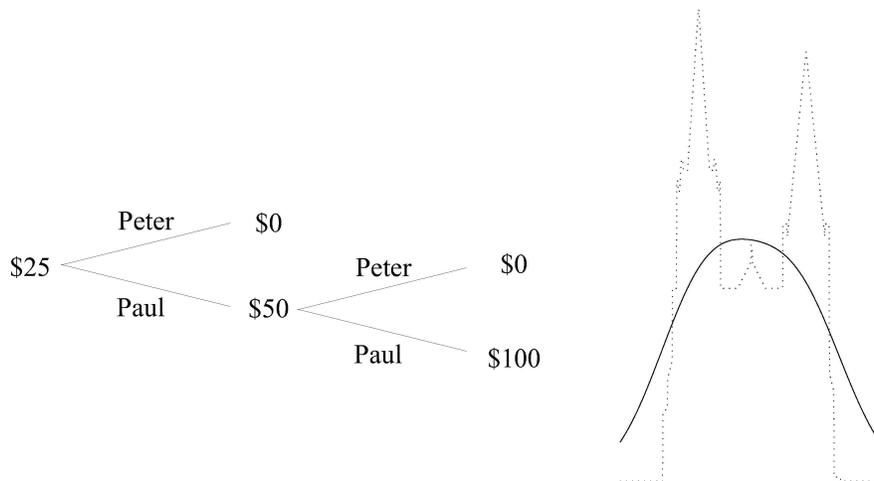# "That's what all the old guys said."
The many faces of Cournot's principle

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

Abstract

Cournot's principle is a family of theses about how mathematical probability can be used to make assertions about facts or events in the world. With one nuance or another, these theses say that this is done by assigning probabilities very close to zero or one to certain facts or events. Events with probabilities close to zero will not happen (or, if they are repeatable, will happen rarely); those with probabilities close to one will happen (or will usually happen).

In this working paper, which remains incomplete, I survey the many ways Cournot's priniciple has been formulated. I quote, with some commentary, scores of prominent authors over several centuries. I include some authors who formulated versions of Cournot's principle only to disagree with them.

I have drawn on this compilation to formulate my own views on Cournot's principle, and I believe that it will also be helpful to others with other views. The compilation can also be useful for historians interested in the contributions of particular authors. The continuity of opinion may be greater than we sometimes think, the originality of particular authors less than we sometimes suppose.

I also sketch a game-theoretic formulation of Cournot's principle that has been studied over the past quarter century by Vladimir Vovk and myself. This formulation applies when mathematical probability is based on a multi-round game in which probabilities are given on each round by one player and another player bets against them. Here the role of an event of small probability is played by the event that the second player multiplies the capital he risks by a large factor, and Cournot's principle becomes the assertion that this will not happen. Because it involves two players, not just one, this formulation can help us understand the diversity of Cournot's principle.

This paper is dedicated to Thierry Martin, whose *Probabilités et critique philosophique selon Cournot* introduced me and others to the many faces of Cournot's principle.

Contents

1   Introduction

Early in 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, then retired but still active at Dartmouth College. We would talk about martingales and about the philosophy of probability. On one occasion I tried to interest Laurie in Émile Borel's dictum that events of extremely small probability are impossible. 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 know something the old guys did not know? What exactly were they saying? Were we really saying it less often, or merely less loudly? These questions had been haunting me long before my conversations with Laurie, and this paper reports on my efforts to answer them.

1.1   Why I cared

In the 1970s, I pondered Jacob Bernoulli's 17th-century concepts of moral (i.e., practical) impossibility and certainty and their roots in yet older thinking about non-numerical probability. Bernoulli used the concepts in his book on probability, *Ars conjectandi* [14, 15], published after his death. There he equated practical certainty with probability close to one, practical impossibility with probability close to zero, and he proposed that the government instruct judges on how close was close enough.

Bernoulli's proposal for using mathematical probability in legal matters seemed quaint, but as a teacher of statistics, I constantly encountered his question of how close to zero is close enough. How small a probability is small enough to reject a hypothesis? When teaching probability, I found myself explaining to students that we are *almost certain* an event will not happen when its probability is close or equal to zero. There was usually a clever student who argued that what happens, when described in detail, always has a small or zero probability. This has been called the *lottery paradox*. I would brush the paradox aside, but I often felt that I was indoctrinating rather than teaching.

Bernoulli's project had been to turn the mathematics of betting in games of pure chance into a calculus that would help us decide what to believe in the larger world — a guide to life, to use the words of the 18th-century theologian Joseph Butler. This project was spectacularly successful in a way Bernoulli had not intended: most mathematicians and scientists now accept without question that the mathematics of probability should follow the rules for fair bets in games of pure chance. Bernoulli's own project was more complicated than this.

The part of Bernoulli's project that mathematicians now celebrate is his theorem concerning how we can learn about a probability of a repeatable event from multiple trials. As an example, Bernoulli considered randomly drawing a ball from an urn that contains a known total number of balls, some black and some white, in unknown proportion. His theorem says that the frequency of

black outcomes in sufficiently many draws with replacement will tell us the exact number of black balls in the urn with probability as close to one as you want. Choose your threshold for moral certainty, and Bernoulli could tell you a number of draws with replacement sufficient to achieve it.[1] Later mathematicians saw that his calculation establishes a more general statement concerning independent trials of an event with probability p: Given any positive numbers ϵ and δ, no matter how small, there is an integer n such that the frequency f with which the event happens in n trials will be within ϵ of p with probability at least $1 - \delta$. This has been called Bernoulli's *law of large numbers.*

But Bernoulli was only incidentally interested in multiple trials of a repeatable event. His goal was to obtain numerical probabilities close to zero or one for singular facts or isolated events, events that cannot be repeated. Did Maevius murder Titius? There are mutliple arguments: Maevius hated Titius, he turned pale when questioned, a sword stained by blood was found in his house, he passed that day on the road where Titius was found slain, Caius testifies that the two had quarreled that day [15, p. 318]. Bernoulli hoped that his law of large numbers could be used to assess the probability of each argument. How often does turning pale indicate guilt? Etc. Then the probabilities from different arguments could be combined by rules that went beyond the mathematics of betting in games of pure chance. As more evidence is gathered and more arguments combined, one might reach moral certainty about the isolated event, which could not be addressed directly by Bernoulli's theorem.[2]

Bernoulli's picture was constructive; we combine arguments until we reach a probability close to one for the answer to our question. There is no lottery here; no distribution of probabilities over many possibilities, each having a small probability. Bernoulli would have agreed that each sequence of blacks and whites in a long sequence of draws from his urn without replacement would have a very small probability, but he was not interested in being morally certain that any particular sequence would not be obtained; he was interested in moral certainty about the number of black balls. And then he was interested in combining this number with other similarly obtained numbers to decide with moral certainty who murdered Gracchus.

Even though Bernoulli's 18th- and 19th-century successors did not adopt his rules for combining arguments, their probability calculus remained relatively constructive. They constructed probabilities for complicated events from simpler probabilities using the rules of total probability (additivity) and compound probability. It was only in the 20th century that functional analysts like Maurice Fréchet, Paul Lévy, and Andrei Kolmogorov proposed that mathematical probability begin with a comprehensive probability measures that often, like the probabilities for tickets in a lottery, involve a small probability for every precise outcome. Only then did the need arise to somehow qualify the statement of

---

[1]See Stephen M. Stigler's *A History of Statistics* [176] for a careful and detailed account of Bernoulli's formulation of his theorem and his proof.

[2]I wrote about Bernoulli's project in 1978 [165]. A few years later, trying my hand as an amateur medievalist, I contributed an article about moral certainty to Wiley's *Encyclopedia of Statistical Sciences* [166].

Cournot's principle to avoid the lottery paradox.

I have devoted the first decade of my scholarly career to challenging the notion that probability theory should rely completely on the mathematics of betting. In *A Mathematical Theory of Evidence*, published in 1976 [164], I advocated a theory for combining arguments similar to Bernoulli's. I called it the theory of *belief functions*; it was later called the *Dempster-Shafer theory*, because it was inspired not only by Bernoulli's work but also by the work of my mentor A. P. Dempster. It lives on in the hands of The Belief Functions Theory and Applications Society.

In the early 2000s, when I was talking with Laurie, my interest in the history of practical certainty had been rekindled by my collaboration with Volodya Vovk on what he and I called game-theoretic probability. This was a new way of developing the mathematics of betting, accompanied by a new way of relating this mathematics to the world of experience.[3]

I saw my collaboration with Volodya as another step in a campaign against the hegemony of a theory of betting masquerading as a general theory of evidence and numerical probability. When we are using a theory of betting, I argued, let's be honest about it. And let's be honest about what is involved in connecting it with the world of experience.

How can we connect the theory of betting to the world of experience? I learned a clear and precise answer to that question from Volodya, or at least through my collaboration with him. To interpret betting odds as statements about the world, as forecasts or predictions if you will, we think of them as offers to bet. Their validity as forecasts lies in the ability or inability of an opponent to make money by selecting which of the betting offers to take up. My probability forecasts are discredited (or I am discredited as a forecaster) to the extent that my opponent multiplies the money he risks by a large factor. His multiplying his money by a large factor is related, in a way that can be made precise, to an event of small probability happening. His multiplying his money by an infinite factor is related to an event of zero probability happening.

Game-theoretic probability is best taught on its own, without too much attention to the usual theory of probability. But most of those who might be interested have already studied the usual theory, and so the connections must also be explained. In particular, we need to explain how success in betting against probabilities is related to an event of small probability, and how the principle that such success discredits probabilities is related to the traditional notions of moral certainty and moral impossibility. Here I have encountered a surprising problem. In spite of their continuing relevance to practice, these traditional notions have been so obfuscated in our teaching that many modern students are not aware of them, and some authorities, especially some who see Bayes's theorem as more important than Bernoulli's theorem, reject them.

I have thus come to believe that acceptance of the cogency of game-theoretic probability by those who now study and teach probability theory requires a

---

[3]My collaboration with Volodya began in the mid 1990s. Our first book on game-theoretic probability appeared in 2001 [169]; our second, much more complete on the mathematical side, appeared in 2019 [172].

fuller account of the history of moral certainty and moral impossibility, a fuller understanding of why so many smart people saw them as important, and also an understanding of why some did not.

1.2   How I learned

Many scholars became interested in the history of probability and statistics in the 1970s. Among them were the members of a seminar on the topic in Paris, and I had the good fortune of coming into contact with this group, especially in the 1990s. Through Bernard Bru, one of the members of the group, I learned about Thierry Martin's work on Antoine Augustin Cournot (1801–1877). In Thierry's 1996 book, *Probabilités et critique philosophique selon Cournot* [129], I learned that Bernoulli's principle of moral certainty had a name; Maurice Fréchet named it *Cournot's principle* in 1949. Fréchet did this in French of course: *principe de Cournot*.

Whereas Bernoulli had written about moral certainty, Cournot wrote about physical certainty. He wrote that a probability close enough to one can be taken as physical certainty. Going beyond Bernoulli, he wrote that this is the only way to connect the mathematical theory of probability to the physical world outside mathematics.

The name *Cournot's principle* enjoyed some currency in the 1950s. It was used less often when I began studying probability in the 1970s, but it is slowly coming back. Google Scholar reports no instances of its use in academic articles in the 1970s, two instances in the 1980s, three in the 1990s, and 86 since 2000.

I have done my part in this revival; 21 of the 86 instances since 2000 were articles I authored or co-authored. Some of these articles were historical. Beginning with the references that I found in Thierry's work, I have been on the lookout for what various authors said about various versions of Cournot's principle, how the principle evolved over time, and perhaps how the understanding of it differed from one national or linguistic tradition to another. My own understanding of this evolution has evolved, of course. I do not stand by everything I wrote about the history of Cournot's principle in earlier articles [170, 168].

1.3   What I report here

As the title of this paper indicates, I am using *Cournot's principle* as Fréchet did, not to name a single well defined dictum but to name a family of principles. Even in cases where authors seem to be saying the same thing, each says it in their own way. This variety of thought is best presented, I think, by quoting many different authors, in the language in which they wrote and also in translation. Accordingly, I devote the next section to quotations from nearly 100 authors, who wrote over many centuries.

In most cases, I provide some context for the quotations. Even with this context, the brief quotations should not be seen as an adequate account of the thought of any particular author. All the authors had a lot more to say, and some changed their minds over time about some of the points quoted.

My purpose here is not to tell all about particular authors but to exhibit the larger flow of thinking. In the end, this can add to more in-depth profiles of particular authors, because all too often studies of a particular author stumble into attributing insights to the author that may have been commonplaces in their time or even much earlier.

1.4   What I think now

Perhaps we can dismiss some of the disagreements among our authors as quibbling about words. But I emerge from pondering their many ways of using words with the conviction that the interpretation of high probability as practical certainty is central to applications of the standard theory of mathematical probability. We see this interpretation used in practice by authors with a variety of philosophies about the meaning of probability; see for example the quotations from William Feller (who thought of probability as frequency), Harold Jeffreys (who thought of probability as rational belief), and A. Philip Dawid (who thinks of probability as subjective belief).

Bruno de Finetti, while conceding that a person's high probability for an event can be called practical certainty insofar as he will take risks as if the event is certain to occur, argued that high (and low) probabilities are not really special, because they enter into calculations of expected utility that describe the person's behavior in the same way that other probabilities do. I am unconvinced by this argument, because I do not accept the supposed principles of rationality that require a person to make their actions conform to a system of probabilities and utilities [167].

I agree with Condorcet that the concept of practical certainty belongs outside the mathematical theory. Buffon, whom Condorcet criticized, is not the only author to try to put it inside the theory; attempts in the 20th century include those by Gillies and Kyburg. They never work.

All those who use the practical certainty, whatever their philosophy, need to resolve the lottery paradox. I think that the resolution varies according to the application. The requirement that the event of high or low probability be chosen in advance is often one ingredient in a resolution. It is especially relevant when we are testing probabilities. But after we have tested a system of probabilities enough to gain confidence in its predictions (the events it asserts with high probability), we want to use it to make new predictions. We want to make calculations and find new events of high probability. We also want the right to take as discrediting surprising events that we do not anticipate or predict in advance. To resolve the lottery paradox, we must recognize limits on our right to explore and our right to surprise. These limits vary with the application. Often we need the insights of Wald, Ville, and Kolmogorov, and others concerning the limited number of simple or "remarkable" events and the limits on our ability to describe events. The number of black balls in Bernoulli's urn with 50 balls in total and the number of black balls drawn in 50,000 draws with replacement are simpler to describe than the actual sequence of 50,000 draws.

Cournot, Borel, and others sometimes insisted that probabilities extremely

close to zero or one should be deemed impossible or certain, not merely practically impossible or practically certain. I find their arguments convincing but easily misunderstood. These authors wanted to understand scientific theories that involve probabilities in the same way as other theories in physics. Physical theories that do not use probabilities do not say "maybe" when they make predictions, they simply make predictions. They say some things will happen; they are certain. They say some things will not happen; they are impossible. I think such statements are reasonable so long as they are understood as statements of what the theory says. When we accept the theory, we make the statements about impossibility and certainty. But of course there is another level of thought where we can concede that the theory might be not quite right after all or that we might have misapplied it. If we demand a higher standard than this for using the words "certain" and "impossible", we might as well ban them from our language.

1.5 Testing by betting

We can better understand the diversity of Cournot's principle if we separate the role of asserting probabilities from the role of testing them. A system of probabilities begins as purely mathematical object. Then someone asserts that the probabilities are valid or reliable in a particular situation. Then an opponent tries to refute this by describing an event to which the system assigns small probability. The requirement that the opponent describe the event means that it has a simple description. In this picture, we see two ways that subjectivity can enter. The probabilities can be considered the subjective opinion of the "player" asserting them, even if this player is merely a preliminary theory we are testing. The assertion of validity may be relative to the information, computational capacity, or cleverness of the opponent or opponents. Yet the system of probabilities gains a claim to objective value when the events of small probability the opponents describe do not happen.

We can make this picture mathematical, add to its flexibility, and bring in the requirement that the tests be chosen in advance by supposing that the asserted probabilities are betting offers and that the opponent tests them by choosing bets

Consider for example a multi-round game protocol with three players, whom we call Forecaster, Skeptic, and Reality. On each round,

Forecaster announces a probability distribution, say P.
Skeptic announces a random variable Y with finite expected value under P.
Reality announces a value y for Y.
Skeptic pays Y's expected value and gets y in return.

Here Skeptic is betting against Forecaster. He discredits Forecaster if he makes a lot of money. More precisely,

successive bets by Skeptic that begin with unit capital and never risk more discredit Forecaster to the extent that the final capital is large.

6

In [172], Vovk and I called this *the game-theoretic version of Cournot's principle*. More recently, I have called it *the fundamental principle for testing by betting*.

If you want a little more notation, we can number the rounds, write K_0 for the initial unit capital, and write $\mathcal{K}_n$ for the capital after the nth round of betting:

$\mathcal{K}_0 = 1$.
For $n = 1, \ldots, N$:
    Forecaster announces P_n.
    Skeptic announces Y_n with finite expected value $E_{P_N}(Y_n)$.
    Reality announces Y_n's value y_n.
    $\mathcal{K}_n := \mathcal{K}_{n-1} + y_n - E_{P_n}(Y_n)$.

The condition that Skeptic not risk more than his initial capital means that he must always choose Y_n so that

$$\mathcal{K}_{n-1} + y_n - E_{P_n}(Y_n) = \text{(capital at beginning of round } n$$
$$+ \text{ net gain or loss on round } n)$$
$$\geq 0$$

no matter what y_n turns out to be. Sometimes we call this *avoiding risk of bankruptcy*.

If many skeptics test a forecaster in this way and none of them obtain large values of $\mathcal{K}_N$, then the forecaster can claim to be reliable. He can claim that his forecasts describe the world. He can claim that they have objective value. Quantum Mechanics has achieved this objective value in the eyes of today's physicists. National weather forecasters are now also making such claims.

When (1) Forecaster's probabilities are dictated by theory or model (such as Quantum Mechanics some model estimated by a statistician), (2) we accept this model as valid, (3) Skeptic has a strategy that avoids risk of bankruptcy and makes $\mathcal{K}_N$ huge if Reality's moves $y_1, \ldots, y_N$ fall in a set E, then we may say that E is objectively impossible. Perhaps *physically* impossible if the theory is a theory in physics and N and $\mathcal{K}_N$ are huge, perhaps *practically* impossible in less extreme cases.

2   Things some old guys said

Because the periods in which the authors listed have 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 birth.

So far as today's theory of mathematical probability is concerned, we could begin our account of moral certainty with Jacob Bernoulli in the 17th century. But the Latin terms that Bernoulli used — *certitudo moralis* and *probabilitas* — go back much earlier. So I begin with a sampling of how earlier authors writing

in Latin used these and related terms. Is it too narrow to discuss only what was written in Latin and in later languages influenced by Latin? Other ancient languages, such as Sanskrit, Chinese, Arabic, and Hebrew, surely have their own ways of talking about practical certainty. But to translate these into Latin or English might involve choices too arbitrary to be helpful here. How would we decide whether an author is talking about degrees of certainty or degrees of probability? For better or worse, our topic here is an intellectual evolution that begins in Latin.

Cicero popularized the word *probabilitas* in Latin, teaching as the later Greek Academicians had that we cannot have certainty — we must be content with probability. But Jean Gerson, a Christian theologian who was chancellor of the University of Paris in the early 15th century, more than a thousand years after Cicero, was responsible for the earliest known use of *certitudo moralis*; see §2.3.

Here is a list of some other chronological milestones, points by which various important ideas have appeared. It should not be taken as a list of "firsts"; saying that a particular person was the first to think or say something is always dangerous and usually wrong.

- Without using numbers, Luis de Molina (1535–1600; §2.6) cited the random drawing from urns as an example where high probability can provide moral certainty.

- Blaise Pascal (1623–1662; 2.12) can be counted as an opponent of the concept of moral certainty. He famously argued that the reward of eternal happiness justifies following the Catholic faith no matter how small you think its probability is.

- Jacob Bernoull (1655–1705; §2.15) used a scale from zero to one for numerical probability and declared that numerical probability close to one provides moral certainty.

- John Arbuthnot (1667–1735; §2.16), who translated Huygens's tract on games of chance into English, used an observed event that had an extremely small probability under a hypothesis of randomness to reject that hypothesis.

- Denis Diderot (1713–1784; §2.19) distinguished three types of certainty, metaphysical, physical, and moral, on the kinds of evidence that underly them.

- Georges-Louis Buffon (1707–1788; §2.17) argued that the distinction between physical and moral certainty is one of degree rather than kind and proposed numerical thresholds for these two types of certainty.

- In his eulogy for Buffon at the French Academy of Sciences, Nicolas de Condorcet (1743–1794; §2.21) argued that Buffon erred in trying to put thresholds for certainty inside the mathematical theory. The principle of moral certainty belongs outside the mathematical theory, bridging it to

our belief, and the threshold should depend on what truth and what action is being considered.

Condorcet was one of the first proponents of Bayes's rule as a principle of probability. The subtle distinction between his version of Cournot's principle and Cournot's own version is that Cournot saw the principle as a bridge between the mathematics and the real world, whereas Condorcet saw it as a bridge between mathematics and belief about the world. Other Bayesians whose have accepted Condorcet's version of Cournot's principle include A. Philip Dawid (§2.90).

- Joseph Fourier (1768–1830; §2.23) proposed a threshold for practical certainty when calculating a large-sample confidence interval based on Laplace's central limit theorem.

- Siméon Denis Poisson (1781–1840; §2.25), who improved Laplace's proof of the central limit theorem, suggested different thresholds for practical certainty in different contexts. He introduced the term "law of large numbers" to name the empirical regularity of certain statistical ratios over time and offered a generalization of Bernoulli's theorem as an explanation of these regularities.

- Antoine Augustin Cournot (1801–1877; §2.27) contended that equating very small probabilities with physical impossibility is the only way to connect mathematical probability with phenomena.

  In his 1843 book, Cournot gave a clear account of how Poisson's law of large numbers can be proven by combining Bernoulli's theorem with the principle that an event of very small probability is impossible.

  Cournot also had a theory about how the phenomena predicted by high probabilities arise in the world; he talked about the intersection of independent causal chains. So while Condorcet saw high probabilities as features of our beliefs, Cournot saw them as features of how the world worked.

- Robert Leslie Ellis (1817–1859; §2.32) rejected Bernoulli's theorem and Bernoulli's use of the notion of moral certainty. Probabilities are simply frequencies, he argued, and no mumbo-jumbo is needed to make the connection. John Venn (1834–1923; §2.33) took a similar view. Later authors who explicitly rejected Cournot's principle in favor of other ways of connecting probabilities with frequencies include Emanuel Czuber (1851–1925; §2.39), Johannes von Kries (1853–1928; §2.40, and Richard von Mises (1883–1953; §2.55),

- Ludwig Boltzmann (1844–1906; §2.36) gave Cournot's principle an important role in physics by showing that the second law of thermodynamics holds with exceedingly high probability. Boltzmann's exceedingly high probabilities were vastly closer to one than those achieved by the law of

large numbers in the statistical applications discussed by Fourier, Poisson, and Cournot.

- By the beginning of the 20th century, the probability theory of Laplace, Fourier, and Poisson, based on the law of large numbers and the central limit theorem, had been discredited in France by its erroneous applications. Beginning in 1906, the mathematician Émile Borel undertook to revive it, emphasizing the certainty associated with statistical physics.

- Writing about Cournot's proof of the law of large numbers in Russian in 1910, Aleksandr Chuprov (1874–1926; §2.51) explained that it used two "lemmas", a mathematical one and a logical one. The mathematical one was Bernoulli's theorem, the logical one being a version of our principle. This version, which he called "Cournot's lemma", said not that events of very small probability do not happen, it said only that they happen rarely.

- The Russian statistician Evgeny Slutsky (1880–1948; §2.54) made Chuprov's formulation more widely known by publishing his own understanding of it in German. Chuprov's and Slutsky's ideas were further popularized in German by Oskar Anderson (1887–1960; §2.59).

- The English statistician R. A. Fisher (1890–1962; §2.61) expressed Cournot's principle using the notion of typicality.

- Richard von Mises (1883–1953; §2.55) saw probability relating to the world in somewhat different ways in statistics and in physics. In statistical physics, he followed Boltzmann in taking a conclusion with exceedingly high probability to be a scientific prediction. In ordinary applications of statistics, he emphasized Bayes's theorem but considered a mathematical probability a feature of a more complicated mathematical object that he called a collective (*Kollektiv*).

- In the 1930s, Abraham Wald (1902–1950; §2.68) introduced the notion of computability into the interpretation of probability and randomness. As he observed, there are only a countable number of computable tests (= computable sets of probability zero); as there union also has probability zero, they call all be ruled out.

- In 1939, Jean Ville (1910–1989; §2.76 gave a betting interpretation of probabilities equal to zero or one and advanced another formulation of Cournot's principle: only probabilities close to zero or one have direct meaning.

- In 1944, Trygve Haavelmo (1911–1999; §2.77) used Cournot's principle to interpret stochastic processes in cases, as in macroeconomics, where the process as a whole is played out only once and cannot be repeated.

- In 1951 Maurice Fréchet completed the naming of Cournot's principle, replacing Chuprov's and Slutsky's "lemma" by "principle", on the grounds

that "lemma" should be reserved for its use in axiomatic mathematics. He stated three versions of the principle and added to the principle the caveat (which goes without saying, he said) that the event of small probability should be specified in advance. He also asserted that the principle supported his own favorite explication of the meaning of probability — that it is a physical quantity like length, that can be assessed only approximately. The assessment is by calculating frequency in repeated trials. This claim seems to overlook the case of a stochastic process that cannot be repeated, where, as Haavelmo had pointed out, Cournot's principle can still be used.

- Bruno de Finetti (1906–1985; §2.73), partly in response to Fréchet, rejected and mocked Cournot's principle as a way of relating probability to the world. According to de Finetti's subjective interpretation of probability, probabilities are related to the world only through real or potential behavior and choices. Yes, a person may treat events to which he gives very low probability as if they will not happen, but this special case is not fundamental. Other subjective Bayesians, including Colin Howson (1945–2020) and Peter Urbach (§2.89) and Persi Diaconis (born 1945) and Brian Skyrms (born 1938) have similarly rejected Cournot's principle (§2.88).

Except where otherwise noted, the translations are mine. Usually I first give the translation and then provide the original in blue.

2.1   John of Salisbury, c. 1115–1180

John of Salisbury was a prominent twelfth-century English cleric; he served as secretary to Thomas Becket and then became Bishop of Chartres in France.

At the University of Paris, where John studied, and at many other Catholic universities, the seven topics of undergraduate study were divided into the *Trivium* (grammar, dialectic, and rhetoric) and the *Quadrivium* (arithmetic, geometry, astronomy, and music). Was the Trivium important, or did it merit the disdain that its detractors attach to the word *trivial*? In 1159, John completed a manuscript in defense of the Trivium. He called it the *Metalogicon* [160].

In the prologue of *Metalogicon*, John aligned himself with Cicero, with these words:

> Being an Academician in matters about which the wise are uncertain,
> I cannot swear to the truth of what I say. Whether such propositions
> be true or false, I am satisfied with mere probability.

> Academicus in his quae sunt dubitabilia sapienti, non iuro uerum
> esse quod loquor, sed seu uerum seu falsum sit, sola probabilitate
> contentus sum.

(See also [82, pp. 233–235].) In Section 14 of Book 2, John took another step towards "morally certain". Here I will quote Daniel McGarry's translation [161, p. 106]:

Probability alone is sufficient for dialectic. . . . There are some things whose probability is so lucidly apparent that they come to be considered necessary; whereas there are others which are so unfamiliar to us that we would be reluctant to include them in a list of probabilities.

Sola enim probabilitas dialectico sufficit. . . . sunt enim quaedam tanta probabilitatis luce conspicua, ut etiam necessaria reputentur. Quaedam autem eo quod opinioni minus familiaria sint, uix ascribuntur probabilibus.

2.2 Thomas Aquinas, 1225–1274

We need not skip over Aquinas. In a passage in the *Summa Theologiae* that he composed in Paris in 1271–1272, Aquinas quoted Aristotle's dictum that we must not expect to find certainty in all things and added that in matters that are contingent and variable, probable certainty suffices (*Et ideo sufficit probabilis certitudo*).[4]

Here is an English translation and the original text, both provided by the Aquinas Institute `https://aquinas.cc/la/en/~sT.II-II.Q70.A2.sC`:

According to the Philosopher (*Ethic.* i, 3), *we must not expect to find certitude equally in every matter.* For in human acts, on which judgments are passed and evidence required, it is impossible to have demonstrative certitude, because they are about things contingent and variable. Hence the certitude of probability suffices, such as may reach the truth in the greater number of cases, although it fail in the minority. Now it is probable that the assertion of several witnesses contains the truth rather than the assertion of one: and since the accused is the only one who denies, while several witness affirm the same as the prosecutor, it is reasonably established both by Divine and by human law, that the assertion of several witnesses should be upheld. . . .

Respondeo dicendum quod, secundum philosophum, in I Ethic., *certitudo non est similiter quaerenda in omni materia.* In actibus enim humanis, super quibus constituuntur iudicia et exiguntur testimonia, non potest haberi certitudo demonstrativa, eo quod sunt circa contingentia et variabilia. Et ideo sufficit probabilis certitudo, quae ut in pluribus veritatem attingat, etsi in paucioribus a veritate deficiat. Est autem probabile quod magis veritatem contineat dictum multorum quam dictum unius. Et ideo, cum reus sit unus qui negat, sed multi testes asserunt idem cum actore, rationabiliter institutum est, iure divino et humano, quod dicto testium stetur. . . .

---

[4]Secunda Secundae, question 70, article 2.

2.3   Jean Gerson, 1363–1429

Jean Gerson was elected Chancellor of the University of Paris at the age of 32, in 1395. He subsequently took a prominent role at the Council of Constance but was forced into exile from Paris as a result of his opposition to the justification of the murder of the Duke of Orléans as tyrannicide. He is said to have been the first to use the term *certitudo moralis* in writing, in a work he completed in exile in 1418, *De consolatione theologiae* [95]; see [82, 162].

This passage, from [95, p. 231], is quoted by Knebel [110, p. 55] and Schüssler [162]:

> Denique certitudo quae moralis dici potest vel civilis tangitur ab Aristotele . . . non enim consurgit certitudo moralis ex evidentia demonstrationis, sed ex probabilibus conjecturis, grossis et figuralibus, magis ad unam partem quam ad alteram.

> . . . the certainty that can be called moral or civil is touched on by Aristotle . . . moral certainty arises not from the evidence of demonstration, but from probable conjectures, broad and figurative, more on one side than on the other.

Sven Knebel cites several 16th-century authors who adopted *certitudo moralis* and Gerson's definition. In 1646, the Council of Trent used *certitudo moralis et probabilis* instead of Aquinas's *certitudo probabilis* [110, p. 55].

2.4   Andreas de Vega, 1498–1549

By the 16th century, the notions of certainty and possibility were involved in the Roman Catholic doctrine of justification by faith. Was it possible for a person to lead a blameless life? Was damnation certain for those not selected by God?

The Spanish Franciscan Andreas de Vega participated in the sixth session of the Council of Trent, which promulgated its decree on the justification of faith in January 1547. In 1548, de Vega published *Tridentini decreti de justificatione expostiion et defensio lib. XV distincta*, which defended the decree and attacked John Calvin's objections to it. Here is a passage, republished in [182, p. 651] and quoted in [110, p. 147]:

> Some things are called morally possible or impossible, others logically or metaphysically. Morally, possible things can be done often and without great difficulty, impossible things cannot be done except very rarely and with great difficulty. . . . Things that are logically or metaphysically possible can nevertheless be done, those that are logically or metaphysically impossible cannot be done in any way.

> Quaedem einim dicunter moraliter possibilia, seu impossibilia, alia logice, seu metaphysicae. Et moraliter ea dicuntur possibilia, quae

saepe et sine magna difficultate fieri possunt. Contra vero impossibilia dicuntur, quae fieri non possunt, nisi rarissime et cum magna difficultate. . . Logicè verò seu metaphysicè possibilia dicuntur, quae utcunq fieri possunt & econtrario impossibilia, quae null modo.

2.5   Bartolomé de Medina, 1527–1580

The Spanish Dominican Bartolomé de Medina taught at the University of Salamanca. Less known than some of his fellow Dominicans there, he is now remembered for the first clear formulation of the doctrine of *probabilism*. In 1577, in a commentary on Aquinas, he acknowledged arguments that one should follow the most probable opinion, but added this remark, translated by Stefania Tutino [179, Ch. 2].

> Certainly these seem very strong arguments, nevertheless I think that if an opinion is probable, it is legitimate to follow it even if the alternative is more probable.

> Certe argumenta videntur optima, sed mihi videtur quod si est opinio probabili [sic], licitum est eam sequi, licet opposita probabilior sit.

2.6   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 [110].

Molina's complicated reconciliation of grace and free will, which became known as *Molinism*, sparked a controversy that continued for a century. See [179] for history of probabilism.

2.7   Thomas Granger, 1578–1627

Many probabilities concurring prevail much. [97]

2.8   Juan de Lugo y Quiroga, 1583–1660

In the early 17th century, the Spanish Jesuits developed a tripartite classification for evidence, certainty and impossibility. The three types were metaphysical, physical, and moral. This tripartite distinction made its way into the 18th-century French encyclopedias, and physical and moral certainty became identified with high probability.

Juan de Lugo, who became a professor of philosophy and theology in Spain and then a cardinal in Rome, expressed the tripartite classification clearly in 1646 [126, Disput. II, Sectio I]:

Thus it is metaphysical evidence when it appears clearly that the thing is cannot possibly be any other way, e.g. two and two make four, nothing can be and not be at the same time, etc. Physical evidence, however, is when the thing is clearly established, although it could be different metaphysically, but not physically, or taking into account the power of physical and natural causes, e.g. fire makes warmth, the substance of bread is present under the appearances of bread, etc. Finally, it is called moral evidence when the contrary is not ruled out metaphysically, nor even physically, that is, taking into account natural causes; yet it appears so clearly and assuming the contrary is so difficult that assuming the contrary is never assumed, or at least such an assumption is never believed.

Quare evidentia metaphysica est, quando clarè apparet, rem nullo modo poffe aliter se habere, v.g. duo & duo esse quatuor; nihil posse simul esse, & non esse, & alia similia. Evidentia autem physica est quando constat clarè rem, licèt metaphysicè posset aliter se habere, non tamen physicè, seu attenta virtute causarum physicarum & naturalium, v.g. ignem applicatum subjecto capaci calefacere, sub accidentibus panis dari panis substantiam & similia. Denique evidentia moralis dicitur, quando licèt metaphysicè non repugnet contrarium, neque etiam physicè, hoc est attenta virtute causarum naturalium; apparet tamen clarè talis, & tanta difficultas, ut ratione illius nunquam contrarium ponatur, vel ponendum credatur in aliquo casu.

De Lugo was not the only or the first Spanish Jesuit to make a metaphysical/physical/moral distinction. Earlier authors in this tradition include Luis de Molina (1535–1600), and Francisco Suárez (1548–1617). Molina, a Jesuit who taught at universities in Spain and Portugal, wrote about metaphysical/physical/moral impossibility. He used dice as an example. This is documented by Sven K. Knebel [110, 392–399]. Saurez, in his Metaphysicarum disputationum of 1597, distinguished metaphysical, physical, and moral arguments for God's existence.

2.9   René Descartes, 1596–1650

Towards the end of the French edition of his *Principles of Philosophy* [62, 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.

... afin que je ne face point de tort à la verité en la supposant moins certaine qu'elle n'est je distingueray icy deux sortes de certitudes. La premiere est apelée morale c'est à dire suffisante pour regler nos mœurs; ou aussi grande que celle des choses dont nous n'auons point coustume de douter touchant la conduite de la vie, bien que nous sçachions qu'il se peut faire, absolument parlant, qu'elles soient fausses. Ainsi ceux qui n'ont jamais esté à Rome ne doutent point que ce ne soit vne ville en Italie, bien qu'il se pourroit faire que tous ceux desquels ils l'ont appris les ayent trompez. Et si quelqu'vn pour duiner vn chiffre écrit auec les lettres ordinaires s'auise de lire vn C par tout où il y aura vn B, & de lire vn C par tout où il y aura vn B; & ainsi de substituer en la place de chaque lettere celle qui la suit en l'ordre de l'alphabet, & que le lisant en cette façon il y trouue des paroles qui ayent du sens, il ne doutera point que ce ne soit le vray sens de ce chiffre qu'il aura ainsi trouué, bien qu'il se pourrait faire que celuy qui la écrit y en ait mis vn autre tout different en donnant vune autre signification à chaque lettre : car cela peut si difficilement arriuer, principalement lors que le chiffre contient beaucoup de mots, qu'ils n'est moralement croyable.

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

In the Latin version of Part IV of his *Principles of Philosophy*, Descartes says that "some things are considered as morally certain, that is, as having sufficient certainty for application to ordinary life, even though they may be uncertain in relation to the absolute power of God". See pp. 289–290 of *The Philosophical Writings of Descartes*, Vol. 1, J. Cottingham, R. Stoothoff, and D. Murdoch (eds.), Cambridge University Press, 1985.

Jesuit Philosophy on the Eve of Modernity Edited by Cristiano Casalini, Brill, Leiden and Boston, 2019

- Chapter 16, Descartes and the Jesuits, pp. 405–425, by Alfredo Gatto. Notes that Descartes studied at the Jesuit College of La Flèche for 8 years.

- Chapter 17, John Locke and the Jesuits on Law and Politics, pp. 426–443 by Elliot Rossiter.

2.10   Antoine Arnauld, 1612–1694, and Pierre Nicole, 1625–1695

The word *probabilité* does appear, along with *certitude morale*, in the later sections of the Port Royal Logic [10], published by Pascal's friends and fellow Jansenists Antoine Arnauld and Pierre Nicole in the year of Pascal's death.

We find this passage in Ch. XIII of Part IV:

> . . . si toutes ces circonstnaces ont telles, qu'il arrive jamais ou fort rarement que de pareilles circonstances soient accompagnées de fausseté, notre esprit se porte naturellemnt čroire que cela est vrai, & il a raison de le faire, sur-tout dans la conduite de la vie, qui ne demande pas une plus grande certitude que cette certitude morale, & qui se doit même contenter en plusieurs rencontres de la plus grande probabilité.

Then, in Ch. XIV, we find this:

> . . . come nous nous devons contenter d'une certitude morale dans les choses qui ne sont pas susceptibles d'une certitude metaphysique, lors aussi que nous ne pouvons pas avoir une entière certitude morale, le mieux que nous puissions faire quand nous sommes engagés à prendre parti, est d'embrasser le plus probable, puisque ce seroit un renversement de la raison d'embrasser le moins probable.

The authors also echo Pascal's argument for belief in God and piety, expressing more clearly than Pascal did the premise of the argument:

> . . . pour juger de ce que l'on doit faire our obtenir un bien, ou pour éviter un mal, il ne faut pas seulement considerer le bien & le mal en soi, mais aussi la probabilité qu'il arrive ou n'arrive pas; & regarder geometriquement la proportion que toutes ces choses ont ensemble . . .

2.11   Christiaan Huygens, 1621–1695

From the preface of his *Treatise on Light*, published in French in 1690:

> Translate.

The French original in on pp. 2–3 of the preface [105]:

> On y verra de ces sortes de demonstrations , qui ne produisent pas une certitude aussi grande que celles de Geometrie, & qui mesme en different beaucoup , puisque au lieu que les Geometres prouvent leurs Proportions par des Principes certains & incontestables, icy les Principes se verifient par les conclusions qu'on en tire ; la nature de ces choses ne souffrant pas que cela se fasse autrement. Il est possible

toutefois d'y arriver à un degré de vraisemblance, qui bien souvent ne cede guère à une evidence entière. Sçavoir lors que les choses, q'on a demontrées par ces Principes supposez, se raportent parfaitement aux phenomenes que l'experience a fait remarquer ; sur tout quand il y en a grand nombre, & encore principalement quand on se forme & prévoit des phenomenes nouveaux , qui doivent suivre des hypotheses qu'on employe, & qu'on trouve qu'en cela l'effet repond à nostre attente. Que si toutes ces preuves de la vraisemblance se rencontrent dans ce que je me suis proposé de traiter , comme il me semble qu'elles font, ce doit estre une bien grande confirmation du succês de ma recherche , & il se peut malaisement que les choses ne foient à peu près comme je les represente.. . .

2.12   Blaise Pascal, 1623–1662

Pascal did not use the word "probabilité" in his work on the mathematics of games of chance. Nor does it appear in his *Pensées*. When he presents his famous probabilistic argument for belief in God in the *Pensées*, he uses the word *certitude* instead:

. . . il ne sert de rien de dire qu'il est incertain si on gagnera, & qu'il est certain qu'on hasarde; & que l'infinie distance qui est entre la certitude de ce qu'on expose & l'incertitude de ce que l'on gagnera égale le bien fini qu'on expose certainement à l'infiny qui est incertain. Cela n'est pas ainsi: tout joüeur hasarde avec certitude pour gagner avec incertitude; & neanmoins il hasarde certainement le fini pour gagner incertainement le fini, sans pécher contre la raison. Il n'y a pas infinité de distance entre cette certitude de ce qu'on expose, & l'incertitude du gain; cela est faux. Il y a à la vérité infinité entre la certitude de gagner & la certitude de perdre. Mais l'incertitude de gagner est proportionnée à la certitude de ce qu'on hasarde selon la proportion des hasards de gain & de perte: & de là vient que s'il y a autant de hasards d'un costé que de l'autre, le parti est à joüer égal contre égal; & alors la certitude de ce qu'on expose est égale à l'incertitude de ce qu'on expose est égale à l'incertitude du gain, tant s'en faut qu'elle en soit infiniment distante. & ainsi nostre proposition est dans une force infinie, quand il n'y a que le fini à hasarder à un jeu où il y a pareils hasards de gain que de perte, & l'infiny à gagner. Cela est démonstratif, & si les hommes sont capables de quelques véritez ils le doivent estre de celle là.

2.13   John Locke, 1632–1704

Locke published his *An Essay Concerning Human Understanding* in 1689 [124]. 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 infalli-
> bly demonstrated, and that our knowledge of them was perfect and
> certain.

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

In 1992 [78], 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.

2.14 Gottfried Wilhelm Leibniz, 1646–1716

Leibniz's musings about probability were mostly unpublished [140]. The Port
Royal Logic was surely part of his intellectual background. But like Pascal, he
did not emphasize the notion of moral certainty. We can count him as original
in two respects:

- Degree of probability was clearly represented by a number in the Port
  Royal Logic, but full certainty was not represented by the number 1. Leib-
  niz took this step.

- Leibniz saw this numerical probability as a degree of *possibility* rather
  than as a degree of a person's certainty.

These are nuances, but they may have influenced Bernoulli.

2.15 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 proba-
  bility 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. . . .

2.16 John Arbuthnot, 1667–1735

[7]; see articles by Shoesmith.

2.17 Georges-Louis Buffon, 1707–1788

Georges-Louis Leclerc, Comte de Buffon, was a distinguished naturalist and a polymath. Like d'Alembert (see §2.20), he saw Cournot's principle as a solution to the St. Petersburg paradox.

In 1777 [36], 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 [125].

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 [37]. 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.

J'ai reçu, Monsieur, et parcouru avec grand plaisir votre savant "Mémoire sur la probabilité des causes par les événements", et, sans avoir le talent, que vous avez la bonté de m'accorder, de savoir remonter aux causes par les événements, du moins par des voies aussi sûres que les vôtres, j'ai senti la beauté de votre travail et je ne puis que vous exhorter, Monsieur, à continuer vos recherches en ce genre, où il faut plus de délicatesse et d'esprit pur que dans aucune autre partie des mathématiques. J'ai trouvé vos idées d'accord avec les miennes jusquà l'endroit où vous parlez du jeu de croix et pile: la différence matérielle de la piece doit en effet influer à la longue sur le nombre des événements pour ou contre, mais ce n'est pas là la vraie cause qui fait qu'une probabilité, qui dans la speculation est infinie, devient néanmoins finie dans la pratique, et qui, au lieu d'un équivalent infini d'écus ou de demi-écus, fait qu'on se ruinerait si l'on donnait seulement six ou sept écus ou demi-écus toutes les fois qu'on voudra jouer ce jeu. Plusieurs géometres, et entre autres Monsieur Fontaine, qui ont voulu résoudre ce problème, en ont tous manqué la solution, faute d'un principe métaphysique et moral qui se combine ici avec le calcul mathématique; *ce principe est que, toutes les fois qu'une probabilité excède 1/1000, elle est relativement à nous,* parfaitement ègale a zéro. Quelque contradictoire que cette proposition paraisse dans son énoncé, je puis néanmoins la démontrer a n'en pouvoir douter; mais nous causerons de cette matière lorsque j'aurai le plaisir de vous revoir.

2.18   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 [103], 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 [104]. 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 igno-
> rance 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.

2.19   Denis Diderot, 1713–1784

The second volume of Diderot's *Encyclopédie* [65], which appeared in 1752, contained an article on *Certitude* on pp. 845–862. Most of the article was written by Jean-Martin de Prades. But the preface to the article, written by Diderot, included this passage reporting on the distinctions the scholastics had made.

> On distingue dans l'Ecole deux sortes de certitude ; l'une de spéculation, laquelle naît de l'évidence de la chose ; l'autre d'adhésion, qui naît de l'importance de la chose. Les Scholastiques appliquent cette derniere aux matieres de foi. Cette distinction paroît assez frivole : car l'adhésion ne naît point de l'importance de la chose, mais de l'évidence ; d'ailleurs la certitude de spéculation & l'adhésion sont proprement un seul & même acte de l'esprit.
>
> On distingue encore, mais avec plus de raison, les trois especes suivantes de certitude, par rapport aux trois degrés d'évidence qui la font naître.
>
> La certitude métaphysique est celle qui vient de l'évidence métaphysique. Telle est celle qu'un Géometre a de cette proposition, que les trois angles d'un triangle sont égaux à deux angles droits, parce qu'il est métaphysiquement, c'est-à-dire, absolument aussi impossible que cela ne soit pas, qu'il l'est qu'un triangle soit quarré.
>
> La certitude physique est celle qui vient de l'évidence physique : telle est celle qu'a une personne, qu'il y a du feu sur sa main, quand elle le voit, & qu'elle se sent brûler ; parce qu'il est physiquement impossible que cela ne soit pas, quoiqu'absolument & rigoureusement parlant, cela pût ne pas être.
>
> La certitude morale, est celle qui est fondée sur l'évidence morale : telle est celle qu'une personne a du gain ou de la perte de son procès, quand son Procureur ou ses amis le lui mandent, ou qu'on lui envoye copie du jugement ; parce qu'il est moralement impossible que tant de personnes se réunissent pour en tromper une autre à qui elles prennent intérêt, quoique cela ne soit pas rigoureusement & absolument impossible.

2.20 Jean Le Rond d'Alembert, 1717–1783

In 1761 [52, 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.*

> Que conclure de ces refléxions? C'est que *quand la probabilité d'un événement et fort petite, elle doit être regardée & traitée comme nulle; & qu'il ne faut point multiplier* (comme on l'a prescrit jusqu'à présent) *cette probabilité par le gain espéré pour avoir l'enjeu ou l'espérance.* Par exemple, que Pierre joue avec Jacques en 100 coups, à cette condition que si Pierre amene *croix* au centiéme coup, & non auparavant , il recevra de Jacques 2^{100} écus: on trouve (en suivant la régle ordinaire) que Pierre devroit donner un écu à Jacques avant le jeu. Or je dis que Pierre ne doit pas donner cet écu; parce qu'il le perdra *certainement*, & que croix arrivera *certainement* avant le centiéme coup, bien qu'il ne doive pas arriver *nécessairement.*

2.21 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 [43, 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 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.

> M. de Buffon proposait d'assigner une valeur précise à la probabilité très-grande, que l'on peut regarder comme une certitude morale, et de n'avoir au delà de ce terme, aucun égard à la petite possibilité d'un événement contraire. Ce principe est vrai , lorsque l'on veut seulement appliquer à l'usage commun le résultat d'un calcul; et dans ce sens tous les hommes l'ont adopté dans la pratique, tous les philosophes l'ont suivi dans leurs raisonnements: mais il cesse d'être juste, si on l'introduit dans le calcul même, et surtout si on veut l'employer à établir des théories, à expliquer des paradoxes, à prouver ou à combattre des règles générales. D'ailleurs, cette probabilité, qui peut s'appeler certitude morale, doit être plus ou moins grande, suivant la nature des objets que l'on considère, et les principes qui doivent diriger notre conduite; et il aurait fallu marquer pour chaque genre de vérités et d'actions, le degré de probabilité où il commence à être raisonnable de croire et permis d'agir.

The following passage, from Condorcet's essay on the probabilities of decisions [42, xiii–xv], gives an account of the difference between complete and physical certainty.

> Il est cependant entre les vérités, regardées comme ayant une certitude entière & les autres, une différence qu'il est essentiel de remarquer.
>
> Pour les premières, nous ne sommes obligés d'admettre qu'une seule supposition fondée sur la probabilité, celle que le souvenir d'avoir eu la conscience de la vérité d'une proposition ne nous ayant jamais trompé, ce même souvenir ne nous trompera point dans une nouvelle occasion: mais pour les autres, le motif de croire eft fondé d'abord sur ce principe, & ensuite sur l'espèce de probabilité propre à chaque objet. La possibilité de l'erreur dépend de plusieurs causes combinées. Si on la suppose la même pour chacune, le calcul montrera qu'elle sera plus que double s'il y deux causes, plus que triple s'il y en a trois, &c. Ainsi nous donnons le nom de certitude mathématique à la probabilité, lorsqu'elle se fonde sur la constance des loix observées dans les opérations de notre entendement. Nous appelons certitude physique la probabilité qui suppofe de plus la même constance dans un ordre de phénomènes indépendans de nous, & nous conservons le nom de probabilité pour les jugemens exposés de plus à d'autres sources d'incertitude.

> Si nous comparons maintenant le motif de croire les vérités que nous venons d'examiner, avec le motif de croire d'après une probabilité calculée, nous n'y trouverons que trois différences; la première, que dans les espèces de vérités que nous avons examinées, la probabilité est inassignable, & presque toujours tellement grande qu'il seroit superflu de la calculer: la seconde, qu'accoutumés dans le cours de la vie à fonder nos jugemens fur cette probabilité, nous formons ces jugemens fans songer à la nature du motif qui les détermine, au lieu que dans les questions soumises au calcul des probabilités, nous y arrêtons notre attention: dans le premier cas, nous cédons fans le savoir à un penchant involontaire; dans le second, nous nous rendons compte du motif qui détermine ce penchant: la troisième, que dans le premier cas nous pouvons savoir seulement que nous avons des motifs de croire plus ou moins forts; au lieu que dans la seconde, nous pouvons exprimer en nombres les rapports de ces différens motifs .

2.22   Pierre Simon Laplace, 1749–1827

In this paragraph, from the introduction to his *Essai philosophique sur les probabilités* [114],[5] 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.

> Lorsqu'un événement simple ou composé de plusieurs événemens simples, tel qu'une partie de jeu, a été répété un grand nombre de fois ; les possibilités des événemens simples , qui rendent ce que l'on a observé, le plus probable, sont celles que l'observation indique avec le plus de vraisemblance : à mesure que l'événement observé se répète, cette vraisemblance augmente et finirait par se confondre avec la certitude, si le nombre des répétitions devenait infini.

---

[5]The paragraph quoted appears on p. 81 of Bernard Bru's critical edition of Laplace's fifth edition, published in 1825. The translation is mine. There are several English translations of the entire *Essai*; the paragraph quoted appears on p. 36 of Andrew Dale's translation [115].

2.23 Joseph Fourier, 1768–1830

Fourier proposed a threshold for practical certainty when calculating what we now call a large-sample confidence interval. 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 of what we now call 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 is equal to $\sqrt{2}$ times from A's standard deviation; some later authors called g the *modulus*. Fourier 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 [80, 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 $6 + 0.246$.

2.24 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* [3], 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.

> En représentant , comme on le fait ordinairement , par l'unité la certitude absolue , celle par exemple qui résulte d'une démonstration rigoureuse , on pourra regarder comme une certitude morale toute fraction variable qui , sans devenir jamais égale à l'unité , peut en approcher d'assez près pour surpasser toute fraction déterminée.

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

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

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 (1), 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.

1*. en écartant les considérations morales qui font varier la valeur de l'argent, suivant les circonstances où se trouvent les joueurs , il ne saurait y avoir aucun désavantage à jouer à jeu égal contre un adversaire également riche , puisque l'un ne peut rien perdre que l'autre ne gagne, et que tout est égal de part et d'autre; 2*, la même chose a lieu entre deux joueurs , de fortunes inégale , s'ils sont décidés à ne faire qu'un nombre de parties déterminé , et assez petit pour que ni l'un ni l'autre ne puisse être dans le cas de perdre tout ce qu'il pcssède ; 3*. il n'en est pas de même lorsqu'il s'agit d'un nombre indéfini de parties : la possibilité de tenir le jeu plus long-temps , donne au plus riche des deux joueurs un avantage d'autant plus grand qu'il y a plus de différence entre leurs fortunes ; 4*. cet avantage deviendrait infini , si l'une des fortunes pouvait l'être , le joueur le moins riche serait alors sûr de se ruiner , et c'est pour cela que c'est courir à une ruine certaine , que de jouer indifféremment contre tous ceux qui se rencontrent dans la société : on doit en effet , dans la théorie , les considérer comme un seul adversaire dont la fortune serait infinie.

2.25 Siméon Denis Poisson, 1781–1840

Poisson advanced Laplace's theory substantially. Beginning in the 1820s, he simplified the proof of Laplace's theorem, making it accessible to many more mathematicians [100, §17.3]. In 1830, he gave straightforward instructions for calculating limits of practical certainty for the difference between two proportions [151].[6] Finally, in 1837, he pulled together his theoretical and applied results on probability in an impressive treatise, *Recherches sur la probabilité des jugements* [152].

Like Fourier, Poisson discussed limits in terms of numbers of moduli. When writing theory, he required 3, 4, or even 5 moduli for practical certainty [152, §§80, 87, and 96]. But when analyzing data, he used less exigent limits. In §89, when dealing with Buffon's data, he gave limits and odds corresponding to 2 moduli. In §111, he reduced this to 1.92 moduli, corresponding to a bet at odds 150 to 1.

An example of a theoretical discussion is found in §87, where Poisson considered the problem of testing whether the unknown probability of an event E has changed between the times two samples are taken. There are μ observations in the first sample; E happens in n of them, and its opposite $F = E^c$ happens in $m = \mu - n$ of them. For the second sample, he uses analogous symbols μ', n', and m'. He gives formulas, under the assumption that the unknown probability has not changed, for the estimated modulus of the difference $\frac{m'}{\mu'} - \frac{m}{\mu}$ and for the probability that this difference will be within u moduli of 0. Then he writes,

> So if we had chosen a number like three or four for u, making the probability $\tilde{\omega}$ very close to certainty (n⁰ 80), and yet observation gives values for $\frac{m'}{\mu'} - \frac{m}{\mu}$ or $\frac{n'}{\mu'} - \frac{n}{\mu}$ that are substantially outside these limits, we will have grounds to conclude, with very high probability, that the unknown probabilities of the events E and F have changed in the interval between the two series of trials, or even during the trials.

> Si donc on a pris pour u un nombre tel que trois ou quatre, qui rende la probabilit'e $\tilde{\omega}$ très approchante de la certitude (n⁰ 80), et si, néamoins, l'observation donne pour $\frac{m'}{\mu'} - \frac{m}{\mu}$ ou $\frac{n'}{\mu'} - \frac{n}{\mu}$ des valeurs qui s'écartent notablement de ces limites,on sera fondé à en conclure, avec une très grande probabilité, que les chances inconnu des événements E et F ont changé, dans l'intervalle des deux séries d'épreuves, ou même pendant ces épreuves.

The closest Poisson came to identifying ± 2 moduli with practical certainty may have been in §135 of the book, where he considered the 42,300 criminal trials

---

[6]In his second memoir on mathematical statistics, in 1829 [81], Fourier had explained how to calculate limits on a function of several estimated quantities, but he had not spelled out how his formulas specialize to the case where this function is simply the difference between two proportions.

in France during the years 1825 through 1830. The defendant was convicted in 25,777 of these trials. So his estimate of the average probability of conviction, which he called R_5, was $(42300/25777) \approx 0.6094$. His estimate of its modulus was 0.00335. He states that if we use 2 moduli,

> ...we will also have
> $$P = 0.9953,$$
> for the probability, very close to certainty, that the unknown R_5 and the fraction 0.6094 will not differ from each other by more than 0.0067.

> ...on aura aussi
> $$P = 0.9953,$$
> pour la probabilité, très approchante de la certitude, que l'inconnue R_5 et la fraction 0,6094 ne différent pas de 0,0067, l'une de l'autre.

2.26   Thomas Galloway, 1796–1851

In his *Treatise on Probability* [93, p. 144], Galloway adopted Fourier's suggestion that we can consider it certain that the error of a least squares estimate will not exceed 3 moduli. 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 recommended it in the book on the philosophy of science that he published in 1892 [148, pp. 177, 180].

2.27   Antoine Augustin Cournot, 1801–1877

In 1833, Cournot published a translation into French of John Herschel's *Treatise on Astronomy*, which had appeared in English that same year. In an appendix to the translation, he discussed the application of probability to astronomical observations [134]. Here we find this statement about practical certainty [47, vol. XI.2, p. 686].

> A probability of 1000 to 1 is almost considered equivalent to certainty, and one can hardly make the same judgement about a probability of 12 to 1.

> La probabilité de 1000 contre 1 est presque réputée équivalente à la certitude, et il s'en faut bien qu'on porte le même jugement d'une probabilité de 12 contre 1.

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 to give probability

objective meaning. He said this in his 1843 book, *Exposition de la théorie des chances et des probabilités* [44, §43]:[7]

> ... *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.

> ... *L'événement physiquement impossible est donc celui dont la probabilité mathématique est infiniment petite;* et cette seule remarque donne une consistance, une valeur objective et phénoménale à la théorie de la probabilité mathématique.

The phrase "objective and phenomenal" refers to Kant's distinction between the noumenon, or thing-in-itself, and the phenomenon, or object of experience [53].

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 ([46], §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. ...

> En pratique d'ailleurs et dans le monde des réalités, ce que les géomètres appellent une probabilité infiniment petite, n'est et ne saurait être qu'une probabilité excessivement petite. La pointe de cette aiguille si effilée n'est pas un point mathématique comme le

---

[7]Oscar Sheynin's English translation of Cournot's 1843 book is available at www.probabilityandfinance.com. A German translation appeared in 1849 and a Russian translation in 1970.

sommet du cône en question. Elle devient une pointe *mousse*, regardée à la loupe. Avec quelque soin qu'on ait poli le plan d'acier ou d'agathe sur lequel on essaie de la faire tenir en équilibre, des expériences très délicates y indiqueront des aspérités et des stries. Il en résulte que la probabilité de réussir à mettre l'aiguille en équilibre n'est plus à la rigueur infiniment petite, qu'elle n'est qu'excessivement petite, comme le serait la probabilité d'amener l'as cent fois de suite avec un dé non pipé: ce qui suffit pour que l'on juge, sans crainte d'être démenti par l'expérience, que l'équilibre est physiquement impossible.

Pareille remarque s'applique aux valeurs vénales des chances mises dans le commerce. ...

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 con-

stitute 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.

Résumons en quelques mots les principaux points de doctrine que nous avons pris à tâche d'établir dans cet essai.

1. L'idée de *hasard* est celle du concours de causes indépendantes, pour la production d'un événement déterminé. Les combinaisons de diverses causes indépendantes, qui donnent également lieu à la production d'un même événement, sont ce qu'on doit entendre par les chances de cet événement.

2. Quand, sur une infinité de chances, il n'y en a qu'une qui puisse amener l'événement, cet événement est dit *physiquement impossible.* La notion de l'impossibilité physique n'est point une fiction de l'esprit, ni une idée qui n'aurait de valeur que relativement à l'état d'imperfection de nos connaissances: elle doit figurer comme élément essentiel dans l'explication des phénomènes naturels, dont les lois ne dépendent pas de la connaissance que l'homme peut en avoir.

3. Lorsque l'on considère un grand nombre d'épreuves du même hasard, le rapport entre le nombre des cas où le même événement s'est produit, et le nombre total des épreuves, devient sensiblement égal au rapport entre le nombre des chances favorables à l'événement et le nombre total des chances, ou à ce qu'on nomme la *probabilité mathématique* de l'événement. Si l'on pouvait répéter l'épreuve une infinité de fois, il serait physiquement impossible que les deux rapports différassent d'une quantité finie. En ce sens, la probabilité mathématique peut être considérée comme mesurant la *possibilité* de l'événement, ou la facilité avec laquelle il se produit. En ce sens pareillement, la probabilité mathématique exprime un rapport subsistant hors de l'esprit qui le conçoit, une loi à laquelle les phénomènes sont assujettis, et dont l'existence ne dépend pas de l'extension ou de la restriction de nos connaissances sur les circonstances de leur production.

4. Si, dans l'état d'imperfection de nos connaissances, nous n'avons aucune raison de supposer qu'une combinaison arrive plutôt qu'une autre, quoiqu'en réalité ces combinaisons soient autant d'événements qui peuveut avoir des probabilités mathématiques ou des possibilités inégales, et si nous entendons par *probabilité* d'un événement le rapport entre le nombre des combinaisons qui lui sont favorables, et le nombre total des combinaisons mises par nous sur la même ligne, cette probabilité pourra encore servir, faute de mieux, à fixer les conditions d'un pari, d'un marché aléatoire quelconque; mais elle cessera d'exprimer un rapport subsistant réellement et

objectivement entre les choses; elle prendra un caractère purement subjectif, et sera susceptible de varier d'un individu à un autre, selon la mesure de ses connaissances. Rien n'est plus important que de distinguer soigneusement la double acception du terme de *probabilité*, pris tantôt dans un sens objectif, et tantôt dans un sens subjectif, si l'on veut éviter la confusion et l'erreur, aussi bien dans l'exposition de la théorie que dans les applications qu'on en fait.

5. La probabilité mathématique, prise objectivement, ou conçue comme mesurant la possibilité des événements amenés par le concours de causes indépendantes, ne peut en général, et lorsqu'il s'agit d'événements naturels, physiques ou moraux, être déterminée que par l'expérience. Si le nombre des épreuves d'un même hasard croissait à l'infini, elle serait, déterminée exactement, avec une certitude comparable à celle de l'événement dont le contraire est physiquement impossible. Quand le nombre des épreuves est seulement très grand, la probabilité n'est donnée qu'approximativement; mais on est encore autorisé à regarder comme extrêmement peu probable que la valeur réelle diffère notablement de la valeur conclue des observations. En d'autres termes, il arrivera très-rarement que l'on commette une erreur notable en prenant pour la valeur réelle la valeur observée.

6. Lorsque le nombre des épreuves est peu considérable, les formules données communément pour l'évaluation des probabilités *à posteriori* deviennent illusoires : elles n'indiquent plus que des probabilités subjectives, propres à régler les conditions d'un pari, mais sans application dans l'ordre de production des phénomènes naturels.

7. Il ne faut pourtant pas conclure de la remarque précédente, que le nombre des épreuves doive toujours être très-grand, pour donner avec une exactitude suffisante et avec un degré suffisant de vraisemblance, les valeurs réelles de la probabilité d'un événement; seulement cette vraisemblance n'équivaudra pas à une probabilité prise dans le sens objectif. On ne pourra pas assigner la chance que l'on a de se tromper, en prononçant que la valeur réelle tombe entre des limites déterminées: en d'autres termes, on ne pourra pas assigner le rapport du nombre des jugements erronés au nombre total des jugements portés dans des circonstances semblables.

8. Indépendamment de la probabilité mathématique, prise dans les deux sens admis plus haut, il y a des probabilités non réductibles à une énumération de chances, qui motivent pour nous une foule de jugements, et même les jugements les plus importants; qui tiennent principalement à l'idée que nous avons de la simplicité des lois de la nature, de l'ordre et de l'enchaînement rationnel des phénomènes, et qu'on pourrait à ce titre qualifier de probabilités *philosophiques*. Le sentiment confus de ces probabilités existe chez tous les hommes raisonnables; lorsqu'il devient distinct, ou qu'il s'applique à des su-

jets délicats, il n'appartient qu'aux intelligences cultivées, ou même il peut constituer un attribut du génie. Il fournit les bases d'un système de critique philosophique entrevu dans les plus anciennes écoles, qui réprime ou concilie le scepticisme et le dogmatisme, mais qu'il ne faut pas, sous peine d'aberrations étranges, faire rentrer dans le domaine des applications de la probabilité màthématique.

Cournot developed his understanding of probability and his broader philosophy of science in a series of books, written during a career as a university professor and administrator. His terminology for statistical inference (*limite de l'écart*, for example) became standard in France and remained so at least until the middle of the 20th century, but his philosophy of probability was less appreciated. His work in economics became widely appreciated after it was discovered by United States economists at the end of the 19th century. But his philosophy of probability has never gained similar traction; as the philosopher Fernand Faure noted in 1905 [73], it is too philosophical for mathematicians and too mathematical for philosophers.

Nevertheless, Cournot's collected works, edited by a team of French philosophers and mathematicians, appeared in fifteen volumes beginning in 1973 [47]. Thierry Martin published an extensive bibliography of work by and about Cournot in 2005 [133]. There are no recent appraisals of his work in English, but relatively recent appraisals in French include those by Martin [129] and Bertrand Saint-Sernin [159]. See also [130, 132, 178, 33, 8, 9].

2.28 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.

2.29 Jules Gavarret, 1809–1890

In his book on medical statistics [94], he adopts Poisson's standard of $2g$.

2.30 William Fishburn Donkin, 1814–1869

Donkin was a British mathematician and astronomer. Here his 1851 paper on probability, [66], 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.

2.31   Jean Baptiste Joseph Liagre, 1815–1891

Belgian statistician, military officer

2.32   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 [72, 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 pro-
> portional 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.

2.33   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 [183, 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 [2]

2.34 Wilhelm Lexis, 1837–1914

Wilhelm Lexis used Fourier 3 moduli in his *Einleitung in die Theorie der Bevölkerungsstatistik* [123, 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.

2.35 Hermann Laurent, 1841–1908

A French statistician with very wide ranging applied interests, including insurance, agriculture, economics, and meteorology, Laurent used probability theory but without appealing to the notion of moral certainty. Comment on [116, 117].

2.36 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[198], p. 80; Seneta 1997[107, 163]).

This comment by Boltzmann, in 1898 [20, §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.

> Natürlich ist aber zu bedenken, dass es eben Wahrscheinlichkeitsgesetze sind. Die Möglichkeit der Abweichung von denselben kommt praktisch nicht in Betracht; doch ist ihre Wahrscheinlichkeit im Falle, dass die Zahl der Moleküle eine endliche ist, wenn auch unvorstellbar klein, so doch nicht Null; ja sie kann sogar in jedem bestimmt gegebenen Falle nach den Wahrscheinlichkeitsgesetzen numerisch berechnet werden und verschwindet nur für den Grenzfall einer unendlichen Zahl der Moleküle.

The small probabilities discussed by Boltzmann were many orders of magnitude smaller than those calculated in statistical testing by Laplace, Fourier, Poisson, Cournot, Gavarret, Lexis, and Edgeworth. This fact, together with the importance of the second law in physics, aroused new interest in Cournot's principle among mathematicians who found the statistical work unconvincing or uninteresting. This was particularly true in France, where statistical testing

was most practiced during the 19th century and most discredited by the end of the century. In the early 20th century, the French mathematicians Jacques Hadamard (§2.45), Émile Borel (§2.49), and Paul Lévy (§2.58) all became interested in probability and proponents of versions of Cournot's principle because of its success in statistical physics.

2.37   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 [127]. 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.

> Le principe fondamental est celui-ci : Entre deux propositions contraires, l'une peu probable, l'autre très probable, l'esprit humain choisit la seconde, librement, mais presque invinciblement, et la déclare pratiquement certaine.
>
> On déduit de là la légitimité logique de la loi des grands nombres et du principe de l'accumulation des probabilités indépendantes.
>
> La loi des grands nombres s'applique d'abord à la question de la ruine du joueur et à ses conséquences, ensuite à la statistique chaque fois qu'elle rencontre des rapports à peu près constants dans les nombres qu'elle rassemble.
>
> Le principe de l'accumulation des probabilités indépendantes est pratiquement, sinon mélaphysiquement, la source de nos certitudes dans les sciences naturelles et historiques qui reposent en dernière analyse sur le témoignage, chaque fois que nous ne sommes pas personnellement inventeur ou témoin.

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 [138]. 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.

2.38   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 3 moduli [71, §137].

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 [70]. 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."

2.39   Emanuel Czuber, 1851–1925

Emanuel Czuber was an Austrian mathematician whose textbooks on probability, error theory, and mathematical statistics were widely used at the beginning of the 20th century. His general textbook on probability theory and its applications appeared in 1903, with a second edition in two volumes in 1908 and 1910, and a third edition of the first volume in 1914 [51]. The following paragraph, in which he explicitly rejects the concept of practical certainty appears in all three editions with only minor variations, on pp. 14–15 in 1903, on pp. 16–17 in 1908 (first volume), and on pp. 18–19 in 1914. I have omitted the footnotes.

> The relationship between different 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 detects a small source for error, he allows us to glimpse the correct position, which Condorcet had already taken before him and that the new philosophy asserts: *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. The same goes for the multifarious attempts that have been made to use logical value to establish intermediate elements or transitions from probability to certainty on one hand 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.

Nicht immer richtig ist das Verhältnis zwischen der verschiedener Grade fähigen Wahrscheinlichkeit und der absoluten Notwendigkeit oder Gewißheit beurteilt worden. Jacob Bernoulli bezeichnet die Wahrscheinlichkeit als einen Grad der Gewißheit und schreibt einem Ereignis von der Wahrscheinlichkeit 3/5 den entsprechenden Bruchteil der Gewißheit zu; den gleichen Standpunkt hat auch noch der erste deutsche Philosoph, der sich mit der Kritik der Prinzipien der Wahrscheinlichkeitsrechnung befaßte, J. J. Fries, eingenommen. Auch bei Laplace findet sich eine Stelle, die hieran anklingt; an die Bemerkung, daß die Wahrscheinlichkeit sich in Gewißheit verwandle und ihr Ausdruck die Einheit werde, wenn alle Fälle dem Ereignis günstig sind, knüpft er die Worte an, daß unter diesem Gesichtspunkte Gewißheit und Wahrscheinlichkeit vergleichbar seien; durch den weiteren Zusatz aber, daß ein wesentlicher Unterschied zwischen den beiden Zuständen des Geistes bestehe, wenn ihm eine Wahrheit streng bewiesen ist, oder wenn er noch eine kleine Quelle des Irrtums wahrnimmt, läßt er schon den richtigen Standpunkt durchblicken, auf den vor ihm schon Condorcet sich gestellt hat und den die neuere Philosophie behauptet: *Wahrscheinlichkeit und Gewißheit* (oder Notwendigkeit) *sind Dinge wesentlich verschiedener Natur*, und es gibt keine Brücke, die von der einen zur andern geschlagen werden könnte. Damit sind auch die mannigfachen Versuche ihrem logischen Werte nach gekennzeichnet, welche unternommen worden sind, um Zwischenglieder·oder Übergänge zwischen Wahrscheinlichkeit und Gewißheit einerseits und Unmöglilichkeit andererseits herzustellen. So unterscheidet schon Jacob Bernoulli zwischen *absoluter* und *moralischer* Gewißheit und Unmöglichkeit, unter letzteren

The "new philosophy" to which Czuber refers is the work of Jacob von Kries
(1853–1928; §2.40), whose theory of objective probability was popular in Ger-
man philosophy at the time, and also the work of Carl Stumpf (1848–1936)
on subjective probability. His footnote to his mention of Condorcet refers the
reader to Condorcet's *Essai*. Czuber's assertion that Condorcet insisted on an
essential difference between probability and certainty is confirmed by the pas-
sage from the *Essai* I quote in §2.21, but that same passage shows Condorcet
just as willing to bridge them as Bernoulli, d'Alembert, Buffon, and De Morgan.

In the 1914 edition, Czuber added a new footnote at the end of the passage
just quoted, citing Borel [23, p. 19]:

> In his *Élements de la théorie des probabilités* (Paris 1909), p. 19,
> E. Borel gives a nice example of how confounding large probability
> with certainty can lead to absurd results.

As we know, Borel was actually a persistent advocate of confounding extremely
high probability with certainty. But he also noted counterexamples. I discuss in
§2.49 the counterexample Czuber cites here, which involves a simple martingale.

2.40   Johannes von Kries, 1853–1928

One of the most influential of the German philosophers who discussed probabil-
ity 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 fol-
lowed 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 erro-
neously derived by combining Bernoulli's theorem with the belief that small
probabilities can be neglected. Having both subjective and objective probabili-
ties 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[8] 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 [108] and the special issue on von Kries published by the *Journal for General Philosophy of Science* in 2016, especially [202].

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 [139]; Reichenbach 1916 [155]). John Maynard Keynes (1921)[109] brought it into the English literature. When asked about the philosophical basis of the classical probability calculus,

---

[8]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.

many philosophers and mathematicians today will think about arguments for a uniform distribution of probabilities before they think about Cournot's principle.

2.41   Henri Poincaré, 1854–1912

Poincaré's used a version of Cournot's principle in his study of the three-body problem[149, 198]. His recurrence theorem, published in 1890[149], 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.

In 1902, in *La science et l'hypothèse* [150, Ch. XI], Poincaré argued that probability is essential to science. How do we know that Newton's law still be true in the next generation? How do we know that some huge object will not soon perturb the solar system? We can only say that these things have little probability. To persuade his reader that this reliance on probability is more than merely practical and subjective, he argued as follows:[9]

> ... A gambler wants to try a coup, and he asks my advice. If I give it him, I use the calculus of probabilities; but I shall not guarantee success. That is what I shall call *subjective probability*. ... But assume that an observer is present at the play, that he knows of the coup, and that play goes on for a long time, and that he makes a summary of his notes. He will find that events have taken place in conformity with the laws of the calculus of probabilities. That is what I shall call *objective probability* .... There are numerous Insurance Societies which apply the rules of the calculus of probabilities, and they distribute to their shareholders dividends, the objective reality of which cannot be contested.

> ... Un joueur veut tenter un coup; il me demande conseil. Si je lui donne, je m'inspirerai du calcul des probabilités mais je ne lui garantirai pas le succès. C'est là ce que j'appellerai la *probabilité subjective*. ... Mais je suppose qu'un joueur assiste au jeu, qu'il en note tous les coups et que le jeu se prolonge longtemps; quand il fera le relevé de son carnet, il constatera que les événements se sont répartis conformément aux lois du calcul des probabilités. C'est là ce que j'appellerai la *probabilité objective* .... Il existe de nombreuses sociétés d'assurances qui appliquent les règles du calcul des probabilités et elles distribuent à leurs actionnaires des dividendes dont la réalité objective ne saurait être contestée.

In §2.49 I will quote Émile Borel's comment on this passage. Both Borel and Paul Lévy adopted Poincaré's formulation concerning the difference between

---

[9]Here I use the translation by published by Walter Scott Publishing in 1905.

subjective and objective probability. I have not found an author before Poincaré writing in quite the same way, but the formulation echoes Buffon's contention that the difference between moral and physical certainty is one of degree (§2.17).

2.42   Andrei Markov, 1856–1922

Markov, Chuprov's neighbor in Petersburg, learned about the growing field of mathematical statistics from Chuprov [146], and we see an echo of Cournot's principle in Markov's textbook, which appeared in Russian in 1900. (The passage is on p. 12 of the German edition, which appeared in 1912 [128], 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.

2.43   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.

2.44   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 [40, 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. [40, 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 [88]. It brought Fréchet to the same understanding of objective probability as Lévy: it is a physical constant that is measured by relative frequency [83, p. 5]; [85, 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".

2.45   Jacques Hadamard, 1865–1963

Hadamard, the preeminent analyst who did pathbreaking work on Markov chains in the 1920s (Bru 2003a)[34], 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)[99]. 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.

2.46   Ladislaus von Bortkiewicz, 1868–1931

Born to a Polish family in the Russian empire, Bortkiewicz studied mathematics in St. Petersburg and then studied statistics with Lexis in Straßburg. He spent most of his career as professor of statistics in Berlin.

In 1894, early in his career, Bortkiewicz made these comments [185, pp. 354–355]:

> Was nun den ersten Punkt betrifft, so glaube ich, daß unter
> den zahlreichen Quellen falscher Rückschlüsse der erwähnten Art

die zu kleinen absoluten Zahlen keine bedeutende Rolle spielen.
Der erfahrene Statistiker — er mag in Wahrscheinlichkeitsrech-
nung gar nicht unterrichtet sein — wird in gesagter Beziehung von
einem gewissen "statistischen Sinne" ("unmethodical wisdom" nach
Edgeworth) in der Regel ziemlich sicher geleitet. Es ist freilich
vorzuziehen, wenn eine gewisse Uebung in Wahrscheinlichkeitsrech-
nung hinzutritt, die zur Schärfung des gesagten sinnes wesentlich
beitragen kann. Hingegen erscheint die Berechnung der Präcision
in jedem konkreten Fall als ein kaum zu rechtfertigender Luxus bei
der statistischen Produktion. Man bedenke nur, daß es hierbei viel
weniger auf den numerischen Wert derjenigen Wahrscheinlichkeit
ankommt, mit der auf das Verhalten der in Rede stehenden All-
gemeinbedingungen geschlossen werden darf, als vielmehr auf den
Umstand, ob jener numerische Wert der Einheit dermaßen nahe
kommt, daß man die erhaltene Wahrscheinlichkeit praktisch als
Gewißheit betrachten kann, oder aber die verlangte Höhe nicht er-
reicht. Es liegt daher entschieden eine Uebertreibung in der Ansicht
Westergaard's vor, der geneigt ist, jede statistische Untersuchung
als dilettantenhaft anzusehen, bei der auf die Präcisionen keine
Rücksicht genommen worden ist.

Draft translation:

As far as the first point is concerned, I believe that among the
numerous sources of false inferences of the kind mentioned, too small
an absolute number does not play a significant role. The experienced
statistician - he may not be fully informed in probability theory - is
usually guided fairly well by a certain "statistical sense" (according
to Edgeworth). It is, of course, preferable to add a certain exercise
to probabilistic theory, which can substantially contribute to the
sharpening of the said sense. On the other hand, the calculation of
precision in each specific case appears as a barely justifiable luxury
in statistical production. It is only necessary to consider that the
numerical value of the probability with which the behavior of the
general conditions in question may be inferred is much less impor-
tant than the fact that the numerical value of the unity comes so
close to the probability that the numerical value of the unit is can
practically regard the probability obtained as certainty or does not
reach the required height. There is, therefore, a definite exaggera-
tion in Westergaard's view, who is inclined to regard any statistical
investigation as dilettante, in which no consideration has been given
to the precision.

See also [186, p. 825] and [186, 187, 188].

2.47 Georg Bohlmann, 1869–1928

Bohmann was a German mathematician who worked on Lie groups, meteorology and eventually life insurance. In 1903, he became chief actuary for the Berlin subsidiary of the New York Mutual Life Insurance Company. His contributions are reviewed in [112].

In 1901, while working at Göttingen, Bohlmann contributed a 56-page entry on life insurance to the encyclopedia of mathematical sciences that the Göttingen mathematicians were assembling at that time [19, p. 861]. Bohlmann began the article by stating axioms for probability, principally the rule of total probability (additivity) and the rule of compound probability. Then, in a section entitled "Principien nach denen die Theorie auf die Erfahrung angewendet wird" ("Principles by which the theory is applied to experience"), he stated this postulate (p. 861):

> *Postulate.* If we observe *one single* value of f, it does not deviate from its expected value by more than v times its standard deviation.
>
> How large a value of ν we choose is arbitrary. If we choose $\nu = 3$, then we are identifying practical certainty with a probability that is always greater than $1 - 1/\nu^2 = 8/9$ and is equal to $\Theta\left(\frac{3}{\sqrt{2}}\right) = 0{,}9973$ when the *Gauss*ian error law holds.

> *Postulat.* Beobachtet man *einen einzelnen* Wert von f, so weicht dieser von seinem wahrscheinlichen Werte f^0 um nicht mehr als das ν-fache von $M(f)$ ab.
>
> Wie gross man ν wählt, ist willkürlich. Wählt man $\nu = 3$, so identificiert man die praktische Gewissheit mit einer Wahrscheinlichkeit, die jedenfalls grösser als $1 - \frac{1}{\nu^2} = \frac{8}{9}$ ist und gleich $\Theta\left(\frac{3}{\sqrt{2}}\right) = 0{,}9973$ ist, wenn das *Gauss*'sche Fehlergesetz gilt.

In footnotes, Bohlmann noted that he was following Emanuel Czuber [50] in his use of the term wahrschienlichen Werte and the symbol $M(f)$, that Hermann Laurent had called ν the *coefficent de sécurité* in 1873 in a French actuarial journal [116, p. 162], and that the inequality giving 8/9 was due to Chebyshev.

Bohlmann's standard for practical certainty was less demanding than the 3 moduli used by Fourier, Lexis, and Edgeworth and obviously more practical for actuarial work. It is roughly comparable to Gavarret's 2 moduli in the case where the Gaussian law (normal distribution) applies but much much less demanding if we can use only Chebyshev's inequality.

2.48 Arthur Lyon Bowley, 1869–1957

As economist at the London School of Economics, Bowley learned about statistical methods from Edgeworth. In the first edition of his *Elements of Statistics*, published in 1901, he followed Edgeworth in taking 3 moduli as significant.

2.49 Émile Borel, 1871–1956

In 1906, in the first volume of *La Revue du Mois* [21, p. 433], the journal that Borel and his wife founded and edited, Borel elaborated as follows on the passage from Poincaré's *La science et l'hypothèse* quoted in §2.41:

> The difference between objective and subjective probability is not a difference in nature but a difference in degree. A result of the probability calculus merits being called objective *when its probability becomes large enough to be practically indistinguishable from certainty*. It then matters little whether it is a matter of predicting future phenomena or of estimating past phenomena; in the one case as much as in the other, we can affirm that the law will be or has been verified.

> Ce n'est pas une différence de nature qui sépare la probabilité objective de la probabilité subjective, mais seulement une différence de degré. Un résultat du calcul des probabilités mérite d'être appelé objectif, *lorsque sa probabilité devient assez grande pour se confondre pratiquement avec la certitude.* Il importe peu alors qu'ils s'agisse de prévoir des phénomènes futurs ou de recenser des phénomènes passés; on peut également affirmer que la loi sera ou a été vérifiée.

In 1909, in his first textbook on probability theory *Eléments de la théorie des probabilités*, Borel wrote as follows [22, p. 19]:

> L'une des principales sources de ces raissonnements paradoxaux sur lesquels nous aurons á revenir (see no. 18) est la suivant: *on considère un évènement futur comme réalisé, sous prétext que l'expérience a prouvé qu'il est extrémement probable.* On commet ainsi une erreur, sans doute très petite, mais l'accumulation répétée de telles erreurs suffit pour conduire á des conséquences entièrement inexactes.

> One of the main sources of these paradoxical arguments . . . is the following: *we consider a future event as having happened, on the pretext that experience has proven it to be very probable.* We comment an error when do this, a very small error no doubt, but the repeated accumulation of such errors is enough to lead to consequences that are completely inexact.

To illustrate this error, Borel imagined a gambler, Pierre, who plays a simple martingale, one franc on heads, say, over and over. Eventually he will be ahead, having won one more franc than he has lost; as Borel says, this is a practical certainty (*certitude pratique*). We suppose that he pockets this one franc and starts over, until he nets another franc. And so he continues, becoming as rich as he wants.

In 1910, in the second edition of his *Éléments*, page 181:

. . . we should greatly distrust our tendency to consider as *remarkable* a circumstance that we have not specified *before the observation*, because the number of circumstances that can seem remarkable, from different points of view, is very considerable.

. . . on doit se défier beaucoup de la tendance que l'on a à regarder comme *remarquable* une circonstance que l'on n'avait pas précisée *avant l'expérience*, car le nombre de circonstances qui peuvent apparaître comme remarquables, à divers points de vue, est très considérables.

In Le Hasard, the book he published in 1914 to popularize probability theory, Borel wrote [24, p. 15]:

. . . *the object of probability theory is to evaluate probabilities of complex events by means of probabilities, which are assumed known, of simpler events.* Its purpose is to predict with an almost absolute certainty, humanly certain we may say, certain events whose probability is such that it is indistinguishable from certainty.

. . . *la thórie des probabilités a pour objet d'évaluer les probabilités d'événements complexes au moyen des probabilités supposeées connues d'autres événements plus simples.* Son but, c'est arriver à prévoir avec une certitude presque absolue, humainement absolue peut-on dire, certains événements dont la probabilité est telle qu'elle se confond avec la certitude.

In the 1933 paper where Neyman and Pearson proved their famous lemma [144, p. 290], they quoted Borel saying that the event of small probability should be "en quelque sorte remarquable" and cited the source as the 1920 edition of *Le Hasard*. https://archive.org/details/lehasard00boreuoft/page/112/mode/2up?q=remarquable Borel's 1920 edition of Le hasard, cited by Neyman in Neyman-Pearson 1933 as the source of "en quelque sorte remarquable" has relevant passage on pages 112-113 and is available on archive.org.

Borel, however, in a later discussion, considered that the method described could be applied with success provided that the character, x, of the observed facts were properly chosen — were, in fact, a character which he terms "en quelque sorte remarquable."

Borel did not use exactly these words, but he makes the point around p. 112 using the word *remarquable*. The same passages were also in the first edition (1914); the 1920 edition seems to be a reprinting.

Also, Borel discusses "valeur pratique" in Section VIII and asks how small is small enough in §89. Reference Bru's discussion and Constance Reid [156, pp. 80–81], where Le Cam and Neyman later could not find the wording Neyman

had remembered. And E. S. Pearson's later denial. The whole debate of memory between Neyman and E. S. Pearson is discussed thoroughly by Eric Lehmann in The Bertrand-Borel Debate and the Origins of the Neyman-Pearson Theory, 1993 [119].

Although he never attributed it to Cournot, Borel stated the principle many times, often in a style more literary than mathematical or philosophical [21, 22, 24, 25]. 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 [26, 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*[27], 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*[28], 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[30]). It was also popularized by Robert Fortet, in his essay in François Le Lionnais's *Les grands courants de la pensée mathématique*[118], first published in 1948[79, 118].

2.50 George Udny Yule, 1871–1951

Page 262–263 of the first edition of Yule's statistics textbook, published in 1911 [201]:

> 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 fluctu-

ations 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.

2.51   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 [173]. With some justice, he considered Cournot the founder of the philosophy of modern statistics [173, 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 [173, 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.

2.52   Felix Bernstein, 1878–1956

By the end of the 19th century, it was commonplace in mathematics to interpret a set having measure zero as meaning that it can be neglected. The German mathematician Felix Bernstein gave this an interesting twist in 1912, p. 419 [16]:

> Axiom. Bezieht man die Werte einer experimentell gemessenen Größe auf die Skala der Werte aller reellen Zalhen, so kann man in der letzteren von vorherein eine beliebige Nullmenge ausschalten und darf nur solche Folgen der beobachtete Ereignisse erwarten, welche bestehen bleiben, wenn der beobachtete Wert durch einen der übrig bleibenden, innerhalb des Beobachtungsintervalles gelegenen Werte reprasentiert wird.

2.53   Maurice Fréchet, 1878–1973

In 1949, the Swiss academic journal *Dialectica* published a special issue on *Warhscheinlichkeitstheorie und Wirklichkeit*.[10] Several of the contributors, including Oskar Anderson, Émile Borel, and Paul Lévy, touched on the importance of the impossibility or rarity of events of very small probability. This topic was also discussed in the session on probability theory at the 18th international congress on philosophy of science held in Paris later that year, by Borel and Lévy and by Padrot Nolfi (1903-1973), who had co-edited the *Dialectica* special issue.

Statistical physics probably played the largest role in Borel's thinking about Cournot's principle. He had just completed the last of his many books on probability, *Probabilité et Certitude* [30], in which he insisted that a sufficiently tiny probability, such as the probability for the second law of thermodynamics, must be interpreted as absolute certainty. Lévy had also first seen the importance of probability theory in its applications to statistical physics. Anderson, on the other hand, was a statistician, a student of Chuprov's, and he was content with the weaker form of Cournot's principle enunciated by Chuprov and Slutsky, which says only that events with small probability happen rarely.

Fréchet, president of the session on probability at the congress, added his own introduction to the section's proceedings, in which he summarized the different views on Cournot's principle and undertook to reconcile them with his own views about the relation between probability and frequency. These proceedings appeared in 1951.[11]

---

[10]Volume 3, Numbers 1/2, pp. 1–172, edited by H. Jecklin and P. Nolfi.

[11]The proceedings of the congress, edited by Raymond Bayer, were published by Hermann (Paris) in their series *Actualités Scientifiques et Industrielles* [11]. Part IV of the proceedings, number 1146 in the series, was devoted to the session *Calcul des Probabilités*. Fréchet's report, "Rapport général sur les travaux du colloque de calcul des probabilités", appeared on pp. 3–21 [86], and Nolfi's report on the preceding issue of Dialectica appeared on pp. 23–47. Other contributors were Robert Fortet, David van Dantzig, Bruno de Finetti, Jerzy Neyman, Émile Borel, Jean Ville, George A. Barnard, Paul Lévy, and Georges Darmois.

It was here that Fréchet suggested that what Oskar Anderson called Cournot's lemma or bridge (*die Cournotsche Brücke*) the should instead be called *le principe de Cournot,* which translates into English as *Cournot's principle.* The following passages appear on pp. 6–8 of Fréchet's introduction.

> *Le principe de Cournot.* — La question soulevée par M. BOREL, d'abord dans *Dialectica*, pp. 24–27, puis dans un rapport à cette Section, ne concerne en aucune façon la théorie axiomatique des Probabilités, mais elle est fondamentale pour l'interprétation concrète à donner à cette théorie et par suite pour les applications. C'est aninsi qu'on la retrouve traitée indépendamment dans leurs rapports par MM. de FINETTI et VAN DANTZIG et dans *Dialectica*, par MM. Paul LÉVY et ANDERSON. Aussi allons-nous lui accorder une particulière attention.
>
> Laissant de coté la question des chiffres à adopter pour les probabilités négligeables, il s'agit essentiellement du principe que nous appellerons le principe[12] de Cournot (bien qu'il semble avoir été déjà plus ou moins nettement indiqué par d'Alembert).
>
> Selon O. Anderson, on peut l'énocer ainsi:
>
> *A) un événement don la probabilité est très petite n'a lieu que très rarement.*
>
> Paul LÉVY écrit (*Dialectica*, p. 56): "Il semble que tout le monde soit d'accord au moins sur un point: nous voulons que l'événement don la probabilité est très petite soit très peu probable au sens vulgaire du mot" ; et plus loin : "Mais les phénomènes très peu probables sont des phénomènes rares". En combinant ces deux opinions de Paul Lévy, on retrouve la formulation de M. Anderson.
>
> Il serait facile de multiplier les citations et de montrer que la plupart des auteurs acceptent explicitement ou implicitement cette formulation ou une autre très voisine.
>
> Un grand nombre d'auteurs vont même un peu plus loin ; dans la vie réelle, un événement extrêmement rare n'est guère discernable d'un événement impossible, on pourra donc dire :
>
> *B) un événement de probabilité très petite est un événment "pratiquement impossible"* ; on refusera de croire à sa réalisation.
>
> Par exemple, on tire, sans le remettre, sept boules d'une urne contenant 25 boules portant respectivement les 25 lettres de l'alphabet ; si les sept boules tirées formaient, par example, le mot MIRACLE, on referait de croire que l'opération ait été effectuée dans des conditions régulières. Pourquoi ? Parce que, sans qu'on ait à faire de calcul précis, notre expérience inconsciente quotidienne nous apprend que la formation d'un mot français dans une telle opération,

---

[12]Fréchet's footnote: Anderson lui donne (entre autres dans *Dialectica*, p. 69) le nom de lemme ; comme il précise lui-même que ce "lemme" résulte directement de l'expérience, nous préférons réserver l'usage du mot lemme à la théorie axiomatique.

— formation qui nous appraîtra simplement comme un accident admissible du hasard dans le cas où l'in tiereait seulement une, deux ou trois boules —, nous semblerait tout à fait impossible quand on en tire sept.

On soul̀eve cependant l'objection suivante : considérons un tirage quelconque de 7 boules. Il aura donné, par exemple, CDATIPS. Personne n'en sera étonné ; pourtant la probabilité de tirer CDATIPS était *plus* petite que celle de tirer un quelconque mot français de 7 lettres ! Mais, il faut distinguer. Avant le tirage, l'attention n'était fixée sur aucune permutation particulière de letteres. L'événement qui n'a pas provoqué de surprise, c'était de voir sortir une permutation *non distinguée d'avance* pari limmensité

$$25 \times 24 \times 23 \times 22 \times 21 \times 20 \times 19 > 10^9$$

des permutations possibles.

Au contraire, dan cette immensité, le nombre des permutations formant des mots français et qui se distinguent ainsi des autres permutations, est — relativement — extrêment petit ; la probabilité d'en tirer un est — absolument — très petite.

Supposez qu'on ait annoncé d'avance la formation CDATIPS ; alors, aussi, la surprise aurait très grande — et l'objection tombe.

. . .

Cependant l'objection n'ètait pas pas inutile, car elle conduit p̀réciser ce qui va sans dire — mais qui va encore mieux en le disant, — que, considérant un événement comme pratiquement impossible quand sa probabilité est extrêmement petite, nous entendons formuler une prédiction au suject d'un événement bien défini avant qu'ait lieur l'épreuve où l'on constatera si l'éévenement s'est ou non réalisé.

Fréchet also addresses the objection that when n is large, and events $E_1, \ldots, E_n$ each have probabilities so small that each is considered impossible, the event that at least one of them happens (i.e., their union $E_1 \cup \cdots \cup E_n$) might have a large probability and then be considered possible or even certain. Along with Borel, Fréchet argues (p. 8) that when n is this large, it would not be possible to carry out the verification:

In order to refute the principle, you must not only *conceive* of trials permitting you to confirm whether at least one of the events $E_1, \ldots, E_n$ happens, but you must actually carry out one of these trials and proceed effectively to these confirmations.

Il faudrait, pour mettre en défaut le principe, qu'on puisse non seulement *concevoir* des épreuves permettant de constater la réalisation de l'un au moins des événements $E_1, \ldots, E_n$, mais même réaliser une de ces épreuves et procéder effectivement à ces constatations. . . .

Fréchet then offers a third formuation of the principle:

> The two formulations A and B of Cournot's principle have a platonic character. In many circumstances, the calculation of a probability is nevertheless not an end in itself but should serve to guide our actions. Thus, taking inspiration from a remark by Mr. DIVISIA, we can give it the following form:
>
> C) *When an event has an extremely small probability, it is appropriate to act as if it will not happen.*
>
> This rule of action can be preferred, because in a sense it is less categorical than formulations B), because it does not explicitly say that the event will not happen, and also because it expresses, in my opinion, a rule that we all follow, consciously or unconsciously.

> Les deux formes A, B, du principe de Cournot ont un caractère platonique. Dans bien des circonstances, le calcul d'une probabilité n'est pourtant pas une fin en soi, mais doit servir à guider nos actions. Aussi, en s'inspirant d'une remarque de M. DIVISIA, on peut lui donner la forme suivante :
>
> C) *Quand un événement est de probabilité extrêment petite, il convient d'agir comme s'il ne devait pas se produire.*
>
> Cette règle d'action peut être préféfée, parce qu'en un sense elle est moins catégorique que la forme B), puisqu'ellle ne dit pas expressément que l'événement ne se produira pas ; et d'autre part, parce qu'elle exprime, croyons-nous, une règle que nous suivons tous consciemment ou inconsciemment.

Fréchet had his own way of explaining how probability theory connects with the world; he was fond of saying that probabilities are physical quantities that are measured by frequencies. On p. 5 of the report I have been quoting, he claimed that this follows from Cournot's principle:

> ... we are led to consider a probability as *a physical quantity attached to an event and a category of trials, which is measured approximately by the frequencies of that event in large number of trials.*

> ... on est conduit à considérer la probabilité comme *une grandeur physique attachée à un événment et à une catégorie d'épreuves et dont les fréquences de cet événement dans un grand nombre d'épreuves sont des mesures approchées.*

I think many mathematicians in the second half of the 20th century shared this conception of probability, but I have not seen it defended in print very often. The obvious difficulty is that that many probabilities (those in stochastic processes, for example; see §2.77) relate to only a single trial, which can be repeated only in the imagination.

Following Fréchet's suggestion, a fair number of mathematicians used the terms "principe de Cournot", "Cournot's principle", and "Cournotsche Princip" in the 1950s [58, 189, 157, 158]. It was not unusual 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 and Kolmogorov-Smirnov [35, 147]. He had second thoughts about giving so much credit to Cournot; when he reprinted his 1949 report as a section in a book in 1955 [87], he replaced "principe de Cournot" with "principe de Buffon-Cournot". But here no one else seems to have followed his example.

2.54   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 [175, 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.[13] 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".[14] This premise posits "the fact of an existing connection between small probability and rarity" by claiming that "events whose probabilities are very small will not often happen".[15]
>
> If you want to consider this statement as a nomological one, you again come into contradiction with probability theory. No matter how small the probability of an event, it *can* still occur any number of times in a row in a series 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 each, a certain relative 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, no matter how much we let the numbers grow." [16] 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

---

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

[14]Slutsky's footnote: Ibid., p. 230.

[15]Slutsky's footnote: Ibid., p. 230, 227.

[16]Slutsky's footnote: J. v. Kries, op. cit. p. 21.

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 statement about the latter frequencies themselves. [17] 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 little 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 surrounding us.[18] 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.[19] 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.

---

[17]Slutsky'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 farther from being clarified in the end than from anything, 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.

[18]Slutsky'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).

[19]Slutsky'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 dispensable.

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 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.

Aehnliche Gedanken treten bisweilen in einem Gewand auf, in dem man sie nicht so leicht erkennen kann. Es würde sich deshalb lohnen die Auffassung des Gesetzes der grossen Zahlen zu analisieren, die von Prof. Al. A. Tschuprow aufgestellt und auf A. Cournot's Ansichten zurürckgefürht worden ist.[20] Das wesentliche hier gipfelt in der Behauptung, dass die Ableitung des Gesetzes der grossen Zahlen nicht lediglich auf den bekannten sätzen der Wahrscheinlichkeitsrechnung (von Bernoulli, Poisson u. s. f.), sondern noch auf einem besonderen Lemma sich begründet, kraft welchem es eigentlich erst möglich wird "aus der Welt der Wahrscheinlichkeiten ob gross, oder klein, sich in the Welt der Häufigkeiten hinüberzutragen".[21] Diese Voraussetzung stellt "die Tatsache des zwischen kleiner Wahrscheinlichkeit und Seltenheit obwaltenden Zusammenhanges" fest, indem sie behauptet, dass "die Ereignisse, deren Wahrscheinlichkeiten sehr klein sind, sich nicht oft wiederholen werden [Ibid., p. 230, 227].[22]

Will man diesen Satz als einen nomologischen betrachten, so kommt man wieder zum Widerspruch mit der Wahrscheinlichkeitslehre. Es sei die Wahrscheinlichkeit eines Ereignisses so klein, wie man will, in einer Reihe *unabhängiger* Versuch *kann* es dennoch beliebig viele Mal hintereinander auftreten. So ist, z. B., die Wahrscheinlichkeit, dass im Roulette zehntausend millionen Mal hintereinander Rot fällt, nicht die Unmöglichkeit, sondern eine zwar äusserst geringe doch von Null veschiedene Wahrscheinlichkeit, und bei einer hinreichend grossen Anzahl von Reihen zu je 10^{10} Würfen darf das relativ so und so häufige Auftreten solcher Reihen, in welchen alle 10^{10} Male Rot fällt, mit grösster Sicherheit erwartet werden. "Auch die minimalste Wahrscheinlichkeit ist von der Unmöglichkeit noch fundamental unterschieden; und diese Kluft können wir nicht überbrücken, mögen wir die Zahlen auch noch so sehr anwachsen lassen."[23] Nur so viel ist wahr dass die Wahrscheinlichkeit dafür, dass ein sehr wenig wahrscheinliches Ereigniss *häufig* auftreten wird, eine sehr kleine Grösse weit höherer Ordnung ist, als

---

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

[21]Slutsky's footnote: Ibid., p. 230.

[22]Slutsky's footnote: Ibid., p. 230, 227.

[23]Slutsky's footnote: J. v. Kries, op. cit. p. 21.

die seines einmaligen Eintretens. Dass eine solche Auffassung keine Aussage über Häufigkeiten selbst enthält, dürfte, m. E., gerade von dem Standpunkt von A. Cournot aus nicht behauptet werden, denn seine Ansichten gehen gerade darauf aus, dass jede Behauptung über Wahrscheinlichkeiten von Häufigkeiten eine Behauptung über dieses letzteren selbst ist.[24] Um das einzusehen braucht man sich in keine physikalischn, bzw., ontologischen Spekulationen einzulassen, sondern nur den schlichten sinn der entsprechenden Sätze der Wahrscheinlichkeitstheorie sich zur Klarheit zu bringen. Dann sieht man, dass bei betreffenden Voraussetzungen (Bernoulli'schen, Poisson'schen, Markoff'schen) eine fast volle Gewisheit besteht, dass "die Ereignisse, deren Wahrscheinlichkeiten sehr klein sind, sich nicht oft widerholen werden". Steicht man hier das Wörschen "fast", so erhält man das Cournot'sche Lemma, deren Unterschied von dem ersten Sätze *nicht also in dem Inhalt, sondern lediglich in der Modalität der Behauptung* besteht: was dieser nur mit *fast* voller Gewissheit behauptet, das will das Lemma für eine absolut sichere Erkenntniss ausgeben. Und das ist bei unseren Voraussetzungen sicher falsch.

Nun zu einer anderen Interpretationsmöglichkeit. Obiges Lemma kann noch als ein idographischer Satz angesehen werden, als eine Behauptung über die tatsäschliche Struktur der Welt, bzw., ihres uns umgebenden Teiles.[25] Es würde dann besagen, dass obgleich es swischen allen möglichen Konstellationen der Weltelemente auch solche gibt, die uns allerseltsamste Ereignisse vorspielen müssten, so dass, z. B., all Zufallspiele wie von einer dämonischen Kraft verfälscht, warme Körper durch kalte erhitzt, Menschengeschicke durch Stern geleitet eerscheinen würden u. s. f., — doch unsere Welt ist kein von diesem Ausnahmewelten, sondern eine so zu sagen *ordinäre Welt.*[26] Man kann diese Voraussetzung für eine viellecicht wohlbegründete halten, doch so viel ist wahr, dass von den Tatsachen der Vergangenheit des chaotischen Geschehens auf

---

[24]Slutsky's footnote: "La probabilité mathematique devient alors la mesure de la *possibilite physique*. . . l'avantage de cellec-ci c'est d'indiquer nettement l'existance d'un rapport, . . . qui subsiste entre les choses même: rapport, que la nature maintient et que l'observation manifeste lorsque les épreuves se répètent assez". A. Cournot, *Essai sur les fondements de nos connaissances*, Nouvelle édition, Paris, 1912, p. 45 (sperrdruck des Verfassers). Ueber den für sein ganzes system charakteristischen, doch nichts weniger als bis zum Ende geklärten Begriffe "d'impossibilité physique" (oder "d'impossibilité de fait") siehe seine *Exposition de la théorie des chances et des probabilités* Paris, 1843, p. 79–80, 437–438. Vergleich J. v. Kries, *Ueber den Begriff der objektiven Möglichkeit*, "Vierteljahrsschrift f. wiss. Philos." 12 Jahrg, 1888.

[25]Slutsky's footnote: Vielleicht kann Al. A. Tschuprow's Standpunkt in diesem sinne verstanden werden: *Abhandlungen aus der Theorie der Statistik*, 2 Aufl., s. 231, (russisch).

[26]Slutsky's footnote: Vergl. Zilsel, *Versuch einer neuen Grundlegung der statischen Mechanik* "Monatshefte für Math. und Physik" Wien, 1921, Bd. XXXI, p. 153–154. Seine "verallgemeinerte Allagodenhypothese" ist der Hypotehse der Ordinarität der Welt äquivalent. Der Verfasser irrt sich aber, sofern er glaubt, dass seine Konstruktionen einen stochastischen Standpunkt enbehrlich machen.

Kolmogorov included this article in the bibliography of his 1933 *Grundbegriffe* [111].

2.55   Richard von Mises, 1883–1953

Von Mises introduced his mathematical foundation for probability in 1919 [190]. In his view, relative frequency was the true meaning of probability, and limiting frequency in an infinite sequence — he called it a collective (*Kollektiv*) — was its proper mathematical representation. He saw this use of infinity as an idealization, just as Euclid's idea of an infinitely precise point is an idealization, and he thought it similarly appropriate for mathematical reasoning. He thought statistical inference, when it was appropriate, should use Bayes's theorem rather than Bernoulli's theorem. When no frequencies are available to use as the prior probabilities in Bayes's theorem, statistical inference would not be appropriate.

When von Mises began his work on probability theory, there was no alternative consensus about its mathematical foundation. Kolmogorov's axiomatic treatment using measure theory was a dozen years in the future. Most introductions to mathematical probability still began, as Bernoulli had begun, with the "classical" notion of equally possible cases, which von Mises considered confusing and circular. He considered the interpretation of Bernoulli's theorem that equates high probability with practical certainty an unjustified attempt to connect equally possible cases with frequency. He made this point at length in 1928, in the first edition of his *Wahrscheinlichkeit, Statistik und Wahrheit* (*Probability, Statistics and Truth*) [191]. Here is a passage from p. 89:

It is even more groundless to declare, concerning the hypothesis that the relative frequency has a limit, that one sees as "certain" what can only qualify as "highly probable" according to Poisson's or Bernoulli's theorem (H. Weyl). As a recent oddity, we can also cite the attempt to concoct a new concept of "stochastic passage to the limit" from the existence of the limit and the existence of the

Bernoulli-Poisson theorem (E. Slutsky). All these confusions arise from starting with the classical definition of probability, which has nothing to do with the course of events, and afterwards using an way of speaking that refers to this course of events ("one can expect with certainty that ...").

Noch unbegründeter ist es, zu behaupten, mit der Annahme, daß die relative Häufigkeit einen Grenzwert besitzt, sehe man das als"sicher" an, was nach dem Poissonschen oder Bernoullischen Satz nur als "höchst wahrscheinlich" gelten kann (H. Weyl). Als eine Sonderbarkeit neueren Datums sei auch noch der Versuch angeführt, aus der Existenz des Grenzwertes und dem Bestehen des Bernoulli-Poissonschen Satzes einen neuen Begriff des "stochastischen Grenzüberganges" zusammenzubrauen (E. Slutsky). Alle diese Verirrungen entstehen daraus, daß man von der klassischen Wahrscheinlichkeitsdefinition ausgeht, die nichts mit dem Erscheinungsablauf zu tun hat, und sich nachträglich einer Ausdrucksweise bedient, die auf diesen Ablauf Bezug nimmt ("es ist mit Sicherheit zu erwarten, daß ....").

For the formulations by Weyl and Slutsky to which von Mises objected, see §2.57 and §2.54, respectively.

In 1931, von Mises published a textbook on probability and its applications to statistics and theoretical physics [192]. It included a section (§1.5) entitled "Das Verhältnis der Theorie zur Erfahrungswelt" ("The theory's relationship with the world of experience").[27]. Here he emphasized the analogy with geometry and mechanics. The theory intersects with reality twice; first one obtains initial data from reality (initial conditions in the case of mechanics, initial probabilities in the case of probability), then after calculation one obtains predictions about reality. He distinguished three applications of probability, games of chance, statistics, and statistical physics, which differ in how the initial probabilities are obtained. In games of chance great effort goes into making these simple, in statistics they are obtained from data, and in statistical physics they are hypothetical.

Von Mises counted himself as a positivist. In 1939, he published a book explaining his positivist philosophy [193]; an English translation appeared in 1951. In the following passage, from the English translation [195, p. 183], he enlarges on how probability is used in statistical physics, in a way that could be characterized as a version of Cournot's principle.

> In classical physical statistics one starts by making certain plausible assumptions, according to the methods of probability calculus,

---

[27]In 1933, Kolmogorov cited this section of von Mises's 1931 book as the model for his explanation of how his own axiomatic theory of probability related to the empirical world; see §2.70

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.

In later editions of his *Wahrscheinlichkeit, Statistik und Wahrheit*, von Mises toned down his criticism of Cournot's principle only a bit. Here is two passages from the final English edition, published in 1957 [196, p. 116]:

If the probability of an attribute within a given collective has a value near to unity, we may express this fact by saying that 'there is great certainty' or 'we are almost certain' that this event will occur on *one* specific trial. This way of expressing ourselves is not too reprehensible so long as we realize that it is only an abbreviation, and that its real meaning is that the event occurs almost always in an infinitely long sequence of observations.

. . .

Those who think that probability can be defined independently of the frequency of occurrence of an attribute in a sequence of experiments believe that the . . . proposition, whereby probability and frequency roughly coincide in a long run of observations, constitutes a 'bridge' between what actually happens and the concept of probability introduced by them. However, we know that this is a delusion. From the definition of probability as the ratio of favourable to equally likely cases, no logical reasoning will lead to the propositions discussed above — neither to the original Bernoulli-Poisson statement nor to Bayes's converse of it. All that we can logically deduce from this premise is propositions concerning such ratios. A gap remains: the manner in which it is to be crossed is arbitrary and logically not justifiable.

Following a tumultuous debate with Joseph L. Doob at Dartmouth in 1941, von Mises had conceded that there was "no contradiction or irreconcilable contrast" between Doob's theory (considered by Doob and others an elaboration of Kolmogorov's) and his own [197, p.]:

. . . both theories are essentially statistical or frequency theories, equally far from the classical conception based on "equally likely cases." In both theories, probabilities are, of course, measures of sets.

Yet as the quotation from 1957 shows, von Mises still, in the last years of his life, saw the "classical" conception based on equally likely cases, not Kolmogorov-Doob, as his competitor. On p. 99 of the 1957 book, he wrote that Kolmogorov's mathematical investigations "do not . . . constitute the foundations of probability

but rather the foundations of the mathematical theory of distributions, a theory which is also used in other branches of science." I have not found any discussion by von Mises of Kolmogorov's own use of a version of Cournot's principle.

We might summarize by saying that for von Mises, probability theory's job in the world was to predict frequencies in long sequences from calculations about limiting frequencies in infinite sequences. He did not explain how this was consistent with his understanding of statistical physics, and so far as I know he did not consider the similar problem in the case of stochastic processes in other fields, such as economics, where there is also only a single observation; see §2.77.

2.56   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 [180, 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. . . .

2.57   Hermann Weyl, 1885–1955

Weyl was a physicist who also wrote on the philosophy of science. In his *Philosophie der Mathematik and Naturwissenschaft*, which appeared in 1927, he discussed Bernoulli's law of large numbers. Here is an excerpt from an English translation [200]:

> In his [Bernoulli's] calculation the individual trials are treated as statistically independent events. This theorem belongs to pure mathematics. It acquires a relation to reality only by the fact that the occurrence of an event is considered practically certain if its probability deviates from absolute certainty by, say, less than one millionth . . .

2.58   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*[120]

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 ([120], pp. 21, 34; see also [121], 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 ([120], pp. 29–30).

 In his 1925 book [120], 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).[28]

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 [121, 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.*

> 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.*

2.59   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

---

[28]Lévy devotes Section 1 to the first principle and Section 2 to the second. In the preface (p. viii), he cites a 1922 article [99] in which Hadamard stated the two principles.

book[5, 4], but the book may have been less influential than an article Anderson contributed to a special issue of the Swiss philosophy journal *Dialectica*[6, 29, 122] in 1949, alongside articles by Borel and Lévy revisiting their versions of Cournot's principle.

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[6]), and Van Dantzig at the meeting in Paris (Van Dantzig 1951[181]). 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)[142]. The German mathematical statistician Hans Richter, also in Munich, emphasized Cournot's principle in his own contributions to *Dialectica* (Richter 1954; von Hirsch 1954)[157, 189] and in his probability textbook (Richter 1956)[158], 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.

2.60 Charlie Dunbar Broad, 1887–1971

Note Broad's review in 1913 of the reprinting of Cournot's 1851 book [32].

2.61 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 [76]. 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 [38, 39]. 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 the following "prefatory note" to an article he published in 1925 [77]; see [1].

> It has been pointed out to me that some of the statistical ideas employed in the following investigation have never received a strictly logical definition and analysis. The idea of a frequency curve, for example, evidently implies an infinite hypothetical population distributed in a definite manner; but equally evidently the idea of an infinite hypothetical population requires a more precise logical specification than is contained in that phrase. The same may be said of the intimately connected idea of random sampling. These ideas have grown up in the minds of practical statisticians and lie at the basis

especially of recent work; there can be no question of their pragmatic value. It was no part of my original intention to deal with the logical bases of these ideas, but some comments which Dr Burnside has kindly made have convinced me that it may be desirable to set out for criticism the manner in which I believe the logical foundations of these ideas may be established.

The idea of an infinite hypothetical population is, I believe, implicit in all statements involving mathematical probability. If, in a Mendelian experiment, we say that the probability is one half that a mouse born of a certain mating shall be white, we must conceive of our mouse as one of an infinite population of mice which might have been produced by that mating. The population must be infinite for in sampling from a finite population the fact of one mouse being white would affect the probability of others being white, and this is not the hypothesis which we wish to consider; moreover, the probability may not always be a rational number. Being infinite the population is clearly hypothetical, for not only must the actual number produced by any parents be finite, but we might wish to consider the possibility that the probability should depend on the age of the parents, or their nutritional conditions. We can, however, imagine an unlimited number of mice produced upon the conditions of our experiment, that is, by similar parents, of the same age, in the same environment. The proportion of white mice in this imaginary population appears to be the actual meaning to be assigned to our statement of probability. Briefly, the hypothetical population is the conceptual resultant of the conditions which we are studying. The probability, like other statistical parameters, is a numerical characteristic of that population.

We only need the conception of an infinite hypothetical population, in connection with random sampling. The ultimate logical elucidation of the one idea implies that of the other. Also, the word infinite is to be taken in its proper mathematical sense as denoting the limiting conditions approached by increasing a finite number indefinitely. I imagine that an exact meaning can be given to all the ideas required by some process such as the following.

Imagine a population of N individuals belonging to s classes, the number in class k being $p_k N$. This population can be arranged in order in $N!$ ways. Let it be so arranged and let us call the first n individuals in each arrangement a sample of n. Neglecting the order within the sample, these samples can be classified into the several possible types of sample according to the number of individuals of each class which appear. Let this be done, and denote the proportion of samples which belong to type j by q_j, the number of types being t. Consider the following proposition.

Given any series of proper fractions $P_1, P_2, \ldots, P_s$, such that $\sum P_k = 1$, and any series of positive numbers $\eta_1, \eta_2, \ldots, \eta_k$, however

small, it is possible to find a series of proper fractions $Q_1, Q_2, \ldots, Q_t$, and a series of positive numbers $\epsilon_1, \epsilon_2, \ldots, \epsilon_s$, and an integer N_0, such that, if

$$N > N_0$$

and

$$|p_k - P_k| < \epsilon_k \text{ for all values of } k,$$

then will

$$|q_j - Q_j| < \eta_j \text{ for all values of } j.$$

I imagine it possible to provide a rigorous proof of this proposition, but I do not propose to do so. If it be true, we may evidently speak without ambiguity or lack of precision of an infinite population characterised by the proper fractions, P, in relation to the random sampling distribution of samples of a finite size n.

It will be noticed that I provide no definition of a random sample, and it is not necessary to do so. What we have to deal with in all cases is a random sampling distribution of samples, and it is only as a typical member of such a distribution that a random sample is ever considered.

Note the word "typical" in the last sentence. The 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.

2.62   Harold Jeffreys, 1891–1989

In his *Theory of Probability*, Jeffreys uses Bayes's theorem to explain "how an inductive inference can approach certainty". Under certain assumptions, he writes [106, 3rd ed., p. 43]:

> ... repeated verifications of consequences of a fact will make it practically certain that the next consequence of it will be verified. This accounts for the confidence that we actually have in inductive inferences.

2.63   Thornton Fry, 1892–1991

Fry's 1928 textbook, *Probability and its Engineering Uses* [92], 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.

2.64   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 [49, 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 suggests that Cramér was a less careful philosopher than Kolmogorov, for its claim that Principle **B** is a particular case of Principle **A** 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**.

2.65   Jerzy Neyman, 1894–1981

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

[143], 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.

2.66   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[6]), and Van Dantzig at the meeting in Paris (Van Dantzig 1951[181]). 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)[142]. The German mathematical statistician Hans Richter, also in Munich, emphasized Cournot's principle in his

own contributions to *Dialectica* (Richter 1954; von Hirsch 1954)[157, 189] and in his probability textbook (Richter 1956)[158], 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.

2.67   Karl Popper, 1902–1994

Popper adopted a form of Cournot's principle in his *Logik der Forschung*, first published in 1935 [153]. 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, he 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*.

2.68   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 ([199], 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."

2.69   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 [177, 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. Mathemati-
> cally we now know that it is a branch of measure theory, which is
> linked with the real world through a few simple principles embody-
> ing 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 probabili-
> ties 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.

2.70   Andrei Kolmogorov, 1903–1987

Cournot's principle was emphasized by many of the Russian and French math-
ematicians from whom Kolmogorov learned about probability theory, including
Markov, Chuprov, Slutsky, Borel, Lévy, and Fréchet [171]. But in the Soviet
context, it was also mandatory to highlight the primacy of mass phenomena.
So perhaps it is not surprising that he mentions both frequency and Cournot's
principle in 1933 in his *Grundbegriffe der Wahrscheinlichkeitsrechnung* [111], his
celebrated monograph on axiomatic foundations for probability.

Kolmogorov entitled the two-page section where he discussed how probability
theory is used outside mathematics "Das Verhältnis zur Erfahrungswelt" (Rela-
tion to the World of Experience). This echoed Bohlmann's section title "Prin-
cipien nach denen die Theorie auf die Erfahrung angewendet wird" ("Principles
by which the theory is applied to experience") in his 1900 encyclopedia article,
and von Mises's section title "Das Verhältnis der Theorie zur Erfahrungswelt"
("The theory's relationship with the world of experience") in his 1931 textbook.
In a footnote, Kolmogorov referred to the section in von Mises as a model:

> In presenting the conditions required for applying the probability
> calculus to the world of real events, the author has largely followed
> Mr. von Mises's model (see particularly [1] p. 21—27: 'The theory's
> relationship with the world of experience"),

> In der Darstellung der notwendigen Voraussetzungen für die An-
> wendbarkeit der Wahrscheinlichkeitsrechnung auf die Welt der
> reellen Geschehnisse folgt der verfasser im hohem Maße den

Ausführungen von Herrn VON MISES (vgl. insbesondere [1] S. 21-27: "Das Verhältnis der Theorie zur Erfahrungswelt").

Here is how Kolmogorov then gives frequency and Cournot's principle their place:

> 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 $\mathfrak{S}$.

> Unter gewissen Bedingungen, auf die wir hier nicht näher eingehen wollen, kann man voraussetzen, daß einem Ereignis A, welches infolge der Bedingungen $\mathfrak{s}$ auftritt oder nicht, eine gewisse reelle Zahl $\mathsf{P}(A)$ zugeordnet ist, welche folgende Eigenschaften besitzt:
>
> A. Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen $\mathfrak{S}$ eine große Anzahl von n Malen wiederholt und dabei durch m die Anzahl der Fälle bezeichnet, bei denen das Ereignis A stattgefunden hat, das Verhältnis m/n sich von $\mathsf{P}(A)$ nur wenig unterscheidet.
>
> B. Ist $\mathsf{P}(A)$ sehr klein, so kann man praktisch sicher sein, daß bei einer einmaligen Realisation der Bedingungen $\mathfrak{S}$ das Ereignis A nicht stattfindet.

If you read what von Mises wrote in 1931 (I summarize it in §2.55), you may have difficulty seeing it as a model for the passage from Kolmogorov just quoted. Apparently Kolmogorov was thinking about the distinction von Mises drew between ordinary statistical work and statistical physics. Condition A tells us how to interpret probabilities as frequencies in ordinary statistical work without using von Mises's cumbersome concept of collectives, while Condition B, even though it is undeniably a statement of Cournot's principle, summarizes how von Mises thought about very small probabilities in statistical physics.

In a letter to Fréchet in 1939 [17], Kolmogorov wrote that he thought the only theory of probability that could reflect experience truthfully was an informal, not mathematically rigorous theory of finite but very large collectives with approximately stable frequencies. The practical value of his axiomatic theory could be derived from this informal theory. He later decided that a mathematically rigorous theory of finite collectives was possible after all, and this led him to his theory of algorithmic complexity [18].

As Cournot emphasized, many events do not have objective mathematical probabilities; we can give them only non-mathematical "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 all events have objective mathematical probabilities. So it is worth noting that Kolmogorov, von Mises, and most other mathematicians of their time who worked with the concept of objective probability agreed with Cournot that only some events have objective probabilities. Kolmogorov 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.

2.71   Carl Hempel, 1905–1997

In 1965, in his *Aspects of Scientific Explanation* [101, p. 387], Hempel quotes Harald Cramér's formulation, quoted here in §2.64, in which Cournot's principle is a special case of Kolmogorov's principle A.

2.72   Hans Freudenthal, 1905–1990

In an expository article on probability published in 1960 in *Synthese* [90, 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 [89, 91].

2.73   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[57] 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 [58, 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 [54].

2.74   William Feller, 1906–1970

If you repeatedly toss a fair coin, can you count on the number of heads and the number of tails becoming equal at some point? In his section on the random walk in the third edition of volume I of his textbook on probability (1968), Feller derived a formula for the probability f_{2n} for the event that the first equalization happens on the $2n$th toss. Then he made this comment [74, p. 78]:

> It follows ...that $f_2 + f_4 + \cdots = 1$. In the coin-tossing terminology this means that an ultimate equalization ...becomes practically certain if the game is prolonged sufficiently long. This was to be anticipated on intuitive grounds, except that the great number of trials necessary to achieve practical certainty comes as a surprise. For example, the probability that no equalization occurs in 100 tosses is about 0.08.

2.75   Joseph Doob, 1910–2004

Doob's most explicit statement of Cournot's principle comes in a historical essay he published in 1976 [68, 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 [67], 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.

2.76   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 sections, an introduction and a conclusion. Borel quickly arranged for the expanded version to be be published as a book [184]. On pp. 9–10 of the introductory section, 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 co-efficients 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 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 unverifiable 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 [137].

4. La methode axiomatique et les grandes probabilités. Dans les traités modernes, la notion de cas également probable n'est employée que pour familiariser le lecteur avec l'idée de probabilité. Après quelques exercices utilisant cette notion primitive, on passe en general à la théorie axiomatique, qui consiste à considérer des événements aléatoires, sans s'occuper des modes d'apparition, et à leur associer des coefficients compris entre 0 et 1, que l'on appelle probabilités. On donne des règles de composition de ces coefficients quand on compose les événements; ces règles etant calquées sur celles relatives aux probabilités définies par la méthode de Laplace ne sont pas introduites artificiellement. La théorie ainsi construite est logiquement correcte, mais il faut interpreter les coefficients ainsi introduits. On utilise pour cela la valeur subjective des grandes probabilités, déjà mise en évidence dans Laplace. De sorte que la base de la théorie axiomatique peut être considerée comme étant la suivante : *Étant donné des événements aléatoires, on peut leur associer des coefficients compris entre 0 et 1, tels que si on les compose d'après les règles posées comme axiomes, les événements ayant des probabilités très voisines de 1 soient pratiquement certains (par consèquent, ceux dont les probabilités sont très petites sont pratiquement impossibles).*

De sorte que nous pouvons dire, avec M. Fréchet : *La probabilité d'un événement, dans une catégorie d'épreuves déterminée, est une constante physique, dependant de l'événement et de la catégorie d'epreuves, dont on obtient une valeur empirique en procédant à un grand nombre d'épreuves independantes et en notant la fréquence de l'événement.*

Valeur empirique signifie valeur qui a peu de chance de s'écarter beaucoup de la valeur vraie. Nous voyons donc revenir constamment, dans l'interpretation, cette notion de "certitude pratique" in-

79

terprétant la prohabilité voisine de 1. La théorie axiomatique est done susceptible de vérifications indirectes.

Nous avons ainsi affaire, dans la théorie axiomatique, à deux sortes de probabilités : celles qui sont voisines de 0 ou de 1, qui ont une signification subjective, quasi-impossibilité ou quasi-certitude, et celles qui ne sont voisines ni de zero ni de 1, qui, prises isolément, n'ont aucune signification subjective. C'est justement ce manque de signification des probabilités "moyennes" qui est gènant dans la théorie axiomatique : une proposition telle que "la probabilité de pile est $\frac{1}{2}$" n'a pas de valeur, prise isolément, et est invérifiable directement. Si l'on répète l'experience, on deduit de propositions invérifiables une proposition vérifiable. Il semble y avoir là un défaut; on est tenté de ne jamais considérer que des suites assez étendues d'expériences, puisque chaque proposition n'est utilisable qu'en combinaison avec un grand nombre d'autres propositions : cela mène a la théorie statistique et à la négation non seulement de la *valeur subjective*, mais encore de l'*existence* de la probabilité d'un événemenlt isolé.

2.77   Trygve Haavelmo, 1911–1999

Haavelmo's article, "The probability approach to econometrics" [98], is often seen as the founding charter of modern econometrics [141]. 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.

2.78   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 [157, 189].

We find this passage in his 1956 probability textbook [158], which helped bring Kolmogorov's axioms to students in postwar Germany.

> Die praktische GewiBheit, die wir im Leben dauernd annehmen und auch annehmen müssen, wird bei dieser Auffassung zu einer Wahrscheinlichkeit, die eben nur so nahe bei 1 liegt, daB wir gewöhnlich darauf verzichten, überhaupt noch von dem Unterschied zu sprechen. Der Begriff der unkontrollierbaren Störung, der in der klassischen Physik eigentlich ein Fremdkörper ist, erhält so eine einleuchtende Beschreibung: Die Störung entspricht dem Defekt $\epsilon > 0$ der Wahrscheinlichkeit eines praktisch sicheren Ergebnisses gegenüber dem Idealwert 1, der klassisch angenommen werden müßte. Bei astronomischen Untersuchungen ist dieser Defekt so klein, daB er gar keine Rolle mehr spielt; bei unseren Beispielen aus dem täglichen Leben hat ϵ aber einen wesentlich höheren Wert, wie die Existenz von Unglücksfallen zeigt. Und doch sind wir gezwungen, auch ein solches ϵ noch praktisch zu vernachlässigen. Wir kommen so zu der folgenden Formulierung, die COURNOTsches Prinzip genannt wird.
>
> Zu vorgegebenem S_0 mit den möglichen Folgesituationen S_{nu}. sei ein $\epsilon > o$ gewählt. Hat ein S_ν, etwa S_1, eine Wahrscheinlichkeit von mindestens $1 - \epsilon$, so sollen wir so handeln, als ob das Eintreten von S_1 gewiss wäre. Das Eintreten von S_1 heißt dann praktisch sicher.

As a result of Richter's textbook, the name *Cournotsche Prinzip* became fairly widely known among probabilists in Germany.

2.79 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 [60, 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.

2.80 Yuri Prokhorov, 1929–2013, and Boris Sevast'yanov, 1923—2013

Yuri Vasilevich Prokhorov and Boris Aleksandrovich Sevast'yanov 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 [154], Prokhorov and Sevast'yanov echoed Jean Ville's statement that only probabilities close to 0 or 1 have direct meaning.

2.81 David R. Cox, 1924–2022, and David V. Hinkley, 1944–2019

The term *repeated sampling principle* was coined by Cox and Hinkley in their 1974 textbook [48, 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.

2.82 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'.

2.83 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* [113]. 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. . .

2.84 Hugh Everett III, 1930–1982

Quotation from Everett's 1957 thesis: Pages 70–71 of [63]

In the language of subjective experience, the observer which is described by a typical element, $\psi'_{ij...k}$, 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.

2.85 Terrence Fine, 1939–2021

Discuss the relevance of Fine's suggestion, in 1976 [75], 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 [172, Ch. 12] and related work by other authors.

2.86 Per Martin-Löf, born 1942

In Martin-Löf's pathbreaking article on the definition of random sequences, published in 1966, we find this passage [136, 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 said that he learned Cournot's principle from Borel rather than from Kolmogorov. See also [18, 135].

2.87 Donald Gillies, born 1944

In a 1973 book entitled *An Objective Theory of Probability* [96], 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)."

2.88 Persi Diaconis, born 1945, and Brian Skyrms, born 1938

In their 2018 book, *Ten Great Ideas About Chance* [64], Diaconis and Skyrms dismiss Jacob Bernoulli's theorem as a basis for finding a probability from repeated trials, calling the notion that it can provide such a basis "Bernoulli's swindle". They then say this about Cournot's principle:

> . . . it is a remarkably persistent fallacy, easy to swallow in the absence of rigorous thinking. We find it in the French mathematician and philosopher Cournot (1843), who holds that small-probability events should be taken to be physically impossible. He also held that this principle (Fréchet named it Cournot's principle) is the one that connects probabilistic theories to the real world. It is taken, as in Bernoulli, as showing that we should identify probability with relative frequency in a large number of (independent? identically distributed?) trials.
>
> This mantra was repeated in the twentieth century by very distinguished probability theorists, including Émile Borel, Paul Lévy, Andrey Markov, and Andrey Kolmogorov. We cannot help but wonder whether this was to some extent a strategy for brushing off philosophical interpretational problems, rather than a serious attempt to confront them.
>
> Cournot's principle, taken seriously, is absurd. Throw a dart at a target. The chance that it hits any specific point is very small. Then are we supposed to conclude that for any point, it is physically impossible that the dart hits that point? Later statements tend to try to get around this by modifying the principle as saying that an event of very small probability, singled out in advance, is physically impossible. So you have to pick out a point in advance. Why should picking it out in advance make it physically impossible?

2.89 Colin Howson, 1945–2020, and Peter Urbach

In the third edition of their book advocating "the Bayesian approach to scientific reasoning", which appeared in 2006 [102, p. 49], the British philosophers Colin Howson and Peter Urbach say this about Cournot's principle:

> . . . without some qualification Cournot's Principle is false, for events with almost infinitesimal probabilities occur all the time without casting any suspicion upon the theories which assign them those

probabilities: the exact configuration of air molecules in a room at any given time has a microscopic probability; so does any long sequence of outcomes of tosses of a fair coin (even with so small a number as twenty the probability of each possible sequence is already around one in a million). Can the Principle be qualified in such a way as to make it tenable? This is just what the well-known and widely-used theories of bivalent statistical tests of R. A. Fisher and Neyman and Pearson (all believers in a long-run frequency account of statistical probabilities), attempt to do in their different ways. Unfortunately ... these attempts also fail.

2.90   A. Philip Dawid, born 1946

Dawid has always advocated a subjectivist interpretation of probability and often calls himself a Bayesian. He is unusual and among Bayesians, however, by his forthright advocacy of Cournot's principle. In the Rutgers Foundations of Probability Seminar in February 2022, for example, he explained that the meaning of a probability model is simply that it asserts events to which it gives high probability.

In 1985, Dawid stated Cournot's principle and mentioned its relevance to the frequency interpretation of probability as follows [55, p. 116]:

> Now consider the following *rule of interpretation* of probability statements (Cournot's Principle, or Borel's 'Single Law of Probability' [1943]).
>
> > *Rule of Interpretation.* If an event A is assigned probability 1 by a distribution P, then P asserts that A is 'morally certain'.
>
> such a rule would probably be acceptable to most frequentist statisticians, even though it is of a subjectivist nature. As pointed out by de Finetti [1936], only with such a rule can we avoid an infinite regress in a frequency theory of probability. We therefore take the rule as basic in all that follows.

The citations are to [28] (Borel) and [56] (de Finetti). Dawid developed his viewpoint further in [54].

Acknowledgments

References

[1] John Aldrich. Burnside's engagement with the "modern theory of statistics". *Archive for History of Exact Sciences*, 63:51–79, 2009. 67

[2] John Aldrich. W. E. Johnson and Cambridge thought on probability. *International Journal of Approximate Reasoning*, 141:146–158, 2022. 38

[3] André-Marie Ampère. *Considérations sur la théorie mathématique du jeu.* Perisse, Lyon, 1802. 27

[4] Oskar Anderson. *Probleme der statistischen Methodenlehre in den Sozialwissenschaften.* Physica, Würzburg, 1954. 67

[5] Oskar Nikolaevich Anderson. *Einführung in die mathematische Statistik.* Springer, Vienna, 1935. 67

[6] Oskar Nikolaevich Anderson. Die Begründung des Gesetzes der grossen Zahlen und die Umkehrung des Theorems von Bernoulli. *Dialectica*, 3(9/10):65–77, 1949. 67, 71

[7] John Arbuthnot. An argument for divine providence. *Philosophical Transactions of the Royal Society of London*, 27(328):186–190, 1710. 20

[8] Michel Armatte. Discussion de l'article de D. Denis [61]. *Journal de la Société Française de Statistique*, 145(4):27–36, 2004. 36

[9] Michel Armatte. Contribution à l'histoire des tests laplaciens. *Mathematics and social sciences*, 44(176):117–133, 2006. 36

[10] Antoine Arnauld and Pierre Nicole. *La logique, ou l'art de penser.* Savreux, Paris, 1662. This book, also known as the *Port Royal Logic*, had many later editions in many languages. In 1981, Pierre Clair and François Girbal published a critical edition based on the 1683 French edition. 17

[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. 54, 95, 99, 102

[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]. 83

[13] John S. Bell. *Speakable and Unspeakable in Quantum Mechanics.* Cambridge, 1987. 83, 89

[14] Jacob Bernoulli. *Ars Conjectandi.* Thurnisius, Basel, 1713. 1, 19, 90

[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. 1, 2, 19

[16] Felix Bernstein. Über eine Anwendung der Mengenlehre auf ein aus der Theorie der säkularen Störungen herrührendes Problem. *Mathematische Annalen*, 71:417–439, 1912. 54

[17] Laurent Bienvenu, Glenn Shafer, and Alexander Shen. Andrei kolmogorov and leonid levin on randomness. In Laurent Mazliak and Glenn Shafer, editors, *The Splendors and Miseries of Martingalse: Their History from the Casino to Mathematics*, pages 405–414. Birkhäuser, 2022. 75

[18] Laurent Bienvenu, Glenn Shafer, and Alexander Shen. Martingales in the study of randomness. In Laurent Mazliak and Glenn Shafer, editors, *The Splendors and Miseries of Martingalse: Their History from the Casino to Mathematics*, pages 225–263. Birkhäuser, 2022. 75, 86

[19] Georg Bohlmann. Lebensversicherungs-Mathematik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, pages 852–917. Teubner, Leipzig, 1901. 49

[20] Ludwig Boltzmann. *Vorlesungen über Gastheorie.* Barth, Leipzig, 1898. 39

[21] Émile Borel. La valeur pratique du calcul des probabilités. *Revue du mois*, 1:424–437, 1906. Reprinted in [31], Volume 2, pp. 991–1004. 50, 52

[22] 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. 50, 52

[23] Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–270, 1909. Reprinted in [31], Volume 2, pp. 1055–1079. 43

[24] 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. 51, 52

[25] 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 [26]. 52

[26] Émile Borel. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, Paris, 1939. 52, 90

[27] Émile Borel. *Le jeu, la chance et les théories scientifiques modernes*. Gallimard, Paris, 1941. 52

[28] 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. 52, 88

[29] Émile Borel. Probabilité et certitude. *Dialectica*, 3(9/10):24–27, 1949. 67

[30] 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. 52, 54

[31] Émile Borel. *Œuvres de Émile Borel*. Éditions du Centre National de la Recherche Scientifique, Paris, 1972. Four volumes. 90

[32] C.D̃. Broad. Review of *Essai sur les Fondements de nos Connaissances et sur les Caractères de la Critique Philosophique*, [45] by a. cournot. *Mind*, 22(87):309–402, 1913. 67

[33] Benard Bru. Remarques sur l'article de D. Denis [61]. *Journal de la Société Française de Statistique*, 145(4):37–38, 2004. 36

[34] Bernard Bru. Souvenirs de Bologne. *Journal de la Société Française de Statistique*, 144(1–2):134–226, 2003. 47

[35] Bernard Bru. Personal communication, 2003b. 58

[36] George-Louis Buffon. Essai d'arithmétique morale. In *Supplément à l'Histoire naturelle*, volume 4, pages 46–148. Imprimerie Royale, Paris, 1777. 20

[37] George-Louis Buffon. Letter to Laplace, dated april 21, 1774. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 188:1019, 1879. 20

[38] William Burnside. On the idea of frequency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22(5):726–727, 1925. 67

[39] 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. 67

[40] Guido Castelnuovo. *Calcolo delle probabilitá*. Albrighi e Segati, Milan, Rome, and Naples, 1919. Second edition in two volumes, 1926 and 1928. Third edition 1948. 46, 47

[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. 53

[42] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des decisions rendues à la pluralité des vois.* LÍmprimérie Royale, 1785. 24

[43] 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. 23

[44] 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 [47]. 31

[45] Antoine Augustin Cournot. *Essai sur les fondements de nos connaissances et sur les caractères de la critique philosophique.* Hachette, Paris, 1851. Reprinted in 1975 as Volume II (J. C. Pariente, editor) of [47]. 91

[46] 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 [47]. 31

[47] Antoine Augustin Cournot. *Œuvres complètes.* Vrin, Paris, 1973–2010. The volumes are numbered I through XI, but VI and XI are double volumes. 30, 36, 92

[48] David R. Cox and David V. Hinkley. *Theoretical Statistics.* Chapman and Hall, London, 1974. Second edition 1979. 83

[49] Harald Cramér. *Mathematical Methods in Statistics.* Princeton University Press, Princeton, NJ, 1946. 70

[50] Emanuel Czuber. Wahrscheinlichkeitsrechnung. In *Encyklopädie der mathematischen Wissenschaften, Band I, Teil 2*, pages 733–767. Teubner, Leipzig, 1900. 49

[51] 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 third edition of the second volume appeared in 1914. 41

[52] Jean d'Alembert. Réflexions sur le calcul des probabilités. In *Opuscule mathématiques*, volume 2, pages 1–25. David, Paris, 1761. 23

[53] Lorraine Daston. How probabilities came to be objective and subjective. *Historia Mathematica*, 21:330–344, 1994. 31

[54] A. P. Dawid. Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science*, 19:44–57, 2004. 77, 88

[55] A. Philip Dawid. Probability, symmetry and frequency. *The British Journal for the Philosophy of Science*, 36(2):107–128, 1985. 88

[56] Bruno de Finetti. Statistica e probabilita nella concezione di R. de Mises. *Supplemento statistico ai nuovi problemi di politica, storia ed economia*, 2:9–19, 1936. Reprinted on pp. 93–104, with an English translation on pp. 353–364, of [59]. 88

[57] 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. 77

[58] 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. 58, 77

[59] Bruno de Finetti. *Probabilità e Induzione; Probability and Induction*. Biblioteca de STATISTICA, CLUEB (Cooperative Libraria Universitaria Editrice Bologna), Bologna, 1993. A collection of Italian articles, all with English translations by Mara Khale and Antonella Ansani. The collection is introduced by review essays by Richard Jeffrey and Dario Fürst. 93

[60] Morris H. DeGroot. A conversation with Charles Stein. *Statistical Science*, 1(4):454–462, 1986. 82

[61] 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. 89, 91

[62] 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. 15

[63] Bryce S. DeWitt and Neill Graham, editors. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton University Press, Princeton, NJ, 1973. 84

[64] Persi Diaconis and Brian Skyrms. *Ten Great Ideas About Chance*. Princeton, 2018. 87

[65] Denis Diderot and Jean le Rond d'Alembert, editors. *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers, etc.* University of Chicago, 1751–1766. ARTFL Encyclopédie Project (Spring 2021 Edition), Robert Morrissey and Glenn Roe (eds), `http://encyclopedie.uchicago.edu/`. 22

[66] 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, 1851. 36

[67] 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 [194] and comments by Doob and von Mises on each other's articles [197]. 78, 103

[68] 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. 77

[69] Antonin Durand and Laurent Mazliak. Revisiting the sources of Borel's interest in probability: Continued fractions, social involvement, Volterra's prolusione. *Centaurus*, 53(4):306–332, 2011. 50

[70] Francis Edgeworth. Methods of statistics. *Journal of the Statistical Society of London*, Jubilee Volume:181–217, 1885. 41

[71] Francis Edgeworth. Probability. In *Encyclopædia Britannica*, volume 22. Cooper, 11th edition, 1911. 41

[72] Robert Leslie Ellis. On the foundations of the theory of probabilities. *Transactions of the Cambridge Philosophical Society*, 8(1):1–6, 1844. 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. 37

[73] Fernand Faure. Les idées de Cournot sur la statistique. *Revue de Métaphysique et de Morale*, 13(3):395–411, 1905. 36

[74] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, New York, third edition, 1968. Previous editions of this volume appeared in 1950 and 1957. 77

[75] 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. 86

[76] Ronald A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London (A)*, 222:309–368, 1922. 67

[77] Ronald A. Fisher. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22(5):700–725, 1925. 67

[78] Richard Foley. The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly*, 29(2):111–121, 1992. 19

[79] Robert M. Fortet. Opinions modernes sur les fondements du calcul de probabilités. In Le Lionnais [118], 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. 52

[80] 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. 26

[81] Joseph Fourier. Second mémoire sur les résultats moyens et sur les erreurs des mesures. In Joseph Fourier, editor, *Recherches statistiques sur la ville de Paris et le département de la Seine*, pages ix–xlviii. Imprimerie royale, Paris, 1829. 29

[82] James Franklin. *The Science of Conjecture: Evidence and Probability before Pascal.* Johns Hopkins, 2nd edition, 2015. 11, 13

[83] 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.* 47

[84] 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.*

[85] 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. 47

[86] Maurice Fréchet. Rapport général sur les travaux du Colloque de Calcul des Probabilités. In Bayer [11], pages 3–21. 54

[87] Maurice Fréchet. *Les mathématiques et le concret*. Presses Universitaires de France, Paris, 1955. 58

[88] Maurice Fréchet and Maurice Halbwachs. *Le calcul des probabilités à la portée de tous*. Dunod, Paris, 1924. 47

[89] Hans Freudenthal. Ist die Mathematische Statistik Paradox? *Dialectica*, 12(1):7–32, 1958. 76

[90] Hans Freudenthal. Models in applied probability. *Synthese*, 12(2/3):202–212, 1960. 76

[91] Hans Freudenthal. Probabilités objectives ou subjectives? *Logique et Analyse*, 8(32):265–276, 1965. 76

[92] Thornton C. Fry. *Probability and its Engineering Uses*. Van Nostrand, New York, 1928. 69

[93] 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. 30

[94] Jules Gavarret. *Principes généraux de statistique médicale, ou développement des règles qui doivent présider à son emploi*. Bechet, Paris, 1840. 36

[95] Jean Gerson. De consolatione theologiae, 1418. Pp. 185–245 of volume 9 of Gerson's *Œuvres complètes*, Paris, 1973. 13

[96] Donald A. Gillies. *An Objective Theory of Probability*. Methuen, London, 1973. 86

[97] Thomas Granger. *Syntagma logicum: or the divine logike*. Jones, London, 1620. 14

[98] Trygve Haavelmo. The probability approach to econometrics. *Econometrica*, 12(Supplement):1–115, 1944. 80

[99] 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. 47, 66

[100] Anders Hald. *A History of Mathematical Statistics from 1750 to 1930*. Wiley, New York, 1998. 29

[101] Carl G. Hempel. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science.* Free Press, New York, 1965. 76

[102] Colin Howson and Peter Urbach. *Scientific Reasoning: The Bayesian Approach.* Open Court, third edition, 2006. 87

[103] David Hume. *A Treatise of Human Nature.* John Noon, 1739–1740. 21

[104] David Hume. *An Enquiry Concerning Human Understanding.* Millar, 1748. 21

[105] Christiaan Huygens. *Traité de la Lumière.* Pieter van der Aa, Leiden, 1690. 17

[106] Harold Jeffreys. *Theory of Probability.* Oxford University Press, Oxford, 1939. Second edition 1948, third 1961. 69

[107] Norman L. Johnson and Samuel Kotz. *Leading Personalities in Statistical Sciences.* Wiley, New York, 1997. 39, 101

[108] Andreas Kamlah. Probability as a quasi-theoretical concept; J. V. Kries' sophisticated account after a century. *Erkenntnis*, 19:239–251, 1983. 44

[109] John Maynard Keynes. *A Treatise on Probability.* Macmillan, London, 1921. 44

[110] Sven K. Knebel. *Wille, Würfel und Wahrscheinlichkeit: Das System der moralischen Notwendigkeit in der Jesuitenscholastik 1550–1700.* Meiner, Hamburg, 2000. 13, 14, 15

[111] 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. 62, 74

[112] Ulrich Krengel. On the contributions of georg bohlmann to probability theory. *Electronic Journal for History of Probability and Statistics*, 77(1), 2011. 49

[113] Henry E. Kyburg, Jr. *Science & Reason.* Oxford, 1990. 84

[114] Pierre Simon Laplace. *Essai philosophique sur les probabilités.* Courcier, Paris, first edition, 1814. The fifth and definitive edition appeared in 1825 and was reprinted in 1986 (Christian Bourgois, Paris) with a commentary by Bernard Bru. Multiple English translations have appeared. 25, 97

[115] Pierre Simon Laplace. *Philosophical Essay on Probabilities.* Springer, New York, 1995. English translation of [114]; translated by Andrew I. Dale. 25

[116] Hermann Laurent. Détermination des pleins qu'un assureur peut garder sur les risques qu'il garantit. *Journal des actuaires français*, 2:79–90, 161–165, 1873. 39, 49

[117] Hermann Laurent. *Statistique mathématique.* Octave Doin, 1908. 39

[118] 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. 52, 95

[119] Eric L. Lehmann. The Bertand-Pearson debate and the origins of the Neyman-Pearson theory. In J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy, and B. L. S. Prakasa Rao, editors, *Statistics and Probability: A Raghu Raj Bahadur Festschrift*, pages 371–380. Wiley Eastern, 1993. Reprinted on pages 965–974 of *Selected Works of E. L. Lehman*, edited by J. Rojo, Springer, 2012. 52

[120] Paul Lévy. *Calcul de probabilités.* Gauthier-Villars, Paris, 1925. 65, 66

[121] Paul Lévy. *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris, 1937. Second edition: 1954. 66

[122] Paul Lévy. Les fondements du calcul des probabilités. *Dialectica*, 3(9/10):55–64, 1949. 67

[123] Wilhelm Lexis. *Einleitung in die Theorie der Bevölkerungsstatistik.* Trübner, Strassburg, 1875. 39

[124] John Locke. *An Essay Concerning Human Understanding.* Basset, London, 1689. 18

[125] Jeff Loveland. Buffon, the certainty of sunrise, and the probabilistic reductio ad absurdum. *Archive for History of Exact Sciences*, 55:465–477, 2001. 20

[126] Juan de Lugo. *De Virtute Fidei Divinae.* Borde and Arnaud, 1646. 14

[127] 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. 40

[128] Andrei Andreevich Markov. *Wahrscheinlichkeitsrechnung.* Teubner, 1912. Translation of second Russian edition. 46

[129] Thierry Martin. *Probabilités et critique philosophique selon Cournot.* Vrin, Paris, 1996. 4, 36, 88

[130] Thierry Martin. *Bibliographie cournotienne.* Annales littéraires de l'Université Franche-Comté, Besançon, 1998. 36, 88

[131] Thierry Martin. Probabilité et certitude. In Thierry Martin, editor, *Probabilités subjectives et rationalité de l'action*, pages 119–134. CNRS Éditions, Paris, 2003. 88

[132] Thierry Martin, editor. *Actualité de Cournot*. Vrin, Paris, 2005. 36, 88

[133] Thierry Martin. *Nouvelle bibliographie cournotienne*. Presses universitaires de Franche-Comté, second edition, 2005. `http://logiquesagir.univ-fcomte.fr/nouvelle_bibliographie_cournotienne/`. 36, 88

[134] Thierry Martin. Les premiers écrits probabilistes de Cournot (1825-1828). In Jose Maria Arribas et al., editor, *Historia de la probabilidad y la estadística VI*, pages 173–186. AHEPE-UNED, Madrid, 2012. 30

[135] 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. 86

[136] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966. 86

[137] Tim Maudlin. The grammar of typicality. In Valia Allori, editor, *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*. World Scientific, 2020. 79

[138] Laurent Mazliak. Belgium and probability in the nineteenth century: The case of paul mansion. *Science in Context*, 34(3):313–340, 2021. 41

[139] Alexius Meinong. *Über Möglichkeit und Wahrscheinlichkeit: Beiträge zur Gegenstandstheorie und Erkenntnistheorie*. Barth, Leipzig, 1915. 44

[140] Chris Meyns. Leibniz and probability in the moral domain. In Lloyd Strickland, Erik Vynckier, and Julia Weckend, editors, *Tercentenary Essays on the Philosophy and Science of Leibniz*, pages 229–253. Springer, 2017. 19

[141] Mary Morgan. *The History of Econometric Ideas*. Cambridge University Press, Cambridge, 1990. 80

[142] Jerzy Neyman. Foundation of the general theory of statistical estimation. In Bayer [11], pages 83–95. 67, 71

[143] Jerzy Neyman. Indeterminism in science and new demands on statisticians. *Journal of the American Statistical Association*, 55:625–639, 1960. 71

[144] Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society (A)*, 36:289–337, 1933. 51

[145] Kh. O. Ondar, editor. *O teorii veroyatnosteĭ i matematicheskoĭ statistike (perepiska A. A. Markova i A. A. CHuprova)*. Nauk, Moscow, 1977. See [146] for English translation. 100

[146] 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 [145] by Charles M. and Margaret Stein. Additional letters between Markov are provided in translation by Sheynin in [174], Chapter 8. 46, 99

[147] 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. 58

[148] Karl Pearson. *The Grammar of Science.* Scott, London, 1892. A second edition appeared in 1900, a third in 1911. 30

[149] Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890. 45

[150] Henri Poincaré. *La Science et l'Hypothèse.* Flammarion, 1902. 45

[151] Siméon-Denis Poisson. Mémoire sur la proportion des naissances des filles et des garçons. *Mémoires de l'Académie royale des sciences*, IX:239–308, 1830. 29

[152] Siméon-Denis Poisson. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités.* Bachelier, Paris, 1837. 29

[153] Karl R. Popper. *Logik der Forschung.* Springer, Vienna, 1934. An English translation, *The Logic of Scientific Discovery*, was published by Hutchinson, London, in 1959. 72

[154] 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. 83

[155] Hans Reichenbach. *Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit.* Barth, Leipzig, 1916. 44

[156] Constance Reid. *Neyman — From Life.* Springer, 1982. 51

[157] Hans Richter. Zur Begründung der Wahrscheinlichkeitsrechnung. *Dialectica*, 8:48–77, 1954. 58, 67, 72, 81

[158] Hans Richter. *Wahrscheinlichkeitstheorie.* Springer, Berlin, 1956. 58, 67, 72, 82

[159] Bertrand Saint-Sernin. *Cournot: Le réalisme.* Vrin, Paris, 1998. 36

[160] John of Salisbury. *Ioannis Saresberiensis episcopi carnotensis Metalogicon.* Oxford, 1929. John completed the original manuscript in 1159. This is the definitive modern edition, edited by Clement C. J. Webb. 11, 101

[161] John of Salisbury. *The Metalogicon of John of Salisbury: A Twelfth-Century Defense of the Verbal and Logical Arts of the Trivium*. University of California, 1955. Translation of [160], with an introduction & notes by Daniel D. McGarry. 11

[162] Rudolf Schüssler. Jean Gerson, moral certainty and the renaissance of ancient scepticism. *Renaissance Studies*, 23(4):445–462, 2009. 13

[163] Eugene Seneta. Boltzmann, Ludwig Edward. In *Leading Personalities in Statistical Sciences* [107], pages 353–354. 39

[164] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976. 3

[165] Glenn Shafer. Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19:309–370, 1978. 2

[166] Glenn Shafer. Moral certainty. In Samuel Kotz and Normal L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 5, pages 623–624. Wiley, 1985. 2

[167] Glenn Shafer. Savage revisited (with discussion). *Statistical Science*, 1:463–501, 1986. 5

[168] Glenn Shafer. From Cournot's principle to market efficiency. In Jean-Philippe Touffut, editor, *Augustin Cournot: Modelling Economics*, pages 55–95. Elgar, 2008. 4

[169] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001. 3

[170] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006. 4, 101

[171] 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*" [170]. 74

[172] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*. Wiley, New York, 2019. 3, 7, 86

[173] 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. 53

[174] 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. 100

[175] Evgeny Slutsky. Über stochastische Asymptoten und Grenzwerte. *Metron*, 5:3–89, 1925. 58

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

[177] Marshall Stone. Mathematics and the future of science. *Bulletin of the American Mathematical Society*, 63:61–76, 1957. 74

[178] Jean-Philippe Touffut, editor. *Augustin Cournot: Modelling Economics*. Edward Elgar, Cheltenham, 2007. 36

[179] Stefania Tutino. *Uncertainty in Post-Reformation Catholicism: A History of Probabilism*. Oxford, 2017. 14

[180] J. V. Uspensky. *Introduction to Mathematical Probability*. McGraw-Hill, New York, 1937. 65

[181] David Van Dantzig. Sur l'analyse logique des relations entre le calcul des probabilités et ses applications. In Bayer [11], pages 49–66. 67, 71

[182] Andreas de Vega. *Tridentini Decreti de Iustificatione expositio et defensio libris XV distincta, totam doctrinam iustificationis complectentibus*. Peter Canisius, Cologne, 1572. 13

[183] 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. 38

[184] Jean Ville. *Étude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939. This differs from Ville's dissertation, which was defended in March 1939, only in that a one-page introduction was replaced by a 17-page introductory chapter. 78

[185] Ladislaus von Bortkiewicz. Kritische betrachtungen zur theoretischen statistik. *Jahrbücher für Nationalökonomie und Statistik*, pages 641–680, 1894. 47

[186] Ladislaus von Bortkiewicz. Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, pages 821–851. Teubner, Leipzig, 1901. 48

[187] Ladislaus von Bortkiewicz. Realismus und Formalismus in der mathematischen Statistik. *Allgemeines Statistishes Archiv. (Deutschen Statistischen Gesellschaft)*, 9:225–256, 1915. 48

[188] Ladislaus von Bortkiewicz. *Die Iterationen. Ein Beitrag zur Wahrscheinlichkeitstheorie*. Springer, Berlin, 1917. 48

[189] Guy von Hirsch. Sur un aspect paradoxal de la théorie des probabilités. *Dialetica*, 8:125–144, 1954. 58, 67, 72, 81

[190] Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919. 62

[191] Richard von Mises. *Wahrscheinlichkeitsrechnung, Statistik und Wahrheit.* Verlag von Julius Springer, Vienna, 1928. The second edition appeared in 1936 and the third in 1951. A posthumous fourth edition, edited by his wife Hilda Geiringer, appeared in 1972. English editions, under the title *Probability, Statistics and Truth*, appeared in 1939 and 1957. 62, 103

[192] Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.* F. Deuticke, Leipzig and Vienna, 1931. 63

[193] Richard von Mises. *Kleines Lehrbuch des Positivismus. Einführung in die empiristische Wissenschaftsauffassung.* W. P. van Stockum & zoon, The Hague, 1939. 63, 103

[194] 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 [67] and comments by von Mises and Doob on each other's articles [197]. 94, 103

[195] Richard von Mises. *Positivism: A study in Human Understanding.* Harvard, Cambridge, Massachusetts, 1951. Translation of [193]. 63

[196] Richard von Mises. *Probability, Statistics and Truth.* George Allen & Unwin, 1957. Translation of the third edition of [191]. 64

[197] Richard von Mises and Joseph L. Doob. Discussion of papers on probability theory. *Annals of Mathematical Statistics*, 12:215–217, 1941. Discussion of [194] and [67]. 64, 94, 103

[198] Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective.* Cambridge University Press, Cambridge, 1994. 39, 45

[199] 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. 72

[200] Hermann Weyl. *Philosophy of Mathematics and Natural Science.* Princeton, 2009. Translation of the German original, *Philosophie der Mathematik und Naturwissenschaft*, 1927. 65

[201] George Udny Yule. *An Introduction to the Theory of Statistics.* Griffin, London, first edition, 1911. 52

[202] 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. This article is part of a special section on *Kries and Objective Probability.* 44

Index