TWO PROBABILISTIC ESSAYS:
1. THE PROBLEM OF THINKING TOO MUCH
2. MYSTERIES OF CARDANO THE PROBABILIST

by

Persi Diaconis

Technical Report No. 2003-3
February 2003

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

TWO PROBABILISTIC ESSAYS:
1. THE PROBLEM OF THINKING TOO MUCH
2. MYSTERIES OF CARDANO THE PROBABILIST

by

Persi Diaconis
Department of Statistics
Stanford University

Technical Report No. 2003-3
February 2003

This research was supported in part by
National Science Foundation grants DMS 0072360

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu

Two Probabilistic Essays:
1. The Problem of Thinking Too Much
2. Mysteries of Cardano the Probabilist

Persi Diaconis

Abstract

The first essay is on the limited utility of quantitative thinking. The second is on the history of probability.

The Problem of Thinking Too Much

Persi Diaconis
Department of Mathematics
And Statistics
Stanford University

I have in mind the image of a centipede who starts thinking about which leg to move and winds up going nowhere. It is a familiar problem: Any action we take has so many unforeseen consequences, how can we possibly choose? Here is a less grand example: I don't like moving the knives, forks and spoons from dishwasher to draw. There seems no sensible way to proceed. I frequently catch myself staring at the configuration hoping for insight. Should I take the tallest things first or just grab a handful and sort them at the draw. Perhaps I should stop thinking and do what comes naturally. Before giving in to "thinking too little" I recall a friend's suggestion: You can speed things up by sorting them as you put them into the dishwasher. On reflection though, this might lead to nested spoons not getting clean. And so it goes.

I'm not brazen enough to try to carefully define 'thinking' in the face of a reasonably well-posed problem. I certainly include mental computation (running scenarios, back of the envelope computations), gathering information (searching memory, or the Web, calling friends), searching for parallels (this problem seems roughly like another problem I know how to solve. Isn't there an easier special case?) and finally, trying to maneuver our mind into places where we are in tune with the problem and can have a leap of insight.

The problem is this: We can spend endless time thinking and wind up doing nothing or worse, getting involved in the minutia of a partially baked idea and believing that pursuing this is the same as making progress on the original problem.

The study of what to do given limited resources has many tendrils. I will review work in economics, psychology, in search theory, computer science and in my own field of mathematical statistics. These aren't of much help but at the end I collect a few rules of thumb that seem useful.

AN EXAMPLE One of the most satisfying parts of the subjective approach to statistics is de Finettis' solution of common inferential problems through exchangeability. Some of us think de

Finetti has solved Humes' Problem: When is it reasonable to think that the future will be like the past? I want to present the simplest example and show how thinking too much can make a mess of something beautiful.

Consider observing repeated flips of a coin. The outcomes will be called heads (H) and tails (T). In a subjective treatment of such problems one attempts to quantify prior knowledge into a probability distribution for the outcomes. For example, your best guess that the next three tosses yield HHT is the number P(HHT). In many situations the order of the outcomes is judged irrelevant. Then P(HHT) equals P(HTH) equals P(THH). Such probability assignments are called exchangeable.

de Finetti proved that an exchangeable probability assignment for a long series of outcomes can be represented as a mixture of coin tossing: For any sequence x, y, ..., z of potential outcomes

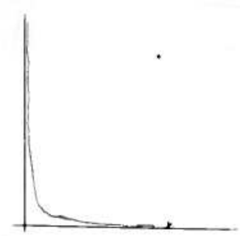
$$P(x, y, \ldots, z) = \int_0^1 p^A (1-p)^B \mu(dp)$$

with 'A' the number of heads and 'B' the number of tails among x, y, ..., z. The right side of this formula has been used since Bayes (1764) and Laplace (1774) introduced Bayesian statistics. Modern Bayesians call $p^A (1-p)^B$ the likelihood and μ the a priori probability. Subjectivists such as de Finetti, Ramsey and Savage (and Diaconis) prefer not to speak about non-observable things such as 'p-The long term frequency of heads'. They are willing to assign probabilities to potentially observable things such as 'one head in the next ten tosses'. de Finetti's Theorem shows that, in the presence of exchangeability, the two formulations are equivalent.

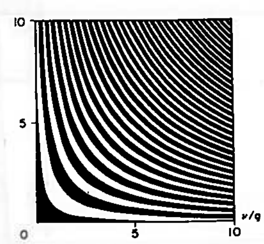
The mathematical development goes further. After observing A heads and B tails, predictions about future trials have the same type of representation with the prior μ replaced by a posterior distribution given by Bayes formula. Laplace and many followers proved that as the number of trials increases, the posterior becomes tightly focused on the observed proportion of heads A/(A+B). Predictions of the future then essentially use this frequency, the prior μ is washed away. Of course, with a small number of trials the prior can matter. If the prior is tightly focused, the number of trials required to wash it away may be very large. The mathematics makes perfect sense of this; 50 trials are often enough. The whole package gives a natural, elegant account of proper inference. I will stick to flipping coins but all of this works for any inferential task from factory inspection of defective parts to evaluation of a novel medical procedure.

Enter Physics Our analysis of coin-tossing thus far has made no contact with the physical act of tossing a coin. We now put in a bit of physics and stir; I promise, a mess will emerge. When a coin is flipped and leaves the hand it has a definite velocity in the upward direction and a rate of spin (say revolutions per second). If we knew these parameters Newton's Laws allow us to calculate how long the coin will take before returning to its starting height and so how many times it will turn over. If the coin is caught without bouncing, we know if it will land heads or tails.

Joe Keller has carried a neat analysis in the 1979 American mathematical monthly. The sketch in Figure 1 shows the velocity/spin plane.



A flip of the coin gives a dot on the figure corresponding to the velocity and rate of spin. For the dot shown, the velocity is high but spin is low. The coin goes up like a pizza and doesn't turn over at all. All the points below the curve correspond to flips where the coin doesn't turn over. The adjacent region contains points where the coin turns over exactly once. It is bounded by a similar curve. Beyond this the coin turns over exactly twice, and so on. The full picture appears in Figure 2



a glance at the figure shows that moving away from the origin, the curves get closer together. Thus, for vigorous flips, small changes in the initial conditions make for the difference between heads and tails.

The question arises: When normal people flip real coins, where are we on this picture? I became fascinated by this problem and have carried out a series of experiments. It is not hard to determine typical velocity. Get a friend with a stopwatch, practice a bit, and time how long the coin takes in its rise and fall. A typical one foot toss takes about half a second (this corresponds to an upward velocity of about 5½ miles per hour). Determining rate of spin is trickier. I got a tunable strobe, painted the coin black on one side and white on the other and tuned the strobe until the coin 'froze' showing only white. All of this took many hours. The coin never perfectly froze and there is variation from flip to flip. In the course of experimenting I had a good idea. I tied a very thin ribbon (actually dental floss) about three feet long to the coin. This was flattened, the coin flipped, the flip timed and then we unwrapped the ribbon to see how often the coin turned over. Based on these experiments, a typical coin turns at 35-40 revolutions per second. Since a flip lasts half a second, typical coin flips rotate between 17 and 20 times.

This is not very much variability and practiced magicians (including the present author) can control coins precisely. Indeed, my colleagues at the Harvard Physics Department built me a perfect coin flipper that comes up heads every time. Most human flippers do not have this kind of control and are in the range of 5½ mph and 35-40 r.p.s. Where is this on Figure 2? In the units of Figure 2, the velocity is about 1/5 – very close to the zero. However, the spin coordinate is about 40 – way off the picture. Thus the picture says nothing about real flips. However, the math behind the picture determines how close the regions are in the appropriate zone. Using this and the observed spread of the measured data allows us to conclude that coin tossing is fair to two decimals but not to three. That is, typical flips show biases such as .495 or .503.

Blending Subjective Probability and Physics. Our refined analysis can be blended into the probability specification. Now, instead of observing heads and tails at each flip we observe velocity-spin pairs. If these are judged exchangeable, a version of de Finetti's Theorem applies to show that any coherent probability assignment must be a mixture of independent and identically distributed assignments:

$$P((v,w) \text{ in } A, \ldots, (v', w') \text{ in } B) = \int F(A) \ldots F(B)\, \mu\,(dF)$$

The meaning of these symbols is slightly frightening, even to a mathematical grown up. On the right, F is a probability distribution on the velocity spin plane. Thus μ is a probability on the space of all probabilities. Here, de Finetti's Theorem tells us that thinking about successive flips is the same as thinking about measures for measures. There *is* a set of tools for doing this but at the present state of development it is a difficult task. It is even dangerous. The space of all probability measures is infinite dimensional. Our finite dimensional intuitions break down and hardened professionals have suggested priors with the following property: As more and more data come in we become surer and surer of the wrong answer. This occurred in the age-old problem of estimating the size of an object based on a series of repeated measurements. Classically, everyone uses the average. This is based on assuming that the measurement errors follow the bell shaped curve. Owning up to not knowing the distribution of the errors, some statisticians put a prior distribution on this unknown distribution. The corresponding posterior distribution became more and more tightly perked about the wrong answer as more and more data came in. A survey of these problems and available remedies can be found in my joint work with David Freedman in the *Annals of Statistics* (1986).

What's the Point? This has been a lengthy example aimed at making the following point. Starting with the simple problem of predicting binary outcomes and then thinking about the underlying physics and dynamics we were led from de Finetti's original, satisfactory solution to talking close to nonsense. The analysis led to introspecting about opinions on which we have we have small hold and to a focus on technical issues far from the original problem. I hope the details of the example do not obscure what I regard as its nearly universal quality. In every walk of academic and more practical study we can find simple examples that on introspection grow to unspeakable 'creatures'. The technical details take over and practitioners are fooled into thinking they are doing serious work. Contact with the original problem has been lost.

I am really troubled by the coin tossing example. It shouldn't be that thinking carefully about a problem, adding carefully collected outside data, Newtonian mechanics and some detailed calculations should make a mess of things.

Thinking About Thinking Too Much

The problem of thinking too much has a prominent place in the age-old debate between theory and practice. Galen's second century attempts to balance between rationalist and empiricist physicians rings true today. A convenient reference is R. Walzer and M. Frode's translation of Galen's three treatises on the nature of science. They translate an opponent of the new theories as claiming "there was a simple way in which mankind actually had made enormous progress in medicine over the ages men had learned from dire experience, by trial and error, what was conducive and what was detrimental to health. Not only did he claim that one should not abandon this simple method in favor of fanciful philosophical theories, which do not lead anywhere, he also argued that good doctors in practice relied on this experience anyway since their theories were too vague and too general to guide their practice". In my own field of statistics, the rationalists are called decision theorists and the empiricists are called exploratory data analysts. The modern debaters make many of the same moves that Galen chronicled.

Economists use Simon's ideas of 'satisficing' and 'bounded rationality' along with more theoretical tools associated with Haryshani's 'value of information'. Psychologists such as Kahneman and Tversky accept the value of the heuristics that we use when we abandon calculation and go with our gut. They have created theories of framing and support that allow adjustment for the inevitable biases. These give a framework for balancing the decision to keep thinking versus getting on with deciding.

Computer science explicitly recognizes the limits of thinking through ideas like N-P completeness. For some tasks, computationally feasible algorithms that can be proved to do reasonably well. Here is a simple example. Suppose you want to pack two suitcases with 'things' of weight A,B, ..., Z. You want to pack them as close to evenly as you can. It can be shown that this is a virtually impossible problem (technically #-P complete). Despite 50 years of effort, we don't know how to find the best packing save for trying all of the exponentially many possibilities. Any progress would give solution to thousands of other intractable problems. Most of us conclude the optimum solution is impossible to find. Undeterred, my friend Ron Graham proposed the following: sort the things from heaviest to lightest (this is quick to do). Then fill the two suitcases by beginning with the heaviest, and each time placing the next thing into the lighter suitcase. Here is an example with five things of weight 3, 3, 2, 2, 2. The algorithm builds up two groups as follows

$$3, \qquad 3/3, \qquad \begin{matrix}2\\3\end{matrix} / 3, \qquad \begin{matrix}2\\3\end{matrix} / \begin{matrix}2\\3\end{matrix}, \qquad \begin{matrix}2\\2\\3\end{matrix} / \begin{matrix}2\\3\end{matrix}$$

this misses the perfect solution, which puts 3,3 in one pile and 2,2,2 in the other. One measure of goodness of a proposed solution is the ratio of the size of the larger pile to the size of the larger pile in the optimal solution. This is 7/6 in the example. Graham proved that in any problem, no matter what the size of the numbers, this 'greedy' heuristic always does at worst 7/6 compared to the optimal. We would be lucky to do as well in more realistic problems.

An agglomeration of economics, psychology, decision theory and a bit of complexity theory is the current dominant paradigm. It advises roughly quantifying our uncertainty, costs and benefits (utility) and choosing the course that maximizes expected utility per unit time. A lively account can be found in I.J. Good's Book, Good Thinking (Don't miss his essay on 'How Rational Should a Manager Be?').

To be honest, the academic discussion doesn't shine much light on the practical problem. As an illustration, some years ago I was trying to decide between moving to Harvard from Stanford. I had bored friends silly with endless discussion. Finally one of them said "you're one of our leading decision theorists. Maybe you should make a list of the costs and benefits and try to roughly calculate your expected utility". Without thinking, I blurted out "come on Sandy, this is serious".

Some Rules of Thumb One of the most useful things to come out of my study is a collection of rules of thumb that my friends use in their decision-making. For example, one of my Ph.D. advisors, Fred Mosteller told me 'other things being equal, finish the job that nearest done'. A famous physicist offered this advice 'don't waste time on obscure fine points that rarely occur'. I've been told that Albert Einstein had the following displayed in his office 'things that are difficult to do, are being done from the wrong centers and are not worth doing'. The decision theorist I.J. Good offers 'the older we become, the more important it is to use what we know rather than learn more'. Galen offered this 'if a lot of smart people have thought about a problem (e.g., God's existence, life on other planets) and disagree, then it can't be decided'.

There are many ways we avoid thinking. I've often been offered the algorithm 'ask your wife to decide' (but never 'ask your husband') one of my most endearing memories of the great psychologist of decision-making under uncertainty Amos Tversky is his way of ordering in restaurants: "Barbara? What do I want?"

Clearly, we have experience gathered over millennia coded into our gut responses. Surely we all hope to call on this. A rule of thumb in this direction is 'trust your gut reaction when dealing with natural tasks such as raising children'.

It's a fascinating insight into the problem of thinking too much that these rules of thumb seem more useful than the conclusions drawn from more theoretical attacks.

In retrospect, I think I should have followed my friend's advice and made a list of costs and benefits. If only so that I could tap into what I was really after along the lines of the following Grook by Piet Hein

A Psychological Tip

Whenever you're called on to make up your mind,
and you're hampered by not having any,
the best way to solve the dilemma, you'll find,
is simply by spinning a penny.
No – not so that chance shall decide the affair
While you're passively standing there moping;
but the moment the penny is up in the air,
you suddenly know what you're hoping.

MYSTERIES OF CARDANO THE PROBABILIST

PERSI DIACONIS

The mystery is this: Cardano wrote the first book on probability, a book that sets the vagaries of chance on a firm mathematical footing. His work was done about 100 years before the celebrated work on Pascal and Fermat[0]. Why doesn't he get credit for the invention of probability?

It is not that Cardano's work is unknown. All standard histories of probability devote some space to Cardano's *Liber de Ludo Aleae*[1]. An English translation with beautifully written commentary by Oysten Ore is widely available[2]. Further, recent years have produced several detailed, scholarly commentaries[3]. Yet, Cardano's work is largely thought of as a crude beginning, without real merit.

As documentation, consider the following. I write as a professor at one of the; great departments of probability and statistics. I took an informal poll of our faculty. All knew Cardano's name and all "knew" that he had somehow gotten it all wrong. My own "knowledge" was more embarrassing. To put things in context, I first came to Cardano through my interest in magic and gambling. His writing contains some of the earliest descriptions of card magic; remarkable for their detail, the tricks described sound like the performance of a 21st century sleight of hand expert. Cardano's descriptions of shaped dice, marked cards and the problems of Kibitzer's signaling in card games also have the ring of truth and modernity. As I grew into the mathematical parts of the subject, I have run year-long seminars on the history of probability and written about history professionally. If anyone should have an informed opinion about Cardano, it should be me! Yet, in reporting on the present project to colleagues I explicitly threw in 'Of course, Cardano got it wrong!'

Lest the reader have my doubts, let me assert that:

Cardano was a remarkable, original and incisive creator of the mathematical foundations of probability.

Having framed the mystery let me sketch its structure. Cardano's character is questionable; he is associated with gambling and cheating. He made his living in part by casting horoscopes and as the author of several popular books on astrology[4]. He 'stole' the mathematical solution of cubic equations from others. At a distance, it seems possible that he was just a hack who never did much of substance. A closer look reveals a much richer portrait and Cardano comes shining through. I will unravel the mystery by discussing Cardano's mathematics and then his probability theory.

CARDANO'S MATHEMATICS

The high point of Cardano's mathematical work is his book *Ars Magna*[5]. This great book is an important milestone in the mathematical tradition beginning with the Greek solution of algebraic equations as exposited by Diophantus and culminating in the work of number theorists in the complete solution of 'Fermat's Last Problem'. Before Cardano, there are only accounts by the Greek mathematician Diophantus and the 9th Century Persian mathematician Al-Khwarizmi. These authors collected a hodgepodge of facts and problems. By contrast, Cardano sees a big picture and puts some order into things. One of his greatest contributions is the first introduction of complex numbers as a crucial tool for understanding and unifying. A second highlight is the first publication of rules for solving cubic and quartic equations.

I will not detail Cardano's mathematical work here except to say that a modern mathematician reading his insightful development will have no doubt that Cardano was a serious mathematician. One point deserves comment. Did Cardano steal someone else's work? The work in question is the general solution of the cubic equation (e.g., find the roots of $x^3 - 13x + 7 = 0$). Depending on the exact form of the equation, 13 cases rose. Cardano (and before him Fiore) could solve several of these cases but many defeated him. Tartaglia, a contemporary of Cardano, managed to solve general cubic equations. Tartaglia communicated his methods to Cardano and Cardano published them in the *Ars Magna*. Cardano gives full credit to Tartaglia, clearly and indeed repeatedly; thus, he did not steal Tartaglia's method in any usual sense. What is in dispute is this: Tartaglia claimed he did not want Cardano to publish the method while Cardano claimed he had permission. The matter has been much discussed and there is merit on both sides of the argument. It does not seem that Cardano's publication should cast any shadow on him as a mathematician.

CARDANO'S PROBABILITY THEORY

The details of Cardano's *Liber de Ludo Aleae* are wonderfully exposited by Oysten Ore. It may help current readers to know a bit more about Ore. He was a distinguished mathematician who spent his career at Yale University working in Algebra and Combinatorial theory. He had a keen interest in history. In addition to his works on Cardano, he wrote on the Norwegian mathematician Niels Abel on the history of Number Theory, and produced a scholarly introduction to Cardano's *Ars Magna*. His style is somewhat popular, very readable, but always truly scholarly. Anyone wishing to understand Cardano must read Ore.

There are many surprises to be found in Ore's portrait of Cardano and his times. Here is an example, viewed in modern light. One of the most interesting things I have ever seen was a high-stakes poker game played over the years 1959 to 1991 or so. The players were world-renowned Bridge experts such as John Crawford and Oswald Jacobi along with a variety of lawyers, manufacturers, and professional gamblers. The stakes were enormous: it was not unusual for a player to win (or lose) tens of thousands of dollars in one evening. The game was table stakes and early on I saw something shocking--two players were left in the pot and one was 'all in', thus there were no further bets. A single round (one card per player) was left to deal. One of the players turned up his cards and said 'I have three sixes, what do you have?' The second player had a four flush and four to an open ended straight. 'What should we do?' asked the first. The two players roughly and rapidly calculated their odds and split the pot in proportion. I was shocked, thinking 'Real men take their chances'. On another occasion, there were three players left, no more betting and two cards left to go in a seven-card stud game. After brief discussion, it became apparent that the odds were too complex to compute in a convincing manner. They divided the pot into four equal parts and dealt a round of two cards for each of the four parts. Each quarter was separately decided. On inquiry, the players allowed that the actual money pressure had no effect on them and it seemed reasonable to take their equity.

At the time, I marveled at their enlightened attitude. I now find much of this in Section 16 of Cardano. Settlements were well known in Cardano's time, at least in the card game of Primero, an early form of poker.

Cardano mentions uneven settlements based on the odds (not just splitting the pot) and cautions that the odds should be computed taking account of cards known to be out of play. I have not found the "deal it several times" variant recorded. This Monte-Carlo approach could have set probability theory well back had it been applied in lieu of mathematics to the problem of points[6].

Popular textbook accounts of early probability often say that the subject started with the Fermat-Pascal correspondence about a dice bet proposed by the Chevaler deMere. Ore wrote a little known history of deMére[7], which expands on the account in his book about Cardano. The dice problem begins by asking how many rolls of two fair dice are needed to have even odds of two sixes coming up at least on one roll. Fermat solved this problem carefully (Answer: 24 rolls give chance about $.4914$ while 25 rolls give chance about $.5055$). His solution uses the basic multiplication rule to determine that the chance of no double six in n rolls is $35/36$ raised to the n^{th} power. Ore points out that Cardano had solved precisely this problem in much the same language in *Liber de Ludo Aleae*, and suggests calling the multiplication rule Cardano's Rule. Ore does a marvelous scholarly task of going through Cardano's probability work one example at a time, clarifying often misunderstood points in loving detail. I largely agree with his analyses and do riot want to repeat them here. I want to make two points not made by Ore.

One reason Cardano's work may not be credited is that it contains some substantial errors. Indeed, the very first calculation reported seems wrong and the error is repeated throughout the text. In Section 9, when Cardano discusses the cast of one die, he writes:

...Thus the chances are equal that a given point will turn up in three throws...

This is wrong. The chance that a given point will turn up in three throws is $1-\left(\frac{5}{6}\right)^3 = .4213$, which is substantially less than $1/2$. Cardano reasons based on the idea that a die has six faces: in six casts each point should turn up once. Sometimes he gets the right answer this way. For example, later in Section 9 he writes, 'that chances are equal that one of three given points will turn up in one throw' (correct). Ore called this argument 'Reasoning on the mean' and gives much further analysis. He does not say it is wrong as clearly as I just have.

A second error seems more serious. In addition to dice, Cardano frequently analyzes games with very irregular dice-like objects--knucklebones or astragal. When tossed these can come up one of four faces. Cardano calculates various chances for multiple rolls <u>assuming that the four faces are equally likely</u>. This last assumption is patently false. The statistician F.N. David reported rough empirical data showing that two of the faces came up about $1/10$ of the time and two of faces each came up about $4/10$ of the time. It's hard to see how a practical gambler like Cardano would miss this point. However, particularly in his Section 31, he calculates using a symmetry analysis[8].

It is important to make three points. First, Cardano has a systematic way of calculating probabilities for equally likely basic events. He clearly lays out what is nowadays regarded as the basic way of calculating: calculate the total number of possibilities and the number of these that are favorable. This ratio is taken as the probability. Cardano carries out this program in fairly complex circumstances and usually gets the correct answer. He also carries out the program for card games. Ore details many of these calculations in modern notation. They are often sophisticated and would defeat many students in a beginning probability course while Cardano gets the right answer.

A second point: Cardano may have meant something that we do not now understand. For example, his 'the chances are equal that a given point will turn up in three throws' may mean that the chance that the point one will turn up in three throws is the same as the chance that the point two will turn up in three throws, and so on. Alternatively, we can take his circuit (the numbers one to six in some random order without repeats) and ask for the chance of a specific point in the first three rolls. The answer is $1/2$. While I think Cardano erred, there are many examples of remarkable calculations being misunderstood as wrong by later generations[9]. To further these difficulties, I work from Gould's English translation. For the point under discussion, the original text reads "tum dimidum semper numeri est aequalitas, ut in tribus iactidus punctum eveniat; nam in sex completur revolutio, aut tria puncta uno lactu." Where I have been able to check, I find Gould's translation good.

The third point: it must be remembered that Cardano's manuscript was published posthumously from a fragment of a larger work. There are many internal signs that what was published were rough notes from work in progress. Cardano corrects some earlier errors later in the text. It seems unfair to charge him for errors when so much is completely new and correct under these circumstances.

My current conclusion is that Cardano's work is amazing! A brilliant, incisive treatment of basic probability, completely original. There are flaws but they are minor and it may well be that later scholarship will explain them. As the rich literature on Cardano confirms, he is a complex character; but among his many successes must be counted the invention of the basic approach and rules of calculation for dealing with probability as it appears in games of chance.

NOTES

0. There is plausible evidence that Cardano wrote a first version of the book in the 1540's (Ore pg. 120-121) and that this was revised to its present form in about 1570 (Ore pg. 122). The careful biography of M. Zollinger (1996). *Bibliographie der Spielbücher des 15. bis 18. Jahrunderts*, Hiersemann, Stuttgart gives 1564 as a plausible date.
1. There are three respected histories of probability. Isaac Todhunter's *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace*, London MacMillian; F.N. Davis *Gods, Games and Gambling*, Griffin, London (1962), Anders Hald *A History of Probability and Statistics and their applications before 1750*, Wiley, New York 1990; all begin with a chapter on Cardano's work. Of course, there is a long, non quantitative prehistory of probabilistic thinking. For entry this literature see James Franklin's *The Scene of Conjecture; Evidence and Probability Before Pascal*, Johns Hopkins Press, Baltimore.
2. Ore, Oysten, *Cardano, The Gambling Scholar*. Princeton University Press. Princeton, NJ (1953). Also available as republished by Dover, New York {1965). The translation of *Liber de Ludo Aleae* by Gould was separately published as *The book on games of chance*, Holt, Rinehart, Winston, New York (1961). I must admit to an unusually close connection to these works. Ore was the advisor of one of my advisors {Andrew Gleason) and the Holt edition has a useful introduction by S.S. Wilks, who was the Ph.D. advisor of my second advisor (Fredrick Mosteller).
3. Tilman Hrischer: Interpretationen zum *Liber de Ludo Aleae*. In Girolamo Cardano: Phiosoph, Naturforscher, Artz. Edited by Eckhard Kessler, Harassowitz. Verlag, Wiesbaden (1994) pg 207-218. Massimo Tabborini: Matematica; Tempo e previsione nel Liber de Ludo Aleae. In Girolamo Cardano, *Le Opere, Le Fonti, La Vita*, Edited by M. Baldi and G. Canziani, Francoangeli, Milan (1997) pg. 227-271.

4. It is difficult for a modern reader to parse astrology. I have been slightly lenient towards early astrology and alchemy, along the lines that astrology begat astronomy and alchemy begat chemistry. Wayne Shumaker (1972) *The Occult Sciences in the Renaissance*, University of California Press, Berkeley argues that this view is off--there were well known early books pointing out the differences between careful scientific study of the stars and astrology. Writers from Ciero to Leonardo pointed to differences between twins and to the long recurrence time between astronomical configurations blocking observational confirmation.

   A definitive study of Cardano through his astrological work appears in Anthony Grafton's (1999) *Cardano's Cosmos: The Worlds and Works of a Renaissance Astrologer*, Harvard University Press, Cambridge. Grafton shows that Cardano earned a substantial part of his living by casting horoscopes for wealthy clients. Cardano wrote a best seller, which contained the first 'celebrity horoscopes' and was not above changing his readings from one edition to the next if a prediction went awry. All of this is pretty seedy stuff and leads to a picture of Cardano as a survivor who would do most anything to keep going. I find the contrast between Cardano's astrology and his probability striking. The cosmos is a complex system and Cardano worked with flexible ad hoc rules; he could (and did) find excuses for failures and point mainly to successes. Rolled dice and shuffled cards are also complex systems and it would be easy to follow the Greeks and others in making up stories and ad hoc rules to account for the vagaries of chance. Instead, Cardano went to the heart of things and invented probability theory.

5. G. Cardano, *Artis Magne Sive de Realis Algebraicis*, (1545, 1572) A good modern translation of the *Ars Magna* with a useful historical introduction by Oysten Ore is T. Richard Willmer (Ed.) (1968) *Cardano the Great Art*, MIT Press, Cambridge. To understand Cardano's achievement and the problems he overcame I strongly recommend Barry Mazur's

forthcoming *Picturing Numbers, Especially* $\sqrt{-15}$. The title refers to Cardano's celebrated coming to terms with imaginary numbers such as $\sqrt{-15}$.

6. To complete my record of modern methods of settling let me add that I have seen players agree to split the pot in casino poker games before the last cards were dealt. Sometimes players will split the pot into pieces, gamble on part of it and settle on part of it. In professionally run poker games, I have seen the house offer 'insurance' that the better hand will not be outdrawn on the last card.

7. Ore, O (1959). *Pascal and the Invention of Probability Theory*, The Colorado College Studies, Spring 1959 Number 3, pg. 11-24. This paper suggests a novel explanation for deMére's 'scandal' over the dice problem based on an inaccuracy in the classical expansion of $\log(1-x)$ when x is not sufficiently small.

8. A correct analysis using four parameters for the face probabilities is tedious but elementary. See K.G. Hagstroem (1932). *Les Preludes Antiques des Terorie de Probabilty*, Stockholm pg.48-51.

9. One favorite example is discussed by Richard Stanley (1997), Hipparchus, Plutarch, Schroder, and Hough. American Mathematical Monthly, 104 344-350. Stanley quotes from Plutarch's 'Table Talk':

   Chrysippus says that the number of compound propositions that can be made from only ten simple propositions exceeds a million (Hipparchus, to be sure, refuted this by showing that on the affirmative side there are $103{,}049$ compound statements, and on the negative side $310{,}952$).

   Plutarch wrote around 100 A.D.. Mathematicians from then to now have been trying to figure out if there was any sense in either of Hipparchus's numbers. In January 1994, half the problem was solved by David Hough. The number $103{,}049$ is the number of distinct ways of bracketing a string of ten letters. By example, there are eleven ways of bracketing four letters, namely

$$abcd, (ab)cd, a(bc)d, ab(cd), (abc)d, a(bcd), ((ab)c)d, (a(bc))d, (ab)(cd), a((bc)d), a(b(cd)).$$

The point of mentioning this here is that early mathematicians had some expertise in solving certain combinatorial problems and many years may be required to sort things out once the secret has been lost.