TEN GREAT IDEAS ABOUT CHANCE

TEN GREAT IDEAS ABOUT CHANCE

PERSI DIACONIS
BRIAN SKYRMS

PRINCETON UNIVERSITY PRESS
Princeton and Oxford

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

Cover image courtesy of Shutterstock

First paperback printing, 2019

Paperback ISBN 978-0-691-19639-8

Cloth ISBN 978-0-691-17416-7

Library of Congress Control Number 2017943311

British Library Cataloging-in-Publication Data is available

This book has been composed in Sabon Next LT Pro text with
Akzidenz Grotesk BQ Display

Printed on acid-free paper. ∞

Printed in the United States of America

CONTENTS

PREFACE

This book grew out of a course that we taught together at Stanford for about 10 years. This was a large, mixed course. There were undergraduates and graduates. There were participants from philosophy, statistics, and a number of other disciplines across the academic spectrum. As the course evolved over time, we came to believe that the story we are telling would be of interest to a larger audience. Our course had as a prerequisite exposure to one course in probability or statistics. Our book retains this level. But for the reader who may have had such a course a long time ago, we have included an appendix designed as a probability refresher.

This is a history book, a probability book, and a philosophy book. We give the history of what we see as great ideas in the development of probability, but we also pursue the philosophical import of these ideas. One reader of an earlier version of this manuscript complained that at the end of the book, he still did not know our philosophical views about chance. We were, perhaps, too evenhanded. This problem has now been fixed. You will see that we are thorough Bayesians, followers of Bayes, Laplace, Ramsey, and deFinetti. Bayesianism is sometimes thought to be opposed to frequencies. We insist that our view does not deny the importance of frequencies or the usefulness of talking about objective chances. Rather, it unifies these considerations within the framework of rational degrees of belief.

At the beginning of this book we are thinking along with the pioneers, and the tools involved are simple. By the end, we are up to the present, and some technicalities have to be at arm's length. We try to ease the flow of exposition by putting some details in appendices, which you can consult as you wish. We also try to provide ample resources for the reader who finds something interesting enough to dig deeper. There is a select annotated bibliography at the end. There are more detailed references in the footnotes.

Persi Diaconis
Brian Skyrms

ACKNOWLEDGMENTS

Many people have helped us in the creation of this book. First of all, we want to thank our ten generations of students for their constructive feedback. Many friends read all or parts of the manuscript, made helpful suggestions, and corrected errors. We especially want to thank Susan Holmes, Christian Robert, Steve Evans, Jim Pitman, Steve Stigler, David Mermin, Simon Huttegger, Jeff Barrett, and Sandy Zabell. Our editor, Vickie Kearn, and the team at Princeton helped turn a preliminary typescript into the finished book that you see.

Gerolamo Cardano

CHAPTER 1

MEASUREMENT

One way to understand the roots of a subject is to examine how its originators thought about it. Some basic philosophical issues are already evident at the very beginning. The first great idea is simply that chance can be measured. It emerged during the sixteenth and seventeenth centuries, and it is something of a mystery why it took so long. The Greeks had a goddess of chance, Tyche. Democritus and his followers postulated a physical chance affecting all the atoms that made up the universe. This is the "swerve" of atoms in Lucretius' *De Rerum Natura*. Games of chance, using knucklebones or dice, were known to Egyptians and Babylonians and were popular in Rome. Soldiers cast lots for Christ's cloak. Greek Skeptics of the later Academy postulated probability (eikos) as the guide to life.[1] Nevertheless, it appears that there was no quantitative theory of chance in these times.[2]

Figure 1.1. Determination of the lawful rood

How do you measure anything?[3] Consider length. You find a standard of equal length, apply it repeatedly, and count. The standard might be your foot, as you pace off a distance. Different feet may not lead to the same result. One refinement, proposed in 1522 for determining a lawful rood (rod), was to line up the feet of 16 people as they emerged from church, as shown in figure 1.1.[4] As the illustration shows, the various folks have very different foot lengths, but an implicit averaging effect was accepted by a group—even though the explicit notion of an average seems to not have existed at the time.

It is worth mentioning a certain philosophical objection at this point. There is a kind of circularity involved in the procedure. We are defining length, but we are already assuming that our standard remains the same length as we step off the distance.

No sensible person would let this objection stop her from stepping off distance. That is how we start. Eventually we refine our notion of length. Your foot may change length; so may the rod; so may the standard meter stick, at a fine-enough precision. Using physics, we refine the measurement of length.[5] So the circularity is real, but it indicates a path for refinement rather than a fatal objection.*

So it is with chance. To measure probability, we first find—or make—equally probable cases. Then we count them. The probability of an event A, denoted by $P(A)$, is then

$$P(A) = \frac{\text{no. of cases in which } A \text{ occurs}}{\text{total no. of cases}}.$$

Note that it follows that

1. Probability is never negative,
2. If A occurs in all cases, $P(A)=1$,
3. If A and B never occur in the same case,

$$P(A \text{ or } B) = P(A) + P(B).$$

In particular, the probability of an event not occurring is 1 less the probability of its occurring:

$$P(\text{not } A) = 1 - P(A).$$

It is surprising how much can be done by ingenious application of this simple idea. Consider the birthday problem. What is the probability that at least two people in a room share the same birthday, neglecting leap years, assuming birthdates are equiprobable and birthdays of individuals in the room are independent (no twins)? If you have not seen it before, the results are a bit surprising.

The probability of a shared birthday in the group is 1 minus the probability that they are all different. The probability that the second person has a different birthday from the first is $(\frac{364}{365})$. If they are different, the probability that the third is different from them is $(\frac{363}{365})$, and so on, for all in the room. So the probability of a shared birthday among N people is

$$1 - \left(\frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365 - N + 1}{365} \right).$$

* What are the paths open for refinement of the notion of equiprobable? They will unfold as we move through the book.

If you are interested in an even-money bet, this formula can be used to find a value of *N* such that the product is close to $\frac{1}{2}$. If there are 23 people in the room, the probability of a shared birthday is slightly greater than $\frac{1}{2}$. If there are 50 people, it is close to 97%.

There are many variations on the birthday problem. These are used for thinking about surprising coincidences. For instance, it is overwhelmingly likely that there are two people in the United States who share a birthday, whose fathers share the same birthday, whose fathers' fathers share this birthday, and so on, four generations back. Useful approximations for working with these variations may be found in an appendix to this chapter. These approximations are, in turn, used to prove de Finetti's representation theorem in an appendix at the end of this book. The point for now is that the basic "equally likely cases" structure has real breadth and strength.

BEGINNINGS

Nothing provides us better candidates for equiprobable cases than vigorous throws of symmetric dice or draws from a well-shuffled deck of cards. This is where the measurement of probability began. We cannot say who was there first, but the idea was clearly there in the sixteenth-century work on gambling by the algebraist, physician, and astrologer Gerolamo Cardano.[6] Cardano, who sometimes made a living at gambling, was quite sensitive to the equiprobability assumption. He knew about shaved dice and dirty deals: ". . . the die may be dishonest either because it has been rounded off, or because it is too narrow (a fault which is easily visible), or because it has been extended in one direction by pressure on the opposite faces. . . . There are even worse ways of being cheated at cards."[7]

In the early seventeenth century Galileo composed a short note on dice to answer a question posed to him (by his patron, the Grand Duke of Tuscany). The Duke believed that counting possible cases seemed to give the wrong answer. Three dice are thrown. Counting combinations of numbers, 10 and 11 can be made in 6 ways, as can 9 and 12. ". . . yet it is known that long observation has made dice-players consider 10 and 11 to be more advantageous than 9 and 12."* How can this be?

Galileo replies that his patron is counting the wrong thing. He counts three 3s as one possibility for making a 9 and two 3s and a 4 as one possibility for making a 10. Galileo points out the latter covers three possibilities, depending on which die exhibits the 4:

<4, 3, 3>, <3, 4, 3>, <3, 3, 4>.

For the former, there is only <3, 3, 3>. Galileo has a complete grasp of permutations and combinations and does not seem to regard it as anything new.

In constructing equiprobable cases, both Galileo and Cardano appear to make implicit use of *independence*. They suppose that for each die, all 6 faces are equally probable and that for throws of 3 dice, all 216 possible outcomes are also equally probable. When we treated the birthday problem earlier, we assumed that different people had independent chances for their birthdays.

With this basic machinery well understood, Pascal and Fermat in their famous correspondence attacked more subtle problems with a different conceptual flavor.

PASCAL AND FERMAT (1654)

The first substantial work in the mathematics of probability appears to be the correspondence between Pascal and Fermat, which began in 1654. We include a discussion for three reasons: (1) It *is* the first; (2) it shows how seemingly complex problems can be reduced to straightforward calculations with equally likely cases; and (3) it introduces the crucial notion of expectation—a mainstay of the subject.

* One strange aspect of the statement of the problem is the comment about long observation. The observation would have had to be long indeed. From Galileo's calculations, the chance of a 9 is $\frac{25}{216}$, about 0.116; the chance of a 10 is $\frac{27}{216}$, about 0.125. The difference between these is 0.009, or about $\frac{1}{100}$. As an exercise, you could calculate how many observations would be required.

the 8 equiprobable cases, player 2 will win the game only if he wins all 3 rounds. His expectation is $\frac{1}{8}$ of the stakes, while player 1 has an expectation of $\frac{7}{8}$ on the stakes. It is fair, then, to divide the stakes in this proportion.

Expectation, computed by counting equiprobable cases, solves the problem. But there may be a large number of equiprobable cases to count. Consider Tartaglia's example. Six points win, and one player has no points and the other, 1 point. Then play must be complete after 10 more rounds. It would be tedious to write out the 1024 possible outcomes. But Pascal had a better way of counting.

To count the cases in which the first player wins, one adds the number of cases in which she gets 6 wins in 10 trials [called 10 choose 6] + the number where she gets 7 wins in 10 trials [10 choose 7] + · · · + the number where she gets 10 wins in 10 trials [10 choose 10]. These numbers are conveniently to be found on the tenth row of Pascal's arithmetical triangle (or Tartaglia's triangle, or Omar Khayyam's triangle[10]), which we show in figure 1.2. The row tells us the number of ways we can choose from a group of 10 objects. Reading from the left, there are 1 way of choosing nothing, 10 ways of choosing 1 object, 45 ways of choosing 2 objects, 120 ways of choosing 3, and so on, to only 1 way of choosing 10.

We want the number of ways of getting 6 wins in 10 trials + the number of ways to get 7 wins in 10 trials + · · · + the number where she gets 10 wins in 10 trials. From row 10 we get

$$210 + 120 + 45 + 10 + 1 = 386$$

for a probability of winning of

$$\frac{386}{1024} \quad \text{(about 38\%)}.$$

Thus a fair division of the stakes gives player 1 (who had no points) $\frac{386}{1024}$ of the stakes and player 2 the rest.

After Pascal and Fermat, the basic elements of measuring probability by counting equiprobable cases, calculating by combinatorial principles, and using expected value are all on the table.

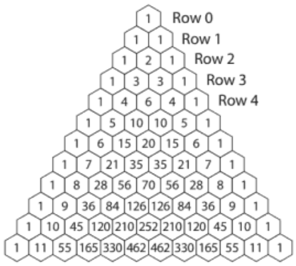

Figure 1.2. Pascal's triangle

HUYGENS (1657)

The ideas in the Pascal-Fermat correspondence were taken up and developed by the great Dutch scientist Christiaan Huygens[11] after he heard about the correspondence on a visit to Paris. He then worked them out by himself and wrote the first book on the subject in 1656. It was translated into English by John Arbuthnot in 1692 as *Of the Laws of Chance*.[12]

Huygens begins his book with a fundamental principle:

> *Postulat*
>
> As a Foundation to the following Proposition, I shall take Leave to lay down this Self-evident Truth: That any one Chance or Expectation to win any thing is worth just such a Sum, as wou'd procure in the same Chance and Expectation at a fair Lay. As for Example, if any one shou'd put 3 Shillings in one Hand, without letting me know which, and 7 in the other, and give me Choice of either of them; I say, it is the same thing as if he shou'd give me 5 Shillings; because with 5 Shillings I can, at a fair Lay, procure the same even Chance or Expectation to win 3 or 7 Shillings.

Huygens assumes that he could, in effect, flip a fair coin to choose which hand to pick.* Then $(\frac{1}{2})3+(\frac{1}{2})7=5$. He then says that the value

* A point made much later by Howard Raiffa against the so-called Ellsberg paradox, which we will visit in our chapter on psychology of chance (chapter 3).

of the wager is the same as the value of 5 for sure. Thus he makes explicit (a special case of) the principle that is implicit in Pascal and Fermat: *expectation is the correct measure of value*.

He then goes on to justify this measure by a fairness argument. Suppose I bet 10 shillings with someone on the flip of a fair coin. This is fair by reasons of symmetry. Now suppose we modify this by an agreement that whoever wins shall give 3 to the loser. This preserves symmetry, so the modified arrangement is also fair. But now the loser nets 3 and the winner retains 7. Any such agreement preserves fairness, including where the winner gives the loser 5, and each has 5 for sure. Huygens then shows how the argument generalizes to arbitrary finite numbers of outcomes and arbitrary rational-valued probabilities of outcomes. It will be a recurring theme that an equality is justified by a symmetry.

NEWTONIAN CONSIDERATIONS

In the preface to the translation of Huygens, Arbuthnot, who was a follower of Newton,[13] makes the following noteworthy remark (L. Todhunter, *A History of the Mathematical Theory of Probability* (Cambridge: Macmillan, 1865); reprinted by Chelsea (New York, 1965), p. 51):

> It is impossible for a Die, with such determin'd force and direction, not to fall on such determin'd side, only I don't know the force and direction which makes it fall on such determin'd side, and therefore I call it Chance, which is nothing but the want of art.

Arbuthnot thus introduces the fundamental question of the proper conception of chance in a deterministic setting. His answer is that chance is an artifact of our ignorance.

Consider tossing a coin just once. The thumb hits the coin; the coin spins upward and is caught in the hand. It is clear that if the thumb hits the coin in the same place with the same force, the coin will land with the same side up. Coin tossing is physics, not random! To demonstrate this, we had the physics department build us a coin-tossing machine. The coin starts out on a spring, the spring is released, the coin spins upward and lands in a cup, as shown in figure 1.3. Because the forces are controlled, the coin always lands with the same side up.

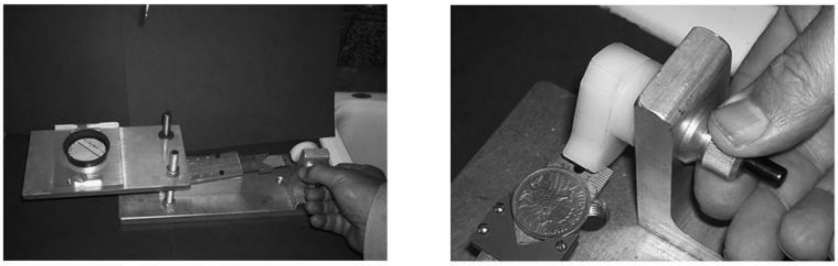

Figure 1.3. A deterministic coin-tossing machine

This is viscerally quite disturbing (even to the two of us). Magicians and crooked gamblers (including one of your authors) have the same ability.

How then is the probabilistic treatment of coin flips so widespread and so successful? The basic answer is due to Poincaré. If the coin is flipped vigorously, with sufficient vertical and angular velocity, there is sensitive dependence on initial conditions. Then a little uncertainty as to initial conditions is amplified to a large uncertainty about the outcome, where equiprobability of outcomes is not such a bad assumption. But the provisos are important. See appendix 2 for a little more on this. We will return to the question in more detail in our chapter on physical chance (chapter 9).

BERNOULLI 1713

In 1713 Jacob Bernoulli's *Ars Conjectandi*[14] was published, 8 years after his death. Bernoulli made explicit the practice of his predecessors. The first part is a reprint, with commentary, of Huygens. The probability of an event is now explicitly defined as the ratio of the number of (equiprobable) cases in which the event happens to the total number of (equiprobable) cases. The probability of being dealt a club from a deck of cards is $\frac{13}{52}$. He also defines the *conditional probability* of a second event (B) conditional on a first (A) as the ratio of the number of cases both happen to the number of cases the first happens:

$$\text{Probability } (B \text{ conditional on } A) = \frac{\text{no. of cases in which } A \text{ and } B \text{ occur}}{\text{no. of cases in which } A \text{ occurs}}.$$

The probability of being dealt a queen given that one is dealt a club is $\frac{1}{13}$.

On the basis of these definitions, he shows that the probabilities of mutually exclusive events add and that probabilities satisfy the multiplicative law, $P(A \text{ and } B) = P(A)P(B \text{ conditional on } A)$. These simple rules form the heart of all calculations of probability.

But Bernoulli's major contribution was to establish a rigorous connection between probability and frequency that had heretofore only been conjectured. He called this his golden theorem.

As an illustration he considers an urn containing 3000 white pebbles and 2000 black pebbles and postulates independent draws with replacement of the pebble drawn. He asks whether one can find a number of draws so that it becomes "morally certain" that the ratio of white pebbles to black ones becomes approximately 3:2. He then chooses a high probability as moral certainty and establishes a number of draws sufficient to provide a positive answer. Then he shows the weak law of large numbers:

> Given any interval around the probability (here $\frac{3}{5}$) as small as you please and any approximation to certainty, $1-e$, as close as you please, there is a number of trials, N, such that in N trials the probability that the relative frequency of draws of white falls within the specified interval is at least $1-e$.

This is a story to which we will return in our chapter on frequency (chapter 4).

SUMMING UP

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

1. Probability is a number between 0 and 1.
2. If A never occurs, $P(A)=0$. If A occurs in all cases, $P(A)=1$.
3. If A and B never occur in the same case, then $P(A \text{ or } B)=P(A)+P(B)$.

Fermat sees clearly that the analysis is the same at any point in the game. Suppose that after the round in question, there will be $n+1$ rounds remaining; give the stakes at this point value 1. Then the value of taking the play is $\frac{1}{6}$ for winning now and $(\frac{5}{6})(1-\left(\frac{5}{6}\right)^n)$ for failing on this throw but possibly eventually winning. The value of taking $\frac{1}{6}$ of the stakes and proceeding with the rest of the game for the diminished stakes is $\frac{1}{6}$ for the cash in hand plus $1-\left(\frac{5}{6}\right)^n$, the probability of eventually, winning times $\frac{5}{6}$ of the the diminished stakes. Pascal immediately agrees with Fermat's analysis.

THE PROBLEM OF POINTS

There is another aspect of Pascal's discussion that is of interest. He starts with the example of a game where two players play for 3 points, where each has staked 32 pistoles ("Pascal and Fermat on Probability," tr. by Vera Sanford in *A Sourcebook in Mathematics*, ed. David Eugene Smith (New York: McGraw Hill, 1929), 546–65. Dover reprint in 1969 available online at https://www.york.ac.uk/depts/maths/histstat/pascal.pdf):

> Let us suppose that the first of them has two (points) and the other one. They now play one throw of which the chances are such that if the first wins, he will win the entire wager that is at stake, that is to say 64 pistoles. If the other wins, they will be two to two and in consequence, if they wish to separate, it follows that each will take back his wager that is to say 32 pistoles.
>
> Consider then, Monsieur, that if the first wins, 64 will belong to him. If he loses, 32 will belong to him. Then if they do not wish to play this point, and separate without doing it, the first should say "I am sure of 32 pistoles, for even a loss gives them to me. As for the 32 others, perhaps I will have them and perhaps you will have them, the risk is equal. Therefore let us divide the 32 pistoles in half, and give me the 32 of which I am certain besides." He will then have 48 pistoles and the other will have 16.

This is not just a calculation of expected value but also a justification of the *fairness* of using it, in terms that are hard for anyone to reject. What you have for sure is yours. For what is uncertain, equal probabilities match equal division. It is a definitive answer to Fra Pacioli's line of thought.

Pascal goes on to show how this reasoning can be further iterated:

> Now let us suppose that the first has *two* points and the other *none*, and that they are beginning to play for a point. The chances are such that if the first wins, he will win all of the wager, 64 pistoles. If the other wins, behold they have come back to the preceding case in which the first has *two* points and the other *one*.
>
> But we have already shown that in this case 48 pistoles will belong to the one who has *two* points. Therefore if they do not wish to play this point, he should say, "If I win, I shall gain all, that is 64. If I lose, 48 will legitimately belong to me. Therefore give me the 48 that are certain to be mine, even if I lose, and let us divide the other 16 in half because there is as much chance that you will gain them as that I will." Thus he will have 48 and 8, which is 56 pistoles.
>
> Let us now suppose that the first has but *one* point and the other *none*. You see, Monsieur, that if they begin a new throw, the chances are such that if the first wins, he will have *two* points to *none*, and dividing by the preceding case, 56 will belong to him. If he loses, they will [be] point for point, and 32 pistoles will belong to him. He should therefore say, "If you do not wish to play, give me the 32 pistoles of which I am certain, and let us divide the rest of the 56 in half. From 56 take 32, and 24 remains. Then divide 24 in half, you take 12 and I take 12 which with 32 will make 44.

This gives us a recursive procedure for fair division. Pascal then projects to games with larger numbers of points, and comes to a general solution of the problem.

APPENDIX 2. PHYSICS OF COIN TOSSING

Drawing balls from an urn, flipping coins, rolling dice, and shuffling cards are basic probability models. How are they connected to their parallels in the real world? Going further afield, these basic models are often used to calculate chances in much more complicated setups; Bernoulli considered the successive scores of two tennis players. Gilovitch, Tversky, and Valone[15] considered the successive hits and misses of basketball players. Shouldn't physics and psychology come into these analyses?

Each of the foregoing examples has its own literature. To give a flavor of this, we consider a single flip of a coin. Afterward, pointers to the analysis of other examples will be given.

Let's take a brief look at a simple version of the physics.[16] When the coin leaves the hand, it has an initial velocity upward v (feet/second) and a rate of spin ω (revolutions/second). If v and ω are known, Newton tells us how much time the coin will take before landing and thus heads or tails are determined. The phase space of a coin in this model is thus as shown in figure 1.4.

A single flip corresponds to a point in this plane. Consider the point in figure 1.4. The velocity is large (so the coin goes up rapidly), but the rate of spin is low. Thus the coin goes up like a pizza tossed in the air, hardly turning. Similarly, a point with v small and ω large may be turning like crazy but never goes high enough to turn over once. From these considerations, it follows that there is a region of initial conditions, close to the two axes, where the coin never turns.

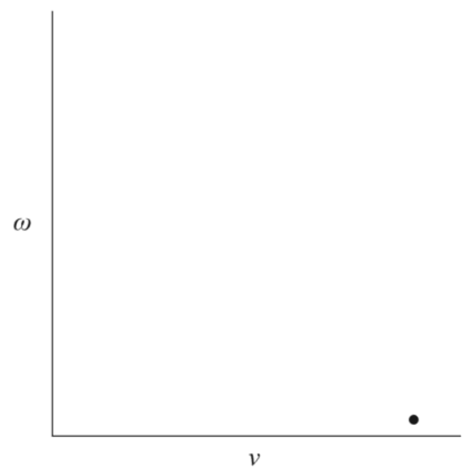

Figure 1.4. The $v\omega$-plane with a single flip

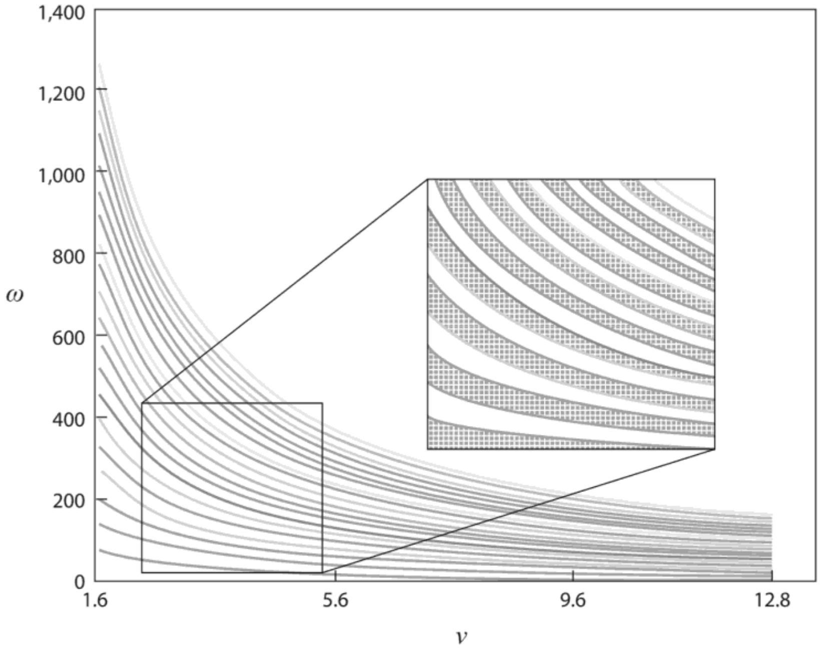

Figure 1.5. The hyperbolas separating heads from tails in part of phase space. Initial conditions leading to heads are hatched, tails are left white, and ω is measured in s^{-1}.

There is an adjoining region where the coin turns once, then a region for two turns, and so on. The full picture is shown in figure 1.5.

Inspection of the picture (and some easy mathematics) shows that regions far from 0 get closer together. So small changes in initial conditions make for the difference between heads or tails.

To go further, one must know the answer to the following question: When real people flip real coins, where are the points on the picture? We have carried out experiments and a normal flip takes about $\frac{1}{2}$ second and turns at about 40 revolutions/second. Look at figure 1.5. In the units of the picture, velocity is about $\frac{1}{5}$, very close to zero. The rate of spin, ω, is 40 units up, however, way off the picture. The math behind the picture says how close the regions are. This coupled with experimental work shows that coin tossing is fair to two decimal places but not to three.

The preceding analysis is in a simple model, which assumes that the coin flips about an axis through the coin. In fact, real coins are more complicated. They precess in amazing ways. A full analysis, with many details, caveats, and full references is in "Dynamical Bias in the Coin Toss,"[17] which concludes that vigorous tosses of ordinary coins are *slightly biased*. The chance of the coin landing the same way it started is about 0.51.

Where does all this analysis leave us? The standard model is a very good approximation. It would take about 250,000 flips to detect the difference between 0.50 and 0.51 (in the sense of giving second-digit accurately). We wish some of the other instances of the standard model were as solidly useful. Similar statements hold for Galileo's dice, but roulette or shuffling cards is another story![18]

If an honest analysis of a simple coin flip leads us into such complications, how much more would be required for an analysis of chances in games of skill or for the application of probability to medicine and law, as envisioned by Leibniz and Bernoulli? Bernoulli appreciated the point (Jacob Bernoulli, *The Art of Conjecturing*, tr. with an introduction and notes by Edith Dudley Sylla (Baltimore: Johns Hopkins University Press, 2006), 327):

> But what mortal, I ask, may determine, for example, the number of diseases, as if they were just as many cases, which may invade at any age the innumerable parts of the human body and which imply our death? And who can determine how much more easily one disease may kill than another—the plague compared to dropsy, dropsy compared to fever? Who, then, can form conjectures on the future state of life and death on this basis? Likewise who will count the innumerable cases of the changes to which the air is subject every day and on this basis conjecture its future constitution after a month, not to say after a year?
>
> Again, who has a sufficient perspective on the nature of the human mind or on the wonderful structure of the body so that they would dare to determine the cases in which this or that player may win or lose in games that depend in whole or in part on the shrewdness or the agility of the players? In these and similar situations, since they may depend on causes that are entirely hidden and that would forever mock our diligence by an

Frank Plumpton Ramsey

CHAPTER 2

JUDGMENT

Our second great idea is that judgments can be measured and that coherent judgments are probabilities. (What exactly is meant by coherence will be made clear.) In the classic gambling games of chapter 1, our judgments were informed by symmetry. Symmetrical cases were judged equiprobable. In this chapter we will see how degrees of belief implicit in judgments about all sorts of cases can also be measured. When they are so measured, coherent judgments turn out to have the same mathematical structure as that discovered by Cardano and Galileo by counting equiprobable outcomes in gambling.

How can we measure the probability that there will be a financial crisis in the next year, that the patient will survive if given this treatment, or that the defendant is guilty? How can we measure the probability that this candidate will win the election, that there will be a depression, or that reckless politics will precipitate a war? Here we do not have a nice set of intuitively equiprobable cases that allowed us to calculate by counting with supposedly perfectly fair dice. But, in fact, the law, politics and medicine were areas where Leibniz and Bernoulli envisioned the most important application of the calculus of probabilities. These probabilities can be no better than our degrees of belief based on the best available evidence.[1] That does not mean that they cannot be measured.

Below, we will be talking about assessing probabilities by betting. There are a number of real world instances in which you can do just that. Perhaps the simplest are prediction markets. These are Web sites in which you can bet for or against well-specified events, things like who will win a football game, or a horse race, or the next election. Prediction markets are not a new invention; in the sixteenth century there were markets for betting on who would be elected Pope.[2] In a typical prediction market, contracts are scaled between 0 and 100. At any time you can see offers to buy and to sell, say at 56.8 and 57.2. If you wish to buy a contract, you can buy instantly at the posted price of 57.2 or post an offer to buy, say, at 57.0, and wait to see whether anyone is willing to sell at that price. If you buy a contract on Clinton winning at 57, that means that for $57, you get a contract that pays $100 if she wins. Of course, the prices fluctuate.

It is very natural to take the current market price as the market's probability. The expected value of a bet that pays off $100 if C; $0 otherwise is $57, if the probability of C is 0.57. If the market prices do not obey the mathematics of probability, then—as we shall see—the market can be arbitraged. We think that *your* prices (I'll buy in for a small amount if the price is below x and sell if the price is above y), are good indications of *your* probability.

There is a healthy emerging literature on prediction markets. Buying stocks, bonds, and insurance are closely related activities. In all of these, the principles we lay out next can be useful.[3]

We divide the body of this chapter in two—between a naive and a sophisticated approach. The initial treatment will assume, like the early gamblers, that for the problems at issue, money is the relevant measure of value. That allows a straightforward way to measure judgmental probability and to infer the mathematical structure of such judgments provided that they are coherent. The second half of the chapter lifts this assumption and gives a more general analysis. This completion of the great idea measures both probability and utility at the same time. The leading ideas of the theory were given in the 1920s by a young genius, Frank Plumpton Ramsey, and fully developed in the 1950s by Leonard Jimmie Savage.[4]

PART I: GAMBLING AND JUDGMENTAL PROBABILITIES

To measure judgmental probabilities, we invert the approach of Pascal and Fermat and follow Huygens. Instead of measuring probabilities to compute expected value, we use expected value to measure probabilities. We measure the expected value imputed to an event by measuring the price that an individual will pay for a wager on that event. Your judgmental probabilities are then the quantities which, when used in a weighted average, give that expected value.

In particular,

The probability of A is just the expected value of a wager that pays off 1 if A and 0 if not.

If you pay a price equal to $P(A)$ for such a wager, you believe that you have traded equals for equals. For a lesser price you would prefer to buy the wager; for a greater price you would prefer not to buy it. So the balance point, where you are indifferent between buying the wager or not, measures your judgmental probability for A.

It is farfetched idealization to assume that people can effortlessly and reliably make such fine discriminations. But taking the first steps of the approximation is perhaps all we need to do for many decisions. How far can we go? There is no clear answer. We proceed to explore the theory that results from the full idealization.

COHERENT JUDGMENTS

Do judgmental probabilities, in general, have the mathematical structure gotten in chapter 1 by counting? Bruno de Finetti showed that if an individual's betting behavior is *coherent*, her judgmental probability, so defined, does indeed have the mathematical structure of a probability. The basic argument can be given very simply.*

Here is our idealized model—not meant to be all of real life but nevertheless meant to be instructive. An individual acts like a bookie—or perhaps like a derivatives trader—and buys and sells bets. She judges a bet as *fair* if her expected value for it is zero, *favorable* if her expected value is positive and disadvantageous if her expected value is negative. She buys fair or favorable bets and sells fair or disadvantageous bets, doing business with all comers. A *Dutch book* can be made against her if there is some finite set of transactions acceptable to her such that she suffers a net loss in every possible situation. We will say that she is *coherent* if she is not susceptible to a Dutch book.

An Example: Suppose you ask me for my judgmental probability that Senator Foghorn will win a second term. After some thought, I say 0.6. Then you ask me for my probability that Bobbie Blowhard will be elected instead. I quickly say 0.1. Then I am asked for the probability that either Foghorn or Blowhard will win, and I say 0.9. If I stick to these judgmental probabilities I am incoherent. You can make a Dutch book against me by buying from me a bet that pays off 1 if Foghorn wins for 0.6 and a bet that pays off 1 if Blowhard wins for 0.1 and then turning around and selling me a bet that pays 1 if either wins for 0.9. You are covered no matter who wins and pocket the profit of 0.2.

If you are kind enough to point this out to me instead of exploiting my incoherence, I may well reconsider my probability judgments. We all make careless judgments that are full of incoherence. Sometimes it doesn't matter much. But what if the stakes were high enough to be important? To take an extreme case, suppose you are a hedge-fund manager and there are other hedge-fund managers in the market. If you were made aware of your incoherence, wouldn't you tend to do a little rethinking?

Aiming for coherence has its roots in a desire for consistency. It applies to logic as well. One of the wisest men we know put it this way: "We all believe inconsistent things. The purpose of rational

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

discussion aims at this: If someone says 'You accept *A* and *B*, but by a chain of reasoning, each step of which you accept, it can be shown that *A* implies not *B*,' you would think that something is wrong and want to correct it."

It is similar with judgments of uncertainty. Of course, there is no bookie, and no one is betting. Still coherence, like consistency, seems like a worthwhile standard.

De Finetti showed that coherence is equivalent to one's judgments having the mathematical structure of probability.

COHERENT JUDGMENTS ARE PROBABILITIES

To say one's judgments have this mathematical structure is just to say that they behave as proportions. They are proportions of partial belief. Proportions have a minimum of 0. A tautology, which is true throughout the whole space of possibilities, has proportion equal to 1. Proportions of a combination of mutually exclusive parts—of jointly inconsistent propositions—add. If there are 20% red beans and 35% white beans in a bag then there are 55% beans that are red or white. As we will see presently, that is all we need for the mathematical structure of probability applied to a finite space of possibilities.[5]

I. Coherence implies probability.

1. Minimum of zero.

   Suppose you give some proposition, *p*, a probability less than 0. Then you will give a bet where you *lose 1 if p, nothing otherwise* a positive expected value.

   You will then suffer a net loss no matter what happens. If *p* doesn't happen, you gain nothing from the bet and lose what you paid for it. If *p* does happen, you have a double loss. You lose what you paid for the bet and you lose the bet as well.

2. Tautology gets probability of 1.

   Suppose that you give a tautological proposition, one that is true in any case, a probability different than 1. It is either greater than or less than 1. If it is greater than 1, you would pay more than 1 for a bet that pays off *1 if p, nothing otherwise*. When bets are settled you would win only 1, for a net loss. If you were to give the tautology probability less than 1, you would sell a bet

Nominated	Elected	B_1	B_2	B_1+B_2	B_3	Total
T	T	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{6}$
T	F	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$\frac{1}{2}$	$-\frac{1}{6}$
F	T	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	0
F	F	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	0

Stage 2: To turn it into a full Dutch book, we hedge. We do this by making an additional even money bet, B4, according to which we pay Fred $\frac{1}{12}$ if she is nominated and he pays us $\frac{1}{12}$ if she isn't. He considers this as fair. His net loss is now $\frac{1}{12}$ no matter what. Fred is subject to a Dutch book. If you want to see the general case worked out, it is in an appendix to this chapter.

P	Q	B_1	B_2	$B_1 + B_2$	B_3	B_4	Total
T	T	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{2}$	$\frac{1}{12}$	$-\frac{1}{12}$
T	F	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{12}$	$-\frac{1}{12}$
F	T	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	$-\frac{1}{12}$	$-\frac{1}{12}$
F	F	$-\frac{1}{3}$	$\frac{1}{3}$	0	0	$-\frac{1}{12}$	$-\frac{1}{12}$

COHERENT UPDATING

So far we have coherence of conditional and unconditional bets at a given time. What about change in probabilities when we get new evidence? Is there a sense of coherent belief change that applies? Suppose that the evidence is some proposition, *e*, which you learn with certainty. Then the standard rule for changing one's judgmental probabilities is to take as one's new probabilities the old probabilities conditional on *e*. This is known as *conditioning on the evidence*. Is there a coherence argument for this rule? We want to emphasize that we have shifted from asking about coherence of degrees of belief to asking about *coherence of rules for changing degrees of beliefs*.

Such an argument is implicit in de Finetti's discussion, and it was made explicit by the philosopher David Lewis.[7] Conditioning on the evidence, Bayesian updating, is the unique coherent rule for updating probabilities in such a situation. Any other rule leaves one open to a Dutch book against the rule—a Dutch book across time, a *diachronic Dutch book*. Here are a model and a precise version of the argument.

The Model

The epistemologist (scientist, statistician) acts as bookie. She has a principled way of updating on the evidence, but we don't presuppose what it is. *Today* she posts her probabilities and does business. *Tomorrow* she makes an observation (with a finite number of possible outcomes, each with positive prior probability.) She updates her probabilities according to her *updating rule*, a *function* that maps possible observational outcomes to revised probabilities. Her updating rule is public knowledge. The day after tomorrow—after her observation—she posts revised fair prices and does business.

A bettor's strategy consists of (1) a finite number of transactions today that the epistemologist considers fair according to her probabilities, and (2) a function taking possible observations to sets of finite transactions the day after tomorrow at the prices the epistemologist *then* considers fair according to her updating rule.

COHERENT BELIEF CHANGE IMPLIES CONDITIONING ON THE EVIDENCE

Let $P(A|e)$ be $P(A \text{ and } e)/P(e)$ and $P_e(A)$ be the probability that the bookie's nonstandard updating rule—the rule at variance with conditioning on the evidence—gives A if e is observed. Suppose $P(A|e) > P_e(A)$[8] and let the discrepancy, δ, be $P(A|e) - P_e(A)$. Here is the bettor's strategy that makes a Dutch book.

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

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

Tomorrow: If e was observed, offer to buy [$1 if A, 0 otherwise] from the bookie at its current fair price: $P_e(A) = P(A|e) - \delta$.

Then, in every possible situation, the bookie loses $$\delta P(e)$.[9]

We could, of course, take a little of our sure winnings and divide them up to sweeten each transaction. *Then the bookie finds every*