recurrent Markov exchangeable stochastic process has a unique representation as a mixture of Markov chains.

MORE PARTIAL EXCHANGEABILITY

134 APPENDIX 1. ERGODIC THEORY AS A GENERALIZATION OF DE FINETTI

The 1931 ergodic theorem of George Birkhoff is a far-reaching generalization of de Finetti's theorem.

What is a symmetry? It is a feature that is invariant under a group of transformations.

135 We see now de Finetti's point of view has a far-reaching generalization to probabilistic symmetries. The ergodic measures can be thought of as surrogates for the physical chance hypotheses.

8 Algorithmic randomness Photo P Martin-Löf

146 Computer generated random numbers: there is a long history of failures.

148 For physical random generators, there are no theoretical guarantees.

. One hopeful development is the use of the logic of complexity theory, as in the scheme of Blum and Michali. These authors offer a generator with the following property: if it fails any polynomial time test, then there is an explicit way to factor that is much faster than any known method. (Thus, if factoring is hard, our numbers are secure. But, of course, if factoring can be done efficiently, all bets are off.) This is close in spirit to the algorithmic complexity accounts treated subsequently in this chapter.

We conclude this practical section with some practical advice on the use of random number generators:

Use at least two generators and compare results. (We recommend Mersenne Twister and one of the generators offered in Numerical Recipes.)

Put in a problem with a theoretically known answer to run along with the other simulations.

Think of using random number generators like driving a car. Done with care, it is relatively safe and useful.

150 ALGORITHMIC RANDOMNESS

In a deeper sense, Church's idea of using computability to define randomness was correct.

151 The source of the problem is not the idea of using computability. Rather, von Mises' idea

of defining randomness by means of place selection functions alone appears to be defective. Church-von Mises randomness requires random sequences to pass only one sort of test for randomness. Sequences can pass this test and fail others. We want random sequences to pass all tests of randomness, with tests being computationally implemented. Per Martin-Löf found how to do this in 1966.

COMPUTABILITY
If you know how to program in any computer language, you already know what computability is. ....

157 RANDOMNESS
Since Church-von Mises randomness proved inadequate, Martin-Löf took a different approach. One could define random sequences by throwing out "atypical" classes-null classes-that were given probability O by a model of flipping a fair coin. The immediate problem is that each individual sequence has probability 0. Here computability comes to the rescue. We throw out only null sets that can be identified by a Turing machine. Since there are a countable number of Turing machines, throwing out all these null sets leaves a set of "typical" random sequences with probability measure one. (This again can be regarded as a way of avoiding the absurdities, discussed in chapter 4, of a literal application of Cournot's principle.)

The nonrandom sequences can be thought of as those failing more and more stringent statistical tests.

159 COMPUTABLE MARTINGALES
The idea of a gambling system for a sequence is made precise in the notion of a martingale.
160 A martingale is a function, CAP, from initial segments to nonnegative reals, such that CAP(s)= [C AP(s followed by 0) +CAP (s followed by 1)] since the odds are fair. A martingale succeeds on a sequence if CAP goes to infinity in the limit.
The martingale is required to be computably enumerable. Then one can define a random sequence in terms of impossibility of a gambling system. A sequence is random if no computably enumerable sequence succeeds in it. In 1971 Schnorr 19 proved that a sequence is random in this sense if and only if it is Martin-Löf random.

KOLMOGOROV COMPLEXITY
In the 1960s Kolmogorov returned to the foundations of probability from a new standpoint.
]
163 Schnorr also proved Kolmogorov random sequence $\equiv$ Martin-Löf.

If randomness is based on computation, then variations on randomness can be based on different computational notions. The theory can be pushed into higher realms of abstraction by equipping a Turing machine with an oracle. .... One can then consider a super-oracle that decides the halting problem for these machines, and so forth, creating a hierarchy of computationally more powerful machines. This leads to a hierarchy of more and more stringent criteria for a random sequences, with Martin-Lof randomness being 1-randomness, the same notion with a Turing machine equipped with an oracle deciding the halting problem being 2-randomness, and so on. The class of N + 1 random sequences is strictly contained in the class of N-random sequences.

In the other direction, weaker kinds of randomness can be gotten by limiting the kind of computation used to test for randomness.

164 P-randomness: using polynomial computability.

9. Physical chance photo Democritus

168 At the end of his discussion, Zermelo concluded that either the second law-as originally formulated-must be given up or the kinetic theory of gases-in its contemporary form-must be given up. Poincare was already an opponent of the kinetic theory and used this as an argument against it. Boltzmann modified his view of the second law. In his replies to Loschmidt and to Zermelo, he says that the second law has only a statistical character.

171 Suppose we have 1000 fleas. Then in the Gibbs picture, the equilibrium distribution piles up close to half on each dog. Once at equilibrium, the ensemble does not change. In the Boltzmann picture , the system spends most of its time close to half on each dog, but fluctuates. And if one waits long enough, it will visit the state where all fleas are on one dog. Both points of view are correct; they are just looking at different things. Everything here is clear, but in what sense is a gas like an Ehrenfest urn?

PROBABILITY, FREQUENCY, AND ERGODICITY

What is the probability at issue, and where does it come from? Boltzmann appears (for the most part) to be a frequentist.

172 This requires ergodicity.

180 Does quantum mechanics need a different notion of probability? We think not. As long as we confine ourselves to the level of experiments and results and eschew extra metaphysics the issues are basically the same as in the classical setting.

184 What caused the trouble was not classical probability, but the assumption of locality.

186 Classical ergodicity implies quantum ergodicity.

34. To even state it correctly requires more machinery than we have in this book. For a review article, see S. Nonnenmacher, "Anatomy of Quantum Chaotic Eigenstates," Seminaire Poincare XIV (2010): 177-220. 35. Thus, the stadium, being ergodic, is quantum ergodic. But the ghost of the exceptional "bouncing ball" orbits remains in the "scarring" visible in some eigenstates. That this scarring is

genuine is proved by A. Hassell. See Hassell's introduction to the subject, "What is Quantum Unique Ergodicity?" Australian Mathematical Society Technical Paper.

10. **Induction** photo D Hume

191 In retrospect, most of what preceded this chapter can be seen as a story of brilliant attempts to come to terms with the problem of induction raised by Hume.

Hume

The classic modern statement of inductive skepticism comes from David Hume, although he
192 reminds us of its ancient sources. How, asks Hume, could we justify inductive reasoning?
Inductive reasoning is not justified by relations of ideas—by mathematical deduction:

> That the sun will not rise tomorrow is no less intelligible a proposition than that it will rise. We should in vain, therefore, attempt to demonstrate its falsehood.

Think of the world as like a movie. You can take one movie, cut it, and splice it to any entirely different one. We could be at the splice. The scenario is not inconsistent. Pure mathematics cannot justify induction. But to try to justify inductive reasoning by inductive reasoning is to beg the question:

> It is impossible, therefore, that any arguments from experience can prove this resemblance of the past to the future, since all these arguments are founded on the supposition of that resemblance.

Then, by Hume's lights, there is no avenue left. This remarkably simple argument has occupied philosophers from his time to the present.

KANT

Kant took Hume seriously and set out to solve the problem of induction. Some philosophers believe that he succeeded. He says that he has. But it is not so easy...

Hume emphasizes that our innate psychology disposes us to causal and inductive formation of expectations (like Russell's chicken). With Kant, this Humean psychology first becomes codified in a priori rules and then somehow transmuted into knowledge of a new kind—the synthetic a priori. Sorry, but we just do not see it. With apologies to Kantians, we agree with the philosopher C. D. Broad:

> There is a skeleton in the cupboard of Inductive Logic, which Bacon never suspected and Hume first exposed to view. Kant conducted the most elaborate funeral in history, and called Heaven and Earth and the Noumena under the Earth to witness that the skeleton was finally disposed of. But, when the dust of the funeral procession had subsided and the last strains of the Transcendental Organ had died away, the coffin was found to be empty and the skeleton in its old place.7

POPPER

One might simply read Hume and give up. This, in fact, is the position taken by Sir Karl

Popper. ... “My own view is that the various difficulties of inductive logic here sketched are insurmountable.”9

195 But it is possible, and sometimes quite reasonable, to be skeptical about some things but not others. Thus, there are grades of inductive skepticism, which differ in what the skeptic calls into question and what he is willing to accept..... We have already been engaged in a discussion of inductive skepticism throughout this book. This discussion began almost contemporaneously with the advent of serious probability theory. Here we review what we have learned from this perspective.

BAYES-LAPLACE

In chapter 2 we saw that Jacob Bernoulli thought that he had solved the problem of inferring chances from data with his law of large numbers:

> [W]hat you cannot deduce a priori, you can at least deduce a posteriori-i.e. you will be able to make a deduction from the many observed outcomes of similar events. For it may be presumed that every single thing is able to happen and not to happen in as many cases as it was previously observed to have happened or not to have happened in like circumstances. 12, 13

Bernoulli had proved that with enough trials it would become “morally certain” that the frequency would be approximately equal to the true chance. If x is approximately equal to y, then y is approximately equal to x. So, after a large number of trials, we can take the true chances to be approximately equal to the observed frequencies.

This informal argument gains an air of plausibility by concealing difficulties behind a cloak of moral certainty and approximate equality— Bernoulli’s swindle.

198 Given their assumptions, Bayes and Laplace show that Bernoulli’s conclusion was right. We can infer the approximate chances a posteriori. And, in this setting, they do give an answer to Hume. They show when, and in what sense, it is rational to believe that the future is like the past. They explain the proper use of past frequencies in calculating “weights” for future events.

But there are assumptions. We can be a little more skeptical and question these assumptions.

199 WHAT ABOUT THE QUANTIFICATION OF IGNORANCE?

Hume has been answered based on a certain kind of model and a certain quantification of ignorance. A more radical skeptic might certainly question both assumptions.

Ignorance is the opposite of knowledge. ... Specification of an ignorance prior is not unique. There are lots of them.

What Laplace showed for the uniform prior holds for all skeptical priors for the biased coin. ... What about the dogmatist (with a peaked prior)? Logic alone does not prevent one from being a dogmatist.

200 There is another big assumption in the Bayes-Laplace model.

WHAT ABOUT THE EXISTENCE OF CHANCES?

The foregoing all takes place within a specific chance model. Perhaps, with Hume, we may believe that "there is no such thing as chance in the world."27 We saw the answer to this enhanced skepticism in chapter 7. It was given by a great admirer of Hume, Bruno de Finetti. .... De Finetti, like Hume, believes that there is no such thing as chance in the world and shows that we can have the virtues of Bayes' analysis without the baggage. If you are skeptical about the existence of chances, the chance model, and the prior over the chances, de Finetti shows how to get them all from your degrees of belief, provided that they satisfy the foregoing condition of exchangeability. Furthermore, you must believe with probability 1 that a limiting relative frequency exists and that with repeated experience you will converge to it.29 If your degrees of belief are exchangeable, you cannot be an inductive
201 skeptic. Your degrees of belief may not be exchangeable. There is no reason that they have to be. What then?

WHAT IF YOU DON'T HAVE EXCHANGEABLE DEGREES OF BELIEF?

Short of exchangeability there may be other symmetries in degree of belief, and such symmetries generally have inductive consequences. A slight weakening to allow for order effects gives Markov exchangeability.

Suppose that you have a space that encapsulates the problem that you are thinking about. You bring to this problem your degrees of belief: a probability measure on that space. Suppose that your degrees of belief have a symmetry-that they are invariant under a group of transformations of the space into itself. For example, the symmetry might be exchangeability-order doesn't matter. Then the transformations that swap sequences with the same frequencies but different orders in initial segments will leave your degrees of belief unchanged. Exchangeability is invariance of your probability under this group of transformations. The transformations represent your conception of a repetition of an experiment. 32 Invariance means that the transformation (or group of transformations) leaves the probabilistic structure unchanged. (We emphasize the subjective nature of the symmetry. You and I may have different conceptions of the repetition of an experiment, since we may have different symmetries in our degrees of belief)...

202 Notice that, in this sense, you cannot be an inductive skeptic. You cannot be a skeptic in the sense of Reichenbach. Such a skeptic doubts the existence of limiting relative frequencies. But you cannot doubt this, provided you are considering a sequence of repetitions of the same experiment, in your sense of same experiment. This is a consequence of the symmetry in your degrees of belief.

As emphasized at the onset, your probabilities and your conception of repetition of the same experiment are up to you. You and I may differ. We may be skeptical about each other but not about ourselves.

WHAT ABOUT THE PREDICATES USED TO DESCRIBE THE WORLD?
In Fact, Fiction and Forecast 36 (1955), the philosopher Nelson Goodman concocts the "grue" hypothesis to show the hopelessness of purely syntactical theories of confirmation. ...

Goodman concludes that regularities in terms of some kinds of predicates are projectible into the future, but regularities in terms of others are not.

204 Goodman's concerns have already been addressed by de Finetti.37 Projectibility is captured in a subjective Bayesian setting by exchangeability or generalizations thereof.
[**C**] In any case some kind of uniformity assumption is required.

WHAT ABOUT UNCERTAIN EVIDENCE?
So far, the envisioned learning experiences have been modeled as conditioning on the evidence, which comes nicely packaged as a proposition. 38 A more radical skeptic may well call even this into question. This is the stance taken in Richard Jeffrey's "Radical Probabilism." 39 Must a radical probabilist perforce be a radical inductive skeptic?...

206 SUMMING UP
Hume remarks that it is psychologically impossible to be a consistent skeptic:

> since reason is incapable of dispelling these clouds, nature herself suffices to that purpose ....

This is, in broad outline, psychologically correct, even though—as we saw in chapter 2—
207 human psychology may exhibit systematic quirks. But, as Hume maintained and Russell emphasized, it is nevertheless logically possible to be a consistent skeptic. Even when we have tried to assume as little as possible, we have still assumed something. One is not logically compelled to believe that one will face a sequence of learning experiments. One may not be coherent or believe that one will remain coherent in the future. One need not believe that there will be a future. Absolute skepticism is unanswerable.

But short of absolute skepticism, there are various grades of inductive skepticism, differing in what the skeptic brings to the table and what he calls into doubt. Some kinds of skeptics may call into question things to which they are implicitly committed. In such a case, reason is capable of dispelling doubts.

In one setting, Bayes, Laplace, and their followers solved the problem of induction. With fewer assumptions, in a setting more congenial to Hume, de Finetti and his heirs solved Hume's problem. It is remarkable the extent to which the logic of coherent belief itself constrains inductive skepticism.