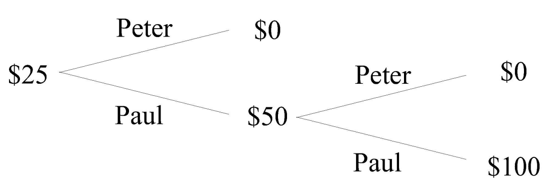
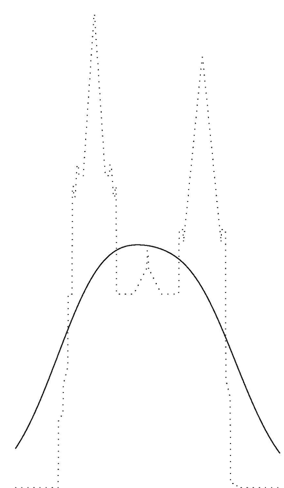
“That’s what all the old guys said”: The many faces of Cournot’s principle

Glenn Shafer
Rutgers University
gshafer@business.rutgers.edu
www.glennshafer.com





The Game-Theoretic Probability and Finance Project

Working Paper #60

First posted January 17, 2022. Last revised January 27, 2022.

Project web site:
http://www.probabilityandfinance.com

Abstract

Cournot's principle is a family of theses about how a system of numerical probabilities can be given an objective interpretation. With one nuance or another, these theses connect a system of numerical probabilities with phenomena by asserting that certain events of high probability will happen and that certain events of low probability will not.

This paper surveys the many ways Cournot's principle has been formulated by quoting scores of prominent authors over several centuries. I include some authors who disagreed with Cournot's principle, sometimes explicitly, and others whose formulations can be thought of either as versions of the principle or as alternatives to it.

The purpose of this compilation is not to decide whether Cournot's principle is right or wrong or to formulate it "correctly" but to show its persistence and its role in the development of the idea of objective numerical probability. The continuity of this development may be greater and the novelty of recent contributions less than we sometimes suppose.

This is an incomplete working paper, with a thousand loose ends. In particular, the compilation of quotations in §3 is incomplete. Additional authors should be listed, quotations from some already listed should be added, and some repetition should be removed.

1 Introduction

Describing a phenomenon with mathematical probability usually involves assigning high mathematical probability to certain events or statements. Physicists tell us that a gas in a confined space will distribute itself uniformly with high probability. Researchers establish the value of a medical treatment by showing that it is more successful than another with high probability. Here high probability is interpreted as practical certainty.

Scholars equated high probability with practical certainty long before they expressed probabilities as numbers. When Jacob Bernoulli undertook to make Christian Huygens's calculus of expectations in games of pure chance into a general mathematical theory of probability and argument, in his posthumously published *Ars conjectandi* (1713, [14]), his readers saw nothing novel in his statement that high probability provides practical certainty. The only novelty was that in Bernoulli's picture, probabilities had become numbers.

Does the use of probability theory to describe phenomena always require that we equate high probability with practical certainty or, equivalently, low probability with practical impossibility? In 1843, Antoine Augustin Cournot asserted that it does. Because he seems to have been the first to make this claim, some authors later called the equation of high probability with practical certainty *Cournot's principle*. This remains a convenient name, but the principle has been formulated in so many different ways and under so many different names that it is best to think of it as a family of ideas rather than a precise assertion.

Around the beginning of the 21st century, when my wife and I were spending our summers in Vermont's Northeast Kingdom, I would occasionally drop by Hanover, New Hampshire, to visit J. Laurie Snell (1925–2011), then retired but still active at Dartmouth College; we would talk about martingales and the philosophy of probability. On one occasion, I tried to interest Laurie in Émile Borel's statement of Cournot's principle. He expressed his disinterest with a wave of his hand and a single sentence: "All the old guys said that." Why did they all say that? Why were we saying it less often in the 21st century? Did we have a way of evading it that the old guys had overlooked? And what exactly were they saying? This paper reports on my effort to answer these questions over the past 20 years.

In the next section of this paper, §2, I survey, from a contemporary viewpoint, the different ways Cournot's principle has been expressed and used. In §3, I document the development of the principle chronologically, quoting when possible, at least in translation, the words different authors used to express it, advocate it, and oppose it.

2 Understanding Cournot's principle

You can equate high mathematical probability with practical certainty even if you see mathematical probability as nothing more than degree of belief; see §3.18

for an example of a contemporary of Cournot's who did this. But Cournot himself wanted to connect mathematical probability with phenomena,not merely with his belief about phenomena. He wanted to interpret probabilities as statements about the world as we experience it, statements that can be confirmed or at least tested by observation. These statements are often statements about frequencies. So we begin, in §2.1, by reviewing the role of Cournot's principle in connecting probabilities with frequencies.

In §2.2, we turn to the problem of qualifying Cournot's principle by limiting the events to which it applies. Any detailed system of probabilities will give small probability to events that actually happen; this has been called the "lottery paradox". So only some events of small probability can be ruled out. Different formulations of the principle specify these in different ways.

How high a probability is needed for practical certainty, how small for practical impossibility? Émile Borel taught that this depends on the context, and various answers have been used by statisticians (§2.3), physicists (§2.5), and pure mathematicians (§2.6).

We conclude in §2.7 with a look at the game-theoretic version of Cournot's principle.

2.1 Resolving the conundrum of Bernoulli's theorem

In his *Logic of Chance*, first published in 1866 [159], the British logician John Venn argued that mathematical probability should be understood simply as frequency—frequency in some reference class. Venn's view has been elaborated in various ways, and it has continued to attract adherents.

Mathematicians who teach and use probability theory encounter a problem if they take Venn's idea seriously. How can they square the idea with Bernoulli's theorem and its many generalizations, which constitute the crown jewel of probability theory as they understand it?

Bernoulli's theorem says that if an event always happens with probability p, the frequency with which it happens in n independent trials will be close to p, as close as you like with probability as high as you like if n is large enough.[1] We can say this with mathematical precision using a few more symbols. Write f_n for the frequency of the event in n independent trials (the number of times it happens divided by n), and choose a small positive number ϵ and a number P close but not equal to 1. Then the theorem says that if n is large enough, the probability of the event $|f_n - p| \leq \epsilon$ is at least P. Yet more precisely: there exists an integer N such that for any integer $n \geq N$, the probability of $|f_n - p| \leq \epsilon$ is at least P. With yet more notation:

$$\mathbb{P}\left(|f_n - p| \leq \epsilon\right) \geq \mathrm{P}. \tag{1}$$

In a sense, (1) identifies the probability p with the frequency f_n. But it does so only approximately and with high probability, a probability greater than P.

[1] This is an abstract, generalized, and somewhat off-center version of the final theorem in Bernoulli's *Ars conjectandi*. See [153] for a careful account of what Bernoulli actually asserted and proved.

If we say that f_n is practically certain to be approximately equal to p, we are using Cournot's principle.

Even this concession to Cournot's principle does not completely save the identification of probability with frequency, because aside from p there is another probability in (1), the probability $\mathbb{P}(|f_n - p| \leq \epsilon)$. Is it also practically certain to be approximately equal to a frequency? We can say yes only if we assume that the whole experiment of n trials is itself repeated many times (all mutually independent) and then apply Bernoulli's theorem and Cournot's principle again. As noted by R. A. Fisher in 1958 [71] and Jacques Bonitzer in 1984 [18, pp. 37–38], there looms here an infinite regress.[2]

The identification of probability with frequency is further weakened, and the role of Cournot's principle is further highlighted, when we consider the generalizations of (1) that are important in mathematical statistics. One direction of generalization arises when the probability p is not constant. This direction was pursued in the mid-19th century by Siméon-Denis Poisson, Jules-Irénée Bienaymé, and Pafnuti Chebyshev. Using results published by Chebyshev in 1846 [40], we can generalize (1) in various ways to the case of a sequence of independent events $E_1, E_2, \ldots$ with unequal probabilities $p_1, p_2, \ldots$. We can say, for example, that for given ϵ and P,

$$\mathbb{P}\left(\left|f_n - \frac{\sum_{i=1}^n p_i}{n}\right| \leq \epsilon\right) \geq \mathrm{P} \tag{2}$$

when $\sum_{i=1}^n p_i(1-p_i)$ is large enough. Here we are identifying a frequency not with a single probability but with an average probability.

In the 20th century, Chebyshev's generalization of the law of large numbers was further generalized to the case of events $E_1, E_2, \ldots$ that are not necessarily independent, so that the probability of E_i after $E_1, \ldots, E_{i-1}$ have been settled depends on which of these earlier events have happened. Using results published by Paul Lévy in 1936 [109], for example, we may generalize (2) to say that

$$\mathbb{P}\left(\left|f_n - \frac{\sum_{i=1}^n \mathbb{P}_{i-1}(E_i)}{n}\right| \leq \epsilon\right) \geq \mathrm{P} \tag{3}$$

when $\sum_{i=1}^n \mathbb{P}_{i-1}(E_i)(1 - \mathbb{P}_{i-1}(E_i))$ is large enough, where $\mathbb{P}_{i-1}(E_i)$ is the probability of E_i happening after $E_1, \ldots, E_{i-1}$ have been settled. This is the *martingale law of large numbers*. It identifies a frequency with an average of probabilities that may themselves depend on what happens.

There are many further generalizations, of course, all appealing to Cournot's principle.

- We can generalize from events $E_1, E_2, \ldots$ to random variables $X_1, X_2, \ldots$, replacing f_n with the average of the first n X_i and replacing p, p_i, and

[2] The problem of this infinite regress has surely long been part of the folklore of mathematical statistics; I learned about it as doctoral student in mathematical statistics in the early 1970s. But Fisher and Bonitzer are the only authors I have found who discuss it in print.

$\mathbb{P}_{i-1}(E_i)$ with expected values and conditional expected values. Conditions that bound the X_i in some way are needed, but the result is that averages of the X_i approximate averages of their expected values or conditional expected values.

- We can assert not only that the frequencies or averages get closer to the theoretical probabilities or expected values but also how the convergence proceeds. For example, as Jean Ville noted in 1939 [160], the frequency f_n in (1) should not be always above or always below p; as n grows, it should sometimes be greater than p, sometimes less. This is one aspect of the law of iterated logarithm.
- And then, of course, there are all the central limit theorems, which add to the laws of large numbers approximate probabilities for just how close the frequencies or averages will be to the theoretical values [68].

If you want to stick with Venn's straightforward identification of probability with frequency, you can simply renounce all this mathematics. Venn's British colleague Robert Leslie Ellis did just this, explicitly rejecting the usefulness and meaningfulness of Bernoulli's theorem [66]. Venn was inclined to agree with Ellis. But this is not the path taken by 20th and 21st century mathematical statisticians and other applied mathematicians who use probability in engineering, computer science, information technology and operations research. They constantly use the law of large numbers and hence Cournot's principle.

2.2 Qualifying Cournot's principle

In most systems of probabilities, not all events of high probability can happen. And what does happen, described in detail, will have negligible or zero probability. (This is the fearsome lottery paradox.) So careful statements and applications of Cournot's principle must limit the events with probabilities close to zero or one that are taken into consideration. Émile Borel insisted that these events be "remarkable in some respect" and specified in advance. Richard von Mises, Abraham Wald, Jean Ville, and Alonzo Church, working in the idealization of an infinite number of trials, limited the events by requiring that they be definable in some language or computable in some sense. Andrei Kolmogorov brought this back to finite reality by giving a rigorous definition of complexity (and therefore simplicity) for elements of finite sets. We predict only those events of high probability that are simply described [16, 124]. In general, conjunctions of descriptions are more complex than their conjuncts; so the conjunction of events that are considered practically certain may not be practically certain.

Though rigorous, Kolmogorov's notion of complexity involves arbitrary constants that become negligible only asymptotically. So in applied settings, the notion of a simple description necessarily remains vague or implicit, and to the extent that the notion is formalized, the formalization may differ greatly from one application to another. Cournot's principle has many faces.

Nearly all our authors before Borel, and many after, ignored the need to qualify Cournot's principle. Why? Were all these brilliant mathematicians simply naive? Did they somehow have the same blind spot? A more reasonable explanation is that they did not have the same concept of event as those who learned mathematical probability after it acquired set-theoretic and measure-theoretic trappings. Now we are taught that probability comes in the form of a measure, and that all measurable subsets are events—even measurable subsets that we could never begin to describe. For earlier authors, an event was implicitly something that could be simply described, by the very fact that it was being discussed.

What level of probability qualifies as high or low enough is also often vague and clearly varies by application. As we will document in §3, some statisticians have been willing to draw conclusions based on probabilities of 1 in 20, while other have required a probability as low as 1 in 20,000. Physicists and pure mathematicians are far more exacting.

2.3 How data scientists use Cournot's principle

A mathematical statistician or other data scientist may use Cournot's principle in three different ways:

Testing Having hypothesized a statistical model (which may be a single probability measure or a parametric or nonparametric family of probability measures), she tests it by checking that various key events to which it assigns probability close to one happen. This procedure is called a "goodness-of-fit" test.

Estimation If the model passes the goodness-of-fit test, she applies Cournot's principle again to get a confidence interval that narrows the model down, essentially to a single probability measure in the ideal case.

Decision-making Finally, she uses the probabilities to make predictions and decisions. Cournot's principle supports the use of the probabilties in decision-making by allowing us to conclude that the average result of a large number of such decisions will be close to optimal.

A statistician can use these methods, of course, without saying that they are applying Cournot's principle. They may advance some other justification for them, or they may ignore the question of justification.

Beginning in the 1970s, most statisticians who use tests and confidence intervals have often called themselves *frequentist*, even when they do not ascribe to any of the frequentist interpretations of probability discussed by philosophers, as in [91, 92]. Other statisticians, who call themselves Bayesians, use Bayes's rule instead of confidence intervals and may or may not use goodness-of-fit tests. They use their probabilities to make decisions, but many rest content that they have maximized their own expected utility in each case, and this does not require any appeal to Cournot's principle.

As I have argued elsewhere [144], clarity would be served by calling the non-Bayesian tests and confidence intervals *Bernoullian* rather than *frequentist*. This would make it clear that the contrast is between two methodologies, not necessarily between two interpretations of probability. Not all statisticians who use Bayes's rule subscribe to a subjective interpretation of probability. Richard von Mises, the most prominent proponent of the "frequency theory of probability" in the early 20th century, emphasized the use of Bayes's rule, and other Bayesians advocate a variety of "objective" interpretations. In general, statisticians and other data scientists have tired of the frequentist vs. Bayes controversy that was so heated during the late 20th century, and we are returning to the classical situation in which the choice between a Bayesian or a Bernoullian method can be a matter of convenience and circumstance rather than philosophy.

Beginning at least in the 19th-century, it was not unusual for mathematical statisticians who preferred Bayesian to Bernoullian calculations, at least in principle, to say that high probability means practical certainty. Examples include Pierre Simon Laplace, Siméon Denis Poisson, Wilhelm Lexis, Francis Edgeworth, and Richard von Mises; see quotations from these authors in §3. More recent Bayesians who combine a preference for Bayesian calculations with an insistence on a purely subjective meaning for probability can also say that high probability means practical certainty, as they give both probability and certainty a subjective meaning; I quote de Finetti making this point in §3.51. Relatively recent Bayesians who explicitly recognize the need for goodness-of-fit tests to check their models' agreement with phenomena include George E. P. Box [30] and Andrew Gelman [87].

2.4 The hypothetical infinite population

The metaphor of an hypothetical infinite population was already common in probability theory in the 18th century. The celebrated 1774 article in which Laplace introduced his version of Bayes's rule [103] begins, for example, with this problem:

> If an urn contains an infinity of white and black tickets in an unknown ratio, and we draw $p + q$ tickets from it, of which p are white and q are black, then we require the probability that when we draw a new ticket from the urn, it will be white.[3]

How is the hypothetical infinite population related to Cournot's principle?

In articles published in the early 1920s [69, 70], R. A. Fisher provided one answer to this question. Leaving the details for our review of Fisher's contributions in §3.40, we can summarize this answer in three points:

- When we say that the hypothetical population is "infinite", we mean that is "very large" or "as large as you wish". This legitimizes the notion that the elements of the population (white and black tickets, for example) can be in definite proportions.

[3] Translation by Stephen M. Stigler.

- When we say that given data is a random sample from the hypothetical population, we mean only that it is typical of a random sample.
- This typicality is tested and confirmed by goodness-of-fit tests (thus implicitly by Cournot's principle), together with the physical fact that the data was sampled from some actual population or was obtained experimentally from independent trials.

This answer is reasonable for the context in which Fisher was working in the 1920s. With some exceptions, such as George Udny Yule's studies of paupers, the British biometric school launched by Karl Pearson emphasized data sets more or less randomly sampled from real populations, not with the analysis of census or economic data.

The application of probability, whether Bernoullian or Bayesian, to census and economic data had been widely discredited by the end of the 19th century, in large part because the assumption of randomness—i.e., independence of observations—was so often unconvincing or even clearly false. But after World War II, the Bernoullian methods of testing, estimation, and prediction again became very popular in economics and statistics. Historians of econometrics (see especially [128]) cite as a turning point the influential 1944 article by Trygve Haavelmo, from which I quote in §3.54. Haavelmo explained Cournot's principle clearly (without mentioning Cournot's name) and argued that it can be applied when observations, instead of forming a random sample, constitute a single observation from a stochastic process. As he further explained, mathematical statisticians were already using stochastic processes in this way in other scientific fields.

Haavelmo was influenced directly by Jerzy Neyman, perhaps the leading mathematical statistician of the time [2]. In 1960 [130, p. 625], Neyman argued that science had entered a period of "dynamic indeterminism,"

> characterized by the search for evolutionary chance mechanisms capable of explaining the various frequencies observed in the development of the phenomena studied.

As we learned when considering the martingale law of large numbers, (3), Neyman's reference to frequencies is consistent with the notion that we have only a single observation from the stochastic process. His term "chance mechanism" evokes, however, a different idea that also became current among mathematical statisticians in the second half of the 20th century. If there is a mechanism at work, can we not imagine it doing its work many times, producing a population of "realizations" in addition to the actual one in the real world? Why not imagine it doings its work infinitely many times, producing a hypothetical infinite population?

This notion of a hypothetical infinite population, or "superpopulation" has been widely used in applied statistics, not only when the model is a stochastic process, but also when the multiple regression is applied to a complete population or a convenience sample [145]. Its fictional nature seems to trouble statisticians, however.

In 1974 [46], David R. Cox and David V. Hinkley proposed what they called "the strong repeated sampling principle":

> ...statistical procedures are to be assessed by their behaviour in hypothetical repetitions under the same conditions. This has two facets. Measures of uncertainty are to be interpreted as hypothetical frequencies in long run repetitions; criteria of optimality are to be formulated in terms of sensitive behavior in hypothetical repetitions.

This proposal has found some favor among Bernoullian statisticians; as of 10 January 2022, it has appeared 130 times in the journals cataloged by Google Scholar.

The repeated sampling principle extracts from the notion of a hypothetical infinite population the aspects that Bernoullian statisticians need. This may diminishes their worry that they are relying on a fiction. But it relies on an implicit appeal to Cournot's principle, because the relevant "behaviour in hypothetical repetitions" happens only with high probability. In the end, perhaps, the repeated sampling principle is merely a fanciful and unnecessarily complicated way of stating Cournot's principle.

2.5 How physicists use Cournot's principle

When analyzing data from experiments, physicists use Cournot's principle just as statisticians and other data scientists do. But they use it in a rather different and purely theoretical way in statistical mechanics.

Statistical mechanics is a theory about the behavior of an astronomical number of interacting particles, in which the law of large numbers applies with probabilities imperceptibly close to one. (See for example [64, Ch. 3].) The reasoning is based on a probability measure under which the positions and and velocities of the myriad particles are all initially independent and then evolve according to deterministic physical laws.

Aside from the strength of its conclusions, the statistical mechanics also differs from the statistical analysis of data in the degree to which the details of its probability model are tested, estimated, or even needed. For example, the assumed initial independence of positions and velocities, while implausible if taken literally, is not essential, because a other probability measures that are absolutely continuous with respect to the one assumed would give the same high-probability conclusions.

2.6 How pure mathematicians use Cournot's principle

When probability theory is treated as pure mathematics, where a premium is put on the precision and simplicity rather than application, it is natural to set aside theorems about finite sequences of events or variables, emphasizing instead infinite sequences, where conclusions can be drawn with probability exactly equal to one.

The earliest example of such a probability-one result in probability theory is André-Marie Ampére's theorem of the gambler's ruin, published in 1802 [4]. A gambler with limited funds who bets against a casino with unlimited funds is sure to be ruined eventually, even if the bets are fair. See §3.12.

In 1909 [21], more than a century after Ampére's work, Émile Borel opened the way for a plethora of probability-one results using the new theory of measure that he had pioneered. *Borel's strong law of large numbers* says that

$$\mathbb{P}\left(\lim_{n\to\infty} f_n = p\right) = 1, \tag{4}$$

where f_n is again the frequency of an event in the first n of an infinite sequence of independent trials. In Borel's framework, the symbol $\mathbb{P}$ now represents a probability measure, (4) implies (1), which we now call a *weak law of large numbers*. The strong law implies the weak law. The generalizations of Bernoulli's theorem mentioned above all have probability-one versions concerning what happens in the limit in an imagined infinite sequence of trials.

By the time Borel published his celebrated strong law, pure mathematicians were already accustomed to treating sets of measure zero as negligible. The phrase *almost everywhere* and its equivalents in French (*presque partout*) and German *fast überall*) were used to describe statements that were true except on a set of measure zero. In the 1930s, after the measure-theoretic framework for probability Borel had pioneered was further formalized by Kolmogorov, "almost everywhere" became "almost sure" or "almost certain" when the measure was a probability measure. Kolmogorov used *fast sicher* in his 1933 German monograph [101], and Paul Lévy used *presque sûrement* in his 1937 monograph [110].

2.7 The betting form of Cournot's principle

In his doctoral thesis, expanded to a book in 1939 [160], Jean Ville brought the foundations of probability back to its roots in betting by showing that in Borel's framework, with its infinite sequence of independent events, say $E_1, E_2, \ldots$, each with probability p, events of probability zero can be characterized in betting terms. Consider a player who is allowed to bet for or against each E_i at the odds $p : (1-p)$ after seeing which of the preceding E_i have happened and which have failed. Suppose A is an event determined by the events' happenings and failings. If the player begins with a finite stake and is never allowed to risk more than his current capital (the initial stake plus his net gain or minus his net loss so far), then A has probability zero if and only if the player has a strategy for which the capital tends to infinity on every sequence of outcomes in A.

More generally, the probability $\mathbb{P}(A)$ of any event A is equal to the least number α such that the player has a strategy that multiplies the capital he risks by $1/\alpha$ whenever A happens.

In [149], Vladimir Vovk and I confirm that Ville's betting picture provides a complete foundation and a generalization of measure-theoretic probability in discrete time. This encourages a reformulation of Cournot's principle in betting terms, which is applied to statistical testing and estimation in [146].

The betting or game-theoretic version of Cournot's principle can be stated this way:

> A system of probabilities (or more generally a system of betting offers or a forecaster whose forecasts are interpreted as prices) *predicts an event* A if there is strategy for betting at the offered odds or prices that multiplies the capital it risks by a large factor unless A happens.

How large is "large" will depend on the context, just as the "small" in the usual statement of Cournot's priniciple does.

One advantage of this game-theoretic version of Cournot's principle is that the word "predicts" makes clear (or at least very natural) the condition that the event A and the corresponding strategy be selected in advance.

By italicizing "predicts" in the game-theoretic statement, we also capture Cournot's point that there is no other way to connect a system of probabilities with phenomena. The statement tells us how a system of probabilities predicts, in a tone that carries the implication that there is no other way for it to predict.

The generalization from probability measures to forecasting, possibly with probabilities that fall short of defining global joint probabilities for successive forecasts or even with betting offers that fall short of defining a probability distribution for each outcome being forecast, is valuable in many different contexts. It provides an objective interpretation for imprecise probabilities. It allows forecasts to be made day by day, it covers decision models where the probabilities are not given for the decisions themselves, and in particular it covers A. Philip Dawid's predictive theory of causality [52, 54].

3 In their own words

The authors listed here were active over a period of four centuries. Because the periods in which they contributed to our topic sometimes overlapped, it is impossible to list them in a way that corresponds perfectly to the chronology of their contributions. For lack of a more perfect ordering, I have listed them in order of their dates of birth.

The brief quotations and comments given here cannot do justice, of course, to the complexity and subtlety of the thought of the individuals quoted. But they give some indication of the historical diversity of thought concerning Cournot's principle. Except where otherwise noted, the translations are mine.

In lieu of a summary, which would be impossible, here are a few ways some of the authors can be roughly grouped—only roughly, because every author has their own way of putting the matter.

- Those who talked about the sufficiency of high probability for moral (i.e., practical) certainty before probability was understood to be numerical: Molina, Descartes, Locke.

- Those who stated that high enough numerical probability is sufficient for practical certainty: Bernoulli, Buffon, d'Alembert, Condorcet, von Mises, Uspensky, Fry, Anderson.
- Those who used the notion that high numerical probability suffices for practical certainty in contexts not involving frequencies: (1) for the St. Petersburg pardox, Buffon; (2) for the gambler's ruin, Ampère.
- Those who stated or suggested specific numerical levels of probability that might be required for practical certainty: Bernoulli, Fourier, Poisson, Gavarret, Lexis, Galloway, Edgeworth.
- Those who denied that high enough probability is sufficient for practical certainty: Pascal, von Kries, de Finetti, Howson.
- Those who explicitly supported Cournot's contention that equating high probability with practical certainty is the only way to connect a system of probabilities with phenomena: Cournot himself, Mansion, Hadamard, Borel, Chuprov, Fréchet, Slutsky, Lévy, Haavelmo, Richter, Dantzig, Freudenthal, Doob, Ville, Stein.
- Those who asserted that equating high probability with practical certainty is *one way* to connect a system of probabilities with phenomena but did not assert that it is the only way: Kolmogorov, Wald, Hempel.
- Those who used Cournot's principle inasmuch as they used significance tests or confidence intervals but did not otherwise comment (so far as documented here) on Cournot's principle: Laplace, Neyman.

We can also note milestones:

- Bernoulli was the first to put the numbers on the principle that sufficiently high probability justifies moral certainty.
- Condorcet, reacting to Buffon and d'Alembert, was the first to insist that this principle should not be put inside the probability calculus (what we now call the mathematical theory of probability).
- Cournot, as I have repeatedly noted, was the first to say that this principle is the only way to relate the mathematical theory to phenomena. The word "phenomena" reflects the influence of Immanuel Kant's *Critique of Pure Reason*, which first appeared in 1781. Whereas Condorcet and Laplace were still working in a time when probability theory could be considered "mixed mathematics" (see Daston [50]), Cournot was beginning to see the theory as pure mathematics, which had to be deliberately connected with phenomena.
- Venn and Ellis were the first to reject Cournot's principle, and Bernoulli's theorem along with it, in favor of identifying probability with frequency. This effort to make probability mixed instead of pure mathematics was not

followed the 20th century mathematical statisticians who called themselves frequentists.

- Fisher was arguably the first to express Cournot's principle using the notion of typicality.
- Wald, who introduced the notion that tests must be computable, was therefore the first to qualify Cournot's principle in a mathematical way.
- Ville was the first to formulate Cournot's principle in terms of betting.

3.1 Luis de Molina, 1535–1600

Molina, a Jesuit who taught at universities in Spain in Portugal, was one of the best known of the late scholastics. He did not measure probability numerically, but he regarded probability as a matter of degree. He insisted that high probability provided moral certainty, and he used games of chance to provide examples. This is documented by Sven K. Knebel [100].

3.2 René Descartes, 1596–1650

Towards the end of the French edition of his *Principles of Philosophy* [59, pp. 482–483], published in 1647, Descartes argues that his system, if not mathematically certain, is at least morally certain. He explains moral certainty this way.

> ...so as to avoid doing harm to the truth by supposing it to be less certain that it is, I will distinguish here between two kinds of certainty. This first is called moral—sufficient, that is to say, for governing our behavior, or as great as that of things affecting the conduct of life that we scarcely ever doubt, even though they could happen, to speak in an absolute sense, to be false. Those who have never been to Rome hardly doubt that it is a city in Italy, even though it could be that everyone they have learned it from had deceived them. And if someone who wants to decode an encoded message written in ordinary letters thinks to read each A as a B, each B as an A, and so on, substituting for each letter the one that follows it in the alphabet, and if when reading it in this way finds words that make sense, he will hardly doubt that he has found the true meaning, even though it could be that the person who wrote it gave it an entirely different meaning by interpreting each letter in some other way: for it would be so hard for this to happen, especially when the message has many words, that it is not morally believable.

Descartes spent eight years of his youth in a Jesuit school.

3.3 Blaise Pascal, 1623–1662

Pascal did not use the word "probability" in his mathematical work on games of chance, but it appears in the final chapters of the Port Royal Logic [10], which are sometimes attributed to him. There we see echoes of Pascal's argument for belief in the Christian God: even the slightest probability for the truth of his existence must be multiplied by the infinite benefit of eternal life. Pascal and his Jansenist friends were fierce opponents of the doctrines of the Jesuits, including, it seems the very notion of moral certainty.

3.4 John Locke, 1632–1704

Locke published his *An Essay Concerning Human Understanding* in 1689 [113]. Chapter XV of Book IV, entitled "Of probability", includes this passage:

> ... most of the propositions we think, reason, discourse—nay, act upon, are such as we cannot have undoubted knowledge of their truth: yet some of them border so near upon certainty, that we make no doubt at all about them; but assent to them as firmly, and act, according to that assent, as resolutely as if they were infallibly demonstrated, and that our knowledge of them was perfect and certain.

Locke famously criticized Descartes's doctrine of innate ideas, and he was scornful of the scholastics. But he is not taking issue with them in this particular passage.

In 1992 [72], the philosopher Richard Foley formulated a thesis of his own:

> ... it is epistemically rational for us to believe a proposition just in case it is epistemically rational for us to have a sufficiently high degree of confidence in it, sufficiently high to make our attitude towards it one of belief.

Foley found this near enough to Locke's views that he called the *Lockean thesis*. The term has been popular in the recent philosophical literature; as of January 11, 2022, it had 678 citations in Google Scholar and 85 in JSTOR.

3.5 Jacob Bernoulli, 1655–1705

Bernoulli's celebrated book on probability, *Ars Conjectandi*, was published posthumously in 1713 [14]. Here are two brief quotations, translated by Edith Sylla [15]:

- From Chapter I of Part IV: Something is *morally certain* if its probability comes so close to complete certainty that the difference cannot be perceived. ...
- From Chapter II of Part IV: Because ... it is rarely possible to obtain certainty that is complete in every respect, necessity and use ordain that

> what is only morally certain be taken as absolutely certain. It would be useful, accordingly, if definite limits for moral certainty were established by the authority of the magistracy. for instance, it might be determined whether 99/100 of certainty suffices or whether 999/1000 is required. ...

3.6 Georges-Louis Buffon, 1707–1788

Georges-Louis Leclerc, Comte de Buffon, was a distinguished naturalist and a polymath. Like d'Alembert (see §3.8), he saw Cournot's principle as a solution to the St. Petersburg paradox.

In 1777 [35], Buffon argued that the distinction between moral and physical certainty was one of degree. An event with probability 9999/10000 is morally certain; an event with much greater probability, such as the rising of the sun, is physically certain [114].

In 1774, Laplace began his work on probability with a remarkable article introducing what was later called inverse probability or the Bayesian method. When he saw the article, Buffon wrote to Laplace urging him to use Cournot's principle. Dated 21 April 1774, the letter was preserved by Laplace's family long enough to be printed in 1879 by the Academy of Sciences [36]. Most of the letter is reproduced here:

> Sir, I received and read with great pleasure your learned "Memoir on the probability of causes from events", and though I lack the talent, which you have so kindly attributed to me, to know how to go from events back to causes, at least not by paths as reliable as yours, I felt the beauty of your work and I can only encourage you, Sir, to continue your research of this kind, which requires more delicacy and pureness of mind than any other part of mathematics. I found your your ideas to be in agreement with mine up until you spoke of the game of heads and tails: the material difference of the coin should indeed have a long-term influence on the number of events for and against, but this is not the true cause that makes a theoretically infinite probability nevertheless become finite in practice and makes it the case that you will go bankrupt if you give only six or seven écus of half-écus every time you play that game, instead of the infinitely many écus or half-écus. Many mathematicians, including Mr. Fontaine, have tried and failed to solve this problem, for lack of a metaphysical and moral principle that joins here with the mathematical calculation; *this principle is that whenever a probability is greater than 1/1000, it is relatively to us* perfectly equal to zero. As contradictory as this proposition seems in its formulation, I can just the same prove it to you beyond any doubt [check original]; but we will talk about this matter when I have the pleasure to see you again.

3.7 David Hume, 1711–1776

Hume's skepticism and his concept of probability did not leave much room for Cournot's principle. Here is a passage in the section entitled "Of the probability of causes" (Book I, Part III, §XII) in *A Treatise of Human Nature*, which appeared in 1739–1740 [95], that illustrates this point.

> ...there is no probability so great as not to allow of a contrary possibility: because otherwise it would cease to be a probability, and would become a certainty. That probability of causes, which is most extensive ... depends on a contrariety of experiments ... An experiment in the past proves at least a possibility for the future.

The skepticism is not expressed in this way in Hume's more mature *An Enquiry into Human Understanding*, which appeared in 1748 [96]. There, in §VI, "Of probability", we find his famous declaration that chance does not exist:

> 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.

He concedes that probability begets *belief* but will not concede that any of "the received systems of philosophy" can justify moral or practical certainty.

3.8 Jean Le Rond d'Alembert, 1717–1783

In 1761 [49, p. 8], when he was a leading French intellectual and the unquestioned leader of mathematics in Paris, d'Alembert gave this account of how Cournot's principle provides a solution of the St. Petersburg paradox.

> ...*when the probability of an event is very small, it should be considered and treated as zero, and we should not multiply* (as has been recommended until now) *this probability by the gain hoped for in order to find the stake or expectation.* For example, if Peter bets with James on 100 tosses of a coin, agreeing that James will give him 2^{100} écus if he get heads on the 100th toss and not before, we find by the usual rule that Peter should give one écu to James before the tosses. I say that Peter should not give that écu, because he will *certainly* lose it. There will *certainly* be a head before the 100th toss, even though it does not happen *necessarily.*

3.9 Nicolas de Condorcet, 1743–1794

In his famous and lengthy eulogy of Buffon, delivered to the Academy of Sciences and published in 1790, we find the following passage [42, pp. 36–37]:

> Mr. de Buffon proposed that we assign a precise value to the very large probability that we can consider moral certainty, and beyond this to ignore the small possibility of a contrary event. This principle

> is true when we only want to make ordinary use of a calculation; and in this sense all men have adopted it in practice and all philosophers have followed it in their reasoning. But it ceases to be correct if we introduce it into the calculus itself, and especially if we want to use to establish theories, to explain paradoxes, and to prove or refute general rules. Besides, this probability, which may be called moral certainty, must be greater or smaller according to the nature of the objects considered and the principles that should guide our conduct; and it would have been necessary to fix the degree of probability at which it begins to be reasonable to believe and allowed to act for each type of truth and action.

This passage seems remarkably original. It is the first source I have seen for two important points: Cournot's principle is outside the probability calculus, and the degree of probability needed depends on the nature of the objects considered.

3.10 Pierre Simon Laplace, 1749–1827

In this paragraph, from the introduction to his *Essai philosophique sur les probabilités*,[4] Laplace combines a statement of Cournot's principle with a likelihood principle:

> When a simple event or an event composed of many simple events, such a round of a game, has been repeated a large number of times, the possibilities of the simple events make what one has observed most likely are those that the observation indicates with the greatest likelihood: as the observed event is repeated, that likelihood increases, finally become indistinguishable from certainty as the number of repetitions becomes infinite.

3.11 Joseph Fourier, 1768–1830

The renowned mathematician Joseph Fourier may be the first to have fixed a level of probability that one can equate with certainty. The level was very high; he allowed a probability of error of only one in 20,000.

In the last decade of his life, in the 1820s, Fourier held a post in the census bureau of the Paris region. In the bureau's report for 1826, he included a manual for using the probability calculus to interpret census results. He used Laplace's normal approximation to the probability distribution of the average A of independent measurements $y_1, \ldots, y_m$ of an unknown quantity H. Instead

[4]The paragraph appears on pp. 81 of Bernard Bru's 1886 reprinting of Laplace's fifth edition, which first appeared in 1825. The following translation is mine. There are several English translations of the entire *Essai*; the paragraph appears on p. 36 of Andrew Dale's translation [106].

of A's standard deviation, he used

$$g = \sqrt{\frac{2}{m}\left(\frac{\sum_{i=1}^{m} y_i^2}{m} - A^2\right)},$$

which differs from A's standard deviation by the factor $\sqrt{2}$. He explained that an interval extending $2.86783g$ above and below the average (about 4 standard deviations) would provide an interval certain to contain H. Here is how he explained the calculation [74, pp. xxi–xxii]:

> To complete this discussion, we must find the probability that H, the quantity sought, is between proposed limits A + D and A + D. Here A is the average result we have found, H is the fixed value that an infinite number of observations would give, and D is a proposed quantity that we add to or subtract from the value A. The following table gives the probability P of a positive or negative error greater than D; this quantity D is the product of g and a proposed factor ∂.
>
> | ∂ | P |
> |---|---|
> | 0.47708 | $\frac{1}{2}$ |
> | 1.38591 | $\frac{1}{20}$ |
> | 1.98495 | $\frac{1}{200}$ |
> | 2.46130 | $\frac{1}{2000}$ |
> | 2.86783 | $\frac{1}{20000}$ |
>
> Each number in the P column tells the probability that the exact value H, the object of the research, is between $A + g\partial$ and $A - g\partial$. Here A is the average result of a large number m of particular values $a, b, c, d, \ldots, n$, ∂ is a given factor, g is the square root of the quotient found by dividing by m twice the difference between the average of the squares $a^2, b^2, c^2, d^2, \ldots, n^2$ and the square A^2 of the average result. We see from the table that the probability of an error greater than the product of g and 0.47708, i.e. greater than about half of g, is $\frac{1}{2}$. It is a 50–50 or 1 out of 2 bet that the error committed will not exceed the product of g and 0.47708, and we can bet just as much that the error will exceed this product.
>
> The probability of an error greater than the product of g and 1.38591 is much less; it is only $\frac{1}{20}$. It is a 19 out of 20 bet that the error of the average result will not exceed this second product.
>
> The probability of an even greater error becomes extremely small as the factor ∂ increases. It is only $\frac{1}{200}$ when ∂ approaches 2. The probability then falls below $\frac{1}{2000}$. Finally one can bet much more than twenty thousand to one that the error of the average result will be less than triple the value found for g. So in the example cited in Article VI, where the average result was 6, we can consider it certain

that the value 6 is not wrong by a quantity three times the fraction 0.082 that the rule gave for the value of g.

The quantity sought, H, is therefore between $6 - 0.246$ and $6 + 0.246$.

3.12 André-Marie Ampère, 1775–1836

Whereas Bernoulli, d'Alembert, and Buffon had proposed selecting some number less than one that would suffice for moral certainty, Ampère realized that he could develop a theory of gambler's ruin with a more demanding concept of moral certainty. In his 1802 book *Considerations sur la théorie mathématique du jeu* [4], he defined this concept on p. 3:

> If we represent absolute certainty, the certainty resulting from mathematical demonstration for example, by unity, as is usually done, then we can consider moral certainty to be any variable fraction that never becomes equal to unity but can get close enough to it as to exceed any particular fraction.

To illustrate his concept of moral certainty, Ampère imagined a man who throws two balanced dice indefinitely many times, for his whole life and beyond if need be, until he gets two sixes. His success is morally certain. The probability of success on each throw is only 1/36, but the probability of eventual success is

$$\frac{1}{36} + \frac{35}{36}\frac{1}{36} + \left(\frac{35}{36}\right)^2 \frac{1}{36} + \cdots = \sum_{n=0}^{\infty} \left(\frac{35}{36}\right)^n \frac{1}{36}, \tag{5}$$

which is equal to one or, as Ampère preferred to say, as close to one as you want.

Ampère similarly imagined a person who starts with a finite fortune and continually bets, each time making a fair bet of one unit at even odds. He may bankrupt many opponents along the way, but there is always another, and so in effect he is betting against an opponent with an infinite fortune. Summing a series, as in (5), Ampère found that the player's own bankruptcy is morally certain.

Here is an English translation of Ampère's explanation:

1. ...leaving aside moral considerations that make the value of money depend on the players' circumstances, there cannot be any disadvantage in playing at equal odds against a player who is equally rich, because the one cannot lose anything the other does not gain, and everything is equal between them;

2. the same is true between two players with unequal fortunes provided they agree to play only a number of rounds small enough that neither can lose everything he has;

3. it is not the same when the number of rounds of play is indefinite: the possibility of staying in the game longer gives the richer of the two an advantage, which increases with the difference between their fortunes;

4. this advantage becomes infinite if one of the fortunes can be infinite, then the less rich player will be sure to be bankrupt, and it is for this reason that a player heads to a certain ruin when he plays indifferently with everyone he encounters in society: in the theory we must in effect treat all these opponents as a single opponent with an infinite fortune.

3.13 Siméon Denis Poisson, 1781–1840

3.14 Thomas Galloway, 1796–1851

Fourier's suggestion that we can consider it certain that the error will not exceed $3g$ was widely repeated in the 19th century. We see it, for example, in Thomas Galloway's *Treatise on Probability* [85, p. 144]. This treatise, published as a book in 1839, first appeared as the article on probability in the 7th edition of the Encyclopedia Britannica. Karl Pearson cited it in the book on the philosophy of science that he published in 1892 [134, pp. 177, 180].

3.15 Antoine Augustin Cournot, 1801–1877

Saying that an event of very small or vanishingly small probability will not happen is one thing. Cournot, as I have repeatedly mentioned, said more. He seems to have been the first to say that this is the only way for probability to gain empirical meaning. He said this in his 1843 book, *Exposition de la théorie des chances et des probabilités* [43, §43]:

> ... *The physically impossible event is therefore the one that has infinitely small probability*, and only this remark gives substance—objective and phenomenal value—to the theory of mathematical probability.

The phrase "objective and phenomenal" refers to Kant's distinction between the noumenon, or thing-in-itself, and the phenomenon, or object of experience [51].

As examples of physically impossible events, Cournot mentioned a cone balancing on its point, the frequency of heads in a long sequence of flips of a fair coin differing too much from one-half, and a loose tile happening to fall on his head from a roof as he walked along a French street. One might suppose that an infinitely small probability is exactly zero, but Cournot and his contemporaries interpreted the idea more broadly. He explained this explicitly in 1875 ([44], §IV.4):

> In practice, moreover, and in the world of realities, what geometers call an infinitely small probability is and can only be an exceedingly small probability. The tip of this very sharp needle is not a mathematical point like the apex of the cone in question. Viewed

> through a magnifying glass, it becomes a *blunt* tip. With whatever care we polish the plane of steel or agate on which we try to balance it, very delicate experiments will show roughness and streaks. It follows that the probability of success in putting the needle in equilibrium is no longer infinitely small, that it is only excessively small, as would be the probability of rolling an ace a hundred times with an unloaded die, which is enough for us to judge, with no fear of being refuted by experience, that the equilibrium is physically impossible.
>
> The same remarks apply to the market value of commercial chances. ...

The concluding section of Cournot's 1843 book summarized its ideas as follows:

> Let us summarize in a few words the main points that we have undertaken to establish in this essay.
>
> 1. The idea of chance is the idea of the concurrence of independent causes to produce a given event. The combinations of different independent causes that all give rise to the same event is what should be meant by the chances of that event.
>
> 2. When only one out of an infinity of chances can produce the event, that event is called *physically impossible*. The notion of physical impossibility is neither a mental fiction nor an idea that has value only relative to the imperfect state of our knowledge. It must figure as an essential element in the explanation of natural phenomena, whose laws do not depend on the knowledge that people might have about them.
>
> 3. When we consider a large number of trials of the same event, the ratio of the number of cases where the event happens to the total number trials becomes practically equal to the ratio of the number of chances favorable to the event to the total number of chances, or to what we call the *mathematical probability* of the event. If we could repeat the trial an infinite number of times, it would be physically impossible that the two ratios would differ by a finite amount. In this sense, the mathematical probability can be considered a measure of the *possibility* of the event, or of the facility with which it happens. By the same token, the mathematical probability expresses a ratio that stands outside the mind that conceives of it, a law to which phenomena are subject, whose existence does not depend on the expansion or narrowing of the our knowledge about their happening.
>
> 4. If, with our imperfect knowledge, we have no reason to suppose that one combination happens more often than another, even though in reality these combinations are events that can have unequal mathematical probabilities or possibilities, and if we understand the *probability* of an event to be the ratio of the number of combinations favorable to the event to the total number of combinations that we put in the same group, this probability can still serve,

when there is nothing better, to fix the terms of a bet or any other risky exchange, but it will no longer express a real and objective relation between things. It will take on a purely subjective character and will be liable to vary from one individual to another depending on their knowledge. Nothing is more important than to carefully distinguish between these two meanings of the term *probability*, one an objective meaning, the other a subjective meaning, if we want to avoid confusion and error, whether in the exposition of the theory or in the applications we make of it.

5. In general, for natural events, whether physical or social, objective mathematical probability, conceived of as measuring the possibility of events arising from the concurrence of independent causes, can only be determined by experience. If the number of trials of the same chance increases to infinity, the probability will be determined exactly, with a certainty comparable to that for an event whose contrary is physically impossible. When the number of trials is merely very large, the probability is given only approximately, but we are still entitled to consider it very unlikely that the real value differs notably from the value derived from observations. In other words, we will very rarely err significantly in taking the observed value to be the real value.

6. When the number of trials is not very great, the usual formulas for evaluating probabilities *a posteriori* become illusory. They no longer give us anything but subjective probabilities, appropriate for determining the terms of a bet but without use with respect to the determination of natural phenomena.

7. Nevertheless, we should not conclude from the preceding remark that the number of trials should always be very large in order to give the real values of the probability of an event with sufficient precision and sufficient confidence. We should conclude merely that the confidence will not be equivalent to a probability in the objective sense. We cannot evaluate the chance we have of erring when we say that the real value falls between certain limits. In other words, we cannot determine the ratio of the number of mistaken judgements to the total number of judgements made in similar circumstances.

8. Independently of mathematical probability, in the two senses considered above, there are probabilities that are not reducible to the enumeration of chances but motivate a host of our judgements, and even the most important ones. These probabilities pertain mainly to our idea of the simplicity of nature's laws, of the order and rational sequence of phenomena, and for this reason we can call them *philosophical* probabilities. All reasonable people have a confused sense of these probabilities. When it becomes distinct or concerns delicate subjects, it belongs only to cultivated intelligences or can even constitute a mark of genius. It forms the basis of a system of critical philosophy, glimpsed in the most ancient schools, that represses or

> conciliates skepticism and dogmatism, but which we must not, for fear of strange aberrations, bring into the domain of mathematical probability.

An English translation of the book by Oscar Sheynin is available at `www.probabilityandfinance.com`. A German translation appeared in 1849 and a Russian translation in 1970.

Cournot further developed his understanding of probability and his broader philosophy of science in a series of later books, written during a career as a university professor and administrator. This work, like his work in economics, became better known after his death. It was influential during the early 20th century among French philosophers. His collected works, edited by a team of French philosophers and mathematicians, appeared in fifteen volumes beginning in 1973 [45]. Thierry Martin published an extensive bibliography of work by and about Cournot in 2005 [121]. There are no recent appraisals of his work in English, but relatively recent appraisals in French include those by Martin [117] and Bertrand Saint-Sernin [142]. See also [118, 120, 156, 31, 8, 9].

3.16 Augustus De Morgan, 1806–1871

From page 396 of De Morgan's entry "Theory of Probabilities", on pages 393–490 of Volume II of Encylopædia Metropolitana, Griffin, London, 1849

> *Mathematical certainty* (a thing perhaps impossible in the strictest sense) is the terminus or limit towards which our impressions approach as our knowledge becomes greater and greater, and is never attained as long as any doubt whatsoever remains. *Practical certainty* is that high degree of probability on which the mind acts at once, without thinking the counter-probabilities sufficiently large to be taken into account; and it depends upon the character of the individual.

3.17 Jules Gavarret, 1809–1890

In his book on medical statistics [86], he adopts Poisson's standard of $2g$.

3.18 William Fishburn Donkin, 1814–1869

Donkin was a British mathematician and astronomer. Here his 1851 paper on probability, [61], serves as an example where high subjective probability is equated with practical certainty.

On the first page of his article (p. 353), Donkin writes,

> It will, I suppose, be generally admitted, and has often been more or less explicitly stated, that the subject-matter of calculation in the mathematical theory of probabilities is *quantity of belief*. A certain number of hypotheses are presented to the mind, along with a certain

> quantity of information relating to them: In what way ought belief be distributed among them?

A few pages later, Donkin discusses the problem of deciding whether a very regular arrangement of objects, say balls in a circle on a table, could have been by purpose or by accident. He calculates a probability for it being by purpose that involves unspecified constants but must be close to 1. His conclusion, on p. 360:

> Thus the mathematical investigation leads, equally with common sense, to a *moral certainty* that the arrangement was designed.

3.19 Robert Leslie Ellis, 1817–1859

Ellis was an accomplished mathematician, but he identified probability with frequency to the extent that he could make no sense of Bernoulli's theorem. He expressed this viewpoint eloquently in short paper that he read to the Cambridge Philosophical Society in 1842 [66, pp. 1–2]:

> ... If the probability of a given event be correctly determined, the event will on a long run of trials, tend to recur with frequency proportional to this probability.
>
> This is generally proved mathematically. It seems to me to be true *à priori*.
>
> When on a single trial we expect one event rather than another, we necessarily believe that on a series of similar trials the former event will occur more frequently than the latter. The connection between these two things seems to me to be an ultimate fact, or rather, for I would not be understood to deny the possibility of further analysis—to be a fact, the evidence of which must rest upon an appeal to consciousness. Let any one endeavour to frame a case in which he may expect one event on a single trial, and yet believe that on a series of trials another will occur more frequently; or a case in which he may be able to divest himself of the belief that the expected event will occur more frequently than any other.
>
> For myself, after giving a painful degree of attention to the point, I have been unable to sever the judgment that one event is more likely to happen than another, or that it is to be expected in preference to it, from the belief that on the long run it will occur more frequently.

Chuprov cited Ellis as being the first to notice that Bernoulli's theorem conflicts with the identification of probability with frequency.

3.20 John Venn, 1834–1923

Twenty years after Ellis expressed his reservations about Bernoulli's theorem, they were echoed by John Venn, another Cambridge scholar who argued that probability should be identified directly with probability. In the first edition of Venn's *The Logic of Chance*, which appeared in 1866, we find this passage [159, pp. 35–36]:

> The reader who is familiar with Probability is of course acquainted with the celebrated theorem of Bernoulli. This theorem, of which the examples just adduced are merely particular cases, is generally expressed somewhat as follows :—that in the long run all events will tend to occur with a frequency proportional to their objective probabilities. With the mathematical proof of this theorem I have nothing to do here; nor, if there is any value in the foregoing criticism, need we trouble ourselves about it, for in that case the basis on which the mathematics rest is faulty, owing to the fact of there really being nothing which we can call the objective probability.
>
> This theorem of Bernoulli seems to me one of the last remaining relics of Realism, which after being banished elsewhere still manages to linger in the remote province of Probability. It is an illustration of the inveterate tendency to objectify our conceptions even in cases where the conceptions had no right to exist at all. A uniformity is observed ; sometimes, as in games of chance, it is found to be so connected with the physical constitution of the bodies employed as to be capable of being inferred beforehand, though even here the connection is by no means so necessary as is commonly supposed; this constitution is then converted into an "objective probability," supposed to develop somehow into the sequence which exhibits the uniformity. Finally, this very questionable objective probability is assumed to exist, with the same faculty of development, in all the cases in which uniformity is observed, however little resemblance there may be between these and games of chance.

The same passage appears, substantially unchanged, on pp. 91–92 of the book's third edition, which appeared in 1888.

The disinterest in Bernoulli's theorem and the other limit theorems of mathematical probability that we see in the writing of Ellis and Venn was a more enduring aspect of thought about probability at Cambridge than their equation of probability with frequency. We find it in later scholars whose "interpretations of probability" were quite diverse: William Ernest Johnson, John Maynard Keynes, Frank Ramsey, and Harold Jeffreys [3]

3.21 Wilhelm Lexis, 1837–1914

Wilhelm Lexis followed Fourier with $3g$ in his *Einleitung in die Theorie der Bevölkerungsstatistik* [112, pp. 98, 100, 106, 144]. For Lexis, it was practically certain (*praktisch die Gewissheit* or *fast mit Gewissheit*, etc.) that an error is less than this quantity.

3.22 Ludwig Boltzmann, 1844–1906

In the second half of the nineteenth century, the principle that an event with a vanishingly small probability will not happen took on a real role in physics, most saliently in Ludwig Boltzmann's statistical understanding of the second law of thermodynamics. As Boltzmann explained in the 1870s, dissipative processes are irreversible because the probability of a state with entropy far from the maximum is vanishingly small (von Plato 1994[166], p. 80; Seneta 1997[97, 143]). Also notable was Henri Poincaré's use of the principle in the three-body problem[135, 166]. Poincaré's recurrence theorem, published in 1890[135], says that an isolated mechanical system confined to a bounded region of its phase space will eventually return arbitrarily close to its initial state, provided only that this initial state is not exceptional. Within any region of finite volume, the states for which the recurrence does not hold are exceptional inasmuch as they are contained in subregions whose total volume is arbitrarily small.

This comment by Boltzmann, in 1898 [17, §40, p. 120], is notable:

> Of course, it should be remembered that these are just laws of probability. The possibility of deviation from the same is practically out of the question; but their probability in case the number of molecules is finite, though unimaginably small, is not zero; indeed, it can even be numerically calculated according to the laws of probability in every given case, and disappears only for the limiting case of an infinite number of molecules.

3.23 Paul Mansion, 1844–1919

On December 16, 1903, the Belgian mathematician Paul Mansion delivered a 60-page discourse, in French, on the objective significance of the probability calculus to the Royal Academy of Belgium [115]. In the conclusion of the discourse, we find this passage:

> The fundamental principle is this: Between two contrary propositions, one little probable, the other very probable, the human mind chooses the second, freely, but almost irresistibly, and declares it practically certain.
>
> From this we deduce the logical legitimacy of the law of large numbers and the principle of the accumulation of independent probabilities.
>
> The law of large numbers applies first to the question of the gambler's ruin and its consequences, then to statistics whenever it encounters nearly constant ratios in the numbers it collects.
>
> The principle of the accumulation of independent probabilities is practically, if not metaphysically, the source of our certainties in the natural and historical sciences, which in the last analysis rely on testimony, every time we are not personally inventor or witness.

Surely we can classify Mansion as a supporter of Cournot's principle. Yet perhaps he differs from Hume, whom I have classified as an opponent, only in tone. Hume emphasized the lack of justification of our mind's irresistible equation of high probability with practical certainty. Mansion does not refute the claim that it is unjustified, but he applauds it.

Laurent Mazliak has reviewed Mansion's career and his role in the history of Belgian mathematics [126]. Mansion was a devout Roman Catholic. One feature of his thought that is interesting for our investigation is his appreciation of the relationship of the role of the Jesuits in the development of probability and his corresponding low opinion of Pascal. He concludes his discourse by equating Laplace's superior intelligence, for whom "nothing is uncertain and the future as well as the past is present to the eye", with God.

3.24 Francis Edgeworth, 1845–1926

Translating Lexis's account into his somewhat idiosyncratic English, Francis Edgeworth called an observed difference *significant of a real difference*, as opposed to accidental, when it differs from zero by more than $3g$ cite[§137]Edgeworth:1911.

Edgeworth apparently first used *significant* in this way in the paper he read at the jubilee meeting of the Statistical Society of London in 1885 [65]. The president of the session reported that when pressed by the Italian statistician Luigi Perozzo on whether his paper contained anything new, Edgeworth had said that "he did not know that he had offered any new remarks, but perhaps they would be new to some readers. He had borrowed a great deal from Professor Lexis."

3.25 Emanuel Czuber, 1851–1925

Czuber's textbooks on probability, error theory, and mathematical statistics were very influential among German literature at the beginning of the 20th century. He was one of the first authors to explicitly criticize the concept of practical certainty. In his 1903 textbook, he singled out Bernoulli, d'Alembert, Buffon, and De Morgan as advocates of the concept and declared them all wrong; the concept, he wrote, violates the fundamental insight that an event of ever so small a probability can still happen [48, pp. 13–15]. See also Meinong 1915 [127, p. 591].

Here is a paragraph from pp. 14–15 of Czuber's 1903 book, which also appears on pp. 16–17 of the 1908 first volume of the second edition.

> The relationship between the various degrees probability can attain and absolute necessity or certainty has not always been correctly assessed. Jacob Bernoulli defines probability as degree of certainty, assigning an event of probability 3/5 the corresponding fraction of certainty. The same view was taken by the first German philosopher who reviewed the principles of probability, J. J. Fries. There is also a

similar sounding passage in Laplace; to the remark that probability turns into certainty and is represented by unity when all cases are favorable to the event, he adjoins the comment that certainty and probability are comparable from this point of view. But by further adding that there is an essential difference between the two states of the mind, the one where a truth has been proven to him rigorously, and the one where he still perceives a small opening for error, he allows us to see the correct position, the position that Condorcet had already taken before him and that the new philosophy adopts: *probability and certainty* (or necessity) *are things of essentially different natures*, and there is no bridge that could be built from one to the other. Proper logical worth is thereby assigned to the multitude of attempts that have been made to produce links or transitions from probability to certainty on one side and impossibility on the other. Thus Jacob Bernoulli already distinguishes between *absolute* and *moral* certainty and impossibility, understanding by the latter very high or very low degrees of probability. Later, in D'Alembert, Buffon, De Morgan, under the names practical certainty, physical impossibility, etc., we find similar conceptions that conflict with the *fundamental* knowledge that an event with probability ever so close to unity does not *have to* happen and that an event with ever so small a probability *can* happen.

3.26 Johannes von Kries, 1853–1928

One of the most influential of the German philosophers who discussed probability in the late nineteenth century was Johannes von Kries, whose *Principien der Wahrscheinlichkeitsrechnung* first appeared in 1886. Von Kries rejected what he called the orthodox philosophy of Laplace and the mathematicians who followed him. As von Kries's saw it, these mathematicians began with a subjective concept of probability but then claimed to establish the existence of objective probabilities by means of a so-called law of large numbers, which they erroneously derived by combining Bernoulli's theorem with the belief that small probabilities can be neglected. Having both subjective and objective probabilities at their disposal, these mathematicians then used Bayes's theorem to reason about objective probabilities for almost any question where many observations are available. All this, von Kries believed, was nonsense. The notion that an event with very small probability is impossible was, in von Kries's eyes, simply d'Alembert's mistake.

Von Kries believed that objective probabilities sometimes exist, but only under conditions where equally likely cases can legitimately be identified. Two conditions, he thought, are needed:

- Each case is produced by equally many of the possible arrangements of the circumstances, and this remains true when we look back in time to earlier circumstances that led to the current ones. In this sense, the relative sizes of the cases are *natural*.

- Nothing besides these circumstances affects our expectation about the cases. In this sense, the Spielräume[5] are *insensitive*.

Von Kries's *principle of the Spielräume* was that objective probabilities can be calculated from equally likely cases when these conditions are satisfied. He considered this principle analogous to Kant's principle that everything that exists has a cause. Kant thought that we cannot reason at all without the principle of cause and effect. Von Kries thought that we cannot reason about objective probabilities without the principle of the Spielräume.

Even when an event has an objective probability, von Kries saw no legitimacy in the law of large numbers. Bernoulli's theorem is valid, he thought, but it tells us only that a large deviation of an event's frequency from its probability is just as unlikely as some other unlikely event, say a long run of successes. What will actually happen is another matter. This disagreement between Cournot and von Kries can be seen as a quibble about words. Do we say that an event will not happen (Cournot), or do we say merely that it is as unlikely as some other event we do not expect to happen (von Kries)? Either way, we proceed as if it will not happen. But the quibbling has its reasons. Cournot wanted to make a definite prediction, because this provides a bridge from probability theory to the world of phenomena—the real world, as those who have not studied Kant would say. Von Kries thought he had a different way of connecting probability theory with phenomena.

Von Kries's critique of moral certainty and the law of large numbers was widely accepted in Germany. For further discussion of his ideas and their influence, see [98] and the special issue on von Kries published by the *Journal for General Philosophy of Science* in 2016, especially [169].

Von Kries's principle of the Spielräume did not endure, for no one knew how to use it. But his project of providing a Kantian justification for the uniform distribution of probabilities remained alive in German philosophy in the first decades of the twentieth century (Meinong 1915 [127]; Reichenbach 1916 [139]). John Maynard Keynes (1921)[99] brought it into the English literature. When asked about the philosophical basis of the classical probability calculus, many philosophers and mathematicians today will think about arguments for a uniform distribution of probabilities before they think about Cournot's principle.

3.27 Andrei Markov, 1856–1922

Markov, Chuprov's neighbor in Petersburg, learned about the growing field of mathematical statistics from Chuprov [132], and we see an echo of Cournot's principle in Markov's textbook. We find the following statement in Markov's

[5]In German, Spiel means "game" or "play", and Raum (plural Räume) means "room" or "space". In most contexts, Spielraum can be translated as "leeway" or "room for maneuver". For von Kries, the Spielraum for each case was the set of all arrangements of the circumstances that produce it.

textbook, which appeared in Russian in 1900. (The passage is on p. 12 of the German edition, which appeared in 1912 [116], p. 12.)

> The closer the probability of an event is to one, the more reason we have to expect the event to happen and not to expect its opposite to happen.
>
> In practical questions, we are forced to regard as certain events whose probability comes more or less close to one, and to regard as impossible events whose probability is small.
>
> Consequently, one of the most important tasks of probability theory is to identify those events whose probabilities come close to one or zero.

3.28 Karl Pearson, 1857–1936

Karl Pearson, in Mathematical Contributions to the Theory of Evolution.—III. Regression, Heredity, and Panmixia. Philosophical Transactions of the Royal Society of London, 1895, series A, vol. 186, pp. 252–318. With respect to Galton's "Special data" on heights:

> ...Thus difference in height is nine times, and the difference in correlation more than six times the corresponding probable error. It is absolutely necessary therefore to conclude that the Essex contribution differs significantly from the remainder of the data.

3.29 Guido Castelnuovo, 1865–1952

Other authors, including Chuprov, enunciated Cournot's principle in its weak form, and this can lead in a different direction. The weak principle combines with Bernoulli's theorem to produce the conclusion that an event's probability will *usually* be approximated by its frequency in a sufficiently long sequence of independent trials, a general principle that has the weak principle as a special case.

This was pointed out by Castelnuovo in his 1919 textbook [39, p. 108]. Castelnuovo called the general principle the *empirical law of chance* (la legge empirica del caso):

> In a series of trials repeated a large number of times under identical conditions, each of the possible events happens with a (relative) frequency that gradually equals its probability. The approximation usually improves with the number of trials. [39, p. 3]

Although the special case where the probability is close to one is sufficient to imply the general principle, Castelnuovo preferred to begin his introduction to the meaning of probability by enunciating the general principle, and so he can be considered a frequentist. His approach was influential at the time. Maurice Fréchet and Maurice Halbwachs adopted it in their textbook in 1924 [80]. It brought Fréchet to the same understanding of objective probability as Lévy: it is a physical constant that is measured by relative frequency [75, p. 5]; [77, pp. 45–46].

The weak point of Castelnuovo and Fréchet's position lies in the modesty of their conclusion: they conclude only that an event's probability is *usually* approximated by its frequency. When we estimate a probability from an observed frequency, we are taking a further step: we are assuming that what usually happens has happened in the particular case. This step requires the strong form of Cournot's principle. According to Kolmogorov (1956),p. 240 of the 1965 English edition), it is a reasonable step only if "we have some reason for assuming" that the position of the particular case among other potential ones "is a regular one, that is, that it has no special features".

3.30 Jacques Hadamard, 1865–1963

Hadamard, the preeminent analyst who did pathbreaking work on Markov chains in the 1920s (Bru 2003a)[33], made the point in a different way. Probability theory, he said, is based on two basic notions: the notion of perfectly equivalent (equally likely) events and the notion of a very unlikely event (Hadamard 1922, p. 289)[90]. Perfect equivalence is a mathematical assumption, which cannot be verified. In practice, equivalence is not perfect—one of the grains in a cup of sand may be more likely than another to hit the ground first when they are thrown out of the cup. But this need not prevent us from applying the principle of the very unlikely event. Even if the grains are not exactly the same, the probability of any particular one hitting the ground first is negligibly small. Hadamard cited Poincaré's work on the three-body problem in this connection, because Poincaré's conclusion is insensitive to how one defines the probabilities for the initial state. Hadamard was the teacher of both Fréchet and Lévy.

3.31 Émile Borel, 1871–1956

Although he never attributed it to Cournot, Borel stated the principle many times, often in a style more literary than mathematical or philosophical [19, 20, 22, 23]. According to Borel, a result of the probability calculus deserves to be called objective when its probability becomes so great as to be practically the same as certainty. He believed that what is negligible depends on the context; in 1939, we wrote that a probability of 10^{-6}, he decided, is negligible at the human scale, a probability of 10^{-15} at the terrestrial scale, and a probability of 10^{-50} at the cosmic scale [24, pp. 6–7].

Borel, sharpened his statement of the principle in the 1940s. In earlier years, he wrote frequently about the practical meaning of probabilities very close to zero or one, but it is hard to discern in these writings the philosophical principle, which we do find in Hadamard and Lévy, that interpreting a very small probability as impossibility is the only way of bringing probability theory into contact with the real world. But in the 1940s, we find the principle articulated very clearly. In his 1941 book, *Le jeu, la chance et les théories scientifiques* *modernes*[25], he calls it the "fundamental law of chance" (la loi fondamentale du hasard). Then, in 1943, on the first page of the text of his "Que sais-je?" volume, *Les probabilités et la vie*[26], he finally coined the name he used thereafter: "the only law of chance" (la loi unique du hasard). This name appears again in the 1948 edition of *Le Hasard* and the 1950 edition of *Éléments de la théorie des probabilités* (see also Borel 1950[28]). It was also popularized by Robert Fortet, in his essay in François Le Lionnais's *Les grands courants de la pensée mathématique*[107], first published in 1948[73, 107].

3.32 George Udny Yule, 1871–1951

Page 262–263 of the first edition of Yule's statistics textbook, published in 1911 [168]:

> We may now turn from these verifications of the theoretical results for various special cases, to the use of the formulae for checking and controlling the interpretation of statistical results. If we observe, in a statistical sample, a certain proportion of objects or individuals possessing some given character—say A's—this proportion differing more or less from the proportion which for some reason we expected, the question always arises whether the difference may be due to the fluctuations of simple sampling only, or may be indicative of definite differences between the conditions in the universe from which the sample has been drawn and the assumed conditions on which we based our expectation. Similarly, if we observe a different proportion in one sample from that which we have observed in another, the question again arises whether this difference may be due to fluctuations of simple sampling alone, or whether it indicates a difference between the conditions subsisting in the universes from which the two samples were drawn: in the latter case the difference is often said to be **significant**. These questions can be answered, though only more or less roughly at present, by comparing the observed difference with the standard-deviation of simple sampling. We know roughly that the great bulk at least of the fluctuations of sampling lie within a range of $\pm$ three times the standard-deviation; and if an observed difference from a theoretical result greatly exceeds these limits it cannot be ascribed to a fluctuation of "simple sampling" as defined in §8: it may therefore be significant. The "standard-deviation of simple sampling" being the basis of all such work, it is convenient to refer to it by a shorter name. The observed proportions of A's in given samples being regarded as differing by larger or smaller errors from the true proportion in a very large sample from the same material, the "standard-deviation of simple sampling" may be regarded as a measure of the magnitude of such errors, and may be called accordingly the **standard error**.

Page 263 of Yule 1911: The deviation observed is 5.1 times the standard error, and, practically speaking, could not occur as a fluctuation of simple sampling.

Page 265: If the observed difference is less than some three times ϵ_{12} it may have arisen as a fluctuation of simple sampling only. [Here ϵ_{12} is the standard error of the difference between two proportions.]

Page 266: As this difference is only slightly in excess of the standard error of the difference, for samples of 34 observations drawn from identical material, no definite significance could be attached to it—if it stood alone.

3.33 Aleksandr Chuprov, 1874–1926

Chuprov, who became professor of statistics in Petersburg in 1910, was the champion of Cournot's principle in Russia. Like the Scandinavians, Chuprov wanted to bridge the gap between the British statisticians and the continental mathematicians [150]. With some justice, he considered Cournot the founder of the philosophy of modern statistics [150, p. 86]. He put Cournot's principle—which he called "Cournot's lemma"—at the heart of this project; in a philosophical book he published in 1910 [41], he called it a basic principle of the logic of the probable. See [150, pp. 95–96].

Kolmogorov included Lévy's book and Slutsky's article in his bibliography, but not Chuprov's book. An opponent of the Bolsheviks, Chuprov was abroad when they seized power, and he never returned home. He remained active in Sweden and Germany, but his health soon failed, and he died in 1926, at the age of 52.

3.34 Maurice Fréchet, 1878–1973

Cournot's principle has many variations. Like probability, moral certainty can be subjective or objective. Some authors make moral certainty sound truly equivalent to absolute certainty; others emphasize its pragmatic meaning.

For our story, it is important to distinguish between the strong and weak forms of the principle (Fréchet 1951, p. 6[78]; Martin 2003[119]). The strong form refers to an event of small or zero probability that we single out in advance of a single trial: it says the event will not happen on that trial. The weak form says that an event with very small probability will happen very rarely in repeated trials.

Borel, Lévy, and Kolmogorov all enunciated Cournot's principle in its strong form. In this form, the principle combines with Bernoulli's theorem to produce the unequivocal conclusion that an event's probability will be approximated by its frequency in a particular sufficiently long sequence of independent trials. It also provides a direct foundation for statistical testing. If the empirical meaning of probability resides precisely in the non-happening of small-probability events singled out in advance, then we need no additional principles to justify rejecting a hypothesis that gives small probability to an event we single out in advance and then observe to happen (Bru 1999[32]).

Other authors, including Chuprov, enunciated Cournot's principle in its weak form, and this can lead in a different direction. See also §3.29.

3.35 Evgeny Slutsky, 1880–1948

The Russian statistician Evgeny Slutsky discussed Cournot's "lemma" in the following passage, translated from his lengthy and influential article on limit theorems, published in German in 1925 [152, pp. 17–19].

> ...It would therefore be worthwhile to analyze the conception of the law of large numbers proposed by Prof. Al. A. Chuprov and traced back to A. Cournot's views.[6] Here the essential point comes down to saying that the derivation of the law of large numbers is based not only the well-known theorems of probability calculus (from Bernoulli, Poisson, etc.), but also on a special lemma, by which it first actually becomes possible "from the world of probabilities, either large or small, to take ourselves over into the world of frequencies".[7] This lemma recognizes "the reality of an existing connection between small probability and rarity" by positing that "events with very small probabilities will not often happen".[8]
>
> If you want to consider this statement nomological, you again come into contradiction probability theory. No matter how small the probability of an event occurring, it can still occur any number of times in a row in a series of independent trials. Make the probability of an event as small as you want, it still *can* still happen arbitrarily many times in a sequence of *independent* trials. The probability of getting red ten thousand million times in a row in roulette, for example, is not an impossibility, but an extremely small but non-zero probability, and with a sufficiently large number of sequences of 10^{10} spins, a certain frequency of occurrences of sequences in which all 10^{10} spins produce red can be expected with the greatest certainty. "Even the smallest probability is still fundamentally different from impossibility; and we cannot bridge this gap, as we can let the numbers grow as much as we like." [9] Only this much is true, that the probability that a very improbable event will occur *frequently* is a very small quantity of far higher order than the probability of its one-time occurrence. In my opinion, it should not be claimed from A. Cournot's viewpoint that such a conception does not contain any statement about frequencies themselves, because his view is precisely that every statement about probabilities of frequencies is a

[6] Author's footnote: Al. A. Tschuprow, *Abhandlungen aus der Theorie der Statistik*, 2 Aufl. 1910 (Russian), p. 227 ff.

[7] Author's footnote: Ibid., p. 230.

[8] Author's footnote: Ibid., p. 230, 227.

[9] Author's footnote: J. v. Kries, op. cit. p. 21.

statement about the frequencies themselves. [10] To see this, you need not to get involved in any physical or ontological speculations, but merely to clarify the simple meaning of the corresponding propositions of the theory of probability. Then you see that under the relevant assumptions (Bernoullian, Poissonian, Markovian) there is almost full certainty that "events whose probabilities are very small will not often happen". If you remove the word "almost", you obtain the Cournotian lemma, which differs from the first statement *therefore not in the content but merely in the modality of the declaration*: what was asserted with only *almost* full certainty, the lemma wants to pass off as absolutely certain knowledge. And given our assumptions, that is certainly wrong.

Now for another possible interpretation. The above lemma can still be seen as an idiographic statement, as a statement about the actual structure of the world, or of the part of the world we are in.[11] It would then mean that although among all possible constellations of the elements of the world there are also ones that would necessarily present us with the very strangest events — so that for example all games of chance would be distorted as if by a demonic force, warm bodies would be heated by cold ones, human fate would seem to be guided by a star, and so forth — yet our world is not one of these exceptional worlds, but an *ordinary world*, so to speak.[12] We may consider this a possibly well-founded assumption, but this much is true: facts about the past of a chaotic event cast no light on its future. As a nomological statement the lemma was wrong, as an idiographic one it is useless. It teaches us nothing about the future fate of the world; it provides us no guarantee against the possibility of jumping into the wonder world of exceptional stochastic states. If we feel no great fear, the reason is simply that we are inclined to grant this possibility an immeasurably small probability. So the hypothesis of the stochastic ordinariness of the world does not justify the law of large numbers. Rather the law of large of

[10] Author's footnote: "The mathematical probability then becomes the measure of *physical probability*...the advantage of this is to clearly indicate the existence of a ratio ...found between the things themselves: a ratio that nature maintains and that observation reveals when trials are repeated enough". A. Cournot, *Essai sur les fondements de nos connaissances*, Nouvelle édition, Paris, 1912, p. 45 (Sperrdruck des Verfassers). On the concept of "physical impossibility" (or "impossibility in fact"), characteristic for his entire system yet no less clarified in the end, see his *Exposition de la théorie des chances et des probabilités* Paris, 1843, p. 79-80, 437-438. Compare J. v. Kries, *Ueber den Begriff der objektiven Möglichkeit*, "Vierteljahrsschrift f. wiss. Philos." 12 Jahrg, 1888.

[11] Author's footnote: Perhaps Al. A. Tschuprow's standpoint can be understood in this sense: *Abhandlungen aus der Theorie der Statistik*, 2 Aufl., S. 231, (Russian).

[12] Author's footnote: Compare Zilsel, *Versuch einer neuen Grundlegung der statischen Mechanik* "Monatshefte für Math. und Physik" Wien, 1921, Bd. XXXI, p. 153-154. His "verallgemeinerte Allagodenhypothese" is equivalent to the hypothesis of the ordinariness of the world. But the author errs insofar as he believes that his construction makes a stochastic standpoint avoidable.

large numbers creates the logical possibility of believing the stochastic ordinariness of the world for the future, since on the basis of all our stochastic-nomological knowledge it acquires a probability practically equivalent to absolute certainty.

Kolmogorov included this article in the bibliography of his 1933 *Grundbegriffe* [101].

3.36 Richard von Mises, 1883–1953

[164, Chapter 15, §4]

> In classical physical statistics one starts by making certain plausible assumptions, according to the methods of probability calculus, about initial probabilities as well as transition probabilities, and derives from them statements about the course of events to be expected with very high probability. The value of this "high" probability is so near to 1 that the statements are practically indistinguishable from those which are called "deterministic". In all cases that can be checked the agreement between observation and calculation proves to be excellent.

3.37 James V. Uspensky, 1883–1947

Uspensky was trained at the University of St. Petersburg and was a member of the Russian Academy of Scientists before emigrating to the United States. He became a professor at Stanford University. We find this passage in his *Introduction to Mathematical Probability*, published in 1937 [157, p. 8]:

> From our experience, we know that events with small probability seldom happen. ... the probability 999,999/1,000,000 may be considered, from a practical standpoint, as an indication of certainty. What limit for smallness of probability is to be set as an indication of practical impossibility? Evidently there is no general answer to this question. Everything depends on the risk we can face if, contrary to expectation, an event with a small probability should occur. Hence, the main problem of the theory of probability consists in finding cases in which the probability is very small or very near 1. ...

3.38 Paul Lévy, 1886–1971

It was Lévy, perhaps, who had the strongest sense of probability's being pure mathematics (he devoted most of his career as a mathematician to probability), and it was he who expressed most clearly in the 1920s the thesis that Cournot's principle is probability's only bridge to reality. In his *Calcul des probabilités*[108] Lévy emphasized the different roles of Hadamard's two basic notions. The notion of equally likely events, Lévy explained, suffices as a foundation for the mathematics of probability, but so long as we base our reasoning only on this notion, our probabilities are merely subjective. It is the notion of a very unlikely event that permits the results of the mathematical theory to take on practical significance ([108], pp. 21, 34; see also [110], p. 3). Combining the notion of a very unlikely event with Bernoulli's theorem, we obtain the notion of the objective probability of an event, a physical constant that is measured by relative frequency. Objective probability, in Lévy's view, is entirely analogous to length and weight, other physical constants whose empirical meaning is also defined by methods established for measuring them to a reasonable approximation ([108], pp. 29–30).

In his 1925 book [108], Lévy developed Jacques Hadamard's idea that probability theory is based on two fundamental notions:

1. equally probable events (événements également probables), and
2. event of very small probability (événement très peu probable).[13]

Whereas the notion of equally probable events expresses probability's subjective starting point, the notion of an event of very small probability allows us to connect probability to objective reality: we predict that the event will not happen. As Lévy further explained in his 1937 book ([110], p. 3),

> We can only discuss the objective value of the notion of probability when we know the theory's verifiable consequences. They all flow from this principle: a sufficiently small probability can be neglected. In other words: *a sufficiently unlikely event can in practice be considered impossible*.[14]

3.39 Oskar Anderson, 1887–1960

Chuprov's enthusiasm for Cournot and the principle was brought from Russian into German by Chuprov's student, Oskar Anderson, who spent the 1930s in Sofia and then moved to Munich in 1942. Anderson called the principle the "Cournotsche Lemma" or the "Cournotsche Brücke"—a bridge between mathematics and the world. We find both phrases already in Anderson's 1935 book[6, 5], but the book may have been less influential than an article Anderson contributed to a special issue of the Swiss philosophy journal *Dialectica*[7, 27, 111] in 1949, alongside articles by Borel and Lévy revisiting their versions of Cournot's principle. Fréchet took these articles as one of his themes in his presiding report at the session on probability at the Congrès International de Philosophie des Sciences[11] at Paris that same year (Fréchet 1951[78]), where

[13] Lévy devotes Chapter 1 to the first principle and Chapter 2 to the second. In the preface (p. viii), he cites a 1922 article [90] in which Hadamard stated the two principles.

[14] *In the original French:* Nous ne pouvons discuter la valeur objective de la notion de probabilité que quand nous saurons quelles sont les conséquences vérifiables de la théorie. Elles découlent toutes de ce principe: une probabilité suffisamment petite peut être négligée; en autre termes: *un événement suffisamment peu probable peut être pratiquement considéré comme impossible.*

he accepted the attribution of the principle to Cournot ("bien qu'il semble avoir été déjà plus ou moins nettement indiqué par d'Alembert") but suggested the appellation "principe de Cournot", reserving "lemma" as a purely mathematical term. It was normal for Fréchet to legislate on terminology; from 1944 to 1948 he had led the effort by the Association Française de Normalisation to standardize probability terminology and notation, putting in place appellations such as Borel-Cantelli, Kolmogorov-Smirnov, etc. (Bru 2003b[34], Pallez 1949[133]). Fréchet had second thoughts about giving so much credit to Cournot; when he reprinted his 1949 report as a chapter in a book in 1955[79], he replaced "principe de Cournot" with "principe de Buffon-Cournot". But here no one else appears to have followed his example.

Both Anderson and the Dutch mathematical statistician David Van Dantzig argued for using Cournot's principle as the foundation for statistical testing: Anderson in *Dialectica* (Anderson 1949[7]), and Van Dantzig at the meeting in Paris (Van Dantzig 1951[158]). Neyman found this view of statistical testing incomprehensible; at the same meeting in Paris he said Anderson was the "only contemporary author I know who seems to believe that the inversion of the theorem of Bernoulli is possible" (Neyman 1951, p. 90)[129]. The German mathematical statistician Hans Richter, also in Munich, emphasized Cournot's principle in his own contributions to *Dialectica* (Richter 1954; von Hirsch 1954)[140, 161] and in his probability textbook (Richter 1956)[141], which helped bring Kolmogorov's axioms to students in postwar Germany. As a result of Richter's book, the name "Cournotsche Prinzip" is fairly widely known among probabilists in Germany.

3.40 R. A. Fisher, 1890–1962

Laplace and Poisson were accustomed to explaining probability in terms of sampling from an urn with infinitely many balls or tickets of different colors. In the 1920s, Fisher similarly used the metaphor of a "hypothetical infinite population"; see for example his celebrated 1922 article on theoretical statistics [69]. What did Fisher mean when he wrote about frequencies in such a population? This question was raised in 1925 by the British mathematician William Burnside [37, 38]. As Burnside pointed out, we can define limiting relative frequencies if we order the elements of a countably infinite set, but the limit depends on the ordering. The limiting relative frequency of natural numbers divisible by 7 among all the natural numbers is $1/7$ if we consider the numbers in their natural ordering, but other orderings give other limits. Fisher responded to Burnside's question in a "prefatory note" to an article he published in 1925 [70]; see [1].

Fisher's prefatory note can be interpreted as conceding that the notion of an infinite hypothetical population is not really needed. All that is needed is the notion that a sample is typical with respect to a particular sampling distribution. In practice, Fisher made typicality operational by means of significance testing. This reduces the picture to Cournot's principle — the principle that a probability model is connected to observed or observable phenomena by the assumption that an event of small probability, selected in advance, has not happened or will not happen.

3.41 Thornton Fry, 1892–1991

Fry's 1928 textbook, *Probability and its Engineering Uses* [84], grew out his teaching at Bell Telephone Laboratories and at MIT.

To illustrate the relation between certainty and high probability, Fry imagined a sequence of urns, the mth one containing one white ball and m black balls. The greater m, the greater the probability that a ball draw from the urn is black. But this probability is never one. As Fry wrote on p. 88, "The limiting condition is certainty, but that limit cannot be reached."

On p. 100, Fry stated and discussed Bernoulli's theorem as follows.

> BERNOULLI'S THEOREM: *If the chance of an event occurring upon a single trial is p, and if a number of independent trials are made, the probability that the ratio of the number of successes to the number of trials differs from p by less than any preassigned quantity, however small, can be made as near certainty as may be desired by taking the number of trials sufficiently large.*
>
> Sometimes the content of a theorem such as this is made clearer by throwing mathematical discretion to the winds and stating it in the form of every-day language. The present appears to be a case of this sort, and therefore we restate the theorem as follows:
>
> *If the probability of an event is p, and if an infinity of trials are made, the proportion of successes is sure to be p.*
>
> ... the statement is as certainly "true" in one sense of the word, as it is *not* "true" in another. ... it fails to stand the test of mathematical rigor, ... It is therefore not a fit foundation for a mathematical theory. but our every-day life is not conducted on such rigorous requirements as to "truth." You say, "Are you sure that he is coming tomorrow?" and receive the answer, "Yes." Both you and your informant understand what you mean: the event is contingent upon his not dying, for example, and perhaps on many other unforeseen circumstances. It is, in fact, not sure at all; it is merely very probable: so probable that the residual doubt is not work expression. Our statement is in the same class. In fact, the residual doubts are even vastly smaller, and may quite properly remain unexpressed.

3.42 Harald Cramér, 1893–1985

Harald Cramér, who felt fully in tune with Kolmogorov's frequentism, repeated the key elements of his philosophy in his 1946 book [47, 148–150]. Cramér expressed Kolmogorov's caution that the theory of probability applies only under certain conditions by saying that only certain experiments are random. In the context of a random experiment $\mathfrak{E}$, Cramér stated Kolmogorov's Principle **A** in this way:

> *Whenever we say that the probability of an event E with respect to*

> *an experiment $\mathfrak{E}$ is equal to P, the concrete meaning of this assertion will thus simply be the following: In a long series of repetitions of $\mathfrak{E}$, it is practically certain that the frequency of E will be approximately equal to P. — This statement will be referred to as the frequency interpretation of the probability P.*

He stated Kolmogorov's Principle **B** as a principle applying to an event whose probability is very small or zero:

> *If E is an event of this type, and if the experiment $\mathfrak{E}$ is performed one single time, it can thus be considered as practically certain that E will not occur.* — This particular case of the frequency interpretation of a probability will often be applied in the sequel.

The final sentence of this passage shows Cramér to be a less careful philosopher than Kolmogorov, for it suggests that Principle **B** is a particular case of Principle **A**, and this is not strictly true. As we noted when discussing Castelnuovo's views, the weak form of Cournot's principle is indeed a special case of Principle **A**. But Principle **B** is the strong form of Cournot's principle, and this is not merely a special case of Principle **A**.

3.43 Jerzy Neyman, 1894–1981

Cite his writings on stochastic processes and frequentism. Look at his two books on Hathi.

[130], p. 625

> The fourth period in the history of indeterminism, currently in full swing, the period of "dynamic indeterminism," is characterized by the search for evolutionary chance mechanisms capable of explaining the various frequencies observed in the development of the phenomena studied. The chance mechanism of carcinogenesis and the chance mechanism behind the varying properties of the comets in the Solar System exemplify the subjects of dynamic indeterministic studies. One might hazard the assertion that every serious contemporary study is a study of the chance mechanism behind some phenomena. The statistical and probabilistic tool in such studies is the theory of stochastic processes, now involving many unsolved problems. In order that the applied statistician be in a position to cooperate effectively with the modern experimental scientist, the theoretical equipment of the statistician must include familiarity and capability of dealing with stochastic processes.

3.44 David van Dantzig, 1900–1959

Both Anderson and the Dutch mathematical statistician David Van Dantzig argued for using Cournot's principle as the foundation for statistical testing: Anderson in *Dialectica* (Anderson 1949[7]), and Van Dantzig at the meeting in Paris

(Van Dantzig 1951[158]). Neyman found this view of statistical testing incomprehensible; at the same meeting in Paris he said Anderson was the "only contemporary author I know who seems to believe that the inversion of the theorem of Bernoulli is possible" (Neyman 1951, p. 90)[129]. The German mathematical statistician Hans Richter, also in Munich, emphasized Cournot's principle in his own contributions to *Dialectica* (Richter 1954; von Hirsch 1954)[140, 161] and in his probability textbook (Richter 1956)[141], which helped bring Kolmogorov's axioms to students in postwar Germany. As a result of Richter's book, the name "Cournotsche Prinzip" is fairly widely known among probabilists in Germany.

3.45 Karl Popper, 1902–1994

Popper adopted a form of Cournot's principle in his *Logik der Forschung*, first published in 1935 [136]. On p. 191 of the English version, published in 1958, we find this passage:

> . . . a physicist is usually quite well able to decide whether he may for the time being accept some particular probability hypothesis as 'empirically confirmed', or whether he ought to reject it as 'practically falsified', *i.e.*, as useless for purposes of prediction. It is fairly clear that this 'practical falsificatio' can be obtained only through a methodological decision to regard highly improbable events as ruled out—as prohibited. But with what right can they be so regarded? Where are we to draw the line? Where does this 'high improbability' begin?
>
> Since there can be no doubt, from a purely logical point of view, about the fact that probability statements cannot be falsified, the equally indubitable fact that we use them empirically must appear as a fatal blow to my basic ideas on method which depend crucially upon my criterion of demarcation. Nevertheless I shall try to answer the questions I have raised—which constitute the problem of decidability—by a resolute application of these very ideas. . .

In the following pages, discusses at length how he proposes to qualify Cournot's principle.

On page 150 of the English edition, we writes in a footnote:

> . . . I now believe that Bernoulli's theorem may serve as a 'bridge' *within* an objective theory—as a bridge from propensities to statistics. See also appendix *ix and sections *55 to *57 of my *Postscript.*

3.46 Abraham Wald, 1902–1950

Wald became a mathematician working with Karl Menger in Vienna and participating in his seminar. Both Menger and Wald fled to the United States as Hitler seized Austria. Menger became a professor at Notre Dame in Indiana; Wald became a professor at Columbia in New York. In February 1941, Wald gave a series of lectures at Notre Dame entitled, “On the principles of statistical inference”. He began with this introduction ([167], pages 1–2, references omitted):

> The purpose of statistics, like that of geometry or physics, is to describe certain real phenomena. The objects of the real world can never be described in such a complete and exact way that they could form the basis of an exact theory. We have to replace them by some idealized objects, defined explicitly or implicitly by a system of axioms. For instance, in geometry we define the basic notions “point,” “straight line,” and “plane” implicitly by a system of axioms. They take the place of empirical points, straight lines, and planes which are not capable of definition. In order to apply the theory to real phenomena, we need some rules for establishing the correspondence between the idealized objects of the theory and those of the real world. These rules will always be somewhat vague and can never form part of the theory itself.
>
> The purpose of statistics is to describe certain aspects of mass phenomena and repetitive events. The fundamental notion used is that of “probability.” In the theory it is defined either explicitly or implicitly by a system of axioms. For instance, Mises defines the probability of of an event as the limit of the relative frequency of this event in an infinite sequence of trials satisfying certain conditions. This is an explicit definition of probability. Kolmogoroff defines probability as a set function which satisfies a certain system of axioms. These idealized mathematical definitions are related to the applications of the theory by translating the statement “the event E has the probability p” into the statement “the relative frequency of the event E in a long sequence of trials is approximately equal to p.” This translation of a theoretical statement into an empirical statement is necessarily somewhat vague, for we have said nothing about the meanings of the words “long” or “approximately.” But such vagueness is always associated with the application of theory to real phenomena.
>
> It should be remarked that instead of the above translation of the word “probability” it is satisfactory to use the following somewhat simpler one: “The event E has a probability near to one” is translated into “it is practically certain that the event E will occur in a single trial.” In fact, if an event E has the probability p then, according to a theorem of Bernoulli, the probability that the relative frequency of E in a sequence of trials will be in a small neighborhood of p is arbitrarily near to 1 for a sufficiently long sequence of trials. If we translate the expression “probability near 1” into “practical certainty,” we obtain the statement “it is practically certain that the relative frequency of E in a long sequence of trials will be in a small neighborhood of p.”

3.47 Marshall Stone, 1903–1989

As recognition for his accomplishments in mathematics, Stone was asked to deliver the Josiah Willard Gibbs Lecture at the meeting of the American Mathematical Society in December 1956. In this wide-ranging lecture, he made the following comments on mathematical statistics [155, p. 71].

> Because of the tremendous scope of its applications, ranging all the way from theoretical physics to the social sciences, mathematical statistics has undergone a rapid and extensive development so that it now enjoys the status of an independent discipline. Mathematically we now know that it is a branch of measure theory, which is linked with the real world through a few simple principles embodying the essence of inductive reasoning. There is, of course, some disagreement as to how these principles should be formulated. It has always seemed to me that they all have to be based on a single rule of thumb, "A sufficiently improbable event may be ignored." In making decisions according to this rule, the role of mathematics is to provide the measure-theoretic calculations of interrelated probabilities and the role of practical insight is to determine for each concrete situation which probabilities are sufficiently small. Why the real world should be amenable to such a rule is, I think, a philosophical question no more—and no less—mysterious than the problem of why it should be amenable to logic.

3.48 Andrei Kolmogorov, 1903–1987

When he wrote his *Grundbegriffe* in 1933, Kolmogorov would have seen many versions of Cournot's principle, by authors whom he respected. He was surely familiar with Markov's textbook.

Cournot's principle was emphasized by many of the Russian and French mathematicians from whom Kolmogorov learned about probability theory, including Chuprov, Slutsky, Borel, Lévy, and Fréchet [148]. He had cited Lévy's 1925 book in his 1931 article on Markov processes and subsequently, during his visit to France, he had spent a great deal of time talking with Lévy about probability. But in the Soviet context, it was also mandatory to highlight the primacy of mass phenomena. So it is not surprising that in his brief explanation of how probability theory can be used in the celebrated monograph he published in 1933 in German [101], he mentions both frequency and Cournot's principle :

> Under certain conditions, that we will not go into further here, we may assume that an event A that does or does not occur under conditions $\mathfrak{S}$ is assigned a real number $\mathsf{P}(A)$ with the following properties:
>
> A. One can be practically certain that if the system of conditions $\mathfrak{S}$ is repeated a large number of times, n, and the event A occurs m times, then the ratio m/n will differ only slightly from $\mathsf{P}(A)$.

B. If $\mathsf{P}(A)$ is very small, then one can be practically certain that the event A will not occur on a single realization of the conditions $\mathfrak{S}$.

As Cournot emphasized, many events do not have objective mathematical probabilities; we can give them only philosophical probabilities. Perhaps not even a superior intelligence could give them objective mathematical probabilities. Today many people think differently; many think, or assume without thinking, that if an event is not determined, then it has an objective mathematical probability. So it is worth noting that Kolmogorov, like many mathematicians of his time who worked with the concept of objective probability, thought that only some events have objective probabilities. He put the matter this way in 1951 in in the *Great Soviet Encyclopedia*:

> Certainly not every event whose occurrence is not uniquely determined under given conditions has a definite probability under these conditions. The assumption that a definite probability (i.e. a completely defined fraction of the number of occurrences of an event if the conditions are repeated a large number of times) in fact exists for a given event under given conditions is a hypothesis which must be verified or justified in each individual case.

3.49 Carl Hempel, 1905–1997

In 1965, in his *Aspects of Scientific Explanation* [93, p. 387], Hempel quotes Cramèr's formulation, in which Cournot's principle is a special case of Kolmogorov's principle A.

3.50 Hans Freudenthal, 1905–1990

In an expository article on probability published in 1960 in *Synthese* [82, pp. 205–206], Freudenthal explained Cournot's principle this way:

> Arbuthnot's statistical inference with its appeal to a model comprising a stochastic device has become exemplary. In the same way D. Bernoulli and Laplace proved that it cannot be by chance that the inclinations of the planetary orbits against the zodiac are as small as they are found by astronomical evidence. Laplace used this as an argument for his cosmogonic theory. The common aim of those statisticians is a statistical reliability of their judgements of nearly 100%. (At the same time the judgements themselves are rather crude, mostly decisions between some 'yes' or 'no'.) Though in modern statistics, we are acquainted with more refined methods, there are still many opportunities to use Arbuthnot's reasoning. Philosophers call it Cournot's principle: if something is proved to be extremely improbable, we are allowed to state that it is impossible. The statement of its impossibility is nearly always stressed by an

> appeal to something like the urn model. The event to be disproved appears to be as improbable as a large sequence of heads or sixes, when tossing a coin or throwing a dice, and so it is impossible.

Additional quotations from [81, 83].

3.51 Bruno de Finetti, 1906–1985

De Finetti participated in the 1949 Paris conference where Fréchet coined the phrase in French: *principe de Cournot*. Shortly afterwards, he brought the name into English, ridiculing it in 1951[55] as "the so-called principle of Cournot".

But while he had no use for Cournot's notion that predicting events of high probability is the only way of connecting a system of probabilities with phenomena, de Finetti had his own way of making sense of the idea that we do predict events when they have high probability. As he explained in a note written in 1951 [56, p. 235] that Fréchet published in 1955 in his *Les mathématiques et le concret*, he did not really disagree with the statement that one should act as if an event with a very small probability should not happen. Rather he took the principle as a tautology, a consequence of the subjective definition of probability, not a principle standing outside probability theory and relating it to the real world; see also Dawid 2004 [53].

3.52 Joseph Doob, 1910–2004

Doob's most explicit statement of Cournot's principle comes in a historical essay he published in 1976 [63, p. 201–202]. There he asks "what principle should be used to translate mathematical probability theorems into real life" and answers thus:

> If one starts with mathematical probability theory the obvious general operational translation principle is that one should ignore real events that have small probabilities. How small is "small" depends on the context, for example, the demands of a client on a statistician. Somewhat more precisely, one first makes a judgment on the possibility of the application of probability in a given context; if so, one then sets up a model and comes to operational decisions based on the principle that hypotheses must be reexamined if they ascribe small probability to a key event that actually happens. (This is, of course a great oversimplification.) ...

In [62], which derives from his discussion with von Mises at Dartmouth in 1940, Doob does not state Cournot's principle directly, but it is suggested by his explanation that practice depends on various forms of the law of large numbers.

3.53 Jean Ville, 1910–1989

Ville was a student of Maurice Fréchet and Émile Borel in Paris. In the path-breaking doctoral thesis that he defended in 1939 he showed that events of probability zero for a sequence of random variables can be identified game-theoretically: an event A has measure zero if and only if there is a strategy for betting on the variables (step-by-step as they are observed) that multiplies the capital it risks by an infinite factor when A happens.

In addition to the official version of the thesis, Ville had written two philosophical chapters, an introduction and a conclusion. Borel quickly arranged for the expanded version to be be published as a book [160]. On pp. 9–10 of the introductory chapter, Ville states a version of Cournot's principle. The passage, quoted here in loose translation from the French, begins with a reference to the standard practice of introducing probability theory by stating axioms for probabilities.

> ... The theory thus constructed is logically correct, but the coefficients thus introduced must be interpreted. For this, we use the subjective value of large probabilities, already highlighted by Laplace. In this way we can take the basis of the axiomatic theory to be the following: *Given a collection of random events, we can associate coefficients between* 0 *and* 1 *with them, such that if we compose these coefficients according to the rules laid down as axioms, the events having probabilities very close to* 1 *are practically certain (and therefore those whose probabilities are very small are practically impossible).*
>
> We can therefore say, with Mr. Fréchet: *The probability of an event in a specified category of trials is a physical constant, depending on the event and the category of trials, for which one obtains an empirical value by conducting a large number of independent trials and noting the frequency of the event.*
>
> Empirical value means a value that has little chance of being far from the true value. So in the interpretation, we constantly come back to the notion of "practical certainty" interpreting the probability close to 1. So the axiomatic theory is can be verified indirectly.
>
> In this way, we deal with two kinds of probabilities in the axiomatic theory: those that are close to 0 or to 1, which have a subjective meaning, quasi-impossibility or quasi-certainty, and those that are close neither to zero nor to 1, which have no subjective meaning when taken in isolation. It is precisely this lack of meaning for "middle" probabilities that is bothersome in the axiomatic theory: a proposition like "the probability of heads is $\frac{1}{2}$" has no value in isolation and is not directly verifiable. If the experiment is repeated, we deduce a verifiable proposition from inverifiable propositions. This seems to be a defect here; we are tempted to consider only sufficiently extended sequences of experiments, because no proposition is usable except in combination with a large number of other propositions: this leads to statistical theory and the negation not only of *subjective value* but even *existence* of probability for an isolated

event.

Ville may have been the first to state so clearly that only probabilities close to zero or one have meaning. This idea was later repeated, with less hesitation, by Kolmogorov's students. It has also been stated in the context of statistical mechanics by philosophers of physics; see [125].

3.54 Trygve Haavelmo, 1911–1999

Haavelmo's article, "The probability approach to econometrics" [89], is often seen as the founding charter of modern econometrics [128]. The article's most fundamental point was Cournot's principle.

As Haavelmo explained, econometricians had been reluctant to adopt probability as a foundation for their work because they incorrectly assumed that probability is applicable only in situations like those studied by the British school of statistics, where a large sample is drawn from a stable population. He made the point as follows (pages 477–478):

> The reluctance among economists to accept probability models as a basis for economic research has, it seems, been founded upon a very narrow concept of probability and random variables. Probability schemes, it is held, apply only to such phenomena as lottery drawings, or, at best, to those series of observations where each observation may be considered as an independent drawing from one and the same 'population'. From this point of view it has been argued, e.g., that most economic time series do not conform well to any probability model, 'because the successive observations are not independent'. But it is *not* necessary that the observations should be independent and that they should all follow the same one-dimensional probability law. It is sufficient to assume that the *whole set* of, say n, observations may be considered as *one* observation of n variables (or a 'sample point') following an n-dimensional *joint* probability law, the 'existence' of which may be purely hypothetical. Then, one can test hypotheses regarding this joint probability law, and draw inferences as to its possible form, by means of *one* sample point (in n dimensions). Modern statistical theory has made progress in solving such problems of statistical inference.
>
> In fact, if we consider actual economic research — even that carried on by people who oppose the use of probability schemes — we find that it rests, ultimately, upon some, perhaps very vague, notion of probability and random variables. For whenever we apply a theory to facts we do not — and we do not expect to – obtain exact agreement. Certain discrepancies are classified as 'admissible', others as 'practically impossible' under the assumptions of the theory. And the *principle* of such classification is itself a theoretical scheme, namely one in which the vague expressions 'practically impossible'

> or 'almost certain' are replaced by 'the probability is near to zero', or 'the probability is near to one'.
>
> This is nothing but a convenient way of expressing opinions about real phenomena. But the probability concept has the advantage that it is 'analytic', we can derive new statements from it by the rules of logic. Thus, starting from a purely formal probability model involving certain probabilities which themselves may not have any counterparts in real life, we may derive such statements as 'The probability of A is almost equal to 1'. Substituting some real phenomenon for A, and transforming the statement 'a probability near to 1' into 'we are almost sure that A will occur', we have a statement about a real phenomenon, the truth of which can be tested.
>
> The class of scientific statements that can be expressed in probability terms is enormous. In fact, this class contains all the 'laws' that have, so far, been formulated. For such 'laws' say no more and no less than this: The probability is almost 1 that a certain event will occur.

Haavelmo went on to explain that a probability law can be tested based on one observation because it makes predictions with very high probability about that one observation, and such predictions are the only kind of prediction science can ever make:

> The class of scientific statements that can be expressed in probability terms is enormous. In fact, this class contains all the 'laws' that have, so far, been formulated. For such 'laws' say no more and no less than this: The probability is almost 1 that a certain event will occur.

3.55 Hans Richter, 1912–1978

The German mathematical statistician Hans Richter, who taught in Munich, emphasized Cournot's principle in his contributions to *Dialectica* in 1954 [140, 161] and in his 1956 probability textbook [141], which helped bring Kolmogorov's axioms to students in postwar Germany. As a result of the book, the name "Cournotsche Prinzip" became fairly widely known among probabilists in Germany.

3.56 Charles Stein, 1920–2016

The following is excerpted from an interview by Morris H. DeGroot, conducted in 1983 and published in *Statistical Science* in 1986 [57, pp. 459–460].

From interview by DeGroot

DeGroot: Let's talk about probability for a moment. You say that the notion of subjective probability is unacceptable to you. What definition of probability do you use?

Stein: Essentially Kolmogorov's. That it is a mathematical system.
DeGroot: Simply any set of numbers that satisfies the axioms of the calculus of probabilities.
Stein: Yes.
DeGroot: But what do these numbers represent in the real world?
Stein: Well, there is no unique interpretation. And of course I'm talking about Kolmogorov's old interpretation of probability and not the complexity interpretation. In his book he mentions briefly two aspects of the interpretation. The first is the traditional relative frequency of occurrence in the long run. And the second is that when one puts forward a probabilistic model that is to be taken completely seriously for a real world phenomenon, then one is asserting in principle that any single specified event having very small probability will not occur. This, of course, combined with the law of large numbers, weak or strong, really is a broader interpretation than the frequency notion. So, in fact, the frequency interpretation in that sense is redundant. This doesn't answer the question, "When I say the probability is 1/6 that this die will come up 6 on the next toss, what does that statement mean?" But then in no serious work in any science do we answer the question, "What does this statement mean?" It is an erroneous philosophical point of view that leads to this sort of question.

3.57 Yuri Prokhorov, 1923–2013

Yuri Vasilevich Prokhorov and Boris Aleksandrovich Sevast'yanov (1923–2013) were both mentored in mathematical probability by Andrei Kolmogorov at Moscow State University in the 1950s.

In their article on probability in the *Soviet Mathematical Encyclopedia* in the 1970s [137], Prokhorov and Sevast'yanov echoed Jean Ville's statement that only probabilities close to 0 or 1 have direct meaning.

3.58 David R. Cox, 1924–2022

The term *repeated sampling principle* was coined by Cox and his junior colleague David Hinkley in their 1974 textbook [46, p. 45]:

> According to the strong repeated sampling principle, statistical procedures are to be assessed by their behavior in hypothetical repetitions under the same conditions. This has two facets. Measures of uncertainty are to be interpreted as hypothetical frequencies in long run repetitions; criteria of optimality are to be formulated in terms of sensitive behaviour in hypothetical repetitions.
>
> The argument for this is that it ensures a physical meaning for the quantities that we calculate and that it ensures a close relation between the analysis we make and the underlying model which is regarded as representing the "true" state of affairs.

3.59 John Stewart Bell, 1928–1990

Page 122 of [13], reprinting [12]:

> ... the *typical track*, if long enough, will serve to test predictions. ... The relevance of this remark is that later we are concerned with theories of the universe as a whole. Then there is no opportunity to repeat the experiment; history is given to us once only. We are in the position of having a single track, and it is important that the theory has still something to say—provided that this single track is not a freak, but a typical member of the hypothetical ensemble of universes that would exhibit the complete quantum distribution of tracks.
>
> ... In the same way as for the α particle track it follows from the theory that the 'typical' world will approximately realize quantum mechanical distributions over such approximately independent components. The role of the hypothetical ensemble is precisely to permit definition of the word 'typical'.

3.60 Henry Kyburg, Jr., 1928–2007

Kyburg, a professor of philosophy and computer of science at the University of Rochester, developed his own concept of practical certainty at length in his 1990 book *Science & Reason* [102]. As he explained on pp. 65–68, he distinguished between practical certainty and evidential certainty, with corresponding *bodies of knowledge*, or sets of propositions:

- an *evidential corpus*, consisting of the propositions "acceptable as evidence in a certain context", and
- and a larger *practical corpus*, consisting of propositions that may be only practically certain.

"The level of practical certainty," he wrote, "is indeed arbitrary, though no more arbitrary than the corresponding values $\alpha = .10$, .05, and .01 so popular in applied statistics."

Kyburg's practical corpus was not closed under conjunction. As he explained,

> ... the set of practical certainties is weakly deductively closed: it contains the deductive consequences of every statement it contains. It is subject to the lottery "paradox" insofar as it may contain each of a set of statements that are jointly inconsistent. But it does not uselessly contain all statements, because it contains no explicitly contradictory statement. Nor does it contain both a statement and its denial, so long as the level of acceptance is chosen to be greater than .5...

3.61 Hugh Everett III, 1930–1982

Quotation from Everett's 1957 thesis: Pages 70–71 of [60]

> In the language of subjective experience, the observer which is described by a typical element, $\psi'_{ij\ldots k}$, of the superposition has perceived an apparently random sequence of definite results for the observations. It is furthermore true, since in each element the system has been left in an eigenstate of the measurement, that if at this stage a redetermination of an earlier system observation (S_l) takes place, every element of the resulting final superposition will describe the observer with a memory configuration of the form $[\ldots, a_i^1, \ldots, \underline{a}_j^l, \ldots, a_k^r, \ldots, \underline{a}_j^l]$ in which the earlier memory coincides with the later—i.e., the memory states are correlated. It will thus appear to the observer which is described by a typical element of the superposition that each initial observation on a system caused the system to "jump" into an eigenstate in a random fashion and thereafter remain there for subsequent measurements on the same system. Therefore, qualitatively, at least, the probabilistic assertions of Process 1 *appear* to be valid to the observer described by a typical element of the final superposition.
>
> In order to establish quantitative results, we must put some sort of measure (weighting) on the elements of a final superposition. This is necessary to be able to make assertions which will hold for almost all of the observers described by elements of a superposition. In order to make quantitative statements about the relative frequencies of the different possible results of observation which are recorded in the memory of a typical observer we must have a method of selecting a typical observer.
>
> ...
>
> The situation here is fully analogous to that of classical statistical mechanics, where one puts a measure on trajectories of systems in the phase space by placing a measure on the phase space itself, and then making assertions which hold for "almost all" trajectories (such as ergodicity, quasi-ergodicity, etc). This notion of "almost all" depends here also upon the choice of measure, which is in this case taken to be Lebesgue measure on the phase space. One could, of course, contradict the statements of classical statistical mechanics by choosing a measure for which only the exceptional trajectories had nonzero measure. Nevertheless the choice of Lebesgue measure on the phase space can be justified by the fact that it is the only choice for which the "conservation of probability" holds, (Liouville's theorem) and hence the only choice which makes possible any reasonable statistical deductions at all.
>
> In our case, we wish to make statements about "trajectories" of observers. However, for us a trajectory is constantly branching

> (transforming from state to superposition) with each successive measurement. To have a requirement analogous to the "conservation of probability" in the classical case, we demand that the measure assigned to a trajectory at one time shall equal the sum of the measures of its separate branches at a later time. This is precisely the additivity requirement which we imposed and which leads uniquely to the choice of square-amplitude measure. Our procedure is therefore quite as justified as that of classical statistical mechanics.

3.62 Terrence Fine, 1939–2021

Discuss the relevance of Fine's suggestion, in 1976 [67], that randomness is what remains when we have made the best predictions we can. This viewpoint has been strengthened by Vovk's work on defensive forecasting [149, Ch. 12] and related work by other authors.

3.63 Per Martin-Löf, born 1942

In Martin-Löf's pathbreaking article on the definition of random sequences, published in 1966, we find this passage [123, p. 616]:

> The interpretation of a probability is currently (e.g., in the Grundlagen by Kolmogorov) governed not only by the clause that the relative frequency in a large number of repetitions of the experiment should be close to it, but also by the following somewhat obscure additional clause. If the probability is very small, we should be practically sure that the event does not occur in a single trial.

Per Martin-Löf has told me that he learned Cournot's principle from Borel rather than from Kolmogorov. See also [16, 122].

3.64 Donald Gillies, born 1944

In a 1973 book entitled *An Objective Theory of Probability* [88], the British author Donald Gillies proposed a philosophical account of significance testing. According to Gillies, the distribution of a random variable ξ is *falsifiable* distribution if ξ's possible values can be partitioned into sets A and C such that

1. ξ's probability of being in C is smaller than some suitably small constant,
2. for each $x \in C$, the ratio $f(x)/f_{\max}$, where f is ξ's probability density and $f_{\max}$ is f's maximum value, is smaller than some other suitably small constant,
3. $f_{\max}$ "is in some sense representative" of f's values for points in A.

Gillies wrote that when a falsifiable distribution follows from a hypothesis H, and "we test H by means of ξ we can be said to be predicting $\xi \in A$, and are regarding our prediction as falsified if if $\xi \notin A$".

Gillies's proposal did not prove appealing to statisticians, at least in part because the ratio $f(x)/f_{\max}$ depends on ξ's scale of measurement. In the continuous case, this ratio will change if ξ is transformed non-linearly. In the discrete case, it will usually change if categories are subdivided. The extent to which Gillies was out of step with statisticians is revealed by his use of "likelihood" to name the ratio $f(x)/f_{\max}$. Statisticians invariably follow Fisher by using "likelihood" for a quantity that is not sensitive to ξ's scale of measurement.

Gillies presented his proposal as a way of squaring statistical testing with Karl Popper's philosophy of falsification. He reviewed the thinking of a number of authors whom I have quoted in this paper. He did not quote Cournot, but he quoted Kolmogorov's conditions A and B. He took as his starting point "the rule of d'Alembert and Buffon", "which stated roughly that we will regard a hypothesis H as falsified if the observed event has a low probability given H (p. 167)."

3.65 Colin Howson, 1945–2020

The British philosopher Colin Howson, an outspoken Bayesian, dismissed Cournot's principle in 1995 with the following argument [94, pp. 16–17]:

> A century and a half ago A. A. Cournot tried to answer this objection by proposing that we treat small probabilities as impossibilities, so that the assumptions of independence etc. become in effect testable hypotheses which can be falsified by observing whether the observed relative frequencies lie within the predicted bands. But a simple consequence of Cournot's rule is that almost every hypothesis of use to statistics is a priori declared false by it. For example, consider the hypothesis which says that a sequence X; of random variables is i.i.d. with a specified probability p strictly between 0 and 1. Then whatever small number ϵ is declared to be the lower limit of physical probability, for some value of n each sequence of outcomes will be assigned a probability smaller than ϵ. Probabilities even of 0 cannot be consistently regarded as impossibilities so long as one countenances the possible truth of hypotheses ascribing continuous distributions to any variate. Gillies [1973] has proposed a more elaborate version of Cournot's rule, but this turns out to be equally unsound (Redhead [1974]). Nor can the apparently more sophisticated Neyman-Pearson theory of statistical tests help out here, for the desirability of minimizing type 1 and type 2 errors explicitly assumes that probabilities approximate long-run relative frequencies.

For Gillies, see §3.64 and [88]. For Redhead, see [138].

4 Acknowledgments

The people who have helped me with this enterprise over the past 25 years are legion. I am particularly indebted, however, to Thierry Martin's pioneering scholarship on Cournot [117, 118, 119, 121, 120] and to the encouragement of all things Cournot by the Centre Cournot and its director, Jean-Philippe Touffut. More recent impetus comes from the colleagues at Rutgers University and around the world who have participated since 2016 in the Rutgers Seminar on the Foundations of Probability. I am especially grateful to the organizers of this seminar, including the philosophers Barry Loewer, Eddy Chen, Isaac Wilhelm, and Snow Zhang and the statisticians Harry Crane and Robin Gong.

References

[1] John Aldrich. Burnside's engagement with the "modern theory of statistics". *Archive for History of Exact Sciences*, 63:51–79, 2009. 37

[2] John Aldrich. The econometricians' statisticians, 1895–1945. *History of Political Economy*, 42:111–154, 2010. 7

[3] John Aldrich. W. e. johnson and cambridge thought on probability. *International Journal of Approximate Reasoning*, 141:146–158, 2022. 24

[4] André-Marie Ampère. *Considérations sur la théorie mathématique du jeu.* Perisse, Lyon, 1802. 9, 18

[5] Oskar Anderson. *Probleme der statistischen Methodenlehre in den Sozialwissenschaften.* Physica, Würzburg, 1954. 36

[6] Oskar Nikolaevich Anderson. *Einführung in die mathematische Statistik.* Springer, Vienna, 1935. 36

[7] Oskar Nikolaevich Anderson. Die Begründung des Gesetzes der grossen Zahlen und die Umkehrung des Theorems von Bernoulli. *Dialectica*, 3(9/10):65–77, 1949. 36, 37, 39

[8] Michel Armatte. Discussion de l'article de D. Denis [58]. *Journal de la Société Française de Statistique*, 145(4):27–36, 2004. 22

[9] Michel Armatte. Contribution à l'histoire des tests laplaciens. *Mathematics and social sciences*, 44(176):117–133, 2006. 22

[10] Antoine Arnauld and Pierre Nicole. *La logique, ou l'art de penser.* Vrin, Paris, 1993. This critical edition, by Pierre Clair and François Girbal, was first issued in 1981. It is based on the 1683 edition of *l'Art de penser*, often called the *Port Royal Logic.* The first edition appeared in 1662, the year of Pascal's death. 13

[11] Raymond Bayer, editor. *Congrès International de Philosophie des Sciences, Paris, 1949; IV: Calcul des Probabilités*. Number 1146 in *Actualités Scientifiques et Industrielles*. Hermann, Paris, 1951. 36, 59, 62, 64

[12] John S. Bell. Quantum mechanics for cosmologists. In C. J. Isham et al., editor, *Quantum Gravity 2: A Second Oxford Symposium*, pages 611–637. Clarendon, Oxford, 1981. Reprinted on pp. 1117–138 of [13]. 49

[13] John S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge, 1987. 49, 54

[14] Jacob Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713. 1, 13, 54

[15] Jacob Bernoulli. *The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis*. Johns Hopkins University Press, Baltimore, 2006. Translation of [14] and commentary by Edith Sylla. 13

[16] Laurent Bienvenu, Glenn Shafer, and Alexander Shen. On the history of martingales in the study of randomness. *Electronic Journal for History of Probability and Statistics*, 5(1), 2009. 4, 51

[17] Ludwig Boltzmann. *Vorlesungen über Gastheorie*. Barth, Leipzig, 1898. 25

[18] Jacques Bonitzer. *Philosophie du hasard*. Messidor, Paris, 1984. 3

[19] Émile Borel. La valeur pratique du calcul des probabilités. *Revue du mois*, 1:424–437, 1906. Reprinted in [29], Volume 2, pp. 991–1004. 30

[20] Émile Borel. *Éléments de la théorie des probabilités*. Gauthier-Villars, Paris, 1909. Third edition 1924. The 1950 edition was translated into English by John E. Freund and published as *Elements of the Theory of Probability* by Prentice-Hall in 1965. 30

[21] Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–270, 1909. Reprinted in [29], Volume 2, pp. 1055–1079. 9

[22] Émile Borel. *Le Hasard*. Alcan, Paris, 1914. The first and second editions both appeared in 1914, with later editions in 1920, 1928, 1932, 1938, and 1948. 30

[23] Émile Borel. Sur les probabilités universellement négligeables. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 190:537–540, 1930. Reprinted as Note IV of [24]. 30

[24] Émile Borel. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, Paris, 1939. 30, 54

[25] Émile Borel. *Le jeu, la chance et les théories scientifiques modernes*. Gallimard, Paris, 1941. 31

[26] Émile Borel. *Les probabilités et la vie*. Presses Universitaires de France, Paris, 1943. Second edition 1946, third 1950, fourth 1958, sixth 1967. The fourth edition was translated into English by Maurice Baudin and published as *Probabilities and Life* by Dover, New York, in 1962. 31

[27] Émile Borel. Probabilité et certitude. *Dialectica*, 3(9/10):24–27, 1949. 36

[28] Émile Borel. *Probabilité et certitude*. Presses Universitaires de France, Paris, 1950. An English translation, *Probability and Certainty*, was published in 1963 by Walker, New York. 31

[29] Émile Borel. *Œuvres de Émile Borel*. Éditions du Centre National de la Recherche Scientifique, Paris, 1972. Four volumes. 54

[30] George E. P. Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society A*, 143(4):383–430, 1980. 6

[31] Benard Bru. Remarques sur l'article de D. Denis [58]. *Journal de la Société Française de Statistique*, 145(4):37–38, 2004. 22

[32] Bernard Bru. Borel, Lévy, Neyman, Pearson et les autres. *MATAPLI*, 60:51–60, octobre 1999. 32

[33] Bernard Bru. Souvenirs de Bologne. *Journal de la Société Française de Statistique*, 144(1–2):134–226, 2003. 30

[34] Bernard Bru. Personal communication, 2003b. 37

[35] George-Louis Buffon. Essai d'arithmétique morale. In *Supplément à l'Histoire naturelle*, volume 4, pages 46–148. Imprimerie Royale, Paris, 1777. 14

[36] George-Louis Buffon. Letter to Laplace, dated april 21, 1774. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 188:1019, 1879. 14

[37] William Burnside. On the idea of frequency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22(5):726–727, 1925. 37

[38] William Burnside. On the "Hypothetical Infinite Population" of theoretical statistics. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(3):670–674, 1926. 37

[39] Guido Castelnuovo. *Calcolo delle probabilitá*. Albrighi e Segati, Milan, Rome, and Naples, 1919. Second edition in two volumes, 1926 and 1928. Third edition 1948. 29

[40] Pafnutii Lvovich Chebyshev. Démonstration élementaire d'une proposition générale de la théorie des probabilités. *Journal für die reine und angewandte Mathematik*, 33:259–267, 1846. 3

[41] Aleksandr Aleksandrovich Chuprov. *Очерки по теории статистики (Essays on the theory of statistics)*. Sabashnikov, Saint Petersburg, second edition, 1910. The first edition appeared in 1909. The second edition was reprinted by the State Publishing House, Moscow, in 1959. 32

[42] Marquis de Condorcet. *Éloge de M. le Comte de Buffon*. Buisson, Paris, 1790. Reproduced on pp. 327–371 of vol. 3 of *Œuvres de Condorcet*, edited by A. Condorcet O'Connor and M. F. Arago, 1847, Paris, Firmin Didot Frères. 15

[43] Antoine Augustin Cournot. *Exposition de la théorie des chances et des probabilités*. Hachette, Paris, 1843. Reprinted in 1984 as Volume I (Bernard Bru, editor) of [45]. 19

[44] Antoine Augustin Cournot. *Matéralisme, Vitalisme, Rationalisme: Études sur l'emploi des données de la science en philosophie*. Hachette, Paris, 1875. Reprinted in 1987 as Volume V (Claire Salomon-Bayet, editor) of [45]. 19

[45] Antoine Augustin Cournot. *Œuvres complètes*. Vrin, Paris, 1973–2010. The volumes are numbered I through XI, but VI and XI are double volumes. 22, 56

[46] David R. Cox and David V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974. Second edition 1979. 8, 48

[47] Harald Cramér. *Mathematical Methods in Statistics*. Princeton University Press, Princeton, NJ, 1946. 38

[48] Emanuel Czuber. *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung*. Teubner, Leipzig, 1903. The preface is dated November 1902. The second edition appeared in two volumes, the first in 1908 and and the second in 1910. A new edition of the second volume appeared in 1914. 26

[49] Jean d'Alembert. Réflexions sur le calcul des probabilités. In *Opuscule mathématiques*, volume 2, pages 1–25. David, Paris, 1761. 15

[50] Lorraine Daston. *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ, 1988. 11

[51] Lorraine Daston. How probabilities came to be objective and subjective. *Historia Mathematica*, 21:330–344, 1994. 19

[52] A. P. Dawid. Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association*, 95:407–448, 2000. 10

[53] A. P. Dawid. Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science*, 19:44–57, 2004. 44

[54] A. P. Dawid. Decision-theoretic foundations of statistical causality. *Journal of Causal Inference*, 9:39–77, 2021. 10

[55] Bruno de Finetti. Recent suggestions for the reconciliation of theories of probability. In Jerzy Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 217–225. University of California Press, Berkeley, 1951. 44

[56] Bruno de Finetti. La notion de "horizon bayesien". In Centre belge de recherches mathématiques, editor, *Colloque sur l'analyse statistique: tenu à Bruxelles le 15, 16 et 17 décembre, 1954*, pages 57–71. Masson, Liège, 1955. 44

[57] Morris H. DeGroot. A conversation with Charles Stein. *Statistical Science*, 1(4):454–462, 1986. 47

[58] Daniel J. Denis. The modern hypothesis testing hybrid: R. A. Fisher's fading influence. *Journal de la Société Française de Statistique*, 145(4):5–26, 2004. 53, 55

[59] René Descartes. *Les Principes de la Philosophie, escrits en Latin, Par René Descartes, Et traduits en François par vn de ses Amis [l'abbé Picot]*. Henry Le Gras, Paris, 1647. 12

[60] Bryce S. DeWitt and Neill Graham, editors. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton University Press, Princeton, NJ, 1973. 50

[61] William Fishburn Donkin. On certain questions relating to the theory of probability. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, (4)1(5–6):353–368, 458–466. 22

[62] Joseph L. Doob. Probability as measure. *Annals of Mathematical Statistics*, 12:206–214, 1941. This article originated as a paper for a meeting of the Institute of Mathematical Statistics in Hanover, NH, in September 1940. It was published together with an article by von Mises [163] and comments by Doob and von Mises on each other's articles [165]. 44, 64, 65

[63] Joseph L. Doob. Foundations of probability theory and its influence on the theory of statistics. In Donald B. Owen, editor, *On the History of Statistics and Probability*, pages 195–204. Marcel Dekker, New York, 1976. 44

[64] Detlef Dürr and Dustin Lazarovici. *Understanding Quantum Mechanics: The World According to Modern Quantum Foundations*. Springer, 2020. 8

[65] Francis Edgeworth. Methods of statistics. *Journal of the Statistical Society of London*, Jubilee Volume:181–217, 1885. 26

[66] Robert Leslie Ellis. On the foundations of the theory of probabilities. *Transactions of the Cambridge Philosophical Society*, 8(1):1–6, 1849. The paper was read on February 14, 1842. Part 1 of Volume 8 was published in 1843 or 1844, but Volume 8 was not completed until 1849. 4, 23

[67] Terrence L. Fine. A computational complexity viewpoint on the stability of relative frequency and on stochastic independence. In William L. Harper and Clifford Alan Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, pages 29–40. Springer, Dordrecht, 1976. 51

[68] Hans Fischer. *A History of the Central Limit Theorem: From Classical to Modern Probabilty Theory*. Springer, New York, 2011. 4

[69] Ronald A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London (A)*, 222:309–368, 1922. 6, 37

[70] Ronald A. Fisher. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22(5):700–725, 1925. 6, 37

[71] Ronald A. Fisher. The nature of probability. *The Centennial Review of Arts & Science*, 2:261–274, 1958. 3

[72] Richard Foley. The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly*, 29(2):111–121, 1992. 13

[73] Robert M. Fortet. Opinions modernes sur les fondements du calcul de probabilités. In Le Lionnais [107], pages 207–215. This volume was reprinted by Blanchard in 1962 and by Hermann in 1998. An English translation, *Great Currents of Mathematical Thought*, was published by Dover, New York, in 1971. 31

[74] Joseph Fourier. Mémoire sur les résultats moyens déduits d'un grand nombre d'observations. In Joseph Fourier, editor, *Recherches statistiques sur la ville de Paris et le département de la Seine*, pages ix–xxxi. Imprimerie royale, Paris, 1826. 17

[75] Maurice Fréchet. *Généralités sur les probabilités. Variables aléatoires.* Gauthier-Villars, Paris, 1937. This is Book 1 of Fréchet (1937–1938). The second edition (1950) has a slightly different title: *Généralités sur les probabilités. Éléments aléatoires.* 30

[76] Maurice Fréchet. *Recherches théoriques modernes sur la théorie des probabilités.* Gauthier-Villars, Paris, 1937–1938. This work is listed in the bibliography of the *Grundbegriffe* as in preparation. It consists of two books, Fréchet (1937) and Fréchet (1938a). The two books together constitute Fascicle 3 of Volume 1 of Émile Borel's *Traité du calcul des probabilités et ses applications.*

[77] Maurice Fréchet. *Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* Gauthier-Villars, Paris, 1938. This is Book 2 of Fréchet (1937–1938). Second edition 1952. 30

[78] Maurice Fréchet. Rapport général sur les travaux du Colloque de Calcul des Probabilités. In Bayer [11], pages 3–21. 32, 36

[79] Maurice Fréchet. *Les mathématiques et le concret.* Presses Universitaires de France, Paris, 1955. 37

[80] Maurice Fréchet and Maurice Halbwachs. *Le calcul des probabilités à la portée de tous.* Dunod, Paris, 1924. 29

[81] Hans Freudenthal. Ist die Mathematische Statistik Paradox? *Dialectica*, 12(1):7–32, 1958. 44

[82] Hans Freudenthal. Models in applied probability. *Synthese*, 12(2/3):202–212, 1960. 43

[83] Hans Freudenthal. Probabilités objectives ou subjectives? *Logique et Analyse*, 8(32):265–276, 1965. 44

[84] Thornton C. Fry. *Probability and its Engineering Uses.* Van Nostrand, New York, 1928. 38

[85] Thomas Galloway. *A Treatise on Probability: Forming the article under that head in the seventh edition of the Encyclopædia Britannica.* Adam and Charles Black, Edinburgh, 1839. 19

[86] Jules Gavarret. *Principes généraux de statistique médicale, ou développement des règles qui doivent présider à son emploi.* Bechet, Paris, 1840. 22

[87] Andrew Gelman and Cosma Rohilla Shalizi. Philosophy and practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66:8–38, 2013. 6

[88] Donald A. Gillies. *An Objective Theory of Probability.* Methuen, London, 1973. 51, 52

[89] Trygve Haavelmo. The probability approach to econometrics. *Econometrica*, 12(Supplement):1–115, 1944. 46

[90] Jacques Hadamard. Les principes du calcul des probabilités. *Revue de Métaphysique et de Morale*, 29(3):289–293, 1922. An expanded version of this article, with the title "Les axiomes du calcul des probabilités", appears on pp. 2161–2165 of the fourth volume of *Oeuvres de Jacques Hadamard.* Although the index of the volume indicates that this longer article was drawn from *Revue de Métaphysique et de Morale*, it is actually drawn from the second volume of Hadamard's *Cours d'analyse* (Hermann, 1930, p. 538f). I am indebted to Bernard Bru for this clarification. 30, 36

[91] Alan Hájek. Fifteen arguments against finite frequentism. *Erkenntnis*, 45:209–227, 1997. 5

[92] Alan Hájek. Fifteen arguments against hypothetical frequentism. *Erkenntnis*, 70(2):211–235, 2009. 5

[93] Carl G. Hempel. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. Free Press, New York, 1965. 43

[94] Colin Howson. Theories of probability. *The British Journal for the Philosophy of Science*, 46(1):1–32, 1995. 52

[95] David Hume. *A Treatise of Human Nature*. John Noon, 1739–1740. 15

[96] David Hume. *An Enquiry Concerning Human Understanding*. Millar, 1748. 15

[97] Norman L. Johnson and Samuel Kotz. *Leading Personalities in Statistical Sciences*. Wiley, New York, 1997. 25, 63

[98] Andreas Kamlah. Probability as a quasi-theoretical concept; J. V. Kries' sophisticated account after a century. *Erkenntnis*, 19:239–251, 1983. 28

[99] John Maynard Keynes. *A Treatise on Probability*. Macmillan, London, 1921. 28

[100] Sven K. Knebel. *Wille, Würfel und Wahrscheinlichkeit: Das System der moralischen Notwendigkeit in der Jesuitenscholastik 1550–1700*. Meiner, Hamburg, 2000. 12

[101] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956. 9, 35, 42

[102] Henry E. Kyburg, Jr. *Science & Reason*. Oxford, 1990. 49

[103] Pierre Simon Laplace. Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie royale des sciences de Paris*, 6:621–656, 1774. Reprinted in Volume 8 of Laplace's *Oeuvres complètes*, pages 27–65, translated into English with commentary by Stigler in [154, 105]. 6, 60

[104] Pierre Simon Laplace. *Essai philosophique sur les probabilités*. Courcier, Paris, fifth edition, 1825. This was the final edition. It was reproduced in 1986 (Christian Bourgois, Paris) with notes and commentary by Bernard Bru. 61

[105] Pierre Simon Laplace. Memoir on the probability of the causes of events. *Statistical Science*, 1(3):364–378, 1986. Translation of [103] by Stigler. 60

[106] Pierre Simon Laplace. *Philosophical Essay on Probabilities.* Springer, 1995. A translation of [104] with commentary and notes based on Bru's. 16

[107] François Le Lionnais, editor. *Les grands courants de la pensée mathématique*. Cahiers du Sud, Paris, 1948. This volume was reprinted by Blanchard in 1962 and by Hermann in 1998. An English translation, *Great Currents of Mathematical Thought*, was published by Dover, New York, in 1971. 31, 58

[108] Paul Lévy. *Calcul de probabilités*. Gauthier-Villars, Paris, 1925. 35, 36

[109] Paul Lévy. La loi forte des grands nombres pour les variables aléatoires enchainées. *Journal de Mathématiques Pures et Appliquées*, (9)15:11–24, 1936. 3

[110] Paul Lévy. *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris, 1937. Second edition: 1954. 9, 36

[111] Paul Lévy. Les fondements du calcul des probabilités. *Dialectica*, 3(9/10):55–64, 1949. 36

[112] Wilhelm Lexis. *Einleitung in die Theorie der Bevölkerungsstatistik*. Trübner, Strassburg, 1875. 24

[113] John Locke. *An Essay Concerning Human Understanding*. Basset, London, 1689. 13

[114] Jeff Loveland. Buffon, the certainty of sunrise, and the probabilistic reductio ad absurdum. *Archive for History of Exact Sciences*, 55:465–477, 2001. 14

[115] Paul Mansion. Sur la portée objective du calcul des probabilités. *Bull. de l'Acad. roy. de Belgique (Classe des sciences)*, 1903(12):1235–1294, 1903. 25

[116] Andrei Andreevich Markov. *Wahrscheinlichkeitsrechnung*. Teubner, 1912. Translation of second Russian edition. 29

[117] Thierry Martin. *Probabilités et critique philosophique selon Cournot*. Vrin, Paris, 1996. 22, 53

[118] Thierry Martin. *Bibliographie cournotienne*. Annales littéraires de l'Université Franche-Comté, Besançon, 1998. 22, 53

[119] Thierry Martin. Probabilité et certitude. In Thierry Martin, editor, *Probabilités subjectives et rationalité de l'action*, pages 119–134. CNRS Éditions, Paris, 2003. 32, 53

[120] Thierry Martin, editor. *Actualité de Cournot*. Vrin, Paris, 2005. 22, 53

[121] Thierry Martin. *Nouvelle bibliographie cournotienne*. Presses universitaires de Franche-Comté, second edition, 2005. `http://logiquesagir.univ-fcomte.fr/nouvelle_bibliographie_cournotienne/`. 22, 53

[122] Per Martin-Löf. Algorithmen und zufällige Folgen. This document, dated April 16, consists of notes taken by K. Jacobs and W. Muller from lectures by Martin-Löf at Erlangen on April 5, 6, 14, and 15, 1966. 51

[123] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966. 51

[124] Per Martin-Löf. The literature on von Mises' Kollectivs revisited. *Theoria*, 35:12–37, 1969. 4

[125] Tim Maudlin. The grammar of typicality. In Valia Allori, editor, *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*. 2020. 46

[126] Laurent Mazliak. Belgium and probability in the 19th century: The case of Paul Mansion. *Science in Context*, pages 00–00, to appear. 26

[127] Alexius Meinong. *Über Möglichkeit und Wahrscheinlichkeit: Beiträge zur Gegenstandstheorie und Erkenntnistheorie*. Barth, Leipzig, 1915. 26, 28

[128] Mary Morgan. *The History of Econometric Ideas*. Cambridge University Press, Cambridge, 1990. 7, 46

[129] Jerzy Neyman. Foundation of the general theory of statistical estimation. In Bayer [11], pages 83–95. 37, 40

[130] Jerzy Neyman. Indeterminism in science and new demands on statisticians. *Journal of the American Statistical Association*, 55:625–639, 1960. 7, 39

[131] Kh. O. Ondar, editor. *O teorii veroyatnosteĭ i matematicheskoĭ statistike (perepiska A. A. Markova i A. A. CHuprova)*. Nauk, Moscow, 1977. See [132] for English translation. 62

[132] Kh. O. Ondar, editor. *The Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. Springer, New York, 1981. Translation of [131] by Charles M. and Margaret Stein. Additional letters between Markov are provided in translation by Sheynin in [151], Chapter 8. 28, 62

[133] A. Pallez. Normes de la statistique, du calcul des probabilités et des erreurs de mesures. *Journal de la Société de Statistique de Paris*, pages 125–133, 1949. 37

[134] Karl Pearson. *The Grammar of Science*. Scott, London, 1892. A second edition appeared in 1900, a third in 1911. 19

[135] Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890. 25

[136] Karl R. Popper. *Logik der Forschung*. Springer, Vienna, 1934. An English translation, *The Logic of Scientific Discovery*, was published by Hutchinson, London, in 1959. 40

[137] Yu. V. Prokhorov and B. A. Sevast'yanov. Probability theory. In *Encyclopaedia of mathematics: An updated and annotated translation of the Soviet "Mathematical Encyclopaedia" (managing editor M. Hazewinkel)*, volume 7, pages 307–313. Reidel, Boston, 1987–1994. 48

[138] M. L. G. Redhead. *On Neyman's Paradox and the Theory of Statistical Tests*, volume 25. 1974. 52

[139] Hans Reichenbach. *Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit*. Barth, Leipzig, 1916. 28

[140] Hans Richter. Zur Begründung der Wahrscheinlichkeitsrechnung. *Dialectica*, 8:48–77, 1954. 37, 40, 47

[141] Hans Richter. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1956. 37, 40, 47

[142] Bertrand Saint-Sernin. *Cournot: Le réalisme*. Vrin, Paris, 1998. 22

[143] Eugene Seneta. Boltzmann, Ludwig Edward. In *Leading Personalities in Statistical Sciences* [97], pages 353–354. 25

[144] Glenn Shafer. How the game-theoretic foundation for probability resolves the Bayesian vs. frequentist standoff, August 2020. GTP Working Paper 56, to appear in *Handbook on Bayesian, Fiducial and Frequentist (BFF) Inference* , edited by Jim Berger, Xiao-Li Meng, Nancy Reid and Min-ge Xie and published by Chapman & Hall. 6

[145] Glenn Shafer. Descriptive probability, September 2021. GTP Working Paper 59. 7

[146] Glenn Shafer. Testing by betting: a strategy for statistical and scientific communication (with discussion). *Journal of the Royal Statistics Society, Series A*, 184(2):407–478, 2021. 10

[147] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006. 63

[148] Glenn Shafer and Vladimir Vovk. The origins and legacy of Kolmogorov's *Grundbegriffe*, April 2013. GTP Working Paper 4. Abridged version published as "The sources of Kolmogorov's *Grundbegriffe*" [147]. 42

[149] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*. Wiley, New York, 2019. 9, 51

[150] Oscar Sheynin. *Aleksandr A. Chuprov: Life, Work, Correspondence. The making of mathematical statistics.* Vandenhoeck & Ruprecht, Göttingen, 1996. Second revised edition, edited by Heinrich Strecker, 2011. 32

[151] Oscar Sheynin. *Aleksandr A. Chuprov: Life, Work, Correspondence. The making of mathematical statistics.* V&R unipress, Goettingen, 2011. Second revised edition, edited by Heinrich Strecker. The first edition appeared in 1996. 62

[152] Evgeny Slutsky. Ueber stochastische Asymptoten und Grenzwerte. *Metron*, 5:3–89, 1925. 33

[153] Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press, Cambridge, MA, 1986. 2

[154] Stephen M. Stigler. Laplace's 1774 memoir on inverse probability. *Statistical Science*, 1(3):359–378, 1986. 60

[155] Marshall Stone. Mathematics and the future of science. *Bulletin of the American Mathematical Society*, 63:61–76, 1957. 42

[156] Jean-Philippe Touffut, editor. *Augustin Cournot: Modelling Economics.* Edward Elgar, Cheltenham, 2007. 22

[157] J. V. Uspensky. *Introduction to Mathematical Probability.* McGraw-Hill, New York, 1937. 35

[158] David Van Dantzig. Sur l'analyse logique des relations entre le calcul des probabilités et ses applications. In Bayer [11], pages 49–66. 37, 40

[159] John Venn. *The Logic of Chance: an essay on the foundations and province of the theory of probability, with especial reference to its logical bearings and its application to moral and social science, and to statistics.* Macmillan, London, 1866. 2, 24

[160] Jean Ville. *Étude critique de la notion de collectif.* Gauthier-Villars, Paris, 1939. 4, 9, 45

[161] Guy von Hirsch. Sur un aspect paradoxal de la théorie des probabilités. *Dialetica*, 8:125–144, 1954. 37, 40, 47

[162] Richard von Mises. *Kleines Lehrbuch des Positivismus. Einführung in die empiristische Wissenschaftsauffassung.* W. P. van Stockum & zoon, The Hague, 1939. 65

[163] Richard von Mises. On the foundations of probability and statistics. *Annals of Mathematical Statistics*, 12:191–205, 1941. This article originated as an address to the Institute of Mathematical Statistics in Hanover, NH, in September 1940. It was published together with an article by Doob [62] and comments by von Mises and Doob on each other's articles [165]. 57, 65

[164] Richard von Mises. *Positivism: A study in Human Understanding*. Harvard, Cambridge, Massachusetts, 1951. Translation of [162]. 35

[165] Richard von Mises and Joseph L. Doob. Discussion of papers on probability theory. *Annals of Mathematical Statistics*, 12:215–217, 1941. Discussion of [163] and [62]. 57, 64

[166] Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*. Cambridge University Press, Cambridge, 1994. 25

[167] Abraham Wald. *On the principles of statistical inference*. University of Notre Dame, 1942. Four lectures delivered at the University of Notre Dame, Februrary 1941. Printed by Edwards Brothers, Lithoprinters, Ann Arbor. 41

[168] George Udny Yule. *An Introduction to the Theory of Statistics*. Griffin, London, first edition, 1911. 31

[169] Sandy L. Zabell. Johannes von Kries's *Principien*: A brief guide for the perplexed. *Journal for General Philosophy of Science*, 47(1):131–150, 2016. 28