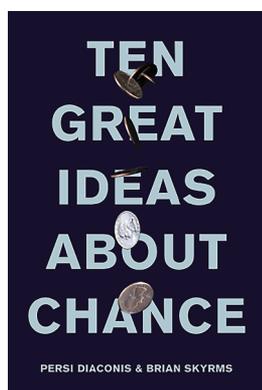




Ten Great Ideas about Chance

A Review by Mark Huber

Princeton University Press, 2018
Hardcover, 272 pages, ISBN: 9780691174167



Ten Great Ideas About Chance
By Persi Diaconis and Brian Skyrms

Most people are familiar with the basic rules and formulas of probability. For instance, the chance that event A occurs plus the chance that A does not occur must add to 1. But the question of why these rules exist and what exactly probabilities are, well, that is a question often left unaddressed in probability courses ranging from the

elementary to graduate level.

Persi Diaconis and Brian Skyrms have taken up these questions in *Ten Great Ideas About Chance*, a whirlwind tour through the history and philosophy of probability. Diaconis is the Mary V. Sunseri Professor of Statistics and Mathematics at Stanford University, and has long been known for his seminal work on randomization techniques such as card shuffling. Skyrms is a professor of philosophy at Stanford and Distinguished Professor of Logic and Philosophy of Science and Economics at the University of California at Irvine. His recent work has been in evolutionary game theory. The book they created is based upon a course they have taught at Stanford for the past ten years, and is aimed at a reader who has already gone through a typical undergraduate course in elementary probability.

Mark Huber is the Fletcher Jones Foundation Professor of Mathematics and Statistics at Claremont McKenna College. His email address is mhuber@cmc.edu.

Communicated by Notices Book Review Editor Stephan Ramon Garcia.

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: <http://dx.doi.org/10.1090/noti1888>

Having had Diaconis as my professor and postdoctoral advisor some two decades ago, I found the cadences and style of much of the text familiar. Throughout, the book is written in an engaging and readable way. History and philosophy are woven together throughout the chapters, which, as the title implies, are organized thematically rather than chronologically.

The story begins with the first great idea: Chance can be measured. The word *probability* itself derives from the Latin *probabilis*, used by Cicero to denote that "...which for the most part usually comes to pass" (*De inventione*, I.29.46, [2]). Even today, modern courtrooms in the United States shy away from assigning numerical values to probabilities, preferring statements such as "preponderance of the evidence" or "beyond a reasonable doubt." Those dealing with chance and the unknown are reluctant to assign an actual number to the chance that an event occurs.

The idea that chance could be measured quantitatively by a number took until the sixteenth century to arise. For instance, consider the problem of finding the probability



Figure 1. A six-sided die.

of rolling a 1 or a 2 on a fair six-sided die (see Figure 1). This classic problem is usually solved by assuming that each outcome of the die is equally likely. Therefore, since there are six sides and probability must sum to one, each has a one in six chance of occurring, and the probability of a 1 or 2 is two out of six, or one third.

Even in the Middle Ages, though, mathematicians such as Cardano were knowledgeable about the cheating ways of gamblers, and considered what might happen with shaved or otherwise altered dice. Gambling drove progress in probability: A historical tidbit offered up in this chapter concerns Galileo being asked by his patron to analyze a particular gambling problem to discover who had the advantage.

A leap forward in understanding how to measure probability came about during a correspondence between Pascal (Figure 2) and Fermat in the 1600s concerning the *problem of points* [4]. In this problem, two equally skilled opponents begin a series of games to decide who obtains a certain amount of prize money. Each winner is assigned a point, and the first to reach a fixed number (set in advance) of points wins the entire prize. Suppose they play the first game and the first player wins. At this point their match is interrupted, and so the question is: How should they split the prize money given this information?

The discussion of Pascal and Fermat is the first time the concept of expected value appears. The *expected value* is the average amount of money that each player would walk away with in their situation, and is how they should split the prize to achieve a fair division. Use of expected value enables the consistent development of probabilities where the outcomes are not all equally likely.

This was the beginning of turning chance into something numerical, something that could be dealt with mathematically rather than as a vague notion of plausibility. This path reaches its culmination in the fifth great idea: the formal definition of a probability measure within a set-theoretic framework. The definition introduced by Kolmogorov is now a standard feature of introductory texts.

These definitions allow for proving theorems and provide a set-theoretic foundation of the field, but they give little insight into why this is the proper way to calculate



Figure 2. Blaise Pascal.

probabilities, or what probabilities actually are. They treat probability as a measure, but they leave unanswered the question of what exactly probability measures.

Now both of the authors are unabashed Bayesians, and upfront about this throughout the work. It is therefore unsurprising that their answers to this question lie with Bayesian philosophy. In the Bayesian approach to chance, what probabilities measure is the degree of belief by an individual that the event will occur. Given an initial set of beliefs, one can update their beliefs as new information about the events comes in. This is the center of the Bayesian philosophy.

The first Bayesian idea, and the second great idea about chance in the book, is judgment, the idea that probabili-

ties arise from individual beliefs about how likely an event is to occur. But how can we ascertain what an individual's belief really is? And how do we derive the logical rules of probability from this point of view?

One way is to use an idea in mathematical finance. Suppose I have a security that pays \$1.00 if event A occurs, and nothing if A does not occur. We use A^C to mean the event that A does not occur, also called the complement of A . How much should I be willing to pay to purchase this security? Well, the expected payoff from owning this security is just the probability that event A occurs. If I judge A to have probability 0.47, then I should be willing to pay \$0.47 to own one share of the security (see Figure 3).

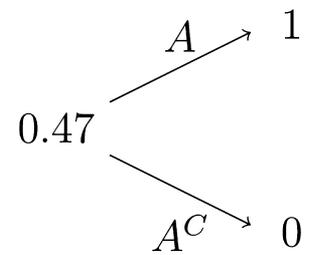


Figure 3. A simple security.

Now suppose I do the same thing for several events. A *Dutch book* is a portfolio of multiple securities that guarantees that the owner gains money. A *book* is the list of odds that a casino or racetrack would post before races. The person making the book is still known as a bookmaker or bookie. The origins of the Dutch part of the term are lost to history.

As an example of a Dutch book, suppose the security for A costs \$0.30 and the security for A^C costs \$0.60. With these prices I could buy one security that pays \$1.00 when A

occurs and nothing when A^C occurs, plus I can buy a second security that pays \$1.00 when A^C occurs and nothing when A occurs. Together, these two securities form a Dutch book that costs $\$0.30 + \$0.60 = \$0.90$. If A occurs, the first security returns \$1.00 and the second security returns nothing. If A^C occurs, then the second security returns \$1.00 and the first security returns nothing. Either way, the Dutch book makes \$1.00, but at a cost of only \$0.90. Therefore I am guaranteed to receive \$0.10 regardless of the outcome. Bruno de Finetti [3] called probabilities associated with such a Dutch book *incoherent* because prices (and probability beliefs) would naturally change if this type of profit taking were possible. Either the price of A or A^C would rise until the possibility of a Dutch book disappeared. This kind of argument leads directly to the notion that the probability of A and A^C should add to 1.

The idea here is not to derive the commonly used definitions of a probability measure, but instead to explain why probabilities behave in the way that they do.

Today, these types of securities actually exist and you can buy and sell them. They are called *prediction markets*, and are a way of gauging the probabilities of future events. One such market is the PredictIt market, created by the Victoria University of Wellington to buy and sell securities based on future events. For instance, as I write this in September of 2018, the PredictIt market is giving a 36% chance that President Trump will be re-elected to his position in 2020. Those who feel the chance is higher should buy this security, while those who feel it is lower should sell.

The sixth great idea is Bayes' Theorem for updating probabilities given new information. Bayes was an English statistician and minister who created the first argument for using prior information about a parameter together with a special case of what is now known as Bayes' Theorem. There are several ways to write this theorem; one such way is by using odds. Suppose that I believe that either theory X or theory Y is true, but not both. Each has equal probability of being true. This means that I give theory X versus theory Y 1:1 odds of being true. Because this is my current belief, this is known as the *prior* in Bayesian thought.

Now suppose that I run an experiment that gives me evidence E . If theory X is true, I can calculate that there was only an 80% chance of seeing evidence E . On the other hand, if theory Y is true, then I only had a 40% chance of seeing evidence E . At this point I know the chance of E given X and E given Y . Can I use that to determine the probability of X given E and Y given E ?

Note that 80% divided by 40% is 2, and so that value is called the Bayes' Factor in favor of theory X . Here is where Bayes' Theorem comes in. To find the new odds of X versus Y , take the original odds (the prior) and multiply by the Bayes' Factor. The original odds for X versus Y were 1:1 (also known as 1/1), so I multiply by the Bayes' Factor of 2 to get the new odds of 2:1 in favor of theory X . In this way, I can use experiments that I run to update the odds and

also the probabilities. This updated probability is called the *posterior*.

If I then ran a second experiment, the posterior from the first experiment would become the prior for the second experiment, and so on. In this way, I could keep multiplying Bayes' Factors and updating my beliefs about the veracity of theory X versus theory Y for as long as I keep getting more evidence.

What about the practitioners who do not hold that probabilities are degrees of belief? Is it a coincidence that they use the same rules and definitions of probability? The seventh great idea provides one answer to this question. This idea is Bruno de Finetti's concept of *exchangeability*.

Exchangeability is a type of symmetry applied to sequences of random variables. For instance, consider a coin that lands heads or tails, H or T , on each flip. Then applying Bayes' Theorem, no matter what your prior is, if the coin is flipped three times, the sequences HHT , HTH , and THH should be exchangeable in the sense that each sequence is equally likely. Whatever probability each of these three sequences has (and that will depend on the prior you choose for the probability of heads), the probability of each of these three sequences will be the same.

Based on this formulation, probabilities can be derived without requiring that chance exist. Whatever probability you assign to these three sequences, you can work backward to figure out what your prior on the chance of heads must have been. In other words, if you can assign probabilities to sequences, then you are behaving as though you had a prior on the probability of heads, even if you did not derive your probabilities in that fashion. This powerful tool can be extended to other processes (such as Markov chains) that have some but not all of the total symmetry found in the independent coin flip example. See [3] for more details.

Frequentist Philosophy takes a different approach, and gets a turn as the fourth great idea of chance. In the frequentist interpretation, probability does not represent the degree of belief that something is true. Instead, there is always exactly one true probability of an event, and that can be found by repeating the identical experiment over and over again and taking the limit of the number of times the event occurs and dividing by the number of times the experiment was tried. This will converge to the probability of the event (for properly defined convergence) and is a special case of the law of large numbers proved by Jacob Bernoulli in *Ars Conjectandi* (1713) [1].

John Venn in *The Logic of Chance* (published 1866) went further, saying that the projected limit of the frequency of an event was actually the probability of the event. They are not just equal numbers: that sequence together with the limit *defines* what chance is. In this approach, I cannot speak of the probability that a single coin flip is heads; instead I must consider an infinite sequence of coin flips, all done identically and independent of one another. Only then can

I take the limit of the number of heads to the number of trials to obtain the true probability.

Of course, the deep question that leaves unaddressed is: What does that infinite sequence of independent coin flips mean? If I already have a probability for my single coin flip, I can calculate for the probability of any finite subset of the coin flips, but Venn is moving in the opposite direction, requiring that the infinite sequence exist before moving to the probability of a single coin.

Later, in 1919, Richard von Mises put forth an answer to this question, not as a direct response to Venn, but rather to David Hilbert. In 1900, Hilbert had proposed ten great problems in mathematics (he later extended these to twenty-three problems), the sixth of which was *"To treat in the same manner, by means of axioms, those physical sciences in which already today mathematics plays an important part; in the first rank are the theory of probabilities and mechanics."*

Von Mises (Figure 4) took up this challenge, and came up with something that he called a *Kollektiv*. This is a sequence that has the properties that are thought of as belonging to a random sequence of independent fair coin flips. For instance, the limiting frequency of H values should be $\frac{1}{2}$, which is a first requirement that the sequence be random. However, a sequence such as HTHHTH... has this property without being random.

Here the problem is if we only consider the places 1, 3, 5, 7, ..., these are all H values, and 2, 4, 6, 8, ... are all T values. We would like it if any subsequence of the sequence had the same $\frac{1}{2}$ limiting frequency, but unfortunately any sequence with limiting frequency $\frac{1}{2}$ will contain a subsequence of all H values.

The eighth great idea, algorithmic randomness, now enters the picture. In the 1930s, logicians such as Church and Turing were developing the idea of what it means to be *computable*. For instance, we could say that a sequence of integers a_1, a_2, a_3, \dots is computable if given the first n terms in the sequence, a Turing machine can calculate the next term in the sequence. There are only a countable number of Turing machines, and so only a countable number of such sequences exist.

It is possible to create a *Kollektiv* such that for any computable sequence, the limiting sequence of H and T values is always $\frac{1}{2}$. However, Jean Ville found a complication in 1939. Ville found a *Kollektiv* such that for all computable

sequences, the limit is $\frac{1}{2}$. However, in the *Ville Kollektiv*, after a while the frequency of H values rises to be at least $\frac{1}{2}$ and stays there. In a true random sequence, the frequency will rise above $\frac{1}{2}$ and drop below $\frac{1}{2}$ infinitely often. In particular, a gambler who was aware of Ville's property could use this fact to create a system with guaranteed returns, something that should not be possible when betting on a truly random sequence.

The remaining great ideas cover different aspects of probability. For instance, despite the laws of probability being well known and commonly understood, people often behave in defiance of these rules. The psychology of chance is the third great idea. This chapter covers the experiments of psychologists that showed that how a question is framed can lead to different decisions.

For instance, Kahneman and Tversky asked the following two questions (mixed among others) in a survey. Both questions had the same background. Suppose that the United States faces an outbreak where without intervention 600 people will die. The first question is to choose between the following two medical options:

1. A treatment where 200 people will be saved.
2. A treatment where there is a $\frac{1}{3}$ chance that all 600 will be saved and a $\frac{2}{3}$ chance that no one will be saved.

The second question also asked participants to choose between two different options.

1. A treatment where 400 people will die.
2. A treatment where there is a $\frac{1}{3}$ chance that no one will die, and a $\frac{2}{3}$ chance that 600 people will die.

They found that almost three-quarters of participants preferred the first treatment in the first question, but in the second question almost three-quarters of participants preferred the second treatment. But it only takes a short while to reflect on the questions and come to the realization that *both treatments are exactly the same*. How a question is framed can affect how we make decisions and deal with chance events.

The ninth great idea is the concept of physical randomness. This type of randomness is often the first type encountered by people as they grow up, and so often it is mistakenly thought to be the only type of randomness. Rolling dice and spinning a cage of lottery balls lead to this style of randomness. Here the final result of a die roll or a coin flip is very sensitive to initial conditions. A very



Figure 4. Richard von Mises.

slight uncertainty in the initial speed and rotation of a coin flip leads to each side being equally likely to come up in practice.

This effect appears even in simple systems like the flip of a coin, but also applies to systems such as gases where a huge number of molecules are continually changing directions and bouncing off of one another and any container. Returning to the Hilbert problems, his sixth question was driven by the need of physicists to understand what randomness is in order to create a well-defined theory of statistical mechanics.

This need grew further with the development of quantum mechanics. The Copenhagen interpretation of quantum mechanics uses probabilities as a way to connect classical thought to the many experiments that show quantum behavior. The authors argue that this does not give rise to a new theory of probability, but rather leads to the same sorts of information that can be incorporated into priors through Bayesian conditioning.

The last great idea presented is induction. The idea of induction is simply that we learn from past experience. Centuries ago, Hume laid out the basic recursive question raised by the use of induction in *A Treatise of Human Nature* [5]. How do we know that inductive reasoning works if the only way we can prove it is to use inductive reasoning? Meeting that challenge is a lens for revisiting the ideas of chance raised by Laplace, Bayes, and de Finetti earlier in the text.

All in all, this is an excellent book for a reader who already understands how to calculate probabilities. The purpose of this book is to consider the questions of what probability is, and what exactly it means. Whether or not the reader has considered these questions before, the book provides a fun and engaging introduction to some of the fascinating ways that probability has been thought about over the centuries, and would make excellent supplemental reading for a probability course.

References

- [1] Bernoulli J (1713), *Ars conjectandi, opus posthumanum. Accedit Tractatus de seriebus infinitis, et epistola gallicae scripta de ludo pilae reticularis*, Basel: Thurneysen Brother, OCLC 7073795
- [2] Cicero, "De inventione," in Works (II), (1976), transl. Hubbell HM, London: Heinemann.
- [3] de Finetti B (1937), "Foresight: Its Logical Laws, Its Subjective Sources," in Kyburg HE and Smokler HEK (eds.), *Studies in Subjective Probability*, Huntington, NY: Robert E. Kreiger Publishing Co.
- [4] Devlin K (2010), *The Unfinished Game: Pascal, Fermat, and the Seventeenth-Century Letter that Made the World Modern*, Basic Books.
- [5] Hume D (1739), *A Treatise of Human Nature*, Oxford: Oxford University Press.



Mark Huber

Credits

Figures 1 and 3 are by Mark Huber.

Figure 2 is courtesy of Palace of Versailles [GFDL (www.gnu.org/copyleft/fdl.html) or CC BY 3.0 (<https://creativecommons.org/licenses/by/3.0>)].

Figure 4 is courtesy of Konrad Jacobs, Erlangen [CC BY-SA 2.0 de (<https://creativecommons.org/licenses/by-sa/2.0/de/deed.en>)], via Wikimedia Commons.

Author photo is ©Mark Huber, 2016.