NBER WORKING PAPERS SERIES

THE "GAMBLER'S FALLACY" IN LOTTERY PLAY

Charles T. Clotfelter

Philip J. Cook

Working Paper No. 3769

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1991

This paper is part of NBER's research program in Taxation. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3769
July 1991

THE "GAMBLER'S FALLACY" IN LOTTERY PLAY

ABSTRACT

The "gambler's fallacy" is the belief that the probability of an event is lowered when that event has recently occurred, even though the probability of the event is objectively known to be independent from one trial to the next. This paper provides evidence on the time pattern of lottery participation to see whether actual behavior is consistent with this fallacy. Using data from the Maryland daily numbers game, we find a clear and consistent tendency for the amount of money bet on a particular number to fall sharply immediately after it is drawn, and then gradually to recover to its former level over the course of several months. This pattern is consistent with the hypothesis that lottery players are in fact subject to the gambler's fallacy.

Charles T. Clotfelter
Box 4875 Duke Station
Durham, NC 27706
and NBER

Philip J. Cook
Box 4875 Duke Station
Durham, NC 27706

THE "GAMBLER'S FALLACY" IN LOTTERY PLAY

Charles T. Clotfelter and Philip J. Cook

June 1991

People expect that a sequence of events generated by a random process will be highly patterned, even when the trials are "objectively" known to be independent. Tversky and Kahneman (1974) note that "Chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction" in order to make the overall sequence representative of the underlying probability distribution. This so-called "gambler's fallacy," which has also been referred to as the principle of the "maturity of chances," was stated eloquently, if unscientifically, by one historian of gambling in 1909 (Ralph Nevill, quoted in Devereux 1949, p. 690n.):

> In games of chance, the oftener the same combination has occurred in succession, the nearer we are to certainty that it will not occur at the next coup. It would almost appear in fact, as if there existed an instant, prescribed by some law, at which the chances become mature and after which they begin to tend again towards equalization.

This belief in what might be called "negative dependence" has been demonstrated in the laboratory (Morrison and Ordeshook 1975) and, more interestingly, at the racetrack (Metzger 1985).[1]

In this paper we offer evidence that the belief in negative dependence influences patterns of play in the numbers game offered by state lotteries. We find that the amount of money bet on a particular three-digit number tends to fall sharply immediately after it "hits," and then gradually recovers to its former level over the course of several months. We document the existence of this pattern for 52 consecutive winning numbers in the Maryland numbers game. A concluding section discusses the gambler's fallacy and an alternative explanation for these results.

Data

The daily numbers games offered by state lotteries since the early 1970s, and by underworld operators for much longer, offer players a variety of bets on the outcome of a daily drawing in which a three-digit number is selected.[2] Players choose their number and the type of bet they wish to make.

We obtained data on a sample of 52 consecutive winning numbers in the Maryland three-digit game, drawn during March and April, 1988. Each number's relative betting frequency was recorded for the day before it was drawn, the day it was drawn, and for the first, second, third, seventh, 28th, 56th, and 84th days after it was drawn. Players in Maryland's numbers game can make a variety of bets on the number drawn, but the two principal types of bets are the straight bet and the box bet. The probability of a correct guess on a straight bet -- three digits in exact order -- is 1 in 1000. Like most states, Maryland pays out $500 on a dollar bet, consonant with its general policy of offering a 50 percent payout rate on lottery games. The box bet allows the player to win if any permutation of his number is drawn. For example, a box based on three different digits yields six possible permutations, for odds of 6 in 1000; Maryland's payout is $80, slightly less than the 50 percent of a "fair" return paid on a straight bet. In the Maryland game, roughly two thirds of all bets are straight bets and the remaining third are box bets.[3]

For analyzing patterns in betting on particular numbers, we need a measure which weights straight bets more heavily than box bets. From the lottery agency's perspective, a straight bet on the number 123 creates 6.25 times (500/80) the prize liability on that number as does a box bet on 123. The total prize liability for a number is one convenient indicator of the amount of betting on that number, one which is computed daily by the Maryland Lottery. In general, liability can be represented:

(1)

$$L_{it} = \sum_j a_j B_{ijt}$$

where a_j is the payout for bet type j and B_{ijt} is the amount of bet type j on number i on day t. In order to adjust for day-to-day variations in the total quantity of betting, we divide the liability generated by betting on this number by the day's total bets to create an indicator Q_{it}

(2)

$$Q_{it} = \frac{L_{it}}{B_t^*}$$

where B_t^* is the total amount bet on day t. Because the prize payouts in Maryland were set to approximate a 50 percent payout rate, the average of this ratio will be close to 50 percent.[4] One way to interpret this ratio is to note that it is proportional to the "market share" for the three-digit number, with 500 as the proportionality constant. If Q_{it} = 50, the market share for number i on day t is 0.1 percent.

Findings

Table 1 presents data on the pattern of betting on the 52 winning numbers in our two-month sample period. The first thing to notice is that there are large and persistent differences in the relative popularity of different numbers. The Q_is on the day the number was drawn range from 14 for the number 899 (on March 5) to 260 for the number 011 (on March 9).[5] Such differences tend to follow predictable patterns. As has been noted elsewhere, there is a strong preference for numbers in the lower half of the distribution, for triples (such as "333") and other simple sequences ("123"), for numbers matching the date, and for numbers corresponding to widely-known events.[6] But the data in Table 1 indicate that the market share of a number changes after it hits. On the day after the numbers were drawn, there was a decline in the amount bet in 37 of 52 cases. After three days the amount bet had declined in <u>every</u> case. However, this drop in betting activity was temporary. After 84 days betting had just about returned to normal, with only slightly more than half of the numbers (28 out of 52) showing lower levels of betting than on the drawing day.

To see how the level of betting varied over time, we calculated the ratio of betting on a winning number t days after it "hit" to the amount bet on it during a two-day baseline period:

(3)

$$\frac{Q_{it}}{\overline{Q_i}} \quad ,$$

where $\overline{Q_i} = (0.5(Q_{i,-1}) + Q_{i0})$.

We examined the median, 25th percentile and 75th percentile values for this ratio at specified intervals following the day it was drawn. These values are graphed in Figure 1. The immediate drop in betting on winning numbers is quite evident. For example, the median ratio fell from 1.00 on the drawing day to 0.64 after three days, then recovered gradually to reach 0.93 after 84 days.

One interesting question is whether popular numbers are more or less subject to the gambler's fallacy. In order to test for this, we estimated regressions of the form

(4)

$$\ln Q_{it} = a + b \ln \overline{Q}_i + u$$

for various periods following drawings. The average proportional degree of decline in betting is a function of pre-drawing popularity:

(5)

$$\frac{Q_{it}}{\overline{Q}_i} = \exp\left(a + (b-1) \ln \overline{Q}_i\right).$$

If b=1, the proportionate decline is unrelated to initial popularity. If b<1, the relative decline is greater for more popular numbers; if b>1, the opposite is true. Table 1 shows that the relative decline was in fact greater for the most popular numbers, with the estimated coefficient for *b* being significantly different from 1 for five of the seven intervals. The differences are not great, however. For example, consider the difference in the predicted decline between the relatively unpopular winning number 044 ($\overline{Q}_i$ = 30) and the relatively popular 303 ($\overline{Q}_i$ = 115). The estimates suggest that the relative betting on the two numbers after seven days would be 0.76 and 0.73, respectively.

Discussion

Our evidence suggests that particular numbers lose market share following a "hit," and then gradually regain it. The obvious conclusion is that some of the bets that would have been placed on that number are diverted to other numbers or not placed at all during the immediate post-hit period. Given that the drawings are random and the reduction in betting is large and consistent, there is little question that a number's being drawn is the cause for the subsequent decline in betting on this number.

We can offer two plausible interpretations of this pattern. First, one could explain the decrease in betting on a drawn number as the result of a widespread perception that when a number is drawn, the probability that it will be drawn again is reduced for several weeks or months -- that the random process is governed by negative dependence. As a result of this fallacious perception, players who ordinarily favor this number will either switch to another number or stop playing for a while. A second and quite different explanation is that when a number hits, the winners who regularly bet that number stop playing the game altogether because of a wealth effect. People who play with the purpose of achieving a certain financial objective, such as obtaining enough cash to make a down payment on a car, may behave in this fashion.

Unfortunately, the available data do not allow us to distinguish between these two explanations directly. There is other evidence, however, that favors the first explanation. First, winners of large lottery prizes usually do <u>not</u> stop playing; one study of million-dollar winners found that they played more on the average than they had before their big win (Kaplan, 1988). Also, the first explanation comports with the advice given in numerous guides for lottery players, which encourage them to study patterns in the sequence of winning numbers to ascertain which ones are "due."[7]

Conclusion

Patterns in numbers play are compatible with one version of the "gambler's fallacy": the probability that a number will win in any one drawing depends on how recently it has won in the past. Unlike the vast bulk of the evidence on this phenomenon, the lottery data have the virtue of being generated in the "real world" outside the laboratory. But this advantage comes at a cost, since it is not possible to rule out a rival explanation. To do that would require daily microdata on betting behavior by individuals.

REFERENCES

Clotfelter, Charles T. and Philip J. Cook. 1989. Selling Hope: State Lotteries in America. Cambridge, MA: Harvard University Press.

Devereux, Edward C., Jr. 1980. Gambling and the Social Structure. New York: Arno Press.

Hogarth, Robert M. 1980. Judgement and Choice. Chichester: John Wiley and Sons.

Kaplan, H. Roy. 1984. "Gambling Among Lottery Winners: Before and After the Big Score." Journal of Gambling Behavior 4 (Fall): 171-182.

Metzger, Mary A. 1985. "Biases in Betting: An Application of Laboratory Findings" Psychological Reports 56(3), 883-888.

Morrison, Rebecca S. and Peter C. Ordeshook. 1975. "Rational Choice, Light Guessing, and the Gambler's Fallacy" Public Choice 22, 79-89.

Reuter, Peter and Jonathan Rubinstein. 1982. Illegal Gambling in New York: A Case Study in the Operation, Sructure, and Regulation of an Illegal Market Washington, DC: USGPO.

Tversky, Amos and Daniel Kahneman. 1974. "Judgement under Uncertainty: Heuristics and Biases." Science 185: 1124-1131.

ENDNOTES

1. She speculated that betting on favorites should be more attractive after a series of races have been won by long shots than after a series of races won by favorites. Her data, based on 11,313 races run on 9-race cards at Thoroughbred tracks in 1978, appear to support this view.

2. Some states also offer two- and four-digit games, but the three-digit version is more popular.

3. On September 4, 1986, for example, 65 percent of bets were straight and 35 percent were box bets. The state offers bettors combinations of straight and box bets as well as combined straight bets on all the permutations of a numbers, but these are all made up of just the two basic types of bets. The only type of bet that is not made up of straight and box bets are exact bets on the front pair and back pair of numbers, which together made up less than one percent of all bets on that date.

4. The ratios for less popular numbers average less than 50 percent while the relatively few very popular numbers average more. For example, in the September 4, 1986 drawing the number 123 had a total liability of $1,885,155, which exceeded the total amount bet on all numbers, $1,494,469. In this case Q_{it} = 131.82. Most numbers have ratios less than 50 percent, as illustrated by the sample of winning numbers shown in Table 1.

5. For all the sample days, the most heavily bet among these winning numbers was 513, which reached a betting index of 470 on May 13, 56 days after it was drawn on March 18. This particular surge in betting provides a dramatic illustration of the popularity of widely-known numbers such as the date. In order to look for evidence of the gambler's fallacy, it is necessary to recognize the vast differences in the average popularity of numbers and to focus on the pattern of betting on a number relative to its usual level.

6. These patterns tend to persist even when the payout on popular numbers is reduced relative to less popular numbers (Clotfelter and Cook 1989; Reuter and Rubinstein 1982).

7. See, for example, Clotfelter and Cook (1989, p. 88).

Table 1

Index of Amount Bet on Three-Digit Numbers Drawn During March and April, 1988 in the Maryland Lottery, Drawing Day and Selected Days Afterwards

		Index of amount bet				
				Days after drawing		
Date of drawing	Winning Number	Day of drawing	1	3	7	56
March 1	295	65	34	24	28	23
March 2	640	32	27	17	17	16
March 3	980	24	20	19	17	23
March 4	957	25	20	18	19	29
March 5	899	14	15	13	17	22
March 7	618	44	23	18	23	26
March 8	639	23	22	18	18	20
March 9	011	260	472	212	180	119
March 10	274	69	44	42	50	68
March 11	472	46	54	37	40	53
March 12	575	23	25	15	17	20
March 14	383	30	31	18	20	21
March 15	277	28	39	24	25	30
March 16	342	82	74	59	65	90
March 17	085	16	17	12	11	14
March 18	513	69	61	39	59	470
March 19	682	33	26	20	24	25
March 21	044	30	30	18	23	23
March 22	820	31	27	26	27	30
March 23	656	26	23	18	21	29
March 24	058	16	16	15	18	16
March 25	908	32	31	26	24	30
March 26	928	40	34	23	30	31
March 28	360	28	32	26	31	41

March 29	202	65	61	38	39	61
March 30	325	126	70	55	48	68
March 31	736	42	29	22	23	29
April 2	950	22	22	14	18	24
April 4	179	57	41	31	28	31
April 5	303	135	95	66	58	64
April 6	837	26	23	15	19	17
April 7	660	30	31	22	25	34
April 8	453	67	56	55	57	81
April 9	972	38	29	21	26	28
April 11	244	41	34	24	27	30
April 12	504	29	20	16	18	15
April 13	718	28	20	17	19	25
April 14	323	134	95	79	81	76
April 15	640	19	20	18	16	20
April 16	957	30	22	20	24	32
April 18	446	19	17	12	20	19
April 19	600	39	29	18	21	21
April 20	830	31	24	21	29	35
April 21	543	119	98	92	111	139
April 22	975	22	23	18	24	26
April 23	077	31	30	24	32	28
April 25	013	47	37	27	33	43
April 26	509	30	25	23	22	29
April 27	383	21	22	17	18	22
April 28	295	30	25	20	26	27
April 29	694	27	22	16	20	30
April 30	667	20	16	12	15	17

Table 2

Regression Estimates

	Equation 1	
Day	a	b
1	.1195	.9254
2	-.1008	.9096**
3	.0007	.8794**
7	.4355*	.7946**
28	.5774*	.7710**
56	.5351	.8165
84	.9730*	.6996**

* Significantly different from 0
** Significantly different from 1

$$\ln Q_{at} = a + b \ln \overline{Q_t}$$

where

$$\overline{Q_t} = 1/2(Q_{t-1,t} + Q_{e,t})$$

Figure 1

RELATIVE BETTING INDEX

Winning Numbers, by date