THE MATHEMATICS OF GAMBLING

By
Dr. Edward O. Thorp

Table of Contents

About the Author

Edward Thorp is adjunct professor of finance and mathematics at the University of California at Irvine, where he has taught courses in finance, probability and functional analysis. He previously taught at the University of California at Los Angeles, the Massachusetts Institute of Technology and New Mexico State University.

Thorp's interest in gambling dates back almost 30 years, while he was still in graduate school at UCLA. It was here that he first formulated his dream of making money from the development of a scientifically-based winning gambling system. His first subject of study was the roulette wheel, which offered him the opportunity to use modern physics to predict the resting place of the ball.

With the roulette work unfinished, Thorp's attention was diverted by the blackjack work of Baldwin, Cantey, Maisel and McDermott. He set to work on this new problem. With the aid of a computer, Thorp developed the basic strategy and the five-count, ten-count and ultimate counting strategies. He used these methods with success in the Nevada casinos. The work was first publicized in a scientific journal and saw broad public exposure in the 1962 book *Beat the Dealer*. The book underwent a revision in 1966 and it is still regarded as the classic early work in the "blackjack revolution" which continues to this day.

In the late 1960s, Thorp developed with Sheen Kassouf a successful method for stock market investing involving warrants that proved so profitable that Thorp turned $40,000 into $100,000 in two years. The strategy was published in *Beat the Market* in 1967. Additionally, using this strategy and further refinements, Thorp manages a multi-million-dollar investment portfolio. He is President of Oakley Sutton Management Corp. and Chairman of the the Board of Oakley Sutton Securities Corp.

Thorp has continued to advance new theories for gambling and other games, as well as the stock market.

Section One

Card Games

Casino card games such as baccarat and blackjack differ significantly from casino games such as craps, roulette, and slot machines in that they are not independent trial processes-that is, the cards that already have been played do affect the odds on subsequent hands.

Consider for a moment the game of blackjack, where the cards used on a round are put aside and successive rounds are dealt from an increasingly depleted pack. The cards are reshuffled before a round if the remaining unused cards would be insufficient to complete a round or earlier, usually at the casino's discretion. What the early research on blackjack (contained in *Beat the Dealer*) showed and what has been confirmed repeatedly in the intervening 23 years is that the end pack provides favorable situations often enough to give the player an overall advantage.

While it is foolish to keep a record of past decisions at craps in order to determine which numbers are "hot" or "cold" (the dice have no memory), an ability to keep track of which cards have been played and knowledge of their relationship to the player's expectation can be beneficial, as long as the cards are not reshuffled after every hand.

The ability to keep track of the cards played does not alone guarantee gambling success at a particular game. Indeed, one of the chief tasks of this section will be to examine the usefulness of card counting strategies in baccarat, considering the bets offered and the nature of the game.

In Chapter 2, we will comment on blackjack systems, as well as statistical methods useful in detecting casino cheating. The latter subject is important to those who play the game seriously, because cheating incidents can erode any small edge the player may gain through the use of basic strategy and card counting.

Chapter 1

Introductory Statement

The casino patron who decides to "try his luck" at the tables and the horse player who wagers at the racetrack confront what seem to be formidable adversaries. The casinos hope to have the advantage on every bet offered and, at the track, the pari-mutuel takeout of 17-25% on every bet assures all but a few will wind up losers.

As soon as he enters the casino, the player must make several important decisions, the first being: What game do I play? Even after this choice is made, most games offer additional options: Do I play individual numbers or the even-money bets in roulette? Do I stand with a pair of eights in blackjack or should I hit or split the pair? Should I bet pass or the one-roll propositions in craps?

The horse player is offered a number of choices as well. He is usually faced with a field of six to 12 horses. He can play one or more horses to win, place, or show, in addition to combining any number of horses in the exotic or "gimmick" wagers.

All of these choices have "right" answers, if the player seeks to maximize his return or minimize his loss. They all can be at least partially solved through the use of mathematical theory. The intelligent player must have a basic understanding of the mathematics behind the game or games he plays if he is to survive financially or actually profit. There *are* situations where the player has the advantage. The most-publicized example, of course, is casino blackjack. The game has become tougher in recent years due to casino countermeasures, but blackjack can still be profitable for the sophisticated player. There could be several other favorable games, as the reader will soon discover.

A familiarity with basic probability will allow the alert gambler to discover those positive expectancy games and exploit them where they exist. A vast knowledge of mathematics is not required. Some of the finest poker players in the country never went to college, but they do have a sense of what makes a good poker hand and what their chances of having the best hand are after all the cards have been dealt.

Mathematical Expectation

I have already made reference to the concept of mathematical expectation. This principle is central to an understanding of the chapters to follow.

Imagine for a moment a coin toss game with an unbiased coin (a coin we assume will produce 50% heads and 50% tails). Suppose also that we are offered an opportunity to bet that the next flip will be heads and the payoff will be even money when we win (we receive a \$1 profit in addition to the return of wager). Our mathematical expectation in this example is:

$$(.5)(1) + (.5)(-1) = 0$$

The *mathematical expectation* of any bet in any game is computed by multiplying each possible gain or loss by the probability of that gain or loss, then adding the two figures. In the preceding example, we expect to gain nothing from playing this game. This is known as a *fair game*, one in which the player has no advantage or disadvantage.

Now suppose the payoff was changed to 3/2 (a gain of \$1.50 in addition to our \$1 bet). Our expectation would change to:

$$(.5)(1.50) + (.5)(-1) = +.25$$

Playing this game 100 times would give us a positive expectation of \$25.

The two examples presented thus far are admittedly simple, but often this type of analysis is all that is needed to evaluate a proposition. Consider the "dozens" bet in roulette. Our expectation for a \$1 bet is:

$$(12/38)(2) + (26/38)(-1) = -.0526$$

As another example, suppose that on the first hand of four-deck blackjack the player bets \$12, he is dealt 6,5, and the dealer then shows an ace up. The dealer asks the player if he wants insurance. This is a separate \$6 bet. It pays \$12 if the dealer's hole card is a ten-value. It pays −\$6 otherwise. A full four-deck pack has 64 tens and 144 non-tens. Assuming the deck is "randomly" shuffled (this means that all orderings of the cards are equally probable), the chances are equally likely that each of the 205 unseen cards is the dealer's hole card. Thus the player's expectation is:

$$(64/205)(12) + (141/205)(-6) = -78/205$$

or about −\$.38. The player should not take insurance.

Different betting amounts have different expectations. But the player's expectation as a *percent of the amount bet* is always the same number. In the case of betting on the Red in roulette, this is 18/38 −20/38 = −2/38 = −1/19 or about −5.26%. Thus, the expectation of any size bet on Red at American double-zero roulette is −1/19 or about −5.26% of the total amount bet. So to get the expectation for any size bet on Red, just multiply by −5.26%. With one exception, the other American double-zero roulette bets also have this expectation per unit bet. The player's expectation per unit is often simply called the player's disadvantage. What the player loses, the house wins, so the *house advantage, house percentage,* or house expectation per unit bet by the player is +5.26%.

A useful basic fact about the player's expectation is this: the expectation for a series of bets is the total of the expectations for the individual bets. For instance, if you bet $1 on Red, then $2, then $4, your expectations are −$2/38, −$4/38, and −$8/38. Your total expectation is −$14/38 or (a loss of) about −$.37. Thus, if your expectation on each of a series of bets is −5.26% of the amount bet, then the expectation on the whole series is −5.26% of the total of all bets. This is one of the fundamental reasons why "staking systems" don't work: a series of negative expectation bets *must* have negative expectation.

Repeated Trials

Expectation is the amount you *tend* to gain or lose on average when you bet. It, however, does not explain the fluctuations from expectation that occur in actual trials.

Consider the fair game example mentioned earlier in the chapter. In a series of any length, we have an expectation of 0. In any such series it is possible to be ahead or behind. Your total profit or loss can be shown to have an average deviation from expectation of about $\sqrt{N}$. Let D = T −E be the difference of deviation between what you actually gain or lose (T), and the expected gain or loss (E). Therefore, for 100 bets, the average deviation from E = 0 is about $10 (in fact, the chances are about 68% that you'll be within $10 of even; they're about 96% that you'll be within $20 of even). For ten thousand $1 bets it's about $100 and for a million $1 bets it's about $1,000. Table 2-1 shows what happens. For instance, the last line of Table 1-1 says that if we match coins one million times at $1 per bet, our expected gain or loss is zero (a "fair" game). But on average, we'll be about $1,000 ahead or behind. In fact, we'll be between +$1,000 and −$1,000 about 68% of the time. (For a million $1 bets, the deviation D has approximately a normal probability distribution with mean zero and standard deviation $1,000.) We call the total of the bets in a series the "action," A. For one series of one million $1 bets, the action is $1,000,000. However (fifth column) D/A=0.001, so

Table 1-1

number of bets N (A = $N)	expected gain E	average size of D is about $\sqrt{N}$	about 68% of time T is between	average size of D/A	about 68% of time T/A is between
100	0	10	10 and - 10	0.1	0.1 and -0.1
10,000	0	100	100 and - 100	0.01	0.01 and -0.01
1,000,000	0	1000	1000 and -1000	0.001	0.001 and -0.001

the deviation as a percent of the action is very small. And about 68% of the time T/A is between −.001 and +.001 so as a percent of the action the result is very near the expected result of zero. Note that the average size of D, the deviation from the expected result E, grows—contrary to popular belief. However, the average size of the percentage of deviation, D/A, tends to zero, in agreement with a correct version of the "law of averages."

For $1 bets on Red at American roulette, the corresponding results appear in Table 1-2. Notice that in the last column the spread in T/A gets closer and closer to E/A = −.0526. This is where we get the statement that if you play a "long time" you'll lose about 5.26% *of the total action.* Note, too, in column 4 that there appears to be less and less chance of being ahead as the number of trials goes on. In fact, it can be shown that in all negative expectation games the chance of being ahead tends to zero as play continues.

Using the concept of *action*, we can now understand the famous "law of averages." This says, roughly, that if you make a *long* series of bets and record both the action (A) and your total profit or loss (T), then the *fraction* T/A is approximately the same as the *fraction* E/A where E is the total of the *expected* gain or loss for each bet. Many people misunderstand this "law." They think that it says the E and T are approximately the same after a long series of bets. This is false. In fact, the difference between E and T tends to get larger as A gets bigger.

Now, the ordinary player probably won't make a million $1 bets. But the casino probably will see that many and more. From the casino's point of view, it doesn't matter whether one player makes all the bets or whether a series of players does. In either case, its profit in the long run is assured and will be very close to 5.26% of the action. With many players, each making some of the 1,000,000 bets, some may be lucky and win, but these will generally be compensated for by others who lose more than the expected amount. For instance, if each of 10,000 players take turns making a hundred $1 bets, Table 1-2 tells us that about 68% of the time their result will be between +$4.74 and −$15.26.

Table 1-2

N (A = $N)	E	D ~ $\sqrt{N}$	about 68% of time T is between	D/A ~ $1/\sqrt{N}$	about 68% of time T/A is between
100	- $5.26	10	+ $4.74 and - $15.26	0.1	+.0474 and -.1526
10,000	- $526.00	100	- $426.00 and - $626.00	0.01	-.0426 and -.0626
1,000,000	-$52631.00	1,000	-$51631.00 and -$53631.00	0.001	-.0516 and -.0536

About 16% of the time the player wins more than \$4.74 ("lucky") and about 16% of the time the player loses more than \$15.26 ("unlucky"). But players cannot predict or control which group they'll be in.

This same "law of averages" applies to more complicated sequences of bets. For instance, suppose you bet \$10 on Red at roulette ($E = -.53$), then bet \$100 on "players" at Baccarat ($E = -\$1.06$), then bet \$10 on a hand in a single-deck blackjack game where the ten-count is 15 tens, 15 others ($E = +\$.90$). The total E is $\$.53 - \$1.06 + \$.90 = -\$.69$. The total A is $\$10 + \$100 + \$10 = \$120$. If you make a long series of bets and record E and A as well as your gains and losses for each one, then just as in the coin matching example (Table 1-1) and the roulette example (Table 1-2), the fraction D/A tends to zero so T/A tends to E/A. That means that over, say, a lifetime, your total losses as a *percent* of your total action will tend to be very close to your total expectation as a percent of your action.

If you want a good gambling life, make positive expectation bets. You can, as a first approximation, think of each negative expectation bet as charging your account with a tax in the amount of the expectation. Conversely, each positive expectation bet might be thought of as crediting your account with a profit in the amount of the expectation. If you only pay tax, you go broke. If you only collect credits, you get rich.

Chapter 2

Blackjack

Blackjack, or twenty-one, is a card game played throughout the world. The casinos in the United States currently realize an annual net profit of roughly one billion dollars from the game. Taking a price/earnings ratio of 15 as typical for present day common stocks, the United States blackjack operation might be compared to a \$15 billion corporation.

To begin the game a dealer randomly shuffles the cards and players place their bets. The number of decks does not materially affect our discussion. It generally is one, two, four, six or eight. There are a maximum and minimum allowed bet.

The players' hands are dealt after they have placed their bets. Each player then uses skill in his choice of a strategy for improving his hand. Finally, the dealer plays out his hand according to a fixed strategy which does not allow skill, and bets are settled. In the case where play begins from one complete randomly shuffled deck, an approximate best strategy (i.e., one giving greatest expected return) was first given in 1956 by Baldwin, Cantey, Maisel, and McDermott.

Though the rules of blackjack vary slightly, the player following the Baldwin group strategy typically has the tiny edge of +.10%. (The pessimistic figure of −.62% cited in the Baldwin's group's work was erroneous and may have discouraged the authors from further analysis.) These mathematical results were in sharp contrast to the earlier and very different intuitive strategies generally recommended by card experts, and the associated player disadvantage of two or three percent. We call the best strategy against a complete deck the *basic strategy*. Determined in 1965, it is almost identical with the Baldwin group strategy and it gives the player an edge of +.13 against one deck and −.53 against four decks.

If the game were always dealt from a complete shuffled deck, we would have repeated independent trials. But for compelling practical reasons, the deck is not generally reshuffled after each round of play. Thus as successive rounds are played from a given deck, we have sampling without replacement and dependent trials. It is necessary to show the players most or all of the cards used on a given round of play before they place their bets for the next round. Knowing that certain cards are missing from the pack, the player can, in principle, repeatedly recalculate his optimal strategy and his corresponding expectation. (The strategies for various card counting procedures, and their expectations, were determined directly from probability theory with the aid of computers. The results were reverified by independent Monte Carlo calculations.)

Blackjack Systems

All practical winning strategies for the casino blackjack player, beginning with my original work in 1961, are based on this knowledge of the changing composition of the deck. In practice each card is assigned a point value as it is seen. By convention the point value is chosen to be positive if having the card out of the pack significantly favors the player and negative if it significantly favors the casino. The magnitude of the point value reflects the magnitude of the card's effect but is generally chosen to be a small integer for practical purposes. Then the cumulative point count is taken to be proportional to the player's expectation.

To a surprising degree, the player's best strategy and corresponding expectation depend only on the fractions of each type of card currently in the pack and only change slowly with the size of the pack. Thus the better systems "normalize" by dividing the cumulative point count by the total number of as yet unseen cards. Most point count systems are initialized at zero cumulative total for the full pack, and the normalized cumulative count is taken to indicate the change in player expectation from the value for the full pack.

The original point count systems, the prototypes for the many subsequent ones, were my five count, ten count, and "ultimate strategy." An enormous amount of effort by many investigators has since been expended to improve upon these count systems. Some of these systems are shown in Table 2-1 (courtesy of Julian Braun).

The idea behind these point count systems is to assign point values to each card which are proportional to the observed effects of deleting a "small quantity" of that card. Table 2-2 (courtesy of Julian Braun, private correspondence) shows this for one deck and for four decks, under typical Las Vegas rules. One must compromise between simplicity (small integer values) and accuracy. My "ultimate strategy" is a point count based on moderate integer values which fits quite closely the data available in the early 1960s. Until recently all the other count systems were simplifications of the "ultimate."

System 1 (Table 2-1) does not normalize by the number of remaining cards. Thus the player need only compute and store the cumulative point count. Normalization gives the improved results of system 2, but requires the added effort of computing the number of remaining cards and of dividing the point count by the number of remaining cards when decisions are to be made. In practice the player can estimate the unplayed cards by eye and use it with system 1 and get almost the results of system 2 with much less effort. Systems 2, 3, 5 and 7 all divide by the number of remaining cards.

Table 2-1

Table 2-1, Braun's simulation of various point count systems. "Bet 1 to 4" means that 1 unit was bet except for the most advantageous x% of the situations, when 4 was bet. To compare systems, x was approximately the same in each case, 21(%).

STRATEGY/SYSTEM	RESULTS OF SIMULATED DEALS-PLAYER'S ADVAN. Flat Bet	Bet 1 to 4
1 Basic Braun + −	.2%	1.4%
2 Braun + −	.7%	2.0%
3 Revere Pt. Ct.	.6%	2.1%
4 Revere Adv. + −	.5%	1.6% to 1.8% +
5 Revere Adv. Pt. Ct. − 71	.6%	2.0%
6 Revere Adv. Pt. Ct. − 73	.8%	2.1% to 2.3% +
7 Thorp Ten Count	.7%	1.9%
8 Hi-Opt	.8%	2.1% to 2.3% +

Table 2-1 (continued)

SYSTEM #	BASIC POINT COUNTS CARD 2	3	4	5	6	7	8	9	T	A	+ N?†	SIDE ACE COUNT?	λ(C) 1 DECK	4 DECK
1	1	1	1	1	1	0	0	0	−1	−1	NO	NO	.963	.961
2	1	1	1	1	1	0	0	0	−1	−1	YES	NO	.963	.961
3	1	2	2	2	2	1	0	0	−2	−2	YES	NO	.979	.977
4	1	1	1	1	1	0	0	−1	−1	0	YES	YES	.885 (.968)*	.884 (.968)*
5	2	3	3	4	3	2	0	−1	−3	−4	YES	NO	.994	.994
6	2	2	3	4	2	1	0	−2	−3	0	YES	YES	.924 (.995)*	.922 (.997)*
7	4	4	4	4	4	4	4	4	−9	−4	YES	NO	.720	.709
8	0	1	1	1	1	0	0	0	−1	0	YES	YES	.874 (.938)*	.871 (.939)*

† Normalized.
* Improvement with ace adjustment.

Table 2-2
Changes in Player Expectation by Removing Individual Cards

One deck Top of Deck Expectancy = 0.10%

Cards removed	A	2	3	4	5	6	7	8	9	10	d*	d/13
Expectation (%)	−.48	.50	.56	.70	.86	.56	.41	.12	−.06	−.39		
Change in Expectation (%)	−.58	.40	.46	.60	.76	.46	.31	.02	−.16	−.49	.31	.024
u_i*	−.604	.376	.436	.576	.736	.436	.286	−.004	−.184	−.514		

Four decks Top of Deck Expectancy = −0.532%

Cards removed	A	2	3	4	5	6	7	8	9	10	d*	d/13
Expectation (%)	−1.130	−.147	−.081	.059	.236	−.078	−.239	−.525	−.714	−1.019		
Change in Expectation (%)	−.598	.385	.451	.591	.768	.454	.293	.007	−.182	−.487	.221	.017
u_i*	−.615	.368	.434	.574	.751	.437	.276	−.010	−.199	−.504		

*See Appendix A.

Systems 4, 6 and 8, which are also normalized, have the first new idea. They assign a point count of zero to the ace for strategy purposes. This is consistent with the evidence: in most instances that have been examined, the optimal strategy seems to be relatively unaffected by changes in the fraction of aces in the pack. However, the player's expectation is generally affected by aces more than by any other card (Table 2-2). Therefore these systems keep a separate ace count. Then the deviation of the fraction of aces from the normal 1/13 is incorporated for calculating the player's expectation for betting purposes.*

The (c) column in Table 2-1 still remains to be explained. It is a numerical assessment of a particular system's closeness to an ideal system based on the change in expectation values contained in Table 2-2. The calculation of the (c) value eliminates the necessity of simulating a large number of hands (say a million) to evaluate a strategy. The computation of these numbers requires some advanced mathematical background, so its explanation is left to the appendix.

Cheating: Dealing Seconds

Various card counting systems give the blackjack player an advantage, provided that the cards are well shuffled and that the game is honest. But many methods may be used to cheat the player. I have been victimized by most of the more common techniques and have catalogued them in *Beat the Dealer.*

One of the simplest and most effective ways for a dealer to cheat is to peek at the top card and then deal either that card or the one under it, called the second. A good peek can be invisible to the player. A good second deal, though visible to the player, can be done so quickly and smoothly that the eye generally will not detect it. Although the deal of the second card may sound different from the deal of the first one, the background noise of the casinos usually covers this completely. Peeking and second dealing leave no evidence. Because these methods are widespread, it is worth knowing how powerful they are.

Does even a top professional blackjack counter have a chance

*See Appendix A, pg. 131.

against a dealer who peeks and deals seconds? Consider first the simple case of one player versus a dealer with one deck. This is an extreme example, but it will illustrate the important ideas.

I shuffle the deck and hold it face up in order to deal practice hands. Because I can see the top card at all times, dealing from a face-up deck is equivalent to peeking on each and every card. I will deal either the first or second card, depending on which gives the dealer the greatest chance to win. I will think out loud as an imaginary dealer might, and the principles I use will be listed as they occur. The results for a pass through one deck are listed in Table 2–3 (pp. 20-21). There were nine hands and the dealer won them all.

On hands one, two, four, six, eight and nine, the dealer wins by busting the player. Because there is only one player, it does not matter what cards the dealer draws after the player busts.

When there are two or more players, the dealer may choose a different strategy. If, for example, the dealer wishes to beat all the players but doesn't want to peek very often, an efficient approach is simply to peek when he can on each round of cards until he finds a good card for himself on top. He then retains this card by dealing seconds until he comes to his own hand, at which time he deals the top card to himself. That strategy would lead to the dealer having unusually good hands at the expense of the collective player hands; because some good hands have been shifted from the players, the player hands would be somewhat poorer than average.

A player could detect such cheating by tallying the number of good cards (such as aces and 10s) which are dealt to the dealer as his first two cards and comparing that total with the number of aces and 10s predicted by theory. In Peter Griffin's book, *The Theory of Blackjack*, he describes how he became suspicious after losing against consistently good dealer hands. Griffin writes that he "...embarked on a lengthy observation of the frequency of dealer up cards in the casinos I had suffered most in. The result of my sample, that the dealers had 770 aces or 10s out of 1,820 hands played, was a statistically significant indication of some sort of legerdemain." Griffin's tally is overwhelming evidence

that something was peculiar. The odds against such an excess of ten-value cards and aces going to the dealer in a sample this size are about four in ten thousand.

Another approach the dealer might select is to beat one player at the table while giving everybody else normal cards. To do this, the dealer peeks frequently enough to give himself the option of dealing a first or second to the unfortunate player each time that player's turn to draw a card comes up. Dealing stiffs to a player so that he is likely to bust is, as we see from the chart in Table 2-3, so easy to do that the player has little chance.

If all dealers peeked and dealt seconds according to the cheating strategy indicated in Table 2-3, I estimate that with one player versus the dealer, the dealer would generally win at least 95 percent of the time. With one dealer against several players, the dealer would win approximately 90 percent of the time. Anyone who is interested can get a good indication of what the actual numbers are by dealing a large number of hands and recording the results.

The deadliest way a dealer can cheat is to win just a few extra hands an hour from the players. This approach is effective because it is not extreme enough to attract attention, or to be statistically significant and therefore detectable over a normal playing time of a few hours. For example, the odds in blackjack are fairly close to even for either the dealer or the player to win a typical hand. Suppose that by cheating the dealer shifts the advantage not to 100 percent but to just 50 percent in favor of the house. What effect does this have on the game?

If we assume that the player plays 100 hands, a typical total for an hour's playing time, and we also assume that the player bets an average of two units per hand, then being cheated once per 100 hands reduces the player's win by one unit on the average. A professional player varying his bet from one to five units would probably win between five and 15 units per hour. The actual rate would depend upon casino rules, the player's level of skill, and the power and variety of winning methods that he employed. Let's take a typical professional playing under good conditions and

Table 2-3

Hand	Top Card	Card Dealt	Comment	Plr. Gets	Plr. Total	Dlr. Gets	Dlr. Total	Result
1	2	First		2	2			
	5	Second (4)	Dlr. trys for good card			4	4	
	5	Second (10)	Dlr. will have 9; Plr. gets stiff	10	12			
	5	First				5	9	
	4	First	Worsens Plr. stiff	4	16			
	5	Second (8)	Prevent Plr. 21	8	24 bust			Dlr. wins
	5	First	Doesn't matter			5	14	
	9	First	Doesn't matter			9	23 bust	
2	4	First		4	4			
	K	First				10	10	
	Q	First	Give Plr. stiff	10	14			
	J	Second (J)	J will bust Plr. (Second turns out to be J, too!)	J	24 bust			Dlr. wins
	J	First	Doesn't matter			J	20	
3	A	Second (3)		3	3			
	A	First				A	1, 11	
	3	First	Bldg. potential stiff	3	6			
	J	First				J	BJ	Dlr. wins
4	2	First		2	2			
	9	First	Would make Plr. 11, so Dlr. takes			9	9	
	K	First	Give Plr. stiff	K	12			
	Q	Second (3)	Guarantees Dlr. win			3	12	
	Q	First		Q	22 bust			Dlr. wins
	8	First				8	20	
5	5	First		5	5			
	J	First				J	10	
	A	Second (A)	Guarantees Dlr. win (Second was A, too!)	A	6,16			
	A	First				A	BJ	Dlr. wins

Table 2-3 (continued)

Hand	Top Card	Card Dealt	Comment	Plr. Gets	Plr. Total	Dlr. Gets	Dlr. Total	Result
6	8	First		8	8			
	Q	First				Q	10	
	6	First	Give Plr. stiff	6	14			
	2	Second (3)				3	13	
	2	First	Give Plr. worse stiff	2	16			
	6	First		6	22 bust			Dlr. wins
	4	First				4	17	
7	7	First		7	7			
	6	Second (A)				A	1, 11	
	6	First		6	13			
	6	Second (K)				K	BJ	Dlr. wins
8	6	First		6	6			
	10	First				10	10	
	K	First	Give Plr. bad stiff	K	16			
	9	Second (2)	Dlr. must win			2	12	
	9	First		9	25 bust			Dlr. wins
	10	First	Doesn't matter	10		10	22 bust	
9	8	First		8	8			
	Q	First				Q	10	
	7	First		7	15			
	10	Second (5)	Dlr. must win			5	15	
	10	First		10	25 bust			Dlr. wins
	7	First				7	22	

assume that his win rate is ten units per hour and his average bet size is two units. Given those assumptions, being cheated ten times per hour or one-tenth of the time would cancel his advantage. Being cheated more than ten percent of the time would probably turn him into a loser.

Cheating in the real world is probably more effective than in the hypothetical example just cited, because the calculations for that example assume cheating is equally likely for small bets and big bets. In my experience, the bettor is much more likely to be cheated on large bets than on small ones. Therefore, the dealer who cheats with maximum efficiency will wait until a player makes his top bet. Suppose that bet totals five units. If the cheat shifts the odds to 50 percent in favor of the house, the expected loss is 2-1/2 units, and just four cheating efforts per 100 hands will cancel a professional player's advantage. A cheating rate of five or ten hands per 100 will put this player at a severe disadvantage.

We can see from this that a comparatively small amount of cheating applied to the larger hands can have a significant impact on the game's outcome. This gives you an idea of what to look for when you are in the casinos and think that something may be amiss.

Missing Cards: The Short Shoe

I have heard complaints that cards have been missing from the pack in some casino blackjack games. We'll discuss how you might spot this cheating method.

In 1962, I wrote on page 51 of *Beat the Dealer*, "Counting the. . . cards. . . is an invaluable asset in the detection of cheating because a common device is to remove one or more cards from the deck." Lance Humble discusses cheating methods for four-deck games dealt from a shoe in his International Blackjack Club newsletter. He says, "The house can take certain cards such as tens and aces out of the shoe. This is usually done after several rounds have been dealt and after the decks have been shuffled several times. It is done by palming the cards while they are being

shuffled and by hiding them on the dealer's person. The dealer then disposes of the cards when he goes on his break." But cheating this way is not limited to the casino. Players have been known to remove "small" cards from the pack to tilt the edge their way. The casino can spot this simply by taking the pack and counting it; the player usually has to use statistical methods.

In the cheating trade, the method is known as the *short shoe.* Let's say the dealer is dealing from a shoe containing four decks of 52 cards each. In 52 cards, there should be 16 ten-value cards: the tens, jacks, queens and kings. Logically, in four decks of 208 cards, there should be 64 ten-value cards. I'll call all of these "tens" from now on. Casinos rarely remove the aces—even novice players sometimes count these.

Suppose the shift boss or pit boss takes out a total of ten tens; some of each kind, of course, not all kings or queens. The shoe is shortened from 64 tens to 54 tens, and the four decks from 208 cards to 198 cards.

The loss of these ten tens shifts the advantage from the player to the dealer or house. The ratio of others/tens changes from the normal $144/64 = 2.25$ to $144/54 = 2.67$, and this gains a little over one percent for the house. How can you discover the lack of tens without the dealer knowing it?

Here is one method that is used. If you're playing at the blackjack table, sit in the last chair on the dealer's right. Bet a small fixed amount throughout a whole pack of four decks. After the dealer puts the cut card back only, let's say, ten percent of the way into the four shuffled decks and returns the decks into the shoe, then ready yourself to count the cards. Play your hand mechanically, only pretending interest in your good or bad fortunes. What you're interested in finding out is the number of tens in the whole four-deck shoe.

Let's say the shift boss has removed ten tens. (Reports are that they seem to love removing exactly ten from a four-deck shoe.) When the white cut card shows at the face of the shoe, let's say that the running count of tens has reached 52. That means mathematically that if all 64 ten-value cards were in the shoe, then, of the remaining 15 cards behind the cut card, as many as 12 of them would be tens, which mathematically is very unlikely. This is how one detects the missing ten tens because the dealer never shows their faces but just places them face down on top of the stack of discarded cards to his right, which he then proceeds to shuffle face down in the usual manner preparatory to another four-deck shoe session.

Although at first the running count is not easy to keep in a real casino situation, a secondary difficulty is estimating the approximate number of cards left behind the cut card after all the shoe has been dealt. To practice this, take any deck of 52 cards and cut off what you think are ten, 15 or 20 cards, commit yourself to some definite number, and then count the cards to confirm the closeness of your estimate. After a while, you can look at a bunch of cards cut off and come quite close to their actual number.

In summary, count the number of tens seen from the beginning of a freshly shuffled and allegedly complete shoe. When the last card is seen and it is time to reshuffle the shoe, subtract the number of tens seen from the number that are supposed to be in the shoe—64 for a four-deck shoe—to get the number of unseen ten-value cards which should remain. If 54 ten-value cards were seen, there should be ten tens among the unused cards. Then estimate the number of unseen cards. You have to be sure to add to the estimated residual stack any cards which you did not see during the course of play, such as burned cards. Step four is to ask whether the number of unseen ten-value cards is remarkably large for the number of residual cards. If so, consider seriously the possibility that the shoe may be short. For instance, suppose there are 15 unseen cards, ten of which are supposed to be ten-values. A computation shows that the probability that the last 15 cards of a well-shuffled four-deck shoe will have at least ten ten-value cards is 0.003247 or about one chance in 308.

Thus the evidence against the casino on the basis of this one shoe alone is not overwhelming. But if we were to count down the same shoe several times and each time were to find the remaining cards suspiciously ten-rich, then the evidence would become very

strong. Suppose that we counted down the shoe four times and that each time there were exactly 15 unseen cards. Suppose that the number of unseen tens, assuming a full four decks, was nine, 11, ten, and 13 respectively. Then referring to Table 2-4, the probabilities to six decimal places are $H(9) = .014651$ to have nine or more unseen tens, and for at least 11, ten, and 13 respectively, the chances are $H(11) = .000539$, $H(10) = .003247$, and $H(13) = .000005$. These correspond to odds of about 1/68, 1/1,855, 1/308 and 1/200,000 respectively. The odds against all these events happening together is much greater still. In this example, the evidence strongly suggests that up to nine ten-value cards are missing. There can't be more than nine missing, of course, because we saw all but nine on one countdown.

If the casino shuffles after only 104 cards are seen, it is not so easy to tell if ten ten-value cards were removed. A mathematical proof of this is contained in the appendix.*

This discussion should make it clear that the method suggested is generally not able to easily spot the removal of ten-value cards unless the shoe is counted several times or is dealt down close to the end.

One of the interesting ironies of the short shoe method of cheating players is that neither the shift boss nor the pitboss—the latter bringing the decks of cards to the dealer's table—need tell the dealer that his shoe is short. Thus, the dealer doesn't necessarily have to know that he's cheating. After all, he's just dealing. It's an open question how many dealers know that they're dealing from a short shoe.

Reports are that the short shoe is a frequent method that casinos use in cheating at blackjack using more than one deck. The tables with higher minimums (say $25) are more tempting candidates for short shoes than those with the lower minimums.

An experienced card counter can improve the method by counting both tens and non-tens. Then he'll know exactly how many unseen cards there are, as well as unseen tens. Table 2-4 can then be used with greater confidence.

In practice, you don't need to count through a shoe while bet-

*See Appendix B, pg. 136.

Table 2-4

K Number Of Unseen Ten-Value Cards	P(K) Probability Of Exactly K Unseen Tens	H(K) Probability At Least This Many Unseen Tens
0	.003171	1.000000
1	.023413	.996829
2	.078818	.973416
3	.160423	.894598
4	.220732	.734176
5	.217437	.513443
6	.158380	.296006
7	.086431	.137626
8	.036132	.050782
9	.011404	.014651
10	.002707	.003247
11	.000475	.000539
12	.000059	.000065
13	.000005	.000005
14	.000000	.000000
15	.000000	.000000

ting (and thus losing money in the process) to find out that the casino is cheating. If you suspect foul play, count while standing behind the player to the dealer's right.

You might easily catch a short shoe by simply counting all the cards that are used, whether or not you see what they are. Then if the remaining cards, at the reshuffle, are few enough so you can

accurately estimate their number, you can check the total count. For instance, you count 165 cards used and you estimate that 31 ± 3 cards remain. Then there were 196 ± 3 cards rather than the 208 expected, so the shoe is short.

A casino countermeasure is to put back a 4, 5 or 6 for each ace or ten-value card removed. Then the total number of cards remains 208, and the casino gets an even greater advantage than it would from a short shoe.

Cards do get added to the deck, and there's a spooky coincidence to illustrate this. On page 51 of *Beat the Dealer*, I wrote in 1962, "One might wonder at this point whether casinos have also tried adding cards to the deck. I have only seen it done once. It is very risky. Imagine the shock and fury of a player who picks up his hand and sees that not only are both his cards 5s, but they are also both spades." And then 15 years later in 1977, a player in a one-deck game did get a hand with two of the same card—the 5 of spades. Walter Tyminski's casino gaming newsletter, *Rogue et Noir News*, reported on page 3 of the June 15, 1977 issue, "What would you do if the player at your right in a single blackjack game had two 5 of spades? Nicholas Zaika, a bail bondsman from Detroit, had that experience at the Sahara in Las Vegas on May 24 at a $5 minimum table.

"Zaika wasn't in the best of humor because he had reportedly lost $594,000 at other Sahara tables, by far the largest loss he has ever experienced. Zaika had the blackjack supervisor check the cards and there were 53 cards in the deck, the duplicate bein the 5 of spades. . .The gamer has engaged the services of Las Vegas attorney George Grazadei to pursue claims he feels that he has against the casino. . .

"The Sahara denies any wrongdoing and says that it is cooperating fully with the investigation. . . Players aren't likely to introduce an extra 5 because the presence of the extra 5 favors the house and not the player."

Suppose instead of just counting tens used and total cards used, you kept track of how many aces, 2s, 3s, queens, kings, and so on were used. This extra information should give the player a better chance of detecting the short shoe. The ultimate proof would be to count the number of each of the 52 types of cards which have been used. Mathematical readers might wish to investigate effective statistical or other ways of using information for detecting shoes in which the numbers of some of the cards have been changed.

Chapter 3

Baccarat

The games of baccarat and chemin de fer are well known gambling games played for high stakes in several parts of the world. Baccarat is said to be a card game of Italian origin that was introduced into France about 1190 A.D. Two forms of the game developed. One form was called baccarat and the other was called chemin de fer. The most basic difference between these two games is simply that three hands are dealt in baccarat (called baccarat en banque in England) and two hands are dealt in chemin de fer (called baccarat-chemin de fer in England and Nevada).

The cards ace through nine are each worth their face value and the cards ten, jack, queen and king are each worth zero points. A hand is evaluated as the sum modulo ten of its cards, i.e., only the last digit of the total is counted. The object of the game is to be as close to eight or nine as possible with two cards, or as close to nine as possible with at most three cards if one does not have eight or nine on his first two cards. Then the high hand wins.

The games of baccarat and chemin de fer became popular in public casinos all over Europe, as well as in private games, about 1830. At the present time, one or both of these games are well known in London, southern France, the Riviera, Germany and the United States. A form of chemin de fer, which we shall call Nevada baccarat, has been played in a few Nevada casinos since 1958.

The rules, structure and format of the three games have strong similarities. I studied Nevada baccarat with William E. Walden most intensively because the casinos where it is played were readily accessible. Our techniques can be carried over to the other forms of baccarat and chemin de fer.

We were originally motivated by the observation that baccarat and chemin de fer have several points of resemblance to the game of blackjack, or twenty-one. The fact that practical winning strategies for twenty-one have been discovered suggested that there might also be practical winning strategies for baccarat and chemin de fer. In contrast to the situation in twenty-one, we found that there are no current practical winning strategies for the main part of the game, i.e., for the money Banker and Player bets.

Rules and Procedures

To begin the Nevada baccarat game, eight decks of cards are shuffled and a joker is placed face up near the end. The cards are then put into a wooden dealing box called a shoe. The first card is exposed, and its value is noted, face cards being counted as tens. Then this number of cards is discarded, or "burned."

The table has twelve seats, occupied by an assortment of customers and shills. A shill is a house employee who bets money and pretends to be a player in order to attract customers or stimulate play. We refer to them indiscriminately as "players." There are two principal bets, called "Banker" and "Players." Any player may make either of these bets before the beginning of any round of play, or "coup."

To begin the evening's play, two of the players are singled out. One is termed The Banker and the other is termed The Player. The seats are numbered *counter*clockwise from one to twelve. Player number one is initially The Banker, unless he refuses. In this case the opportunity passes counterclockwise around the table until someone accepts. The Player is generally chosen to be that player, other than The Banker, who has the largest bet on the Player. We have not noticed an occasion when there were no bets on The Player. When we played, there were shills in the game and they generally bet on The Player (except when acting as The Banker, when they generally bet on The Banker).

The Banker retains the shoe and deals as long as the bet "Banker" (which we also refer to as a bet on The Banker) does not lose. When the bet "Players" (which we also refer to as a bet on The Player) wins, the shoe moves to the player on the right. This player now becomes The Banker. If the coup is a tie, the players are allowed to alter their bets in any manner they wish. The same Banker then deals another coup.

To begin a coup, The Banker and The Player are dealt two cards each. As we noted above, the cards ace through nine are each worth their face value and tens and face cards are each worth zero points. Only the last digit in the total is counted.

After The Banker and The Player each receive two cards, the croupier faces their hands. If either two-card total equals 8 or 9 (termed a natural 8 or a natural 9, as the case may be), all bets are settled at once.

If neither The Player nor The Banker have a natural, The Player and The Banker then draw or stand according to the set of rules in Table 3-1.

The high hand wins. If the hands are equal, there is a tie and no money changes hands. Players are then free to change their bets in any desired manner. If the coup being played is complete when the joker is reached, the shoe ends and the cards are reshuffled. Otherwise the coup is first played out to completion. Then the shoe ends and the cards are reshuffled. However, the casino may reshuffle the cards at any time between coups.

Table 3-1

Player having		
0-5	draws a card	
6-7	stands	
8-9	turns cards over	
Banker having	draws when The Player draws	does not draw when The Player draws
0	none, 0-9	
1	none, 0-9	
2	none, 0-9	
3	none, 0-7, 9	8
4	none, 2-7	0, 1, 8, 9
5	none, 4-7	0-3, 8, 9
6	6, 7	none, 0-5, 8, 9
7	stands	stands
8	turns cards over	turns cards over
9	turns cards over	turns cards over

The Main Bets

Two main bets against the house can be made. One can bet on either The Banker or The Player. Winning bets on The Player are paid at even money. Winning bets on The Banker are paid 0.95 of the amount bet. The five percent tax which is imposed on what otherwise would have been an even-money pay-off is called "vigorish." For eight complete decks, the probability that The Banker wins is 0.458597, and the probability of a tie is 0.095156.

The basic idea of the calculation of these numbers is to consider all possible distinct six-card sequences. The outcome for each sequence is computed and the corresponding probability of that sequence is computed and accumulated in the appropriate register. Numerous short cuts, which simplify and abbreviate the calculation, were taken.

The house advantage (we use advantage as a synonym for mathematical expectation) over The Player is 1.2351 percent. The house advantage over The Banker is 0.458597X5 percent—1.2351 percent or 1.0579 percent, where 2.2930 percent is the effective house tax on The Banker's winnings. If ties are not counted as trials, then the figures for house advantage should be multiplied by 1/0.904844, which give a house advantage per bet that is not a tie, over The Banker of 1.1692 percent and over The Player of 1.3650 percent. The effective house tax on The Banker in this situation is 2.5341 percent.

We attempted to determine whether or not the abnormal compositions of the shoe, which arise as successive coups are dealt, give rise to fluctuations in the expectations of The Banker and The Player bets which are sufficient to overcome the house edge. It turns out that this occasionally happens but the fluctuations are not large enough nor frequent enough to be the basis of a practical winning strategy. This was determined in two ways. First, we varied the quantity of cards of a single numerical value. The results were negative.

We next inquired as to whether, if one were able to analyze the end-deck perfectly (e.g. the player might receive radioed instructions from a computer), there were appreciable player advantages on either bet a significant part of the time. We selected 29 sets of 13 cards each, each set drawn randomly from eight complete decks. There were small positive expectations in only two instances out of 58. Once The Player had a 3.2% edge and once The Banker had a 0.1% edge.

We next proved, by arguments too lengthy and intricate to give here, that the probability distributions describing the conditional expectations of The Banker and The Player spread out as the number of unplayed cards decreases. Thus there are fewer advantageous bets of each type, and they are less advantageous, as the number of unplayed cards increases above 13. The converse oc-

Table 3-2

Bacarrat (Eight Decks)

Subtract one Card of Value	Change In Player Bet	Advantage of: Banker Bet	Relative Point Values (Banker)	An Approximate Point Count
0	−0.002%	+0.002%	2	0
1	−0.004%	+0.004%	4	1
2	−0.005	+0.005	5	1
3	−0.007	+0.007	7	1
4	−0.012	+0.012	11*	2
5	+0.008	−0.008	−8	−1

Table 3-2 (continued)

Subtract one Card of Value	Change In Player Bet	Advantage of: Banker Bet	Relative Point Values (Banker)	An Approximate Point Count
6	+0.011	−0.011	−11	−2
7	+0.008	−0.008	−8	−1
8	+0.005	−0.005	−5	−1
9	+0.003	−0.003	−3	0

**Arbitrarily reduced from 12 to 11 so points in pack total zero* curs as the number of unplayed cards decreased below 13.

The observed practical minimum ranged from eight to 17 in one casino and from 20 up in another. The theoretical minimum, when no cards are burned, is six. Thus the results for 13 unplayed cards seem to conclusively demonstrate that no *practical* winning strategy is possible for the Nevada game, even with a computing machine playing a perfect game.

To see why, consider the accompanying Table 3–2 (pp. 34-35), based on Table 2 of Walden's thesis.

From this Table we see the effect of removing one of any card from the eight-deck baccarat pack. Proceeding in the way we developed the theory of blackjack, we get relative point values which are listed in the next to the last column. The last column gives a simpler approximate point count system.

We would now like to know how powerful a point count system in baccarat is compared with point count systems in blackjack. To do this we compute the root mean square (RMS) value of the column called "Change in Advantage of Banker Bet."

We do this by squaring each of those numbers, counting the square for zero-value cards four times because there are four times as many. Then we add these squares, divide by 13 and take the square root. The resulting root mean square or RMS value is .0064%. That measures how fast the deck shifts from its base starting value for a full pack.

Taking one card of a given rank (we think of there being 13 ranks) changes the fraction of any of these 13 ranks by an amount 32/416-31/415 which equals .00222. If we divide the RMS value by this value we get .0288 as a measure of how rapidly this advantage of the two bets shifts from the starting value as the composition of the deck changes.

Now we are going to compare this with the situation in blackjack. Table 3-3, for one-deck blackjack, can be treated in the same way to see how fast the advantage changes in blackjack.

The Table is from Peter Griffin's book, *Theory of Blackjack Revised*, page 44. We get RMS value of 0.467%. The corresponding change in the fraction of a single rank when one card is drawn

Table 3-3
One-Deck Blackjack

Subtract One Card of Value	Change In Advantage of Player
A	−0.61%
2	0.38%
3	0.44%
4	0.55%
5	0.69%
6	0.46%
7	0.28%
8	−.00%
9	−.18%
10	.51%

is 4/52-3/51 or .0181. The ratio of RMS value to this value is .258%. If we divide this by the corresponding result in baccarat we get 8.97, which tells us that as the true count in blackjack varies the change in player advantage or disadvantage shifts nine times as fast in blackjack as it does in baccarat.

Note that dividing the RMS value by the "change in fraction" of a single pack adjusts the one-deck blackjack figures and the eight-deck baccarat figures so they are comparable. If we had used, for example, an eight-deck blackjack table instead, we still would have had a final ratio of about nine times.

This allows us to translate how well a point count in baccarat works compared with one in blackjack. In baccarat we start out with more than a 1% disadvantage and with eight decks. Imagine a blackjack game with an eight-deck pack and a 1% disadvantage. Now imagine play continues through the blackjack deck. The blackjack deck advantage from a − 1% starting level might, on very rare occasions, shift 9% to a + 8% advantage.

As often as this happens in blackjack would be the approximate frequency with which we would get 1/9th as much shift in baccarat; meaning from a − 1% advantage to a 0% advantage or break even for the banker bet. Since there are two bets, banker and player, the player bet would also be break even or better about as often.

The conclusion is that you might expect to break even or better in eight-deck baccarat about twice as often as you would expect to have an 8% edge in eight-deck blackjack. How often would you have a 1% advantage in eight-deck baccarat? About twice as often as you would get a 17% edge in blackjack. The obvious conclusion is that advantages in baccarat are very small, they are very rare and the few that occur are nearly always in the last five to 20 cards in the pack.

The Tie Bet

In addition to wagers on the Player or Banker hands, the casinos offer a bet on "ties." In the event the Banker and Player hands have the same total, this bet gains nine times the amount bet. Otherwise the bet is lost. The probability of a tie is 9.5156%, hence the expectation of the bet is −4.884%.

It is clear, however, that the probability and thus the expectation of a tie depends on the subset of unplayed cards. For instance, in the extreme and improbable event that the residual deck

consists solely of ten-value cards, the probability of a tie is equal to one and the expectation is nine. Thus card counting strategies are potentially advantageous.

Using computer simulation, random subsets of different sizes were selected from a complete 416-card (eight deck) pack. The results were disappointing from a money-making perspective—the advantages which occur with complete knowledge of the used cards are limited to the extreme end of the pack and are generally not large. Practical card counting strategies are at best marginal, and at best precarious, for they are easily eliminated by shuffling the deck with 26 cards remaining.

Section Two

The Wheels

It doesn't require an extensive mathematical background to look at the 38 identically-sized spaces on an American roulette wheel (note the 35-1 payoff on a single number) and conclude that the game is unbeatable. With a 1/38 chance of having a number come up on the next spin and the 35-1 payoff, it is easy to calculate the often-quoted expectancy of the player of −5.26. The odds for other wheels, especially the Wheel of Fortune, appear even more against the player.

The unbeatability of the roulette wheel is based on the mechanical perfection of the wheel—such a conclusion is based on the assumption that the ball has an equal chance of landing in each pocket. This may or may not be true, although Allan Wilson, in *The Casino Gambler's Guide*, and others give fairly convincing evidence for the existence of biased wheels—wheels sufficiently biased to overcome the house advantage.

The very mechanical perfection of the wheel, however, would suggest the applicability of the laws of physics to prediction of the next number, whether the game is roulette or the Wheel of Fortune. Just as the future position of a planet can be predicted quite accurately, so can an understanding of the physical laws at work minimize the uncertainty surrounding the resting place of the ball or the final position of the wheel.

It is not possible, of course, to obtain an exact prediction. But this is not absolutely necessary to assure a profit. As Marvin Karlins has pointed out in his book *Psyching Out Vegas,* "Simply being able to predict which half of the wheel the ball will plunk into would give the player such a whopping edge that he could go for the chandeliers. . .and make it."

The following two chapters investigate the promise of this approach to beating the wheel as well as discussing some of the difficulties that might arise implementing such a strategy in the casino environment.

Chapter 4

Roulette

It was the spring of 1955. I was finishing my second year of graduate physics at U.C.L.A. In the course of the next year I would make three decisions that would shape my life for the next 28 years. I married (my present wife, Vivian), I changed my field of study from physics to mathematics, and I began to toy with the fantasy that I could shatter the chains of poverty through a scientifically-based winning gambling system.

I was living in Robison Hall, the student-owned cooperative. For $50 a month and four hours work a week, we got our room and board. I had lived in the co-ops for nearly six years of undergraduate and graduate work, on a budget of about $100 a month. Part of this came from scholarships and, in the early years, I got some help from home. But I was basically self-supporting like most of the other 200 or so co-op residents.

I attended classes and studied from 50 to 60 hours a week, generally including Saturdays and Sundays. I had read about the psychology of learning in order to be able to work longer and harder. I found that "spaced learning" worked well: study for an hour, then take a break of at least ten minutes (shower, meal, tea, errands, etc.). One Sunday afternoon about 3 p.m., I came to the co-op dining room for a tea break. The sun was streaming through the big glass windows. (Robison, designed by Richard Neutra in the '30s, was very radical for that time. It had so many big sheet glass windows that it was often called "the glass house.") My head was bubbling with physics equations, and several of my good friends were sitting around chatting.

In our mutual poverty the conversation readily turned to fantasies of easy money. We began to speculate on whether there was a way to beat the roulette wheel. In addition to me, the group included math majors Mel Rosenfeld and Andy Bruckner (now professors of mathematics at U.C. Santa Barbara), Tom Scott, and engineering major Rick Rushall. After all these years it's hard to be sure of exactly who said what, but we began the discussion by acknowledging that mathematical systems were impossible. I'll demonstrate this in a future chapter.

Then we kicked around the idea of whether croupiers could control where the ball will land well enough to significantly affect the odds. I will show later that this is impossible under the usual conditions of the game. (The incredible thing is that logical reasoning could even be used to settle such a question.) It was a short brainstorming step to wondering whether wheels were imperfect enough to change the odds to favor the player. Those in the group who "knew" assured me that the wheels are veritable jeweled watches of perfection, carefully machined, balanced and maintained. This is false. Wheels are sometimes imperfect enough so they can be beaten. I had no experience with gambling, or with casinos, or with roulette wheels, so I accepted the mechanical perfection of roulette wheels.

But mechanical perfection, for a physicist, means predictability. You can't have it both ways, I argued. If these wheels are very imperfect the odds will change enough so we can beat them. If they are perfect enough we can predict (in principle) approximately where the ball will land. Suddenly the orbiting roulette ball seemed like the planets in their stately and precise, predictable paths. In

my mind there was that intuitive "click" of discovery that I would experience again and again. Unknowingly, I had just taken the first step on a long journey in which I would discover winning systems such as those for blackjack and for the options market, and I would accumulate a wealth I never imagined.

One side argued that it is a long way from prediction in principle to practical prediction. My group said that, over and over, the story of science has been a rapid leap from a theoretical vision ($E=MC^2$) to an unexpected practical result (nuclear power plants). By now our initial group of people agreed that the idea had merit and might well work. The novel debate attracted listeners, some of them cynical. They challenged us to prove the idea worked. The ten minute "study break" had run into a couple of hours. We adjourned with the half definite idea of "doing something."

In the following weeks the idea kept coming back to me: measure the position and velocity of the roulette ball at a fixed time and (maybe) you can then predict its future path, including when and where the ball will spiral into the rotor. (The rotor is the spinning circular central disc where the ball finally comes to rest in numbered pockets.) Also measure the rotor's position and velocity at a (possibly different) fixed time and you can predict the rotor's rotation for any future time. But then you will know what section of the rotor will be there when the ball arrives. So you know (approximately) what number will come up!

You can see that the system requires that bets be placed *after* the ball and rotor are set in motion and somehow timed. That means that the casinos have a simple, perfect countermeasure: forbid bets after the ball is launched. However, I have checked games throughout the world, including Reno, Las Vegas, London, Venice, Monte Carlo, and Nice. Only in a few cases were bets forbidden after the ball was launched. A common practice instead was to call "no more bets" a revolution or two before the ball dropped into the center.

The simple casino countermeasure meant that there were two problems: (1) find out whether exact enough predictions could be made to get a winning edge, first in theory and then in the casino itself, and (2) camouflage the system so the casinos would be unaware of its use. If we could solve the prediction problem, the camouflage was easy. Have an observer standing by the wheel recording the numbers that came up, as part of a "system." Many do this so it doesn't seem out of place. But the observer also wears a concealed computer device with timing switches. His real job is to time the ball and rotor. (Much later we settled on toe-operated switches, leaving both hands free and in the open.) The computer would make the prediction and transmit it by radio to the bettor. The bettor, at the far end of the layout, would appear to have no connection to the observer-timer. The bettor would have a poor view of ball and rotor and would not pay much attention to them. To further break any link between timer and bettor, I would have several of each, with identical devices. They would each come and go "at random."

The important bets have to be placed after the ball is launched. A bettor who only bet then, and who consistently won, would soon become suspect. To avoid that, I planned to have the bettor also make bets before the ball was launched. These would be limited so their negative expectation didn't cancel all the positive expectation of the other bets. I became a radio amateur (W6VVM) when I was 13 (back in 1945 when there weren't easy novice-class tests), so I thought I could build the radio link and other electronic gadgetry.

This left me with the prediction problem to solve. More than a year passed without much time for roulette: I got my Master's degree in Physics (June 1955) and wrote the first part of my Ph.D. thesis on nuclear shell structure (Mayer-Jensen theory). The mathematical problems that I ran into led me in the fall of 1955 to take graduate math courses. I needed so many that I got my Ph.D. in math instead! And early in 1956 I got married. I had been working as a tutor and one of my "students" was T.T. Thornton. He was an independently wealthy, knowledge-loving bachelor of about 45, who had degrees in English and chemistry. Now he was getting a degree in mathematics, just for the pleasure

of it. He was an excellent student who didn't need a tutor but had hired me simply to learn faster and more efficiently.

We shared bits and pieces of our hopes, dreams, and enthusiasms.

After I had mentioned the roulette project, I was surprised and touched by his gift of a half-sized wheel. It was black plastic (bakelite?) made in France. I learned later that it cost the enormous sum of $25. Though I had thought about the roulette system off and on, the gift of this wheel (sometime in 1958, I recall) got me to work more seriously on it. My first idea was to use a home movie camera to film the orbiting ball. I then plotted the amount the ball had traveled versus the number of the frame of the film. I expected that the pictures were taken at a uniform rate of 24 (?) frames per second so I could plot (angular) distance traveled versus time as in Figure 4-1. Instead of a smooth graph like the solid line in Figure 4-1, my first film showed a peculiar wavy structure, like the dashed line.

After thinking about this, I guessed that this was because the camera did not run at uniform speed. By taking a movie of a stopwatch that timed in hundredths of a second, I found that the camera did vary in speed. Photo stores confirmed this. The distortion of the curve in Figure 4-1 is analogous to the way a musical tone is distorted by a phono turntable whose speed varies slightly.

My next move was to take a movie of the rotating ball and the stopwatch. This gave me an accurate time for each frame. (I still have a roll of these pictures, postmarked January 16, 1959.) But there was still some "ripple" to the curves. (I later learned that even a slight tilt would cause this.) Worse, I found that the curves were not consistent from spin to spin. The situation was something like Figure 4-2. This meant the ball behaved differently from spin to spin. This meant that the distance it traveled varied even with the same initial velocity. This doomed predictability on my wheel.

I found with further experiments that my half-sized wheel was really very irregular. The track was curved like a tube and the ball

Figure 4-1

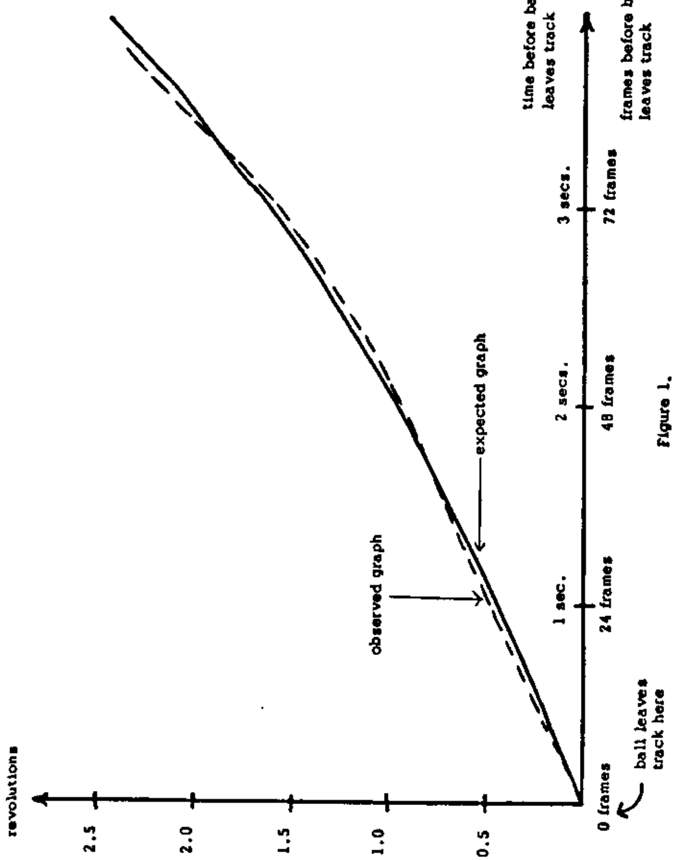
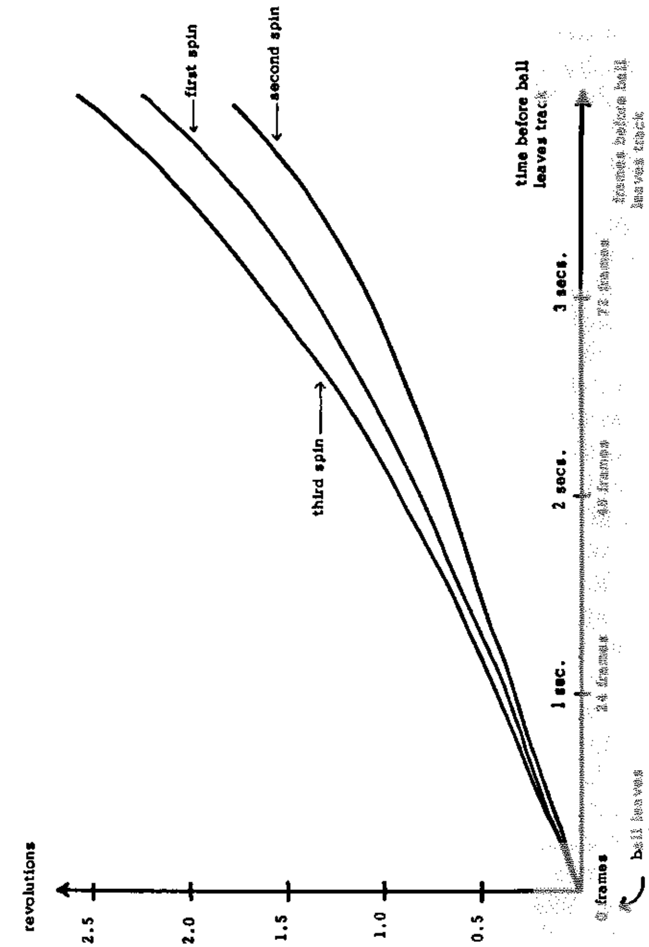


Figure 4-2



"rattled around" erratically, up and down, as it orbited. The slick bakelite surface was moulded, not machined. The ball also skidded and bounced. And there was a horizontal junction which added irregularities to the track.

But full-sized wheels were not like that. In December 1958, I made my first visit to the casinos. I observed several regulation wheels and found that the ball moved smoothly in its track. Also the track was a pair of flat-beveled, carefully-machined surfaces, not a tube. When I saw how good the casino wheels were, I was more convinced than ever that prediction was possible. But I needed a full-sized wheel and some good laboratory equipment to continue. How could I pay for it? I got my Ph.D. in June of 1958 and was teaching at U.C.L.A. Though my wife was finally able to stop working, we had no savings and I barely supported us. I couldn't ask her to go back to work to buy me a roulette wheel and to finance my pipe dream.

But I persisted. I simulated the study of the problem of whether the roulette ball would, for the same starting velocity, travel about the same distance along the track. I set up a little vee-shaped inclined trough. I would start a marble from a fixed height (a mark on the trough) and measure how far across the floor it rolled. I was encouraged but not surprised to find that the distance the marble went could be predicted closely from the starting height.

One memorable evening when my in-laws were due for dinner, I ran overtime on a marble experiment. They came into the kitchen wondering why I hadn't come to greet them at the door. They found me rolling marbles down a little wooden trough and across the floor. All over the floor were little distance markers and pieces of tape.

In early 1959 Vivian and I spent time with Mel and Judy Rosenfeld, working on a radio link for the casino test of my yet to be completed roulette system. We took model airplane radio control equipment and altered it somewhat. We succeeded in getting a workable but somewhat inconvenient radio link.

Then around March or April of 1959, I pushed the roulette project aside. Twelve man years of blackjack calculations arrived, courtesy of Baldwin, Cantey, Maisel and McDermott. I had convinced myself (as described in *Beat the Dealer)* that I could devise a winning blackjack card counting system and now I set to work on this intensely. The impractical marble roller now said he could beat the casinos at blackjack. What next?

I wrote my blackjack computer programs in the summer and fall of 1959. Testing, then debugging followed, and then from late 1959 through early 1960 my computer production runs produced the basic results that gave me the five-count system in early 1960. Then during 1960 I worked out most of the ten-count system and the ideas for the ultimate strategy. I also made the computer runs and worked out the methodology so that all of today's so-called "one parameter" blackjack systems could be readily devised by anyone versed in the use of computers. In December 1960, The Notices of the American Mathematical Society carried the abstract of my upcoming talk, "Fortune's Formula: The Game of Blackjack." Life would never be the same again. The intense professional and public interest aroused by the abstract, even before the talk, led me to seek quick publication in a scientific journal. I chose to try the Proceedings of the National Academy of Sciences. I needed a member of the Academy to communicate (i.e. approve and forward for recommended publication), so I sought out the one mathematics member of the Academy at M.I.T., Claude Shannon.

Claude Shannon: Genius

Shannon, then in his early forties, was and is one of the most famous applied mathematicians in the world. As one genius among many, he was relatively unnoticed as a graduate student—until he handed in his master's thesis. It developed the mathematical theory of switching electrical networks (e.g. telephone exchanges) and became the landmark paper in the subject. After receiving his doctorate, Shannon worked at Bell labs for several years and then became world-famous for papers establishing the mathematical foundations of information theory.

I was able to arrange a short appointment early one chilly December afternoon. But the secretary warned me that Shannon was only going to be in for a few minutes, not to expect more, and that he didn't spend time on subjects (or people) that didn't interest him (enlightened self-interest, I thought to myself).

Feeling both awed and lucky, I arrived at Shannon's office for my appointment. He was a thinnish alert man of middle height and build, somewhat sharp featured. His eyes had a genial crinkle and the brows suggested his puckish incisive humor. I told the blackjack story briefly and showed him my paper. We changed the title from "A Winning Strategy for Blackjack" to "A Favorable Strategy for Twenty-One" (more sedate and respectable). I reluctantly accepted some suggestions for condensation, and we agreed that I'd send him the retyped revision right away for forwarding to the Academy.

Shannon was impressed with both my blackjack results and my method and cross-examined me in detail, both to understand and to find possible flaws. After my few minutes were up, he pointed out in closing that I appeared to have made the big theoretical breakthrough on the subject and that what remained to be discovered would be more in the way of details and elaboration. And then he asked, "Are you working on anything else in the gambling area?"

I decided to spill my other big secret and told him about roulette. Several exciting hours later, as the wintery sky turned dusky, we finally broke off with plans to meet again on the roulette project. Shannon lived in a huge old three story wooden house on one of the Mystic Lakes, several miles from Cambridge. His basement was a gadgeteer's paradise. It had perhaps a hundred thousand dollars worth of electronic, electrical and mechanical items. There were hundreds of categories, like motors, transistors, switches, pulleys, tools, condensors, transformers, and on and on.

Our work continued there. We ordered a regulation roulette

wheel from Reno and assembled other equipment including (most important) a strobe light and a large clock with a second hand that made one revolution in one second. The dial was divided into hundredths of a second and still finer time divisions could be estimated closely. We set up shop in "the billiard room," where a massive old dusty slate billiard table made a perfect solid stable mounting for the roulette wheel.

Analyzing the Motion

My original plan was to divide the various motions of ball and rotor into parts and analyze each one separately. They were:

• The ball is launched by the croupier. It orbits on a horizontal track on the stator until it slows down enough to fall off this (sloped) track towards the center (rotor). Assume at first that (a) the wheel is perfectly level, and (b), the velocity of the ball depends on how many revolutions it has left before falling off. Referring to Figure 4-2, (b) means that every spin would produce the same curve, not different ones like my half-sized wheel. Put another way, this means that if you timed one revolution of the ball on the stator, you could tell how many more revolutions and how much more time until the ball left the track. If these assumptions turned out to be poor, we would attempt to modify the analysis.

• Next analyze the portion of the ball orbit from the time the ball leaves the track until it crosses from the stator to the rotor. If the wheel is perfectly level and there are no obstacles, then it seems plausible that this would always take the same amount of time. (We later learned that wheels are often significantly tilted. This tilt, when it occurs, can affect the analysis substantially. We eventually learned how to use it to our advantage.) There are, however, vanes, obstacles, or deflectors on this portion of the wheel. The size, number, and arrangement vary from wheel to wheel.

On average, perhaps half the time these have a significant effect on the ball. Sometimes they knock it abruptly down into the rotor, tending to cause it to come to rest sooner. This is typical of "vertical" deflectors (ones approximately perpendicular to the ball's path). Other times they "stretch out" the ball's path, causing it to enter the rotor at a more grazing angle and to come to rest later, on average. This is typical of "horizontal" deflectors (ones approximately parallel to the ball's path).

• Assume the rotor is stationary (not real), and beat that situation first. Reasoning: if you can't beat a stationary rotor, you can't beat the more complex moving rotor. Here the uncertainty is due to the ball being "spattered" by the frets (the dividers between the numbered pockets). Sometimes a ball will hit a fret and bounce several pockets on, other times it will be knocked backwards. Or it may be stopped dead. Occasionally the ball will bounce out to the edge of the rotor and move most of a revolution there before falling back into the inner ring of pockets. Thus, even if we knew where the ball would enter the rotor, the "spattering" from the frets causes considerable uncertainty regarding where it finally stops. This tells you that there is no possible reliable "physical" method for predicting ahead of time which pocket the ball is going to land in, unless the wheel is grossly defective or crooked. That makes the roulette method "used" in the movie "The Honeymoon Machine," where the players forecasted the exact pocket, an impossibility. It also tells you that successful physical prediction can at most forecast with an advantage which sector of the wheel the ball will end in.

• Assume now that the rotor is moving. Generally the ball and rotor move in opposite directions; increasing the velocity of the ball relative to the rotor. We'll assume this is always the case. I've never seen or heard of a casino spinning ball and rotor in the same direction. If this were done, the relative motion of ball and rotor would be even less than with a stationary rotor and prediction would be easier yet. With a moving rotor, the amount of ball "spattering" is increased and predictability is further reduced. Note that this change depends on the rotor velocity. Since that varies from time to time and from croupier to croupier, this adds further complexity. It turns out that the velocity of the rotor changes very slowly, so it is possible to predict with high accuracy

which part of the rotor will be "there" at the predicted time and place that the ball leaves the stator.

I will now take you through a simplified version of what we first tried to do. Later, with that overview to guide us, I'll explain some of the modifications we had to make and describe our casino experiences.

First, let's consider part 1, the motion of the ball on the track. The actual function $x(t)$, which describes the number of remaining revolutions x versus the remaining time t, is theoretically very complex.*

Our first problem, and the key one, was to predict when and where on the stator the ball would leave the track. This problem was key because once we knew this, everything else except rotor velocity was a "constant." And rotor velocity is easy to measure in advance and incorporate into the prediction, as we shall see. Our method was to measure the time of one ball revolution. If the time were short, the ball was "fast" and had a long way to go. If the time were "long," the ball was "slow" and would soon fall from the track.

We hit a microswitch as the ball passed a reference mark on the stator. This started the electronic clock. This was at time t_1 (to go) with x_1 revolutions to go. (There are many such "marks" available on all actual casino wheels.) When the ball passed the reference mark the second time we hit the switch again, stopping the electronic clock. That was at a time t_0 (left to go) before the ball left the track) with x_0 revolutions left. The clock measured $t_1 - t_2$, the time T for one revolution (*so* $x_1 - x_0 = 1$.)*

Movie Experiments

The function $x(t)$ which we are using in this illustration is not the actual one. The actual $x(t)$ can be determined by a "movie experiment" like the ones I described earlier which I did in 1959 on my half-size wheel. To do this experiment today, get a full-size roulette wheel, a large clock which reads accurately in hundredths of a second or better, and a video camera or movie camera. Then take a movie of the orbiting ball. The successive frames give

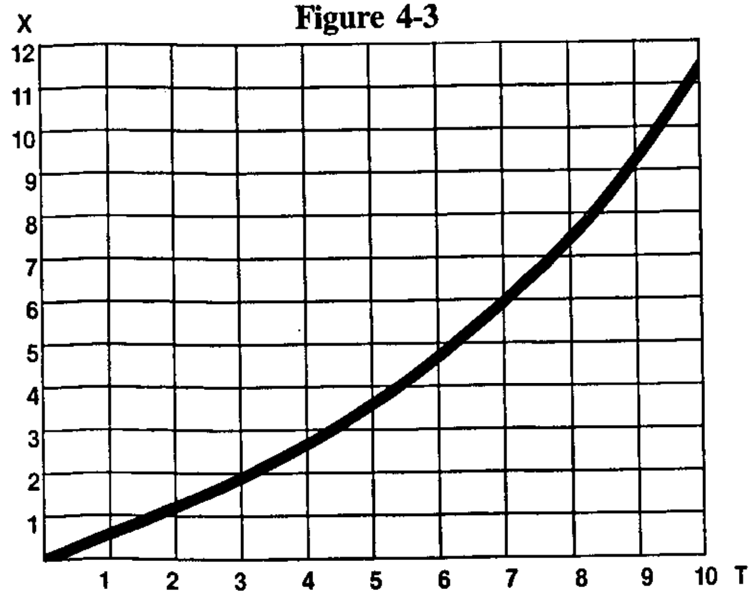
*See Appendix C, pg. 137.

*See Appendix D, pg. 137.

values for t and $x(t)$, which can be plotted to get an $x(t)$ curve like that of Figure 4-3. Several movies should be made to see how much the $x(t)$ curve varies from one spin to another. This uncertainty is a source of errors in determining T, that I'll discuss later on. These $x(t)$ errors can be incorporated into the theory in the same way as the timing errors. They each cause some uncertainty in the predicted $X_0(T)$ value. The data from the movie experiment can be improved if the camera frames are synchronized to a strobe so that the motion of both ball and clock is "stopped" rather than blurry. I didn't do this in my original movies, so I got a short blurry arc, instead of a ball, in each frame.

If an appropriate clock is not available, you can use a high quality phonograph turntable instead. These rotate at very uniform speeds which can be verified for your turntable with a strobe. Now get a stiff paper disc and mark the edges in equal small units. Number these units (much as you would a "circular" ruler) for ease in reading. Now place a thin fixed pointer just

Figure 4-3



above the disc. When the disc rotates, you have a very accurate clock whose hand is fixed and whose face moves. If you use a paper disc of polar coordinate graph paper (glued, perhaps, to an old record), there will be 360 equally spaced degree marks.

At 33⅓ r.p.m., each mark is 1/200 sec. At 45 r.p.m., each mark is 1/270 sec., and at 78 r.p.m., each mark is 1/468 sec. On a 12-inch disc, the 360 marks will be spaced about a tenth of an inch apart so additional marks can be used or the pictures can simply be read to a fraction of an interval. Record test discs with equally spaced "spokes," for use with a strobe for testing turntables, are also available and can be used.

Timing Errors

Shannon and I used the switch which measured T to flash a strobe as well as start and stop the clock. We discovered the lights and the strobe flash "stopped" the ball at each of the two instants the switch was hit. This allowed us to see how much the ball was off the reference mark. Since we knew approximately how fast the ball was moving, we could tell about how much in time we were early or late in hitting the switch. This enabled us to correct the times recorded on the clock, thereby making the data much more accurate. We also learned from the visual feedback how to become much more accurate at timing.

Here's an illustration. Suppose the track of the wheel was 25 inches in diameter. (I don't have any of this equipment now so I'm remembering back over 20 years and recalling about what the sizes, velocities, etc. seemed to be. They'll be close enough to be representative and good enough to show you how to do it all again, better for you if you want to.) Suppose the ball is ¾ inch in diameter and T, the time for one revolution, is 0.8 seconds. Then the track is 78.54 inches in length, or 98.17 ball diameters. If the ball center is one diameter away from the reference mark when the strobe flashes, then the timing error is about 1/98.17 of T or about 8/1000 of a second. There will be one of these errors when the switch is first hit and another when it is hit the second time. With practice we were able to reduce each error to a typical (root mean square) size of one ball diameter or about 8/1000 seconds. According to the theory of errors, the two errors together give a typical (root mean square) size of $\sqrt{2}/1000$ or about 11.2/1000 seconds.

These errors would be unobservable in casino play, so we couldn't correct for them there. The critical question is how do they affect the prediction?*

A Simple Casino Countermeasure

It should be clear that for this method to work, we have to time the ball (and rotor) before placing our potentially winning bets. (Earlier bets are losing, on average, so are only camouflage.) Thus, the casino must allow us to continue to bet for a time after the ball is launched. I have observed roulette wheels all over the world: Monte Carlo (our final goal), Nevada, Puerto Rico, Nice, Venice, and London. The practice has been, generally but not always, to allow bets until the ball was almost ready to fall off the track. This was much longer than we needed. Be warned again, though; all the casino needs to do to prevent our method is to forbid bets once the ball is launched. That simple perfect countermeasure is the Achilles heel of the system and a major reason why I never made a total effort to implement it. (People who use the system in casino play say the casinos don't catch on and don't use the countermeasure. But if the player is not really careful, I would expect the casino to catch on.)

The ball timing errors cause errors in predicting both the time and place the ball leaves the track. Even if the spiral path of the ball down the stator into the rotor is always the same in time and distance, this still yields errors in predicting when and where on the rotor the ball enters.*

*See Appendix E, pg. 138.

*See Appendix F, pg. 138.

Error Analysis

We have a long list of sources for errors in the prediction of the ball's final position. They are:

E1 Rotor timing—use 1.4 pockets to illustrate.

E2 Ball timing—use 5.5 pockets to illustrate.

E3 Variations in ball "paths" on rotor (see Figure 4-1). Error size is unknown, call it *X*.

E4 Ball path down stator: error due primarily to "deflectors" and varies with the type and placement. Use seven pockets to illustrate.

E5 Variation in distance ball travels on rotor: error due primarily to frets between pockets "spattering" ball, plus occasional very long paths along the rim of the rotor "outside" the pockets. Use six pockets to illustrate.

E6 Tilted wheel. (We didn't know about this yet.)

For illustrative purposes, assume the errors approximately obey the normal probability distribution. Then the standard deviation (typical size) of the sum of several errors is the square root of the sum of all the squared errors. For instance, using "pockets" as our unit, combined errors E4 + E5 have typical size $\sqrt{(6^2 + 7^2)} = \sqrt{85} = 9.2$ pockets. Now add on the timing errors: $E_1 + E_2 + E_4 + E_5$ have typical size $\sqrt{(1.4^2 + 5.5^2 + 6^2 + 7^2)} = \sqrt{117.21} = 10.8$ pockets. Thus the timing errors in this example cause very little additional error: just 10.8 −9.2, or 1.6 pockets.*

Of course, we haven't added in *E3* yet and, if X is big enough, it could ruin everything. Possible variations in the ball orbit behavior on the stator were difficult for us to measure because we found it hard to tell at exactly what point the ball lost contact with the outer wall of the wheel. We also learned from both our own lab experiences and from watching in the casinos why the orbit varied somewhat. Once a drunken, cigar-smoking bettor knocked his ash onto the track. This was hard to clear out. It got on the ball and spread out on the track. That immediately changed the ball's

*See Appendix G, pg. 140.

behavior. Skin oil from our fingers or the croupier's would slowly "poison" ball and track and seem to affect the orbit behavior.

If we or the croupier gave the ball lots of axial "spin" (in the sense of tennis or ping pong), it could take several revolutions around the track before this abnormal spin energy was converted to orbit energy. (We named this effect after the famous quantum mechanics concept of " spin-orbit coupling.") On the other hand, the ball might be launched with no spin or backspin, so it would skid for a while before spin and orbit got "into synch."

Advantage Versus Error

Obviously, the greater the error, the less the advantage. If we assume the total prediction error *E* is (approximately) normally distributed, then we can construct a table showing the player's expected gain or loss as a function of *E*.

Table 4-4 gives the results for a bet on the best pocket and also for a bet on the best "octant." The best octant is a set of five pockets, two on each side of the best pocket.

The Table shows that, when the prediction error is normally distributed, the typical forecast error (standard deviation) must be 16 pockets or less, in order for the bettor to have an advantage. This is 16/38, or about 0.42 revolutions. This is true both for bets on the best pocket and the best octant. Since the best octant includes four pockets that aren't quite as good as the best, the advantage is somewhat less for a given typical error *E*. However, as we will see later in discussing the Kelly-Breiman system for money management, it is generally better for a small to medium-sized bankroll to bet the best octant.

Kimmel and the Dealer's Signature

Stephen Kimmel asserted that a dealer who works eight hours a day, 50 weeks a year, tends to spin the ball and rotor in a habitual, regular way. This would make possible accurate predictions—a bet on ten pockets, Kimmel contended, would have a 50% chance of success. His views were contained in an article "Roulette and Randomness" in the December, 1979 issue of *Gambling Times*.

Table 4-4

Typical Error E (No. of Pockets)	Percent Advantage Betting on Best		Typical Error E (No. of Pockets)	Percent Advantage Betting on Best	
	Pocket	Octant		Pocket	Octant
0	3500.00	620.00	16	0.46	0.30
1	1278.53	611.06	17	−1.62	−1.72
2	610.69	467.86	18	−3.01	−3.07
3	376.52	328.65	19	−3.90	−3.94
4	258.12	236.98	20	−4.46	−4.49
5	186.76	175.71	21	−4.81	−4.82
6	139.09	132.62	22	−5.01	−5.02
7	105.00	100.89	23	−5.13	−5.13
8	79.41	76.65	24	−5.19	−5.19
9	59.54	57.60	25	−5.23	−5.23
10	43.77	42.38	26	−5.24	−5.25
11	31.19	30.18	27	−5.25	−5.25
12	21.24	20.52	28	−5.26	−5.26
13	13.54	13.03	29	−5.26	−5.26
14	7.73	7.37	30	−5.26	−5.26
15	3.47	3.24	∞	−5.26	−5.26

I don't believe Kimmel's approach works. Here's why: there are three important conditions that must remain roughly constant throughout play for the player to take advantage of the regularity of the dealer's signature. These conditions are (1) the rotor velocity should be approximately the same each time the ball is spun, (2) the spinning ball should make approximately the same number of revolutions each time, and (3) the initial position of the rotor when the dealer launches the ball should be approximately the same each time. This third condition, which is not mentioned in Kimmel's article, is crucial.

By way of illustration, suppose that the rotor velocity was exactly the same each time and that the dealer spun the ball exactly the same number of revolutions in each instance. Suppose further that the ball spun exactly eight revolutions and the rotor four revolutions during this time. Given those assumptions, the ball would land about 12 revolutions beyond the point where it was launched. In other words, if the number 13 was passing the ball as the dealer released it, the ball would arrive 12 revolutions later, *relative to the spinning rotor,* at approximately the number 13. You can see, however, that if the number 2 on the rotor was closest to the ball at the instant it was released, the ball would then end up near that number 12 revolutions later.

If the dealer releases the ball without regard to which number on the spinning rotor is closest to the launch point, the ball would randomly fall on the rotor 12 revolutions later. In this case, there would be no predictability whatsoever, even though the rotor velocity is absolutely fixed and the number of ball revolutions constant. Any variance in rotor velocity or number of ball revolutions would further guarantee a random outcome. Because Kimmel did not discuss variations in the point of release, I do not believe in his method.

There is a better approach to this statistical analysis of roulette. Watch a dealer and count the number of revolutions the ball makes on the stator from the time of release until it crosses onto the rotor. Note how constant that number of revolutions is. The results of your observations can be statistically stated as some average number of revolutions plus an error term.

Next, count the number of revolutions the rotor makes during the time the ball is on the stator. This will give you another average for the number of rotor revolutions, plus a second error term. Finally, count how far the ball travels on the rotor after it has crossed the divider between the rotor and stator. You can summarize these results as some average number of revolutions or pockets plus an error term.

In order for this approach to work, it is necessary that the square root of the sums of the squares of the error terms be less

than 17 pockets. The proof of this appears in Table 4-4 which shows what the rate of return is, given various root mean square errors. That table demonstrates that a positive return is possible only when that root mean square error is less than 17 pockets.

Now for the improved method. In the unlikely event that the root mean square error is less than 17 pockets, then—and only then—you have a chance to win. The key lies in using the position of the rotor when the ball is launched as your starting point for predicting where the ball will fall out on the wheel.

For example, suppose you find that for a certain dealer the ball travels eight revolutions with a root mean square error of five pockets. Suppose also that during this time, the rotor travels four revolutions, with a root mean square error of six pockets. And suppose still further that once the ball is on the rotor, it travels 13 pockets with a root mean square error of eight pockets. Given these suppositions, you can predict that the ball will travel eight revolutions plus four revolutions plus 13 pockets from the launch position, or 13 pockets beyond that point. The root mean square error is the square root of five squared plus six squared plus eight squared. This turns out to be 11.2 pockets, well within the required error of less than 17 pockets. In this case, the prediction system would work.

However, I think you will find that when you collect this data, the errors at each stage are several times as large as I have used in this example. My own observation is that the dealer error in the number of revolutions for the ball spin is about 20 pockets for the more consistent dealers; it is much larger with a less consistent one. I also noticed that the rotor velocity is not nearly as constant as Kimmel would like. That is because the dealer gives it an extra kick every few spins to rebuild its velocity.

It is also true that the deflecting vanes on the sides of the rotor add considerable randomness to the outcome, as do the frets or spacers between the pockets. The upshot is that I don't believe that any dealer is predictable enough to cause a root mean square error of less than 17 pockets. I'm willing to examine proof to the contrary, but I would be very surprised if anyone could ever produce it.

If a dealer dutifully practiced spinning the ball a fixed number of revolutions, and if a motor drive spun the rotor at a constant velocity, and if we have a very good way of deciding exactly which number is opposite the ball just as it is released, it might be barely possible to gain a small prediction advantage. I consider even that very unlikely.

In closing, I'll give you the perfect casino countermeasure to the strategy of the dealer's signature, pretending for the moment that the strategy worked. First, the casino halts the betting before the dealer spins the ball. Second, the dealer closes his eyes or looks away from the wheel when he releases the ball so that he has no knowledge of which number on the rotor is closest to the ball when it is launched. Then, for the reasons explained above, the result will be perfectly random.

Chapter 5

The Wheel of Fortune

In the last chapter, I described a system for winning at roulette based on physical prediction. That system was developed largely in 1961 and 1962 in collaboration with Claude Shannon at MIT. One by-product was an even simpler system for physical prediction of the Wheel of Fortune. A story about me and blackjack card-counting in *Life* magazine, March 27, 1964, reported on this in a section entitled "Beating the Wheel of Fortune with the Big Toe."

While I was at the Fifth Annual Conference on Gambling and Risk Taking at Caesars Tahoe in October of 1981, I collected data on a Wheel of Fortune at Caesars. I wanted to see whether their wheels could still be predicted in the same way.

My Casio C-80 watch has a digital stop watch feature which times to 1/100 of a second. I used it to time one revolution of the wheel and then recorded how many revolutions it went. I collected the data in Table 5-1 at the Wheel of Fortune nearest to Caesars' cashier cage.

To see how predictable the Wheel was, I looked for a

Table 5-1

Wheel of Fortune Data
Caesars Tahoe

Observation Number	Time T	Further Raw Data	Revs R Decimal	Prediction P	Error P—R	Error In "Pegs"	Expected Time E	Time Difference T—E (Sec)
1	5.11	3.5 + 22p	3.907	3.856	−.051	−2.8	5.08	+.03
2	5.33	3.5 − 2p	3.463	3.527	+.064	3.5	5.38	−.05
3	5.09	4 − 6p	3.889	3.889	.000	0.0	5.09	.00
4	4.83	4 + 15p	4.278	4.345	.067	3.6	4.87	−.04
5	5.87	3 − 9p	2.833	2.876	.043	2.3	5.91	−.04
6	5.67	3 + 10p	3.185	3.095	−.090	−4.9	5.59	+.08
7	7.14	2 − 6p	1.889	1.901	.012	0.6	7.16	−.02
8	5.78	3 + 3p	3.056	2.972	−.084	−4.5	5.70	.08
9	4.66	4.5 + 8p	4.648	4.687	.039	2.1	4.68	−.02
10	5.16	4 − 9p	3.833	3.778	−.055	−3.0	5.12	.04
11	6.37	2.5 − 9p	2.333	2.419	.086	4.6	6.48	−.11
12	6.93	2 + 3p	2.056	2.024	−.032	−1.7	6.88	.05

mathematical curve which would best fit these data points. A curve which worked well was R = A times T to the B power where A = 121.545 and B = −2.11153. In the equation, T is the time for the wheel to make one revolution and R is the number of additional revolutions which it then travels. Intuitively, if T is short, the wheel did one revolution quickly so it will go far and R will be large. But if T is long, the wheel was slow and will stop soon so R will be small.

The letter p in the third column of the Table ("raw data") stands for "pegs." The wheel has pegs separating the payoff numbers. As the wheel rotates, the pegs push past a flexible "flapper." This gradually slows the wheel. When the wheel stops, the winning number is the one with the flapper between its pegs.

The raw data column gives 3.5 + 22p for observation number 1. This means that the wheel traveled 3.5 revolutions plus 22 pegs or further numbers. Since there are 54 numbers in all, it went 3.5 + 22/54 or 3.907 revolutions in all. That is shown under "decimal" in column 4.

The prediction P is made from the equation. The "error" P-R is the amount the prediction is off from what actually happened. Strictly speaking, what I am calling a prediction is only a fit to the data. The fit approaches a "true" fit more closely as more data is included. However, there is generally a difference between the "true" fit and the actual fitted equation.

New data tends to cluster around this slightly different unknown true fit, so it will tend to deviate from the actual fit to the data by this extra amount. Thus, we expect future data to be predicted by the equation not quite as well as the data in Table 5-1.

The error P − R has a standard deviation ("typical size") of .0587 revolutions, or 3.2 numbers. The true curve location (standard deviation of the curve) is probably within .0169 revolutions or 0.9 numbers, on average. Considering this and the greatest positive and negative values in the column, error in "pegs" suggests that the prediction will almost always be within five "pegs" or positions of the actual outcome.

Table 5-2 shows the actual arrangement of numbers on the wheel. They are listed in order, clockwise, as seen by the player. Each number gives the profit per unit bet. Thus, a player who bets on 2 wins $2 for each $1 bet. The number marked 40A, and called Caesars, pays 40 to 1 and the number 40B, called Cleo, also pays 40 to 1. A bet on one of them does not win if the other one comes up.

There are 24 "ones" in Table 5-2. Thus, if each of the 54 numbers comes up once, "one" wins 24 times and loses 30 times for a loss of 6 units in 54 unit bets, or an expected loss rate of $-6/54 = -1/9 = 11.1\%$. Similar calculations lead to Table 5-3. For the player who doesn't predict, the house edge is enormous. This is a game to avoid.

Table 5-2

2	1	40A	2	1	2
1	2	1	10	1	5
1	2	1	20	1	2
1	5	2	1	10	1
2	5	1	2	1	40B
1	2	1	5	2	1
10	1	5	1	2	1
20	1	2	1	5	2
1	10	1	2	5	1

Table 5-3

Number	House Edge	
1	6/54	11.1%
2	9/54	16.7%
5	12/54	22.2%
10	10/54	18.5%
40A	13/54	24.1%
40B	13/54	24.1%

Now let's see what the player advantage might be from predictions. Suppose for the sake of discussion that the final wheel position is always within five numbers of the predicted wheel position. For any prediction in the eleven number strip centered around 40A, we should bet on 40A. In 54 spins where each final position occurred once, we will place 11 bets on 40A and win one of them for a gain of $40 - 10 = 30$ units.

The discussion is the same for 40B. For any prediction in either of the eleven number strips surrounding each 20, twenty-two numbers in all, we bet on 20. In twenty-two bets we expect to win 20 units twice and lose one unit twenty times for a net gain of

twenty units. This leaves 54-44 or ten predicted positions where we need instructions.

There are four 10s in this left-over set of ten positions. Suppose we bet the 10 each time one of these positions is predicted. It seems plausible to suppose that we would win ten units four times and lose 1 unit six times for a net gain of $40 - 6 = 34$ units. (Actually, since the 10s in this case are either the predicted number or within one position of the predicted number, we expect to do better still.

Finally, in 54 unit bets we net 30 units from 40A, 30 units from 40B, 20 units from the two 20s, and 34 units from the four 10s, for a total of 114 units/54 units or a 211% rate of return.

It may be possible to improve both the timing procedure and the method of exploiting predictability. This would improve the results.

We see now that the Caesars wheel can be predicted well enough so that we can beat it if the casino will let us put down bets after the wheel has been set in motion.

Section Three

Other Games

While we have so far concerned ourselves solely with casino games, some of the principles we have used are equally applicable to other gambling situations. In this section, we apply mathematical theory to horse race betting and backgammon.

Horse racing is the number one spectator sport in America and a large amount of its success in this regard can be attributed to the wagering opportunities. The racegoer becomes a participant in the spectacle. While we offer no surefire system, we do suggest an approach that shows promise for the gambler-investor.

Backgammon is an exceedingly complicated game from a mathematical point of view. Because of the possibility of repeated restarts by the counters, the game is potentially infinite. This impedes analysis, but we offer several insights into the end game and the doubling cube, parts of the game where optimal strategies can be computed.

Chapter 6

Horse Racing

While so far I have limited myself to discussing casino games, the concepts presented apply equally to other gambling games, such as horse racing. At the racetrack, one is offered a variety of different wagering possibilities. The player can wager on one or more horses to win, place, or show, as well as combine any number of horses in the various exotic bets (daily double, exactas, quinellas, etc.) The goal of the gambler at the racetrack is to isolate in each race those bets that yield a positive return, after considering the pari-mutuel takeout and breakage.

One approach with applications in horse racing involves the use of a technique called hedging. Hedging, often used in the securities and financial markets, involves taking two or more investment positions simultaneously. The risks should cancel out and an excess rate of expected return should remain.

Why "Hedge?"

In the securities and finance markets, to hedge is to take two or more investment positions simultaneously. The risks should cancel out and an excess rate of expected return should remain.

In a real horse race (or any pari-mutuel contest for that matter), the true probabilities are not known. If we knew the true probabilities or had better estimates than the pari-mutuel pool offers, we might find horses with a positive expectation. Then we could simply bet directly on those horses instead of developing the following method for the daily double.

There is a plausible argument which upholds the pari-mutuel pool's estimate of the true horse winning probabilities: "If there were a method of predicting horse winning probabilities, and these probabilities differed enough from the pari-mutuel pool's estimate to give the predictor an advantage, then he would place bets and by so doing would cause the pari-mutuel pool odds to shift in such a way as to reduce that advantage. With many bettors and much information and available computing power the overall effect is to reduce such advantages so they are small or even become disadvantages."

In other words, "If you could beat the casinos at blackjack, then they would change the game so you couldn't. Thus, there isn't any system for beating them."

If we assume that pari-mutuel pool probabilities are true probabilities then the horse hedge system does not improve our edge over the track take! You might think that makes the horse hedge idea useless, but this is not true. Consider the daily double pool: The payoffs should be consistent with the probabilities in the individual race win pool; but in general, they aren't consistent. Thus, we have a chance to use the probabilities based on the individual race win pools.

The Daily Double

Let's apply the horse hedge idea to the daily double bet. The same idea, with some modifications, also applies to exactas, pick six and similar bets and to exactas, quinellas and trifectas in jai alai.

For a little background on the daily double, I quote from the book *Science in Betting: The Players and the Horses,* by E. R. Da Silva, and Roy M. Dorcus:

In daily double betting, any horse in the first race can be combined with any horse in the second race, and to win the bettor must successfully select the winners of both races. Some bettors combine all of the horses in the first two races. If there are ten horses in each race, in order to cover all possible combinations of horses, one would have to buy one hundred tickets at $2.00 each. If by chance long-odds horses won both races, it would be possible to make a profit on that single daily double. However, such a situation is not common throughout a week or a season. One daily double $2.00 ticket at Del Mar recently paid $2,878.60, another paid $685 and there were in addition three others during this season which paid over $200—yet actual returns for this season were only $6,808. The average number of horses was eleven in the first race and ten in the second. To have combined all these horses in all the daily doubles for this season would have cost $200 per race, and since there were forty-two days in this season, the total cost would have been $9,240, producing a loss of several thousand dollars.

Notice that they consider betting equal amounts on each horse. From one season at Del Mar, they found that $9,240 in total bets were returned; $6,808 for a payback fraction of 0.74 or a loss of 26% of the amount bet. Thus, betting equal amounts on each combination did not work.

Illustrated in Table 6-1 is the horse hedge method for daily doubles in a real race. The Table lists the winning probabilities based on odds for the first race at Del Mar on August 13, 1980. The horses are listed according to post position in the first column. The second column has the handicapping odds given in the *L.A. Times* on the morning of race day. The third column is obtained from these odds by taking the right hand number in the second column and dividing by the sum of the two numbers.

For example, 30-1 gives a probability of 1/31 = 0.0323. For the horse in the 13th post position, 7-2 gives a probability of 2/9 = 0.2222. When there is no track take, the probabilities calculated this way must add up to 1.00.

Table 6-1

First Race

Horse [p.p.]	A.M. Odds [h.c.]	Prelim. Probs.	Corrected Probs.	Final Odds: 1	Prelim. Probs.	Corrected Probs.
1	30-1	0.0323	.0246	114.80	.0086	.0068
2*	4-1	0.2000	.1526	3.10	.2439	.1922
3	2-1	0.3333	.2544	0.70	.5882	.4635
4	20-1	0.0476	s	s	s	s
5	5-1	0.1667	.1272	7.30	.1205	.0949
6	6-1	0.1429	.1090	12.10	.0763	.0601
7	10-1	0.0909	.0694	24.00	.0400	.0315
8	30-1	0.0323	.0246	72.60	.0136	.0107
9	30-1	0.0323	s	s	s	s
10	8-1	0.1111	.0848	6.40	.1351	.1065
11	15-1	0.0625	.0477	52.50	.0187	.0147
12	20-1	0.0476	.0363	82.50	.0120	.0094
13	7-2	0.2222	s	s	s	
14	8-1	0.1111	s	s	s	
15	10-1	0.0909	.0694	81.20	.0122	.0096
	sum	1.7237				
After scratches		1.3105	1.0000	1.2691	1.0000	
		Estimated track take			21.20%	

When there is a track take, the probabilities calculated from the final payoff odds at race time will equal more than 1.00. In fact, they add to 1/K, where K is the fraction of the pool, which is returned to the bettors. This rule is not quite exact due to the irregular effects of breakage, but the effects are generally small and not worth discussing.

In order to correct for probabilities that do not add up to 1.00, we add them, deducting horses which may have been scratched. We then use the final total and divide it into the preliminary probabilities so that it equals 1.00. (Corrected probabilities appear in column four.)

Column five gives the final odds on various horses. Column six has corresponding uncorrected probabilities and column seven lists corrected probabilities. Notice that column six adds to 1.2691; by dividing this into 1.00 we get 0.7880 which corresponds to a track take of 21.30% for this particular race.

Column three equaled 1.7237 before deducting the horses which were later scratched, making the track take too large. The sum for Del Mar is typically about 1.20; therefore the handicapper's setting of the odds was not consistent. On average, the odds were set too low in this race for the horses. When four horses were scratched, the odds on the remaining horses gave probabilities which equaled 1.3105. That is the typical sum at Del Mar.

Table 6-2 presents the probability calculations for the second race. The fourth column appears to equal 0.9999, but shows 1.0000, because the entries have been rounded off to four places.

The final outcome of the daily double: horse 2 won the first race; horse 1 won the second race; and a winning $2 ticket paid back $38.60 or $19.30 per unit bet. The amount bet on each of the 15 × 8 or 120 combinations is proportional to the product of the corresponding probabilities.

For example, if we use the corrected probabilities based on the morning odds, we have .1526 for horse 2 in the first race and .1725 for horse 1 in the second race. The product of these two numbers is .0263. That means we bet .0263 of our total unit bet on the combination which actually won the daily double.

Table 6-2

Second Race

Horse [p.p.]	A.M. odds [h.c.]	Prelim. Probs.	Corrected Probs.	Final Odds: 1	Prelim. Probs.	Corrected Probs.
1*	7-2	.2222	.1725	2.60	.2778	.2313
2	15-1	.0625	.0485	18.80	.0505	.0421
3	3-1	.2500	.1941	5.90	.1449	.1207
4	6-1	.1429	.1109	13.30	.0699	.0582
5	5-2	.2857	.2218	3.80	.2083	.1735
6	6-1	.1429	.1109	15.80	.0595	.0496
7	10-1	.0909	.0706	2.80	.2632	.2192
8	10-1	.0909	.0706	6.90	.1266	.1054
	SUM	1.2880	1.0000		1.2007	1.0000
		Estimated track take			16.72%	

Therefore, we have a return of \$19.30 × this probablity or .5080 of a unit which means we lost 49.2% of our bet. If we had used the final odds, the probabilities are .1922 and .2313. Their product is .0445 and we would receive this amount × \$19.30 or .8580 of a unit, or a loss of 14.2%.

On page 127, Da Silva and Dorcus warn you that:

> In doing any statistical work on daily doubles, the reader must be careful not to use the actual closing odds of the horses, as listed the day following the races in result charts from newspapers or from the Form or the Telegraph, since these last-minute odds are not available to the daily double bettor for either the first or the second races. The bettor must rely only upon the probable odds for statistical study of daily double betting, odds which are given in the Morning Line at the tracks, in the Form or Telegraph under different handicappers such as Sweep, Analyst, Trackman, or given in the track programs.
>
> Furthermore, in dealing with these probable odds, the bettor must remember that they may or may not correspond to the last-minute closing odds on the toteboard.

(For the first race only, the actual odds that we would use in practice may be fairly close to these final odds if we were actually at the track watching the toteboard.) At this point, we can see difficulties with the horse hedge idea as it relates to the daily double.

For example, there is a minimum \$2 bet. In order to approximate the various probabilities of the typical one hundred or so combinations, we have to make several hundred \$2 bets which requires a substantial bankroll. Another problem is that the final pari-mutuel pool odds are unknown. Even if we did know the odds on the individual races, the true probabilities of the individual horses winning in their respective races would still be unknown. Therefore, we don't know if the horse hedge method will give us an advantage over the track take.

Even if it does give us an advantage, we don't know if we can gain enough to overcome the track take for an overall advantage.

This is the reason why this system needs further development.

One way to get around the difficulties is to keep a record of the final odds and the corresponding probabilities and bet accordingly. If pari-mutuel odds are a fair estimate of the true odds, then this indicates the sort of gain to be had from horse hedging. If the gain is large enough to produce a substantial advantage, then there might still be an advantage if we use good odds that are available to us at the time we place our bets.

To show you how to keep this sort of record, I will use one average figure to correct of the track take. Table 6-3 shows the sum of the uncorrected probabilities for the first five races on three consecutive days. The days are August 13, 14 and 15, 1980 at Del Mar. The two entries followed by question marks suggest that there may be data errors or newspaper misprints. Except for the two questionable figures, the uncorrected probability sums are close to 1.20. The average, of the 13 remaining races in Table 6-3 works out to be 1.2033.

Table 6-3

Uncorrected Probability Sums

Race #	8/13/80	8/14/80	8/15/80
1	1.2691?	1.1970	1.2177
2	1.2007	1.1968	1.2104
3	1.2064	1.1973	1.2026
4	1.1960	1.2049	1.2059
5	1.2082	0.9966?	1.1986

To simplify, I shall use 1.20 in my computations in Table 6-4. The fractions estimate the investment returned for each day the horse hedge system is used at Del Mar. If you want to construct a similar table, get extensive racing records from your track, and determine whether the method works over a past sample.

Table 6-4

Del Mar

Race Date	Race 1 Winner Odds: 1	Prelim. Prob.	Corrected Prob.	Race 2 Winner Odds: 1	Prelim. Prob.	Corrected Prob.	Product of Probs.	Daily Double Payback for 1	Fraction Returned
8/13/80	3.10	.2439	.2033	2.60	.2778	.2315	.0470	19.30	.9080
8/14/80	2.50	.2857	.2381	.50	.6667	.5556	.1323	6.50	.8600
8/15/80	0.70	.5882	.4902	1.90	.3448	.2874	.1409	5.50	.7748
8/16/80	1.80	.3571	.2976	2.30	.3030	.2525	.0752	12.90	.9695
8/17/80	5.50	.1538	.1282	6.70	.1299	.1082	.0139	50.30	.6979
8/18/80	1.60	.3846	.3205	9.50	.0952	.0794	.0254	41.60	1.0582
8/20/80	0.90	.5263	.4386	5.30	.1587	.1323	.0580	17.80	1.0327
8/21/80	2.20	.3125	.2604	6.40	.1351	.1126	.0293	45.50	1.3343
8/22/80	1.70	.3704	.3086	11.60	.0794	.0661	.0204	36.90	.7532
8/23/80	7.40	.1190	.0992	1.10	.4762	.3968	.0394	27.60	1.0865

Ten races: estimated average fraction returned 0.9476

Table 6-4 shows the idea at Del Mar. The second, third, and fourth columns list the corrected probabilities based on the final odds of 2-1 for the winning horse in the first race. The fifth, sixth, and seventh columns do the same thing for the second race. The eighth column is the product of these probabilities (the pari-mutuel estimate of the probability of a pair of horses winning the daily double). For the last column, multiply the payback on $1 which is the fraction of the unit be returned to us.

In our sample of ten races, we get an estimated payback of 94.76%, or a loss of 5.24%. We are estimating the average effective track take as 16.67% so the system does better than average but still does not win.

For a clear explanation of daily double betting, exactor or exacta betting, odds, and trifecta betting, I refer you to the appendix of *Harness Racing Gold,* by Prof. Igor Kusyshyn, published by International Gaming Inc., 1979 ($14.95).

The New York Racing Association takeout is currently 17% although it has been 14%. The California takeout is 15.75%. Of course the effect of breakage is to increase the average takeout somewhat beyond these figures.

Readers who want to know more about the calculation of winning probabilities based on the pari-mutuel odds should read Chapter 3 in *Horse Sense,* by Burton P. Fabricand, published by David McKay and Co., 1965. The book is hard to obtain, but I believe you can find it in the larger libraries.

Fabricand takes a sample of 10,000 races, with 93,011 horses and 10,035 winners (some dead heats). He finds that the average loss, from betting on the favorites (high pari-mutuel probability of winning), is considerably smaller than the average loss from betting the long shots (low pari-mutuel probability of winning).

For extreme favorites, the sample showed a profit and for horses with a pari-mutuel win probability of 30% or more the average loss was just a few percent. It ranged gradually higher as the odds lenghtened for horses with odds of 20 to 1 or more and pari-mutuel probabilities averaging about .025; the average loss to the bettors was 54 percent.

This indicates that the odds, from the pari-mutuel pool for winners, are systematically biased; they can be improved by incorporating a correction factor based on a data sample similar to Fabricand's. The correction would increase the probabilities assigned to the favorites and decrease the probabilities assigned to the long shots systematically.

A more readily available source for the same information is Fabricand's latest book *The Science of Winning*, published by Van Nostrand Reinhold in 1979. On page 37, a table shows how a player's expectaton varies with the odds. The sample has 10,000 races with 10,035 winners because of dead heats.

In 1984 a book by Ziemba and Hausch, *Beat the Racetrack: A Scientific Betting System*, appeared with a practical winning method at the track. I went to Hollywood Park with the author William T. Ziemba, Ph.D., and used the system successfully. The idea of true win probabilities discussed in this chapter is used by them to check the place and how pari-mutual pools. When horses in these pools are significantly under bet, they offer positive expectations. Good bets appear on average about once per two races.

Chapter 7

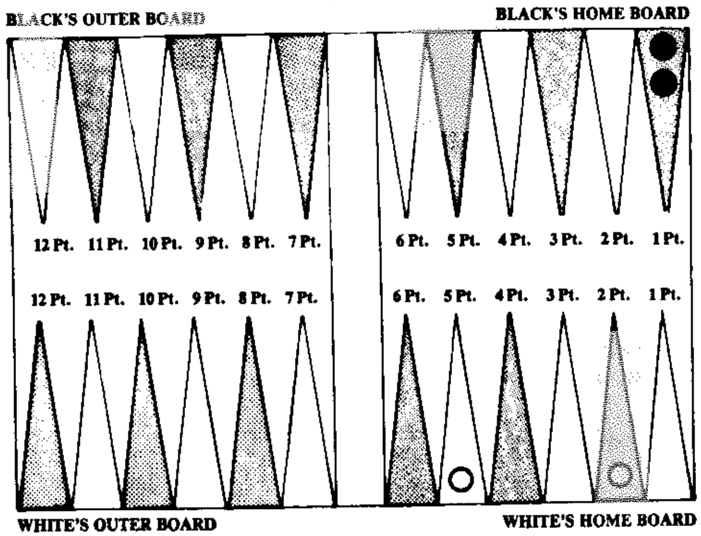
Backgammon

Backgammon has taken its place alongside bridge as a favorite pastime of sophisticated gamers. It is essentially a racing game, where each player tries to get his pieces off the board first. But, thanks to the doubling cube, it's also a gambling game, played for high stakes in clubs across the country. The basics of this intriguing board game are really very easy to master. (See *The Rules of Backgammon*, pp. 84-85.)

This chapter will focus on several aspects of backgammon that can be solved mathematically. You will learn useful but simple odds for bearing off with only two men left. Most good players already know this. But don't go away, good players. Later you will learn facts about backgammon that few in the world are aware of.

As an introduction to end positions, suppose you are White and it is your turn to roll in the position of Diagram 1. The doubling cube is in the middle.

Diagram 1



Questions:

1. What is your chance to win?
2. Should you double?
3. How much do you gain or lose by doubling?
4. If you double, should Black accept?
5. How much does Black gain or lose by accepting your double?

White wins only if he bears off on his next roll. So to help us solve end positions of this type, we calculate a table of chances to take off men in one roll. The exact result is given in Table 7-1, and the chances to the nearest percent are given in Table 7-2.

As you can see from Table 7-1, the exact chances of winning if you have a man on the five point and a man on the two point are 19 in 36 or .5277. . . Table 7-2 gives your chances to the nearest percentage, or 53%. Now you have the answer to question 1.

To see how Table 7-1 is calculated, recall that there are 36 *equally likely* outcomes for the roll of two dice. These are listed in Table 7-3.

Table 7-1

a man on the	0 pt	1 pt	2 pt	3 pt	4 pt	5 pt	6 pt
0 pt	off	36	36	36	34	31	27
1 pt	36	36	36	34	29	23	15
2 pt	36	36	26	25	23	19	13
3 pt	36	34	25	17	17	14	10
4 pt	34	29	23	17	11	10	8
5 pt	31	23	19	14	10	6	6
6 pt	27	15	13	10	8	6	4

Table 7-2

a man on the	0 pt	1 pt	2 pt	3 pt	4 pt	5 pt	6 pt
0 pt	off	100%	100%	100%	94%	86%	75%
1 pt	100%	100%	100%	94%	81%	64%	42%
2 pt	100%	100%	72%	69%	64%	53%	36%
3 pt	100%	94%	69%	47%	47%	39%	28%
4 pt	94%	81%	64%	47%	31%	28%	22%
5 pt	86%	64%	53%	39%	28%	17%	17%
6 pt	75%	42%	36%	28%	22%	17%	11%

Table 7-3

second die shows ▸ / ▾ first die shows	1	2	3	4	5	6
1	1-1	1-2	1-3	1-4	1-5	1-6
2	2-1	2-2	2-3	2-4	2-5	2-6
3	3-1	3-2	3-3	3-4	3-5	3-6
4	4-1	4-2	4-3	4-4	4-5	4-6
5	5-1	5-2	5-3	5-4	5-5	5-6
6	6-1	6-2	6-3	6-4	6-5	6-6

Think of the two dice as labelled "first" and "second." It might help to use a red die for the "first" die and a white one for the "second" die. Then if the red (first) die shows 5 and the white (second) die shows 2, we call the outcome 5-2. If instead the first die shows 2 and the second die shows 5, this is a different one of the 36 rolls and we call it 2-5. Outcomes are named x-y where x is the number the first die shows and y is the number the second die shows.

To see that White has 19 chances in 36 to win in the situation presented in Diagram 1, we simply count winning rolls in Table 7-3. If either die shows at least 2 and the other shows at least 5, White wins. He also wins with 2-2, 3-3, and 4-4. This gives the 19 (shaded) winning outcomes in Table 7-3.

As another example, suppose the two men to bear off are both on the six point. Then if two different numbers are rolled, White can't come off in one turn. Of the six doubles, only 3-3 or higher works. This gives four ways in 36 or 11%, in agreement with Tables 7-2 and 7-3. This simple counting method produces all the numbers in Table 7-1.

Now we are ready to answer question 2. Should White double, in Diagram 1? The answer is yes, and here's why. We have seen

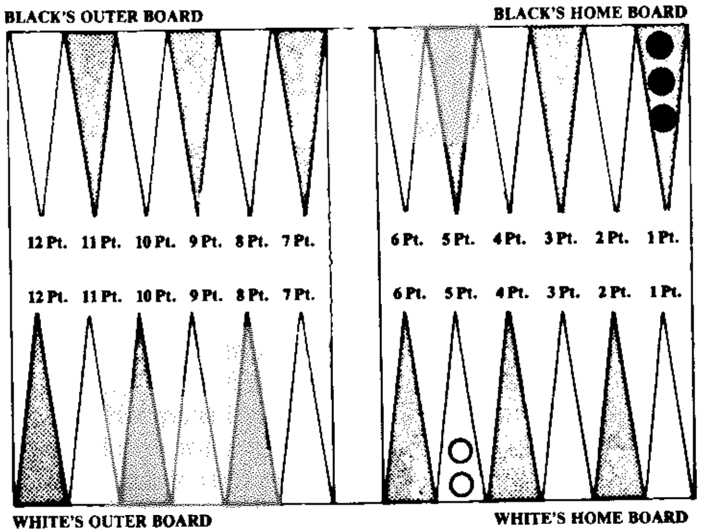
that White wins, on average, 19 times in 36. If we say the stake is one unit, then if he does *not* double, in 36 times he wins one unit 19 times and loses 1 unit 17 times for a gain of two units/36 times = 1/18 = 0.055. . . If White *does* double, Black can either accept or fold. Suppose Black accepts. Then the stakes are two units and a calculation like the previous one shows White gains an average of four units /36 times = 1/9 = 0.111. . . unit per time. White gains twice as much by doubling as by not doubling. If Black folds instead, then White wins one unit at once, which is even better.

This also answers the rest of the questions. In answer to question 3, White gains an extra 5.55% of a unit, on average, by doubling. Answer to question 4: Black should accept. He loses 1/9 unit on average by accepting and one unit for sure by folding. This answers question 5: if Black makes the error of folding, he loses an extra 8/9 unit or 89%.

The usefulness of Table 7-2 is generally limited to situations where you have just one or two rolls left before the game ends. But it is surprising how often the Table is valuable. Here are some more examples to help alert you to these situations. In Diagram 2, Black has the doubling cube. White has just rolled 2-1. How does he play it?

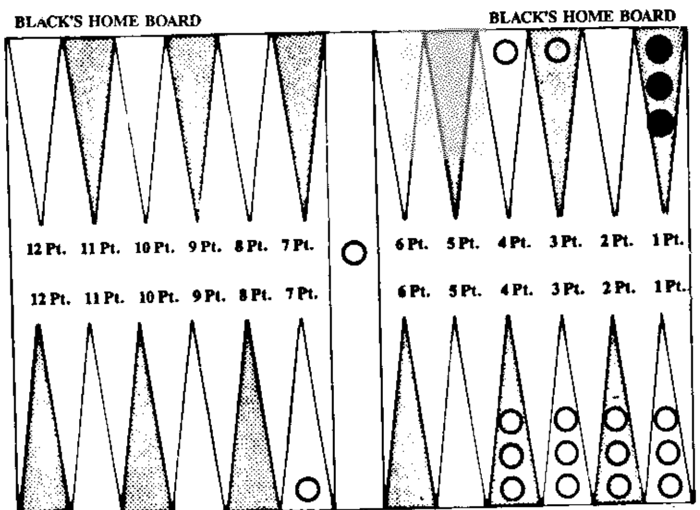
If Black rolls double on the next turn, he wins at once and it won't matter what White did. So White only needs to consider the case where Black does not roll doubles. Then White will have one more turn, and he wants to leave himself with the greatest chance to bear off on that turn. White can move one man from the 5 point to the 4 point and one man from the 5 point to the 3 point. By Table 7-2, this gives him a 47% chance to win if Black does not roll doubles. Or, White can move one man from the 5 point to the 2 point, leaving the other man on the 5 point. This gives him a 53% chance to win if Black does not roll doubles, so this is the best way to play the 2-1.

Diagram 2



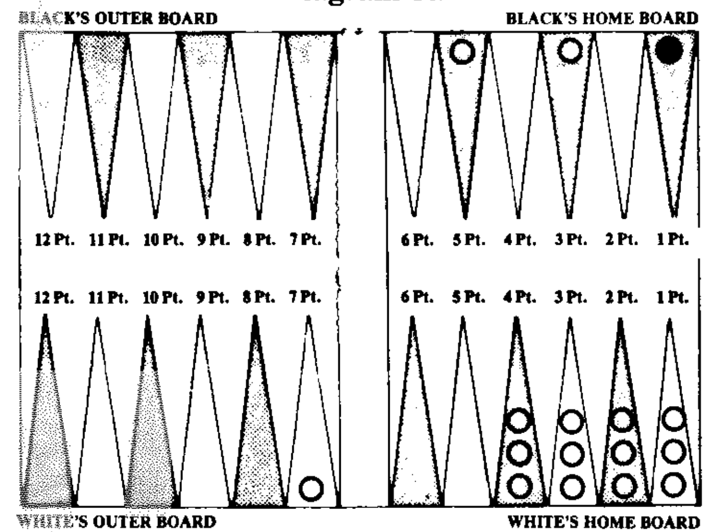
In Diagram 3, White's problem is to avoid a backgammon: if Black wins before the White men escape from Black's home board, Black will win 3 units. Otherwise, he will only gammon White for two units.

Diagram 3



White has just rolled 4-1. He must use the 4 to move the man on the bar to the Black 4 point (dotted circle). White can then move this man on to the Black 5 point, in which case, if Black does not roll doubles, White's situation on his last turn is shown in Diagram 3a.

Diagram 3a



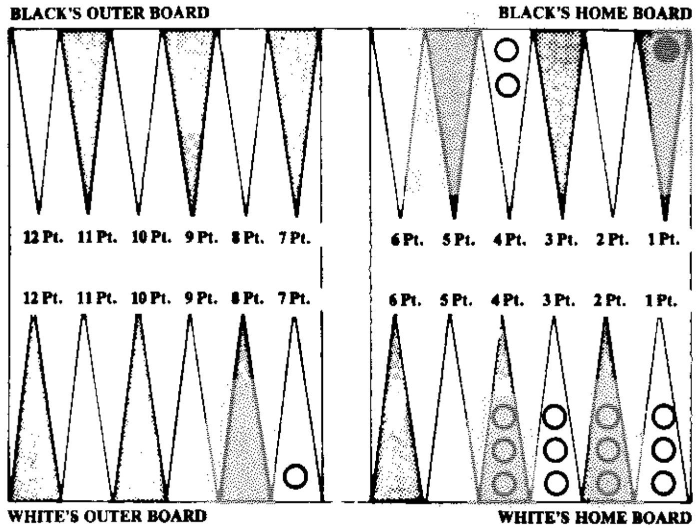
The chance for White to remove both men from Black's home board on the next roll is the same as the chance to bear off both men when one is on the 4 point and the other is on the 2 point. According to Table 7-2, this is 64%.

Suppose instead White plays both men to the Black 4 point. Then Diagram 3b shows the board if he survives Black's next roll.

His chance to save himself from backgammon is the same as bearing off two men from the 3 point in one roll. Table 7-2 gives 47%. Therefore, the play in Diagram 3a is best.

If instead White rolled 4-2 in Diagram 3, he could enter on 2 and move his other man to the 7 point, giving an 86% chance (Table 7-2, man on 5 point and man on 0 point) to escape Black's home board on the next roll. Or he could play to leave his two

Diagram 3b



back men on the Black 5 and 4 points. This gives only a 69% chance and is the inferior choice.

An outstanding reference work is *Backgammon* by Paul Magriel, The New York Times Publishing Company, 1977, $20. Most of Table 7-1 appears there on page 404. A convenient reference for practical play is the "Backgammon Calculator," Doubleday, 1974, $1.95. This handy cardboard wheel has most of Table 7-2 on the back.

Here are some questions to check your understanding. Refer to Diagram 2, assuming Black has the doubling cube and White has just rolled 2-1.

1. Should Black double after White makes the best move?

2. How much would Black gain or lose by so doubling?

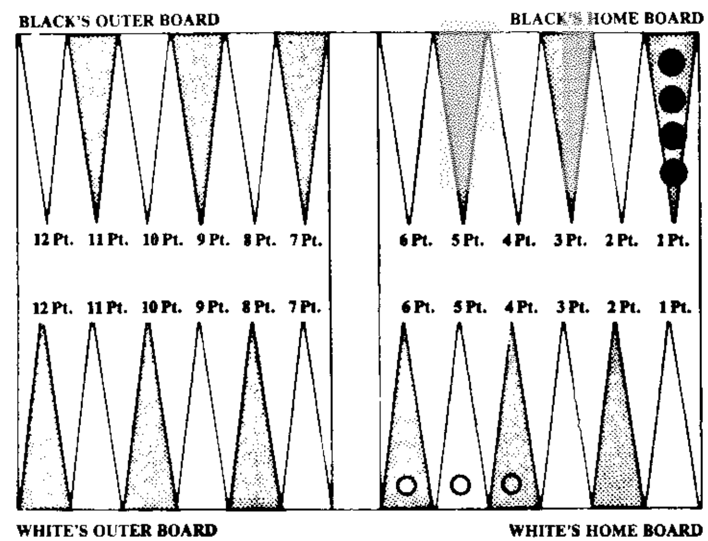
3. Should White accept a Black double? If he does, instead of folding, how much does he gain or lose?

4. What is the best way for White to play 3-2 in Diagram 2?

We will now present the complete, exact solution to all backgammon positions when each player has only one or two men left in his own home board. Don Smolen and I calculated it in 1975 and kept it to ourselves for several years.

I realize that it is often not practical or desirable to use the tables I provided during the game. Fortunately, many of these situations are covered by a handy rule that appeared in a "Sheinwold on Backgammon" column in the *Los Angeles Times*. Sheinwold considers the situation in Diagram 4. The problem is whether White, having rolled 6-2, should play the 2 so that he leaves his two men on 5 and 2 or 4 and 3.

Diagram 4



We solved this same problem earlier when discussing Diagram 2. We saw then from Table 7-2 that leaving men on 5 and 2, is best because it gives White a 53% chance to get off on the next turn, whereas leaving men on 4 and 3 gives only a 47% chance. Now consider the general question: If you have to leave one or two men after your turn, what is the best "leave"? Assuming that the positions between which you must choose have the *same pip count*, the correct rules, which Sheinwold gives, is:

Rules for Leaving One or Two Men

(1) If possible, leave one man rather than two.
(2) If you must leave two men, leave them on different points, if possible.
(3) If you still have a choice, move off the 6 point.
(4) If you are already off the 6 point, move the man on the lower point.

It is easy to prove these rules correct by using Table 7-2. This is shown again here in condensed form as Table 7-4.

Table 7-4

a man on the	1 pt	2 pt	3 pt	4 pt	5 pt	6 pt
0 pt	100% 1 pip	100% 2 pips	100% 3 pips	94% 4 pips	86% 5 pips	75% 6 pips
1 pt	100% 2 pips	100% 3 pips	94% 4 pips	81% 5 pips	64% 6 pips	42% 7 pips
2 pt		72% 4 pips	69% 5 pips	64% 6 pips	53% 7 pips	36% 8 pips
3 pt			47% 6 pips	47% 7 pips	39% 8 pips	28% 9 pips
4 pt				31% 8 pips	28% 9 pips	22% 10 pips
5 pt					17% 10 pips	17% 11 pips
6 pt						11% 12 pips

To check the rules, we simply check Table 7-4 for each pip count to see if it always tells us which of two "leaves" to pick. For example, with a pip count of 6, part (1) of the rule says correctly that 0 pt. -6 pt. is best. Then (2) says correctly that among the three remaining two-men positions, 3 pt. -3 pt. is worst. In a similar way the rule is verified in turn for positions with pip counts of 4, 5, 6, 7, 8, and 10. There's nothing to check for pip counts of 1, 2, 3, and 9 because the choices are equally good for these pip counts. There's nothing to check for counts of 11 and 12 because for these pip counts there is only one choice of position.

More examples illustrating the rule appear in *How Good are You at Backgammon: 75 Challenging Test Situations* by Nicolaos and Vassilios Tzannes, Simon and Shuster, 1974. You can use these rules to solve at once test situations 40, 41, 42, and 43. The authors give a rule (page 94) but it is neither as clear nor as simple as ours.

We proved the rule for leaving one or two men just for the case where you will have at most one more turn to play. In that case,the percentages in Table 7-4 let us compare two positions to see which is better. What if there is a chance that you'll have more than one turn? This could happen, for instance, if we change Diagram 4 so that Black has five men on the one point instead of four. Then Black could roll non-doubles on his next turn, leaving three men on the 1 point; White could roll 1-2 on his next turn, reducing his 5 pt. -2 pt. position to one man on the 4 point: Black could roll non-doubles again, leaving one man on the 1 point; and White then gets a second turn. It turns out that the rule gives the best choice against *all* possible positions of the opponent, not just those where you will have at most one more turn to play. (Note: There is one possible, unimportant exception that might arise, but the error is at most a small fraction of a percent.)

Now we return to the Thorp-Smolen solution of all end games with just one or two men in each home board. We will label home board positions as follows: 5 + 3 where there is a man on the 5 point and a man on the 3 point, with the largest number first. With both men on, say, the 4 point, we call the position 4 + 4.

With only one man on the 5-point we write 5 + 0. Think of the 0 as indicating that the second man is on the 0 = "off" point.

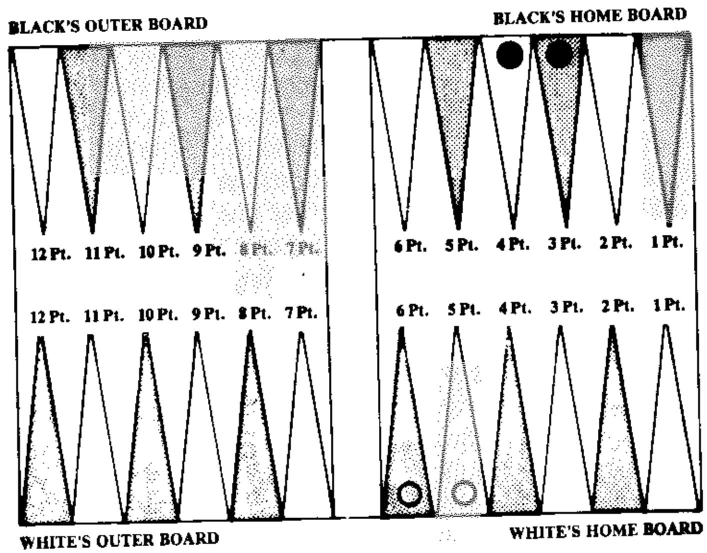
There are six home board positions with one man, namely 1 + 0, 2 + 0, . . . 6 + 0. There are 21 home board positions with two men. Thus there are 27 one- or two-man positions for each player.*

Table 7-5 gives the first part of our solution. It tells Player One's "expectation," rounded to the nearest percent, if One has the move and Two owns the cube. By One's expectation we mean the average number of units One can expect to win if the current stake is "one unit" and if both players follow the best strategy. Of course, if a player doesn't follow the best strategy, his opponent can expect on average to do better than Table 7-5 indicates.

The A above 6 + 0 means this column also applies to any count of up to 3 pips: 1 + 0, 2 + 0, 1 + 1, 3 + 0, or 2 + 1. The C above 6 + 0 means that this column also applies to 4 + 0, 3 + 1, 5 + 0, or 4 + 1. The A for Player One means the same as for Player Two.

We will illustrate the use of the Table with Diagram 5.

Diagram 5



*See Appendix H, pg. 140.

Table 7-5

Two has → / One has ↓	A,C 6+0	2+2	3+2	4+2 5+1	5+2	3+3 4+3	6+1	5+3	6+2	4+4	6+3	5+4	6+4	5+5	6+5	6+6
2+1 A	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
3+1, 4+0	89	90	90	91	94	95	95	96	96	97	97	97	98	98	98	99
5+0	72	74	75	78	85	87	88	89	90	92	92	92	94	95	95	97
4+1	61	63	65	70	78	82	84	85	86	88	89	89	91	94	94	96
6+0	50	53	56	61	72	76	79	81	82	85	86	86	89	92	92	94
2+2	44	48	51	57	69	74	77	78	80	83	85	85	88	91	91	94
3+2	39	42	46	52	66	71	75	76	78	81	83	83	86	90	90	93
4+2, 5+1	28	32	36	44	60	66	70	72	74	78	80	80	34	88	88	92
5+2	06	11	16	27	48	55	60	63	66	71	73	73	79	84	84	89
3+3	-06	00	06	18	41	50	56	59	62	68	71	71	77	82	82	88
4+3	-06	00	06	18	41	50	56	59	61	67	70	70	76	82	82	88
6+1	-17	-10	-04	09	35	45	51	54	57	64	67	67	74	80	80	87
5+3	-22	-15	-09	05	32	41	48	51	54	61	64	64	71	78	78	85
6+2	-28	-21	-14	01	29	38	45	49	52	59	63	63	70	77	77	84
4+4	-39	-31	-23	-08	22	33	40	44	48	55	59	59	67	74	75	82
6+3	-44	-36	-28	-12	18	28	36	40	44	51	55	55	63	71	71	79
5+4	-44	-37	-30	-14	17	28	35	39	43	51	54	55	62	70	70	78
6+4	-56	-48	-40	-23	06	17	25	29	33	42	46	46	55	63	63	72
5+5	-67	-59	-51	-35	-02	10	19	24	28	37	41	42	51	59	60	70
6+5	-67	-59	-51	-36	-07	03	11	15	19	28	32	33	43	51	53	63
6+6	-78	-71	-64	-50	-25	-17	-09	-05	-01	07	12	12	23	32	36	48

It is White's turn to move so he becomes Player One. Player Two, or Black, has the cube. We look along the row 6 + 5 and the column 4 + 3. Table 7-5 shows Player One's (White's) expectation as 03, so White has a 3% advantage. He expects to win on average 3% (more exactly, 2.54%) of the current stake. If the current stake is \$1,000, White should accept a Black offer to "settle" the game if Black offers more than \$25.40. If Black offers less, White should refuse.

Table 7-6 gives the expected gain or loss (to the nearest percent) for Player One when he has the move and the doubling cube is in the middle.

Unlike Table 7-5, in this case One has the option of doubling before he moves. If One does not double, Two will be able to double on his turn. If One doubles, Two then has the choice of accepting the double or folding. If Two accepts, play continues with doubled stakes and Two gets the cube. If Two folds, he loses the current (undoubled) stakes and the game ends.

Table 7-7 gives the expected gain or loss for Player One when he has the move and the doubling cube. The columns for 6 + 4, 5 + 5, 6 + 5, and 6 + 6 are the same as for Table 7-6 so they have been omitted.

In this case, One has the option of doubling before he moves. However, in contrast to Table 7-6, if One does not double, he keeps the cube so Two cannot double on his next turn. If One does double, Two can accept or fold. If he accepts, the stakes are doubled, play continues, and Two gets the cube. If instead Two folds, he loses the current (undoubled) stake and the game ends. Table 7-8 also tells whether One should double and whether Two should accept when One has the cube.

Doubling strategy is the same whether One has the cube or it is in the middle, except for the shaded region. If he makes the mistake of doubling, Two should accept. When the cube is in the middle, One should double for positions in the shaded regions and Two should accept.

We will now show how to use the tables to play *perfectly* in any of the 27 × 27 = 729 end positions covered by the tables. We'll

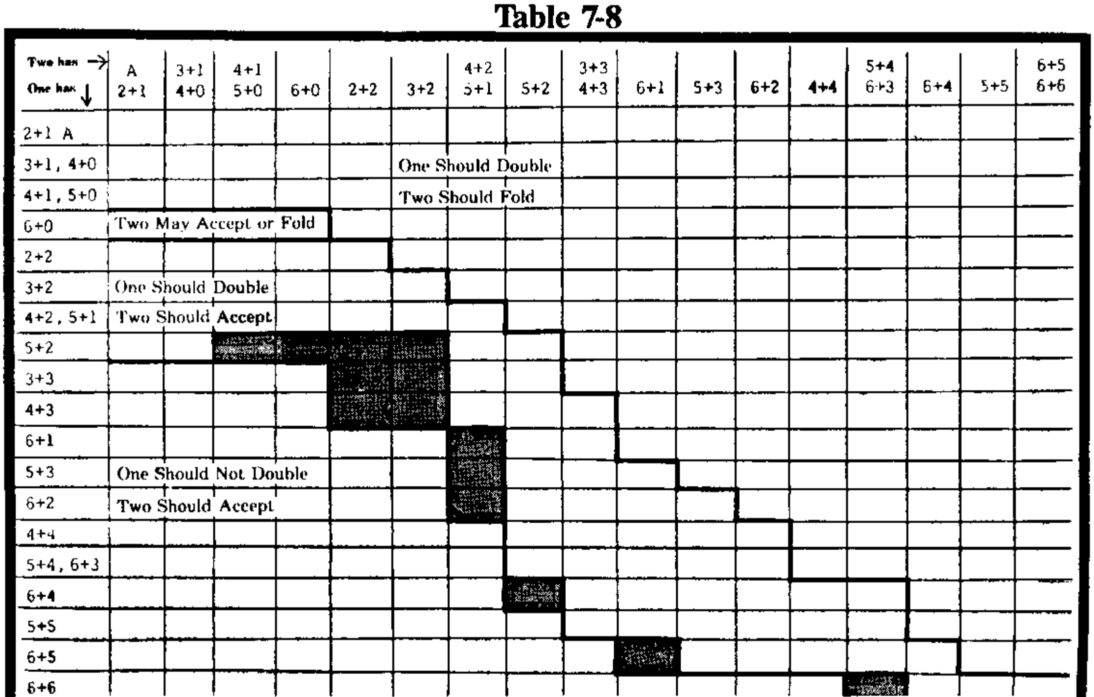
Table 7-6

One has ↓ / Two has →	A,C 6+0	2+2	3+2	4+2 5+1	5+2	3+3 4+3	6+1	5+3	6+2	4+4	6+3	5+4	6+4	5+5	6+5	6+6
6+0 A,C	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
2+2	89	95	100	100	100	100	100	100	100	100	100	100	100	100	100	100
3+2	78	85	91	100	100	100	100	100	100	100	100	100	100	100	100	100
4+2, 5+1	56	64	72	88	100	100	100	100	100	100	100	100	100	100	100	100
5+2	11	22	32	53	95	100	100	100	100	100	100	100	100	100	100	100
3+3, 4+3	-06	01	12	36	83	100	100	100	100	100	100	100	100	100	100	100
6+1	-17	-10	-04	19	70	89	100	100	100	100	100	100	100	100	100	100
5+3	-22	-15	-09	10	63	82	95	100	100	100	100	100	100	100	100	100
6+2	-28	-21	-14	01	57	77	91	98	100	100	100	100	100	100	100	100
4+4	-39	-31	-23	-08	44	66	81	88	96	100	100	100	100	100	100	100
6+3	-44	-36	-28	-12	36	57	72	80	88	100	100	100	100	100	100	100
5+4	-44	-37	-30	-14	34	55	71	78	86	100	100	100	100	100	100	100
6+4	-56	-48	-40	-23	13	34	50	58	67	83	91	92	100	100	100	100
5+5	-67	-59	-51	-35	-01	21	39	47	56	74	83	83	100	100	100	100
6+5	-67	-59	-51	-36	-05	08	21	30	38	56	65	66	86	100	100	100
6+6	-78	-71	-64	-50	-22	-10	-02	03	07	16	24	25	46	65	71	96

Table 7-7

Two has → / One has ↓	A 2+1	3+1 4+0	5+0	4+1	6+0	2+2	3+2	4+2 5+1	5+2	3+3 4+3	6+1	5+3	6+2	4+4	6+3	5+4
6+0 A,C	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
2+2	89	89	89	89	89	95	100	100	100	100	100	100	100	100	100	100
3+2	78	78	78	78	78	85	91	100	100	100	100	100	100	100	100	100
4+2, 5+1	56	56	56	56	56	64	72	88	100	100	100	100	100	100	100	100
5+2	11	11	19	24	29	32	34	53	95	100	100	100	100	100	100	100
3+3, 4+3	-06	00	09	15	21	24	27	36	83	100	100	100	100	100	100	100
6+1	-17	-10	-00	06	13	16	19	25	70	89	100	100	100	100	100	100
5+3	-22	-15	-05	02	08	12	15	22	63	82	95	100	100	100	100	100
6+2	-28	-21	-10	-03	04	08	11	18	57	77	91	98	100	100	100	100
4+4	-39	-31	-20	-12	-04	-00	04	11	44	66	81	88	96	100	100	100
6+3	-44	-36	-24	-16	-08	-04	00	08	36	57	72	80	88	100	100	100
5+4	-44	-37	-25	-17	-09	-05	-01	07	34	55	71	78	86	100	100	100
6+4	-56	-47	-35	-26	-18	-14	-10	-01	16	34	50	58	67	83	91	92
5+5	-67	-58	-45	-36	-27	-23	-18	-10	08	21	39	47	56	74	83	83
6+5	-67	-58	-45	-37	-29	-24	-20	-12	05	14	22	30	38	56	65	66
6+6	-78	-70	-57	-49	-41	-37	-33	-25	-08	00	08	13	17	25	30	30

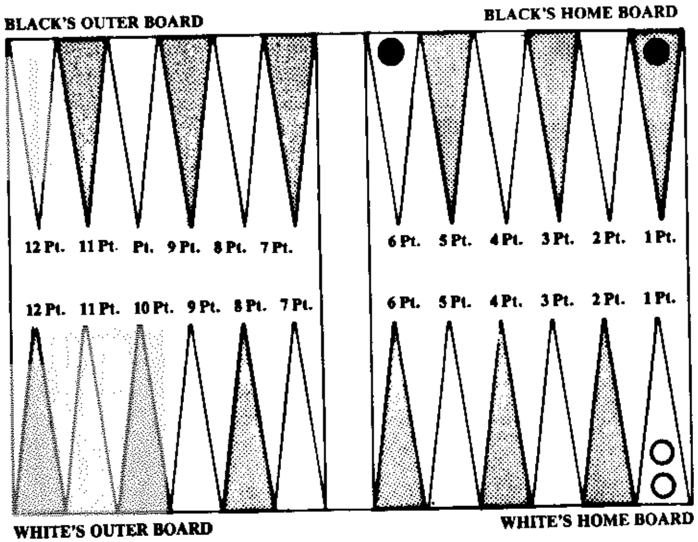
Table 7-8



run through sample end games step by step, showing player expectation, doubling strategy, and the best way to play each roll.

I earlier referred to a book entitled *How Good Are You at Backgammon: 75 Challenging Test Situations* by Nicolaos and Vassilios Tzannes, Simon and Shuster, 1974. Consider first Situation 74 from the Tzannes' book. This is shown in Diagram 6.

Diagram 6



It is Black's turn so he is Player One. Black doubles. Should he? If he does, should White accept? The cube is in the middle. We look in Table 7-8, row $6+1$, column $1+1$. Black should not double. If he does, White should accept. (This is correctly recommended by Tzannes' book.) Table 7-6 shows that Black's expectation under best play, which means not doubling, is -17%. If instead Black has the cube, we use Tables 7-7 and 7-8. In this example we get exactly the same answer. This isn't always the case, though, as we will see.

This example is also easy to analyze directly. If Black bears off in his next turn he will win. The chances are 15/36 (Table 7-1). If he does not bear off at once, White will win and Black will lose. So if the current stake is 1 unit, and Black does not double, Black's expected gain is $+1$ unit X $15/36 - 1$ unit X $21/36 = -6/36 = -16\ 2/3\%$. Now suppose Black doubles and White accepts. Then Black's expected gain is $+2$ units X $15/36 - 2$ units X $21/36 = -12/36 = -33\%$. On average Black will lose an extra 16 2/3% of a unit if he makes the mistake of doubling and White accepts.

It's easy to see from this type of reasoning that if Player One has any two-man position and Player Two will bear off on the next turn, then Player One should not double (if he can) when his chance to bear off in one roll is less than 50%. If his chance to bear off is more than 50%, he should double. Referring to the same Table 7-1 proves this rule which the Tzannes cite for these special situations:

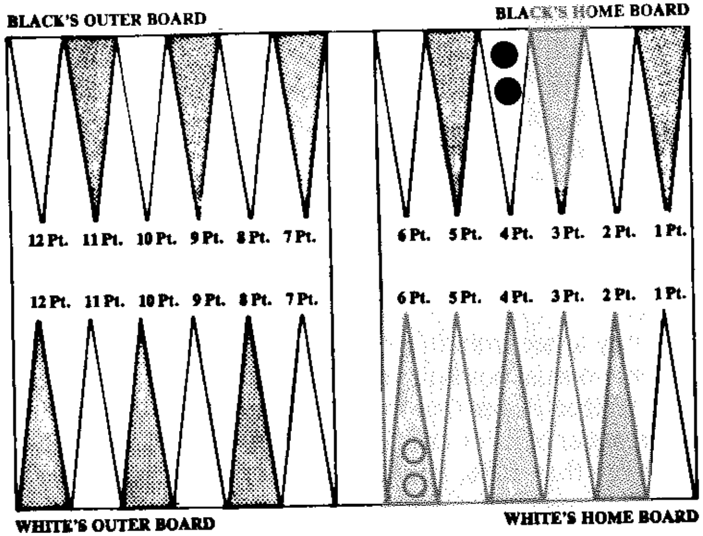
> *With double three, six-one, six-two (or*
> *anything worse)*
> *Keep dumb, hope for the best. Anything*
> *better, don't delay,*
> *Double the stakes with zest.*

The Tzannes' Situation 73 is similiar.

Here is a trickier situation that I don't think you could figure out without help from the tables. Suppose White has $6+6$, Black has $4+4$, White is to roll and the doubling cube is in the middle. This is shown in Diagram 7. Should White double? How does the game proceed for various rolls?

White is Player One. He consults Table 7-6 and sees his expectation is 16%. But Table 7-8 tells White not to double. We now show how to use that table to play optimally for a sample series of rolls. Suppose White rolls 3-1. How does he play it? He can end up with $6+2$ or with $5+3$.

Diagram 7



The rule stated earlier says that 5+3 looks better because it gives him a greater chance to bear off on the next turn. This is proven by the tables as follows: after White plays, it will be Black's turn. Black will be Player One with 4+4, White will be Player Two with either 5+3 or 6+2. The cube will be in the middle. Which is best for White? Consult Table 7-6. We find Player One (Black) has an expectation of 88% if White has 5+3 whereas Black has 96% if White has 6 + 2. White wants to keep Black's expectation down so he plays to leave 5 + 3.

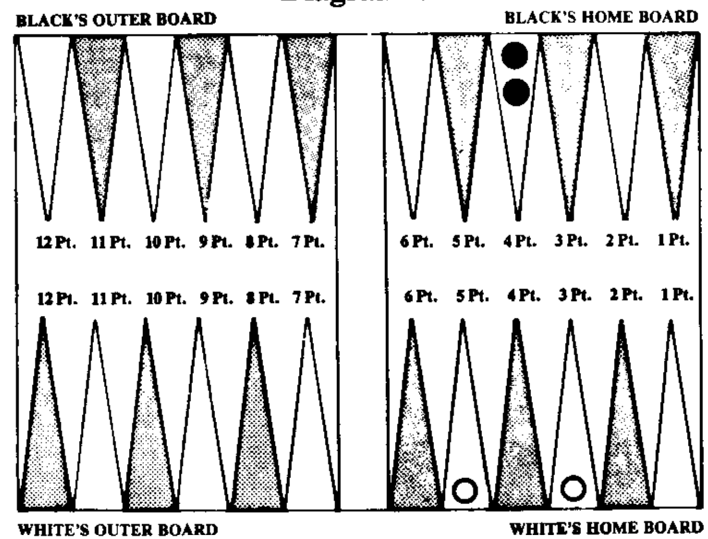
The situation after White makes this move is shown in Diagram 8.

Black is to roll and the cube is in the middle. Should Black double? Should White accept? Table 7-8 says Black should double and White should accept. Table 7-6 says Black's expected gain is 88% of the one-unit stake.

Next Black rolls 2-1. He can leave 4+1 or 3+2. The rule mentioned says 4+1 is better. To confirm this, note that after Black moves, White will be Player One with 5+3, Black will be Player

Two with either 4+1, or 3+2 and White will have the cube. Therefore we consult Table 7-7, not Table 7-6. If Black leaves 4+1, White's expectation is 2% of the current two-unit stake. If Black leaves 3+2, White's expectation is 15%. Therefore Black leaves 4+1.

Diagram 8



It is now White's turn. The situation is shown in Diagram 9.

The stake is 2 units, White's expectation is 2% of 2 units or .04 unit and White has the cube. What should he do? Table 7-8 tells us White should not double.

White now rolls 5-2, leaving 1+0. Black does not have the cube. Table 7-5 gives his expectation as 61% of 2 units or 1.22 units. He wins or loses on this next roll.

The tables show certain patterns that help you to understand them better. For instance, for a given position it is best for Player One to have the cube. It is next best for Player One if the cube is in the middle and it is worst for Player One for Player Two to have the cube. Therefore for a given position, Player One's expectation is greatest in Table 7-7, least in Table 7-5, and in between in

Diagram 9

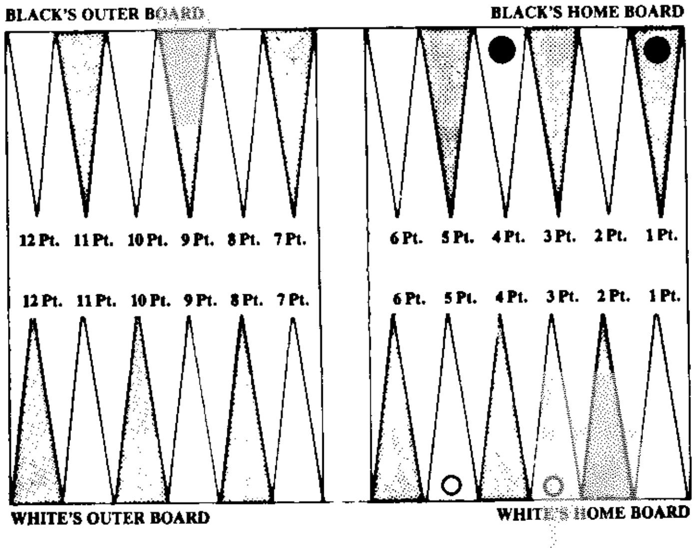


Table 7-6. For instance, with Player One having 6 + 6 and Player Two having 4 + 4, Player One's expectation is 25% if he has the cube, 16% if it is in the middle, and 7% if player Two has the cube.

Sometimes two or even all of the expectations are the same. For instance, if Player One has 6+6 and Player Two has 6+5, Player One's expectation is 71% if he has the cube or if it's in the middle. If Player Two has the cube Player One's expectation drops to 36%.

Examination of the doubling strategies in Table 7-8 shows that the positions where Player One should double and Player Two should fold are the same whether Player One has the cube or the cube is in the middle. Although this happens for the two-man end positions we are analyzing here, it is not always true in backgammon. The positions where Player One should double and it doesn't matter if Player Two accepts or folds also are the same in Table 7-8. But some of the positions where Player One should

double and Player Two should accept are different. If Player One has the cube, Table 7-8 shows that he should be more conservative. Intuitively, this is because if he has the cube and does not double, he prevents Player Two from doubling, whereas if the cube is in the middle, Player Two cannot be prevented from doubling.

Table 7-8 leads to an example that will confound the intuition of almost all players. Suppose Player One has 5 + 2 and has the cube. Consider two cases: (a) Player Two has 1 + 0 and (b) Player Two has 6 + 0. In which of these cases should Player One double? Clearly 6 + 0 is a worse position than 1 + 0. And the worse the position the more likely we are to double, right? So of the four possible answers (double 1 + 0 and 6 + 0, double 1 + 0 but not 6 + 0, double 6 + 0 but not 1 + 0, don't double 1 + 0 or 6 + 0) we "know" we can eliminate "double 1 + 0, don't double 6 + 0," right? WRONG. The only correct answer, from Table 7-8 is: double 1 + 0 but don't double 6 + 0. Try this on your expert friends. They will almost always be wrong. If they do get it right they probably were either "lucky" or read this chapter. In that case if you ask them to explain why their answer is correct, they probably won't be able to.

You may think that the loss would be slight by doubling 6 + 0 erroneously. But you have an expected gain of 29% by not doubling (Table 7-7) whereas by doubling it can be shown that your expectation drops to only 11%.

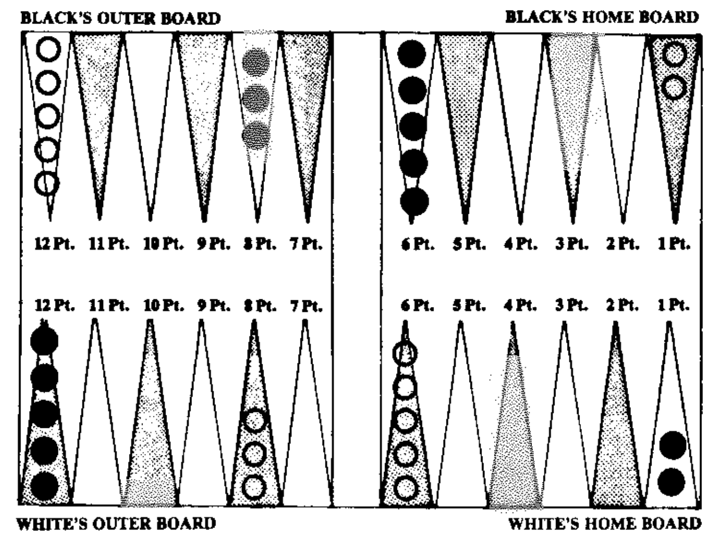
The exact explanation is complex. The basic idea, though, is that if Player One doubles Player Two, Player Two accepts, and Player One doesn't win at once, Player Two can use the cube against Player One with great effect at Players Two's next turn.

Jacoby and Crawford discuss what is essentially the same example (they give Player Two 4 + 1 instead of 6 + 0) on pages 116-117 of their excellent *The Backgammon Book,* Viking Press, New York, 1970. Table 7-8 shows that essentially the same situation occurs when Player One has 5 + 2 and Player Two has 4 + 1, 5 + 0, 6 + 0, 2 + 2 or 3 + 2 and for no other two-man end positions.

Tables 7-5, 7-6, 7-7, and 7-8 present the complete, exact solutions to two-man end games in backgammon. The tables were calculated by a general method I have discovered for getting the complete exact solution to all backgammon positions that are pure races (i.e. the two sides are permanently out of contact). The intricate and difficult computer programs for computing Tables 7-5 through 7-8 were written by Don Smolen so Tables 7-5 through 7-8 are our joint work. Don was a computer scientist at Temple University. He is now trading stock options on the floor of the American Stock Exchange. A skilled backgammon player, he won the 1977 American Stock Exchange tournament.

* * *

The Rules of Backgammon



The backgammon board is divided into two rectangles by the *bar*. Each side of the board contains six *points* of alternating colors. The game is played with light and dark pieces called *stones*, with the lighter color designated "White" and the darker color "Black." Each player has fifteen stones.

The stones are initially arranged as shown in the diagram above. In this case, the Black player would be seated at the bottom of the board, while White would be at the top. The points are numbered on the diagram for the sake of clarity; no numbers appear on an actual backgammon board. The six points in the upper righthand corner constitute Black's *inner table*. The six points at the lower right are White's *inner table*. The object of the game is to move your stones around the board until they are all in your inner table. In this case, White would move his stones counterclockwise and Black would progress clockwise. Once your stones are all in your inner table, you begin to *bear them off*, and the first player to remove all of his stones from the board wins the game.

Backgammon is played with two dice, which are shaken in and rolled from a cup. To begin play, each player rolls one die, and the high roll gets the first turn. The first move is determined by the two numbers which the opponents rolled. From then on, each player rolls two dice when it is his turn to move. The two numbers rolled dictate the players' moves as follows. Suppose the numbers on the dice are 5 and 3. The player may a.) move one stone five points and then three more, b.) move one stone three points then five more, or c.) move one stone five points and another stone three. When a double number is rolled, the player may make four moves. A roll of 3-3 would allow moving one stone 12 points, four stones 3 points each, or any other pattern involving groups of three.

When moving his stones around the board, no player may land on a point occupied by two or more *opposing* stones. Such a point is considered *made* by the opponent and often interferes with the way in which a player intended to take his turn. On the other hand, a point occupied by a single stone is a *blot*. This point is vulnerable to being *hit* by an opposing stone that lands on it. When a stone is hit, it is sent off the board and onto the bar, where it must remain until it can be entered on the opponent's inner table. The player whose stone has been hit is forbidden to make any other moves until he has entered his stone from the bar.

In order to enter, the player must roll a number which allows him to move to a point on his opponent's inner table which is not *made*. For example, say White has hit one of Black's blots, sending it to the bar. White has made points 3, 5, and 6 on his inner table. In this situation, Black must roll a 1, 2, or 4 in order to enter his stone.

Once a player has succeeded in moving all fifteen of his stones into his inner table, he may begin *bearing off*. This consists of removing stones from the points and off the board according to the numbers rolled. For example, a roll of 3-4 means that a player may remove stones from positions 3 and 4. If he has no stones on one or both of these points, he bears off from the next lowest point. The "race" to bear off continues until one player has taken

all his stones off the board. He, of course, is the winner.

A player is credited with having won a double game, or a *gammon*, if he bears off all his stones before his opponent has borne off any. If a player wins a game while his opponent still has a stone on the winner's inner table or on the bar, he has made a *backgammon* and wins triple the stakes.

If you play backgammon for money, a *doubling cube* is used. This cube bears the numbers 2, 4, 8, 16, 32, and 64. At any point during the game, one player may double the stakes. His opponent must either accept the double or forfeit the game. If he accepts, the opponent gains possession of the cube and may "turn the cube" back at *his* opponent whenever he feels he has the upper hand in the game. It's not hard to see that high stakes game can result very easily from a game in which the lead changes hands frequently.

Section Four

Money Management

The importance of money management and bet sizing has been stressed increasingly in recent years and rightly so. For even if the player has discovered a favorable betting situation, how he wagers determines his success or failure. Ultimately, it is the "bottom line" on which a gambler's performance is judged. It is fine, of course, to describe the favorable situation to a friend or business associate, but the next question is likely to be "How much money are you making from this situation?"

The problem for the gambler is that much of the advice on money management is conflicting or confusing, or simply based on false premises. There are hundreds of schemes designed to overcome the house edge in roulette and craps based solely on manipulating the size of one's bets. As will be seen, all such attempts are futile.

Even assuming the player has discovered a favorable game (i.e., one offering a positive expectation), the question naturally arises: How does one best use a limited amount of capital to exploit this positive expectation? Wager too boldly and the player risks losing his entire bankroll, even though he or she may have made only favorable bets. This is commonly known as gambler's ruin. On the other hand, betting too conservatively the player severely limits his opportunity to make a good return on his capital.

Fortunately for the player, there exists a mathematical theory for committing resources in favorable games. This will be discussed in Chapter 9.

Chapter 8

Mathematical Systems

Before looking at the optimal strategy for exploiting a positive expectancy situation, it may be worthwhile to evaluate what I refer to as mathematical systems. Although here I use roulette as an example, the principles apply equally to craps and the Wheel of Fortune.

By a "mathematical system" I mean a system where the player decides which bet to make using only the following information:

(1) a record of what numbers have come up on some number of past spins, and

(2) a record of the bets he has made, if any, on those spins.

We assume here that when the player bets, for him all numbers are equally likely to come up on each spin of the wheel. This means not using biased wheels or physical prediction method.

Roulette has long been the prototype of unbeatable gambling games. It is normally regarded as a repeated independent trials process which generates at each trial precisely one from a set of random numbers. Players may wager on particular subsets of random numbers (e.g., the first dozen, even, an individual number, etc.), winning if the number which comes up is a number of the chosen subset. A player may wager on several subsets simultaneously and each bet is settled without references to the others. To fix the discussion, let's consider the standard American wheel. This has thirty-eight numbers, namely 0, 00, 1, 2,. . .36.

The mathematician's *assumption*, that each of these numbers is equally likely beforehand to come up on any spin of the ball and wheel, seems plausible. The wheels are carefully machined and balanced by the manufacturer. They are checked from time to time by the casinos. When they show signs of wear they may be thoroughly reconditioned. Even if the wheel has irregularities which make some numbers more favored than others, if the player does not know this and his system is not designed to exploit this, then mathematical reasoning—based on the assumption that all numbers are equally likely to come up—gives correct conclusions about that player's system.

The Doubling-up System

One more assumption must be made to properly evaluate mathematical systems. We must also assume there is a smallest allowable house (minimum) bet (such as $1) and a greatest allowable house (maximum) bet (such as $1000). Casinos need to fix a maximum bet in order to stop the simple mathematical system of "doubling up." To see why, imagine we've found a casino with no maximum. We bet $1000, because Red pays even money or 1 for 1. If we lose, we double and bet $2000 on the second turn. If that wins, we net $1000 on the two turns. If the second bet loses, we double again and bet $4000 on the third turn. Having lost $3000 on the first two turns, a win of $4000 on the third turn nets $1000 on the cycle of three turns. We continue doubling our bet after each loss. Finally, when we win, we have a net gain of $1000. We put this $1000 safely aside and start a new cycle of doubling until we win with a bet of $1000 on the Red. Each completed cycle wins another $1000 net. Table 8-1 illustrates this cycle.

The doubling-up system in Table 8-1 with no casino limit on bets is being discussed *not* because anyone would ever be allowed

Table 8-1

turn #	amount bet	total profit if cycle ends on this turn	chance cycle ends on or before this turn: exact	decimal approximation
1	$ 1,000	$1,000	1-(20/38)	0.4737
2	$ 2,000	$1,000	$1-(20/38)^{2}$	0.7230
3	$ 4,000	"	$1-(20/38)^{3}$	0.8542
4	$ 8,000	"	$1-(20/38)^{4}$	0.9233
5	$ 16,000	"	$1-(20/38)^{5}$	0.9596
6	$ 32,000	"	$1-(20/38)^{6}$	0.9787
7	$ 64,000	"	$1-(20/38)^{7}$	0.9888
8	$ 128,000	"	$1-(20/38)^{8}$	0.9941
9	$ 256,000	"	$1-(20/38)^{9}$	0.9969
10	$ 512,000	"	$1-(20/38)^{10}$	0.9984
11	$1,024,000	"	$1-(20/38)^{11}$	0.9991
31	$1000 x 2^{30} or about a trillion	"	$1-(20/38)^{31}$	0.999,999,997,7
36	$1000 x 2^{35} or about 34 trillion	"	$1-(20/38)^{36}$	0.999,999,999,9
100	about $6 x 10^{32}	"	$1-(20/38)^{100}$	
n	$1000 x 2^{n-1}	$1,000	$1-(20/38)^{n}$	

to do it, but to illustrate ideas we will be using. To see how ridiculous the system would be, note that if the first ten turns of a cycle have lost, on the eleventh turn the player bets 1,024 times his initial bet. His initial bet was $1,000, so he bets $1,024,000. Of course the chance is small that this will happen. The last column shows a chance of 0.9984 that the cycle ends on or before the tenth turn, hence that the eleventh bet is never made. Thus, the chance of reaching the eleventh turn is only 1 −0.9984 = 0.0016 or 0.16% or about one chance in 613. But if the doubling-up system is used long enough, it will happen.

With 30 losses in a row, the player is supposed to bet about one trillion dollars on the thirty-first turn. This is about the net worth of the New York Stock Exchange. On turn 36, the bet is about $34 trillion. This exceeds the net worth of the world! (The net worth of the U.S.A. is about 6 trillion current dollars. I'd guess the net worth of the world to be about $30 trillion.) The player should arrange from the start to have unlimited credit, *reasonably* pointing out that since he must eventually win he is sure to pay off!

Real casinos don't go for this. They have house limits (which they may increase sometimes under special circumstances) and credit limits. So this "sure-fire winning system" is never used. But players for centuries have used modified doubling-up systems in actual casino play. An illustration is given in Table 8-2. Here the player starts by betting $1 on Red. He keeps doubling his bet until he wins. Then he starts the cycle over with a $1 bet on Red. Each cycle produces a $1 profit *unless*—and here is the catch—he loses ten times in a row and then wants to bet $1024 on the eleventh turn of the cycle. The house limit prevents that and prevents further doubling if the player loses on his eleventh turn.

Notice from Table 8-2 that if the player wins after nine or fewer losses, he wins $1 and successfully completes the cycle. But if he loses ten times in a row, he can bet only $1000 on the eleventh turn. If he then wins, he loses "only" $23 on this cycle. But if he loses on the eleventh turn, he loses $2023 on the cycle, for a major disaster. Of course, the chance of ever reaching the eleventh turn of a cycle is as we saw before, only about one chance in 613.

Is this system any good, or do the chances of loss on the eleventh turn ruin it?

We are going to find out that the "house percentage advantage" on Red is not changed *in the slightest* by the doubling-up system. In fact, the disaster of the eleventh turn is *exact* compensation to the casino for the high chance the player has of winning $1 per cycle. We will show this by a computation. But what is perhaps truly amazing is that this is also true for all mathematical systems, no matter how complex, including all those that can ever

Table 8-2

turn #	amount bet	total $ losses before bet	net profit if cycle ends, this turn	chance of this result: exact	decimal approximation
1	1	0	1	$18/38$	0.4737
2	2	1	1	$20/38 \times 18/38$	0.2493
3	4	3	1	$(20/38)^2 \times 18/38$	0.1312
4	8	7	1	$(20/38)^3 \times 18/38$	0.0691
5	16	15	1	$(20/38)^4 \times 18/38$	0.0363
6	32	31	1	$(20/38)^5 \times 18/38$	0.0191
7	64	63	1	$(20/38)^6 \times 18/38$	0.0101
8	128	127	1	$(20/38)^7 \times 18/38$	0.0053
9	256	255	1	$(20/38)^8 \times 18/38$	0.002789
10	512	511	1	$(20/38)^9 \times 18/38$	0.001468
11	1000	1023	-23	$(20/38)^{10} \times 18/38$	0.000773
			or -2023	$(20/38)^{11}$	0.000858
				total = 1	

be discovered. Since there are an infinite number of such systems, we cannot prove this by computation (an infinite amount of time would be needed to do the required infinite number of computations). Instead, I will indicate how the mathematician, by logic (like the logic of, say, plane geometry with its axioms, theorems and proofs) can show that none of this infinite number of systems is any good.

A lot of what I'm saying is easier than it sounds. For instance, to see that there are an infinite number of systems for roulette, all, I have to do is give you any endless list of systems. Here is one such list (always bet on Red); System 1. Bet $1 on Red if Red came up one turn ago; if it didn't come up one turn ago, bet $2. System 2. Always bet $1 on Red if it came up two turns ago; if it did not come up two turns ago, bet $2. And so on for systems 3, 4, . . . etc.

I didn't say my list of systems would be interesting, only that it would be endless!

The doubling-up system can be good for some fun even if it doesn't alter the house edge. Suppose you're in Las Vegas with your spouse or your date. It's almost dinner time and you say casually, "Dinner for two will run us about thirty dollars. Why don't we eat for free? I'll just pick up $30 at this roulette wheel. It'll only take a few minutes." If you have $2100 in your pocket and the house limits are from $1 to $1000 on Red, you can use the doubling-up system. You need to complete 30 cycles without ever having a string of eleven losses. You will win $1 per cycle, for a total of $30, and be off to dinner.

How safe is this scheme? What are your chances? Table 8-1 says that the chance a cycle lasts 10 turns or less, and therefore you win $1, is 0.9984. The chance that you do this 30 times in a row turns out to be 0.9984^{30} or 0.9522, so the chance you will succeed is over 95%. If you set your sights lower, say $20 or $10, then the chances of success go up to 96.79% and 98.38%, respectively. But be warned: if you fail, you can lose as much as $2023.

An important factor in determining the risk of failure is the ratio of the house maximum bet on Red to the minimum bet. To illustrate, suppose instead of $1 to $1000 for a ratio of 1000, the betting limits were $2 to $500, for a ratio of 500/2 = 250. Then if we start a cycle with a $2 bet, we hit the house limit on the ninth spin, after eight losses. (To see this, use Table 8-2 and double all the numbers in the second, third and fourth columns, because we *start* with a $2 bet rather than a $1 bet, as before.) Now the chance the cycle ends in eight turns or less is (from the last column of Table 8-1) 0.9941. Thus to win $30 you need to complete 15 cycles, the chance of which is 0.9941^{15} or 0.9152. If you try this in a roulette game with better odds, say single-zero European style, the chance of success increases.

The doubling-up system is one of a class of systems that are sometimes called martingales. The origin of the term is given in the American Heritage Dictionary, New College Edition, which is the most informative definition I have seen on this. The word evolved from a similarly named village of Martigues in the Provence district of southern France, whose residents were viewed as peculiar and were roundly ridiculed with Gallic expertise. Their bizarre behavior included such things as gambling with the doubling-up system and lacing up their pants from behind. To use the doubling-up system became known as gambling "al la martigalo" (fem), "in the Martigues manner," i.e., "in a ridiculous manner."

There are many other popular "mathematical" systems. "Tripling up," where the player bets 1,3,9,27, etc. until he wins, then repeats, is like doubling up, but it wins faster and runs into trouble (in the form of the house limit) faster.

If you want to know more about "mathematical systems," consider these books:

The book *Casino Gambling, Why You Win, Why You Lose,* by Russell T. Barnhart (Brandywine, N.Y., 1978, $12.95). Barnhart is a skilled magician and a longtime student of gambling. He has gambled extensively all over the world so he knows both the theory and practice of his subject. The book has 50,000 spins from an actual wheel and an elaborate discussion of mathematical or "staking" systems.

Allan Wilson's classic *Casino Gambler's Guide* has considerable material on systems and their fallacies. His treatment of biased roulette wheels may be the best ever written.

Richard Epstein's engaging treatise, *The Theory of Gambling and Statistical Logic, Revised,* (Academic Press, 1977) is a landmark in the subject. Much of it requires a university-level mathematics background. However, it is the best single reference work in print on the general subject of games and gambling, and even the general reader can glean much from browsing through it.

Now I'll explain why mathematical systems like the doubling-up system, cannot reduce the casino percentage.

The Problem with Doubling Up

One reason I chose roulette to illustrate mathematical systems is that it is easy to understand the odds and probabilities.

One correct version of the so-called "law of averages" says that in a "long" series of bets, you will *tend* to gain or lose "about" the total expectation of those bets. This means that a series of "bad" bets is also "bad," and that systems don't help.

Applying these ideas to the doubling-up system, let's calculate the player's expectation for one *cycle*. Think of a complete cycle as a single (complicated-looking) bet. Now refer to Table 8-2. The fifth column gives the probability that the cycle ends on turn #1, #2, etc. and the fourth column gives the gain or loss for each of these cases. Multiply each entry in the fourth column by the corresponding entry in the fifth column. Then add the results:

$\$1 \times 18/38 + \$1 \times 20/38 \times 18/38 + \ldots + \$1 \times (20/38)^9 \times 18/38 - \$23 \times (20/38)^{10} \times 18/38 - \$23 \times (20/38)^{10} \times 18/38 - \$2023 \times (20/38)^{11}$ which simplifies to $1 - 24 \times (20/38)^{10} - 2000 \times (20/38)^{11} = 1 - 0.0391\ldots - 1.7168\ldots = -\$.7560266578\ldots\ldots$ Thus, the expected loss to the bettor is about $-\$.76$ per cycle.

Now let's calculate the expected (or "average") amount bet on one cycle. Referring again to Table 8-2, we see that if the cycle ends on turn #1, the total of all bets is \$1, if it ends on turn #2, the total of all bets is \$1 + \$2, if it ends on turn #3, the total is \$1 + \$2 + \$4, etc. If the cycle ends on turn #11, the total amount bet is \$2,023. (To get these totals as of the end of any turn, add columns two and three.) Then multiply these total amounts bet by the chances in column five to get $\$1 \times 18/38 + \$2 \times (20/38) \times (18/38) + \$4 \times (20/38)^2 \times (18/38) + \ldots + \$512 \times (20/38)^9 \times (18/38) + \$2023 \times (20/38)^{10} \times (18/38) + \$2023 \times (20/38)^{11}$ which simplifies to $\$2 \times (18/38) \times ((40/38)^{10} - 1)/(40/38 - 1) + \$2024 \times (20/38)^{10} - \$1 = \$14.3645065$. If we divide the expected loss by the average bet per cycle we get $-\$.756\ldots \div \$14.36\ldots 1/19$ exactly or -5.26%.

These calculations are tedious, and for each system the details are different, so they have to be done again. And there are an infinite number of gambling systems, so calculations can never check them all out anyhow. Clearly this is not the way to understand gambling systems. The correct way is to develop a general

mathematical theory to cover gambling systems. That has been done and here's how it works. First we define the *action* in a specified set of bets to be the total of all bets made. From what we have said, your expected (gain or) loss is your action (i.e., the total of all your bets) times the house edge. For example, if you bet \$10 per hand at blackjack and play for 10 hours, betting 100 hands per hour, you have made a thousand \$10 bets, which is \$10,000 worth of "action." If you are a poor blackjack player and the casino has a 3% edge over you, your expected loss is $\$10{,}000 \times 3\% = \$300$. Your actual loss may be somewhat more or somewhat less.

If Nevada casino blackjack grosses a total of \$400 million per year and the average casino edge over the player is 2% of the initial wager, then we can determine the total action (A) per year: $.02A = \$400{,}000{,}000$ so $A = \$20$ billion. Thus from these figures we would estimate \$20 billion worth of bets are made per year at Nevada blackjack. The 2% figure might be substantially off. We could get a fairly accurate idea of the true figure by making a careful statistical sampling survey. If, instead, the figure is 4%, then $A = \$10$ billion. With 1%, $A = \$40$ billion per year.

Guidelines for Evaluating Systems

The general principles we have discussed apply to almost all gambling games, and when they apply, they guarantee that systems cannot give the player an advantage.

To help you reject systems, here are conditions which guarantee that a system is worthless:

I. Each individual bet in the game has negative expectation. (This makes *any series* of bets have negative expectation.)

II. There is a maximum limit to the size of any possible game. (This rules out systems like the no-limit doubling up system discussed.)

III. The results of any one play of the game do not "influence" the results of any other play of the game. (Thus, in roulette, we assume that the chances are equally likely for all of the numbers on each and every future spin, regardless of the results of past spins.)

IV. There is a minimum allowed size for any bet. (This is necessary for the technical steps in the mathematical proof. Most people would take for granted that there is such a minimum, namely some multiple of the smallest monetary unit. In the U.S.A., the minimum allowed bet is some multiple of one cent. In West Germany, it may be some multiple of the pfenning, and so forth.)

Under these conditions, it is a mathematical fact that every possible gambling system is worthless in the following ways:

(1) Any series of bets has negative expectation.

(2) This expectation is the (negative) sum of the expectations of the individual bets.

(3) If the player continues to bet, his total loss divided by his total action will tend to get closer and closer to his expected loss divided by his total action.

(4) If the player continues to bet it is almost certain that he will:

(a) be a loser;

(b) eventually stay a loser forever, and so never again break even;

(c) eventually lose his entire bankroll, no matter how large it was.

To give you an idea of how valuable this result is for spotting worthless systems, here are some examples of systems which cannot possibly give the player an advantage:

1. All the roulette systems I have ever heard of, except the following two types. (a) Biased wheels, in which condition (I) may be violated; the numbers are no longer equally likely, so bets on some numbers may have positive expectation. (b) Physical prediction methods, in which the position and velocity of ball and rotor are used to predict the outcome.

2. All craps systems I have ever heard of, except possibly those using either crooked dice or physical "control" of dice.

(Note: While at the Fifth Annual Gambling Conference at Lake Tahoe, I saw a dice cheat control the dice, at a private showing. I then saw him win at a casino. I heard he did this regularly. His badly mutilated body was found in the Las Vegas area a year later.

3. Any systems for playing keno, slots and chuck-a-luck.

As a further illustration, consider the book *Gambling Systems That WIN,* published by Gambling Times, 1978, paperback, $2. Of the fourteen systems given there, our result applies at once to eight. (The other six are one blackjack system, four racing systems, and a basketball system.)

(In the case of sports bets, it is generally difficult to determine whether condition I is satisfied. In the case of blackjack, condition I fails if the player counts cards, and there are, in fact, some winning systems, as most of you know.)

This leaves eight systems in *WIN:* four craps systems, one baccarat system, two roulette systems, and a keno system.

Conditions I through IV hold for all eight systems so none of them are winning systems. Nor do any of them reduce the house edge in the slightest. However, they may be entertaining. Also, in games like keno, craps, and roulette, where the expectation may vary from one game to another or from one type of bet to another, some ways to bet are "smarter" (translation—less dumb; more accurate translation—less negative expectation but still losing) than others.

For those who are prepared to lose, but want to lose more slowly, such systems may be of interest.

In most cases, the basic information is a list of the various bets in the game and their expectation. Then, if you must play, choose only bets with the least negative expectation. The "system" complexities and hieroglyphics are not essential.

It may amuse you to see why condition IV is needed. Suppose, instead, that there is *no* minimum bet and that we are playing Red at roulette. Our first bet is $1,000. There is an 18/38 chance that we win $1,000 and a 20/38 chance we lose $1,000. Now suppose that the second bet is $0.90, the third bet is $0.09, the fourth bet is $0.009, the fifth bet is $0.0009, etc. (Remember: *no* minimum.) Then the total of all bets from the second on is $0.99999... = $1.00.

The total gain or loss on these bets is between—\$1.00 and +\$1.00. The total action on all bets is \$1,000+\$1=\$1,001.

If we won the first bet, our total winnings (T) will always be between \$999 and \$1,001. This happens with probability 18/38. Therefore, conclusions 4(a), 4(b), and 4(c) fail. Also, our total action is \$1,001 so T/A is always between \$999/\$1,001 and \$1,001/\$1,001. But our expected gain (E) is negative so E/A is less than 0. Therefore, if we win the first bet, T/A does not tend to get closer and closer to E/A. Therefore, conclusion 3 also fails.

Conclusion 4(c) also deserves some comment. Actually, there is an insignificantly small chance the player can win the casino's bankroll before losing his. But even for moderate-size casino bankrolls,this possibility is so tiny as to be negligible, no matter how large the player's bankroll! We will discuss this in the next chapter. It is also discussed at some length in the 1962 edition of my book *Beat the Dealer,* and in Feller's great *An Introduction to Probability and its Applications, Vol. I,* Wiley. Thus, a more exact version of conditions I-IV would include information about the size of the casino bankroll. Then conclusion 4 would include information about the tiny chance that 4(a), (b), and (c) don't happen.

As far as I know, the most elementary mathematical proof ever given for all this is in my textbook, *Elementary Probability,* available from Robert E. Krieger Publishing Co., Inc., 645 New York Avenue, Huntington, New York 11743. The proof is outlined on pp. 84-85, exercises 5.12 and 5.13. It requires no calculus and can be followed by a good high school mathematics student if he or she works through pp. 1-85.

We now have a powerful test for showing that a system doesn't win. This keeps us from wasting our money and time buying or playing losing systems. It also helps us in our search for systems that do win, by greatly narrowing the possibilities.

Chapter 9

Optimal Betting

It is somewhat ridiculous to discuss an optimal money management strategy when the player has a negative expectancy. As indicated in Chapter 8, with an enforced house maximum and minimum wager, there is no way to convert a negative expectation into a positive expectation through money manipulation. Any good money management plan says not to wager in such a situation. Players facing a negative expectancy should look elsewhere for a gambling game or, at the very least, bet insignificant amounts and write off in their mind the expected loss as "entertainment."

After the gambler has discovered a favorable wagering situation, he is faced with the problem of how best to apportion his limited financial resources. There exists a rule or formula which you can use to decide how much to bet. I will explain the rule and tell you the benefits that are likely if you follow it.

Let's begin with a simple illustration that I deliberately exaggerated to better get the idea across. Suppose you have a very rich adversary who will let you bet any amount on heads at each toss of a coin and that you both know that the chance of heads is some number p greater than ½. If your bet pays even money, then you have an edge. Now suppose $p = 0.52$, so you tend to win 52 percent of your bets and lose 48 percent. This is similar to the situation in blackjack when the ten-count ratio is about 1.5 percent. Suppose too that your bankroll is only $200. How much should you bet? You could play safe and just bet one cent each time. That way, you would have virtually no chance of ever losing your $200 and being put out of the game. But your expected gain is .04 per unit or .04 cents per bet. At 100 one cent bets an hour, you expect to win four cents per hour. It's hardly worth playing.

Now look at the other extreme where you bet your whole bankroll. Your expected gain is $4 on the first bet, more than if you bet any lesser amount. If you win, you now have $200. If you again bet all of it on your second turn, your expected gain is $8 and is more than if you bet any lesser amount. You make your expected gain the biggest on each turn by betting everything. But if you lose once, you are broke and out of the game. After many turns, say 20, you have won 20 straight tosses with probability. $.52^{20} = 0.000002090$ and have a fortune of $104,857,600, or you have lost once with probability 0.999997910 and have nothing. In general, as the number of tosses increases, the probability that you will be ruined tends to 1 or certainty. This makes the strategy of betting everything unattractive.

Since the gambling probabilities and payoffs at each bet are the same, it seems reasonable to expect that the "best" strategy will always involve betting the same fraction of your bankroll at each turn. But what fraction should this be? The "answer" is to bet $p - (1 - p) = 0.52 - 0.48 = 0.04$, or four percent of your bankroll each time. Thus you bet $4 the first time. If you win, you have $104, so you bet 0.04 × $104 = $4.16 on the second turn. If you lost the first turn, you have $96, so you bet $0.04 \times \$96 = \$3.84$ on the second turn. You continue to bet four percent of your bankroll at each turn. This strategy of "investing" four percent of your bankroll at each trial and holding the remainder in cash is known in investment circles as the "optimal geometric growth

portfolio" or OGGP. In the 1962 edition of *Beat the Dealer,* I discussed its application to blackjack at some length. There I called it the Kelly system, after one of the mathematicians who studied it, and I also referred to it as (optimal) fixed fraction (of your bankroll) betting.

Why is the Kelly system good? First, the chance of ruin is "small." In fact, if money were infinitely divisible (which it can be if we use bookkeeping instead of actual coins and bills, or if we use precious metals such a gold or silver), then any system where you never bet everything will have zero chance of ruin because even if you always lose, you still have something left after each bet. The Kelly system has this feature. Of course, in actual practice coins, bills or chips are generally used, and there is a minimum size bet. Therefore, with a very unlucky series of bets, one could eventually have so little left that he has to bet more of his bankroll than the system calls for. For instance, if the minimum bet were $1, then in our coin example, you must overbet once your bankroll is below $25. If the minimum bet were one cent, then you only have to overbet once your bankroll falls below 25 cents. If the bad luck then continues, you could be wiped out.

The second desirable property of the Kelly system is that if someone with a significantly different money management system bets on the same game, your total bankroll will probably grow faster than his. In fact, as the game continues indefinitely, your bankroll will tend to exceed his by any preassigned multiple.

The third desirable property of the Kelly system is that you tend to reach a specified level of winnings in the least average time. For example, suppose you are a winning card counter at blackjack, and you want to run your $400 bankroll up to $40,000. The number of hands you'll have to play on average to do this will, using the Kelly system, be very close to the minimum possible using any system of money management.

To summarize, the Kelly system is relatively safe, you tend to have more profit, and you tend to get to your goal in the shortest time.

Blackjack Money Management

The Kelly system calls for no bet unless you have the advantage. Therefore, it would tell you to avoid games such as craps and keno and slot machines. However, if you have the knowledge and skill to gain an edge in blackjack, you can use the Kelly system to optimize your rate of gain. The situation in blackjack is more complex than the coin toss game because (1) the payoff on a one-unit initial bet can vary widely, due to such things as dealer or player blackjacks, insurance, doubling down, pair splitting, and surrender, and (2) because the advantage or disadvantage to the player varies from hand to hand.

However, we can apply the coin toss results to blackjack by making some slight modifications. First, let's see where the coin toss example's best fixed fraction of four percent came from. The general mathematical formula for the Kelly system is this: In any (single) favorable gambling situation or investment, bet that fraction of your bankroll which maximizes $E\ ln\ (1\ +\ f)$, where E is the expected value and ln is the natural logarithm (to the base $e=2.71828\ldots$). This ln function is available on most hand calculators. In the case of coin tossing, the best fraction, which I call f^*, is given for a favorable bet by $f^*=2p\ -\ 1$, where p is the chance of success on one toss, and $f^*=0$ if $p=1/2$, i.e., if the game is either fair or to your disadvantage. Note too that $f^*=2p\ -\ 1$ is coincidentally your expected gain per unit bet.

Now your expected gain in blackjack varies from hand to hand. If we think of successive hands as coin tosses with a varying p, then we should bet $f^*=2p\ -\ 1$ whenever our card count shows that the deck is favorable. When the deck is unfavorable, we "should" bet zero. Uston-type team play approximates this ideal of betting zero in unfavorable situations. You can also achieve this sometimes by counting the deck and waiting until the deck is favorable before placing your first bet. But it is impractical to bet zero in unfavorable situations, so we bet as small as is discreet. Think of these smaller, slightly unfavorable bets as a "drain" or "tax" which "water down" the overall advantage of the

favorable bets. To compensate for this reduced advantage, f^* should generally be "slightly" smaller than the $2p\ -\ 1$ computed above. Another effect of the small, slightly unfavorable bets is to increase the chance of ruin a little.

The most important blackjack "correction" to the f^* computed for coin tossing is due to the greater variability of payoff. Peter Griffin calculates that the "root mean square" payoff on a one-unit blackjack bet is about 1.13. It turns out then that f^* should be corrected to about $(2p\ -\ 1)/1.27$ or about .79 times the advantage. Shade this to .75 because of the "drain" of the small, unfavorable bets and we have the farily accurate rule: For favorable situations at blackjack, it is (Kelly) optimal to bet a percent of your bankroll equal to about 3/4 percent advantage. For instance, with a $400 bankroll and a one percent advantage, bet 3/4 of one percent of $400, or $3.

The Kelly System for Roulette

In general in roulette, the house has the edge, and the Kelly system says, "don't bet." But in my chapter on physical prediction at roulette, I described a method where we (Shannon and I), with the aid of an electronic device, had an edge of approximately 44 percent on the most favored single number. That corresponds to a win probability of $p=0.04$, with a payoff of 35 times the bet, and a probability of $1\ -\ p\ =0.96$ of losing the bet. It turns out that $f^*=.44/35=.01257$. The general formula for f^* when you win A times a favorable bet with probability p and lose the bet with probability $1-p$, is $f^*=e/A$ where $e=(A+1)p-1>0$ the player's expected gain per unit bet or his advantage. Here $A=35$, $p=.04$, and $e=0.44$. In the coin toss example, $A=1$, $p=.52$, and $e=.04$.

Using any fixed betting function f, the "growth rate" of your fortune is $G(f)=p\ ln\ (1+Af)+(1-p)ln(1-f)$. After N bets you will have approximately $exp[N\ \ G(f)]$ times as much money, where exp is the exponential function, also given on most pocket calculators.

For the roulette single number example, using my hand calculator (an HP65) gives $G(f^*) = 0.04 \ln(1 + 35f^*) + 0.96 \ln(1 - f^*) = .04 \ln(1.44) + 0.96 \ln(0.98743) = .04 \times .36464 + 0.96 \times (-0.01265) = 0.1459 - .01215 = .00244$. After 1,000 bets, you will have approximately $\exp[2.44] = 11.47$ times your starting bankroll.

Notice the small value of f^*. That's because the very high risk of loss on each bet makes it too dangerous to bet a large fraction of your bankroll. To show the advantages of diversification, suppose instead that we divide our bet equally among the five most favored numbers, as Shannon and I actually did in the casinos. If one of these numbers come up, we win an amount equal to $(35 - 4)/5$ of our amount bet, and if none come up, we lose our bet. Thus $A = 31/5 = 6.2$ The other four numbers are not quite as favored as the best number. However, to illustrate diversification, suppose that the five-way bet has the same .44 advantage. This corresponds to $p = 0.20$. Then $f^* = .44/6.2 = 0.07097$, so you bet about seven percent of your bankroll and $G(f^*) = 0.20 \ln(1 + 6.2f^*) + 0.80 \ln(1 - f^*) = 0.01404$. This growth rate is about 5.75 times that for the single number. After 1,000 bets, you would have approximately 1.25 million times your starting bankroll. Such is the power of diversification.

What is the price of deviating from betting the optimal Kelly fraction f^*? It turns out that for bet payoffs like blackjack, which can be approximated by coin tossing, the "performance loss" is not serious over several days play. But for the roulette example, the performance loss from moderate deviations from the Kelly system is considerable.

APPENDICES

APPENDIX A.

Suppose point count systems which are "closer" to the relative u_i values of Table 2-2 are likely to be "better." To test this we require a precise meaning for "better" and a precise measure of "closeness." We begin by basing the definition of "better" on the notions of probabilistic dominance, and of risk, used in mathematical finance.

Definition 1. Let F and G be probability distribution functions. Then F *probabilistically dominates* G if $F(x) \leq G(x)$ for all x. If in addition $F(x_0) < G(x_0)$ for at least one x_0 then F *strictly probabilistically dominates* G. If F and G arise from random variables X and Y, respectively, or from probability measures u and v, respectively, then the defined terms apply to these pairs if they hold for F and G.

That F probabilistically dominates G is equivalent to $P(X \geq x) \geq P(Y \geq x)$ for all x. If X is the player expectation from point count system A and Y is the player expectation from system B, then this means that the chance of finding expectations of x or more is always at least as good as using A as it is by using B. One can show that this means that a player following A has at least as great an expected return as B with "the same risk level."

However, probabilistic dominance is inadequate as a definition of "better" because the typical situation is that F is "spread out" more in both directions from the mean full deck expectation $E_0=0$. Thus F dominates G for $x > E_0$ and G dominates F for $x < E_0$. In fact G is (to a good approximation) a convex contraction of F. More precisely, if E_F and E_G are the respective means of F and G, we will find $E_F \geq E_G \geq E_0$ with $Y-E_G$ a convex contraction (this is equivalent to the notion "less risky than" of portfolio theory); of $X - E_F$. Thus F is both "spread out more" than G and translated in the positive direction more. The reason why E_F, $E_G \geq E_0$ is because E_0 is the expectation using the basic strategy and constant bets, equivalent to the full pack expectation. When (advantageous) counting systems are used, the strategy for playing hands is improved whenever the player has seen any cards other than the ones he and the dealer use on the first round. Since this generally happens with positive probability, we then have E_F, $E_G > E_0$.

Definition 2. Point count system A is *better than* system B if E_F E_G and also $P(X \geq x)$ $P(Y \geq x)$ for $x \geq E_G$.

Typically count systems satisfy $E_F \geq E_G \geq E_0$ and $X-E_F=a(Y-E_G)$, $a \geq 1$ (a special case of convex contraction). These conditions imply A is better than B.

Assume that the betting systems $b(E)$ are numerical functions of the expectation E. Further assume $b(E)=1$ if $E \leq 0$ and $b(E) \geq 1$ if $E > 0$. These are the ones generally considered. The popular fallacious systems such as the martingales (e.g. "doubling up"), and the La Bouchere which incorporate past results, are of no interest here.

Theorem 3. With the preceding notation and assumptions, if A is *better than* B, then for any betting system $b_B(E)$ based on the B point count, there is a betting system $b_A(E)$ based on the A point count such that the return R_A per unit bet by A (approximately) probabilistically dominates R_B. Further, R_A and R_B have approximately the same risk. In fact $R_A=R_B+c$, where $c \geq 0$.

Proof: If F and G are continuous, define b_A by $b_A(F^{-1}(G(E)))=b_B(E)$. Then note that the first unit of each bet has expectation E_A for A and E_B for B. The remainder of the bet is non-zero only if $E \geq E_F$. Then for corresponding percentiles of the respective distributions, A places the sames bets as B. But $F(E) \leq G(E)$ if $E \geq E_F$ so A has in each instance at least as great expectation, hence has at least as great expectation overall. Thus the total expected return to A is at least as large as for B. Also $R_A \geq R_B$ per unit since the bets placed have the same distribution.

In reality F and G are not continuous; instead they are finite. But they may be arbitrarily closely approximated by continuous distributions so the result extends, with one qualification. If F or G is discontinuous, extend the graphs of F and G by adding vertical segments at the discontinuity points so that the extensions $\tilde{F}$ and $\tilde{G}$ have inverses defined on (0,1). Then for those E' such that G is discontinuous at E' or F is discontinuous at $F^{-1}(G(E'))$ it may be necessary to define $b_A(F^{-1}(G(E')))$ "probabilistically", so it is multiple-valued, each value occurring with specified probabilities.

To show that $R_A=R_B+c$, which implies the same risk, it suffices to assume that at each percentile level y for the distributions F and G we have the conditional distributions given y satisfying $F(x|y)=G(x-f(y)|y)$ where $f(y) \geq 0$. Since this only holds approximately in practice, we have $R_A=R_B+c$.

Now we turn to the problem of measuring "closeness" of a given count to the "ultimate" strategy. We shall assume that point count strategies are of the form $C=(c_1,c_2,\ldots c_{13})$ where c_1 is the value assigned for an ace, $c_2,\ldots,c_9$ are the point counts for ranks 2 through 9, and $c_{10}=\ldots=c_{13}$ are the point counts for tens, jacks, queens, and kings respectively. In practice these are lumped together and only ten point count values are specified. By writing C with 13 components we gain a symmetry which yields substantially simpler proofs. Note that C and aC, a=0, are equivalent and will be identified.

Definition 4. If $\Sigma\Delta E_i=0$ the *ultimate strategy* $U=(u_1,\ldots u_{13})$ is the one given by $u_i=\Delta E_i$ where ΔE_i is the change in expectation from removing one ith card from the complete pack. If $d=\Sigma\Delta E_i=0$ then U is given by u_i-d/13.

In Table 2-2, we have d for one deck is .024 and d for four decks is .017. The u_i rows are calculated in Table 2-2 from Definition 4.

It is tempting to think of U as representing to good approximation the direction of the gradient E at $f_1=\ldots=f_{13}=1/13$ of the player's expectation $E(f_1,\ldots,f_{13})$ as a function of the fraction f_i of the cards from i=1 to 13. Then we calculate $(C)=C\bullet U/\|C\|\bullet U/\|C\|\bullet\|U\|$, i.e. the projection of C in the E direction. The numerator is the inner or scalar product and $\|C\|=(\Sigma c_i^2)^{1/2}$.

Next we claim that λ (C) gives the approximate ratio of the spread of the C distribution F_c about E_c to the U distribution F_u about E_u. Then λ (C) is the desired measure of closeness. In particular, for approximately the same risk per unit, and the same distribution of the bet sizes, it would follow that $E(R_u) \doteq E(R_c)/(C)$. Then C_1 and C_2 are arbitrary strategies $E(R_{c1})/E\ (R_{c2}$ $\doteq \lambda\ (C_1)/\ \lambda\ (C_2)$ for the same risk level and distribution of bet sizes. Thus the "power" of a strategy C is proportional to its λ (C).

This conclusion is true but the argument must resolve two obstacles:

(1) In the preceding discussion we treated C, U, ∇E, etc. as though they were given in Cartesian coordinates when in fact they are not.

(2) The probability distribution of $E(f_1,\ldots,f_{13}$ must be considered in reaching the conclusion and in general will invalidate it.

Note further that both U and C are linear approximations to an in general curved "surface". Also in the real case the domain is a large finite subset of points of the possible $(f_1,\ldots,f_{13})$, each of positive probability. (The original discovery of winning blackjack systems [Thorp, 1961], was motiviated by this model.) First I introduced the $E(n,\ldots,n_{13})$ "surface", where n_i is the number of cards remaining of denomination i. Intuitive arguments "convinced" me that the E surface should have substantial deviations from E_0, the full deck expectation. The next step was to approximate by "the" $E(f_1,\ldots,f_{13})$ "surface", and then to "linearize" the problem by assuming that $E(f_1,\ldots,f_n) \doteq E_0 + \Sigma_{ki}\ \Delta\ f_i$, where $\Delta\ f_i=f_i$-1/13.) Thus there is the approximation of a discrete problem by a continuous one. Nonetheless, we shall show:

Theorem 5. If the probability distribution of $(f_1,\ldots,f_{13})$ is approximately rotationally symmetric about (1,...,1)/13 then the relative power of any point count system C is proportional to $(C)=C.U/\|C\|\bullet\|U\|$. The powers of two count systems which exploit the count information equally (e.g. if one normalized by the number of as yet unseen cards so does the other; if one carries a side ace count for betting and sets the ace equal to 0 for strategy, so does the other, etc.) are approximately proportional to their λ's.

Proof.

APPENDIX B.

Suppose (Hypothesis I) that the shoe really has four complete decks. Then the number X of unseen ten-value cards among the 104 cards (two decks) not seen will average 32. In the general case with U unseen cards, T tens in the whole pack, and N non-tens in the whole pack, the average value A of X is given by $A = UT/(N + T)$. In our example, $U=104$, $T=64$ and $N=144$, so we get $A=104 \times 64/208=32$. But there will be a fluctuation around this number. Mathematicians use the standard deviation S to measure this fluctuation. The formula $S^2 = [UTN/T + N)^2](1 - (U - 1)/N + T - 1)]$.

For our example, $S^2 = (104 \times 64 \times 144/208^2)[(1 - 103/207)(= 11.1304$, so $S = \sqrt{11.1304} = 3.3362$. To a good approximation, X is "normally distributed" with mean $A = 32$ and standard deviation $S = 3.3362$.

Now, suppose instead (Hypothesis II) that the deck has ten ten-value cards removed. Then $U=94$, $T=54$ and $N=134$. If Y is the number of unseen cards, we have the real $A=25.6364$, but we think there are ten more ten-value cards. So assuming incorrectly that no ten-values are gone, the number that we deduce for Y has an average of $A + 10=35.6364$. The real S^2 for Y is $94 \times 54 \times 134/198^2$ $(1 - 93/197) = \sqrt{9.1593}$, so $S=3.0264$.

What we want to know is whether to believe Hypothesis I ("null hypothesis") or Hypothesis II. This is a classic statistics problem. It turns out that in order for us to have a good chance to believe the correct hypothesis, the A value for X and Y need to be at least two and preferably several S units apart. In this example, they differ by only $35.6364 - 32=3.6364$ which is about one S unit. Of course, repeated countdowns of this same shoe will again increase our ability to tell whether the shoe is short.

APPENDIX C.

For this first simple discussion, let's suppose $\underline{x}(t)=a \exp(bt) + c$, where a, b, and c are constants and *exp* is the exponential function. This is one of the simplest mathematical functions that has the right "shape." (Note: Mathematical readers may wish to redo this discussion using the quadratic $\underline{x}(t)=at^2 + bt + c$ to see the difference.)

I recall that the ball velocity at the point where it fell from the track was about 0.5 revolutions per second (r.p.s.) and that ten revolutions earlier it was about 2 r.p.s. Using this and the choice $t=0$ when the ball leaves the track gives $a=10/3$, $b=3/20$, and $c=-10/3$. Thus, $x(t)=10(\exp(3t/20) - 1)/3$ in r.p.s., and this gives an angular velocity v in r.p.s. of $v(t)=½\exp(3t/20)$. Figure 4-1 shows a graph of $x(t)$.

APPENDIX D.

A calculation shows, for our illustrative $x(t)$ function, that $x_0(T)=1/\exp(3T/20) - 1) - 10/3$. Thus, from T we can predict the number of revolutions until the ball leaves the track. For instance, if $T=1$ *sec.*, we predict the ball will leave the track in $x_0(1)=1/(\exp(3/20) - 1) - 10/3=2.85$ revolutions after the switch is hit the second time. If instead $T=½$ *sec.*, then we predict $x_0(½)=9.51$ revolutions.

APPENDIX E.

Math readers: $dx_0(T)/dT = -(3x_0(T) + 10)^2/60$. It can be shown that for the $x(t)$ of this example, the error $\Delta x_0 T$ in the prediction of $x_0(T)$ due to an error ΔT in measuring T, is given by $\Delta x_0(T) = -(3x_0(T) + 10)^2 T/60 = -3 T/(20[exp(3T/20) - 1]^2)$. For instance, if $T = 0.8$ sec. and $\Delta T = 0.012$ sec., we have a prediction error of $\Delta x_0(0.8) = 0.11$ revs or 4.2 numbers on the wheel. In our illustration $T = 0.8$ sec. means $x_0(T) = 4.51$ revolutions to go. The time to go is $(20/3)log_e(3x_0(t)/10 + 1)$ or 5.70 sec. We have somewhat less time than this to bet.

APPENDIX F.

In our example, the equation for $t_0(T)$ is $t_0(T) = (20/3)log_e(3/10)/exp(3T/20) - 1)) = (20/3)log_e(3x_0(T)/10 + 1)$. The error is approximately $\Delta t_0(T) = -(\Delta T)exp(3T/20) /(exp(3T/20) - 1)$. Thus again, if $T = 0.8$ sec. and $\Delta T = 0.012$ sec., $\Delta t_0(T) = -0.106$ sec. With a rotor speed of 0.33 r.p.s., this causes a rotor prediction error of 0.036 rev. or 1.3 pockets. In our example then, we measured T too large by 0.012 sec. This led us to believe the ball would leave the track at a point about 4.2 pockets before where it did. Therefore, we forecast impact on the rotor 4.2 pockets early. It also led us to believe the ball would leave the track sooner in time. Thus, we thought the rotor wouldn't revolve as far as it did. This made us forecast impact another 1.3 pockets early, for a total error of 5.5 pockets early. There are other important sources of error, so our final predictions were not this good. But they were good enough.

In summary, note that an error where ΔT is positive, i.e., we think T is bigger than it really is because we hit the switch early the first time or late the second time, leads us to think the ball

is slower than it is. That makes us think $x_0(T)$ is shorter. Thus, we expect the ball at the rotor too soon and forecast impact on the rotor ahead of where it tends to occur. Conversely, if T is negative (last on the first switch or early on the second), we think T is smaller, the ball is faster, and mistakenly forecast $x_0(T)$ and $t_0(T)$ as too big. Then we predict impact behind where it tends to occur.

The rotor angular velocity, followed a law close to $r(t) = Aexp(-bt)$. A typical value for A was 0.33 rev./sec. The "decay" or "slowing down" constant b was very small. The rotor is massive and spins on a well-oiled bearing (on our casino wheel, it was the pointed end of a sturdy steel shaft). In the course of a minute or two, the slowing was hardly perceptible. (Note: Stroboscopic "beat frequency" techniques, plus an accurate clock, can quickly and easily give a very precise measurement of b and the slowing down.)

Let's take $b = -log_e(10/11)/120$ or 0.000794/sec., which corresponds to a slowing down from 0.33 rev./sec. to 0.30 rev./sec. in two minutes. This seems like the right order of magnitude. To put the rotor position into the tiny computer we were going to build, we planned to hit a rotor timing switch once when the zero passed a reference mark on the wheel, and then hit the switch again when the zero passed the reference mark a second time. Since the rotor velocity was small and nearly constant, this was a less "sensitive" measurement. Therefore, we planned to do it first, shortly before the ball was spun.

How much error in the ball's final position (pocket) comes from rotor timing errors? Assume for simplicity that the rotor makes one revolution in about three seconds (.33 rev./sec.) and that we can neglect the slowing down of the rotor. Then, as in the ball timing, we might expect a typical (root mean square) size of about 11.2/1,000 seconds for the combined effect of the two errors. If the rotor really makes one revolution in 3.000 seconds, and we think it takes 3.0112 seconds, then in 30 seconds we think the wheel will travel 9.9628 revolutions whereas it really travels 10.0000 revolutions. Thus, the rotor goes .0372 rev. or 1.4 pockets farther than expected. Similarly, if we think the rotor takes 2.9888 seconds for one revolution, then in 30 seconds the rotor goes .0375 rev. or 1.4 pockets less than we expected.

APPENDIX G.

I am using the normal approximation for the statistical discussion. I think it is very nearly an accurate description of what happens and that this approximation only slightly affects the discussion.

APPENDIX H.

In general, there are exactly $(5+r)!/5!r!$ home board positions with exactly r men. There are exactly $(6+r)!/6!r!-1$ home board positions with from one to r men. Thus, since $r=15$ is possible in the actual game, there are a total of $21!/6!\ 15!\ -1=54{,}263$ different home board positions for one player. The symbol r!, read "r factorial," means $1\times2\times3\times\ldots\times r$. Thus $1!=1$, $2!=2$, $3!=6$, $4!=24$, etc.

Scholarly References

For those readers who are especially interested in the technical work behind the material in this book and other work by Professor Thorp, here is a list of some of his related scholarly publications.

Game Theory

1. "A Favorable Strategy for Twenty-One," *Proceedings of the National Academy of Sciences 47* (1961). 110-112. (MR #B846).

2. BOOK: *Beat the Dealer: A Winning Strategy for the Game of Twenty-One,* Random House, New York, 1962; Vintage paperback also, 1966; revised edition, 1966.

3. "A Partial Analysis Of Go," (with W. E. Walden), *The (British) Computer Journal* 7:3 (1964), 203-207. (MR33 #2424), (Z *123*, 340).

4. "A Favorable Side Bet in Nevada Baccarat," (with W. E. Walden), *Journal of the American Statistical Association 61,* Part I (1966), 313-328.

5. "Repeated Independent Trials and a Class of Dice Problems," (Mathematical Note), *American Mathematical Monthly* (August-September 1964), 778-781. (Z 123, 364).

6. "Optimal Gambling Systems for Favorable Games," *Review of International Institute of Statistics* 37:3 (1969), 273-293. (Z *191*, 497).

7. "Solution of a Poker Variant," *Information Sciences* 2/2 (1970), 299-301. (MR S42 #8952), (Z *205*, 232).

8. "A Computer-Assisted Study of Go on M x N Boards," (with W. E. Walden), Lecture Notes in Operations Research and Mathematical Systems in *Theoretical Approaches to Non-Numerical Problem Solving,* Vol. 28, R. Banerji and M. D. Mesarovic, eds., Springer-Verlag, New York, 1970, 303-343; later revised version, *Information Sciences* 4:1 (1972), 1-33 (MR45 #8684), (Z 228.68028).

9. "Non-Random Shuffling With Applications to the Game of Faro," *Journal of the American Statistical Association,* 842-847, December 1973. Much expanded version appears in *Gambling and Society,* edited by W. Eadington, Charles C. Thomas, Springfield, Illinois, 1975, as: Probabilities and Strategies for the Game of Faro, pp. 531-560.

10. "The Fundamental Theorem of Card Counting With Applications to Trente et Quarante and Baccarat," (with W. E. Walden) *International Journal of Game Theory* 2 (1973), 109-119. (Z 258.90054)

11. "Backgammon: Part I, The Optimal Strategy for the Pure Running Game," Proceedings of the Second Annual Conference on Gambling, Lake Tahoe, 1975. ms. 42 + pp. See news report on this paper in: "Beating the Game," an article by Dietrick E. Thomsen, *Science News,* Vol. 107, March, 1975.

12. "Blackjack Systems," Proceedings of the Second Annual Conferrence on Gambling, Lake Tahoe, 1975. ms. 15 + pp. See news report on this paper in: "Beating the Game," an article by Dietrick E. Thomsen, *Science News,* Vol. 107, March, 1975.

Probability, Statistics

13. BOOK: *Elementary Probability,* Wiley, New York, 1966.

Mathematical Finance

14. BOOK: *Beat the Market,* (with S. Kassouf), Random House, New York, 1967.

15. "Portfolio Choice and the Kelly Criterion," *Proceedings of the 1971 Business and Economics Section of the American Statistical Association* 1972, 215-224. To be reprinted in *Investment Decision-Making,* edited by J. Bicksler. Reprinted in *Stochastic Optimization Models in Finance,* Academic Press, edited by W. T. Ziemba, S. L. Brumelle, and R. G. Vickson, 1975, pp. 599-620.

16. "The Capital Growth Model: An Empirical Investigation," (with James Bicksler), *Journal of Financial and Quantitative Analysis,* March 1973, Vol. VIII, No. 2, pp. 273-287.

General Interest

17. "A Professor Beats the Gamblers", *The Atlantic Monthly,* June, 1962. Reprinted in "The Gambler's Bedside Book," John K. Hutchens, ed., Tapliniger, New York, 1977, pp. 166-177.