AN INVESTIGATION OF STRATEGIES IN BASEBALL

George R. Lindsey

SACLANT ASW Research Center, La Spezia, Italy

(Received November 20, 1962)

The advisability of a particular strategy must be judged not only in terms of the situation on the bases and the number of men out, but also with regard to the inning and the score. Two sets of data taken from a large number of major league games are used to give (1) the dependence of the probability of winning the game on the score and inning, and (2) the distribution of runs scored between the arrival of a new batter at the plate in each of twenty-four situations and the end of the half-inning. The first shows that the runs adding most to the probability of winning are the one tying the score and the one putting the batting team one ahead. All other runs are slightly less valuable early in the game, and become increasingly less valuable as the game proceeds. By combining the two sets of data, the situations are determined in which an intentional base on balls, a double play allowing a run to score, a sacrifice, and an attempted steal are advisable strategies, if average players are concerned. An index of batting effectiveness based on the contribution to run production in average situations is developed.

IN THE course of most baseball games a number of occasions arise in which the managers of the competing teams are called upon to select one out of two or more possible strategies.

For example, when the batting team has one or more runners on base, alternative strategies are:

(a) have the runner(s) try to steal a base, or
(b) rely on a safe hit by the batter to advance the runner(s).

If there are less than two out, a third possible strategy would be:

(c) have the batter try to sacrifice, thus advancing the runner(s) at the expense of another out.

For the fielding team, a choice always available, but most attractive with runners on second and third base but first base empty, and a dangerous batter at the plate, is:

(a) to issue an intentional base on balls to the batter, instead of
(b) pitching to the batter.
Another choice of strategies faces the fielding team when the batting side has men on first and third, or first, second, and third, with none out. They can

(a) place their infield close, and, if the batter hits a ground ball to an infielder, throw home in an attempt to put out the runner and prevent a run, or

(b) place the infield deep, and attempt to convert a ground ball into a second-to-first double play, allowing one run to score.

By collecting statistics from a large number of baseball games it should be possible to examine the probability distributions of the number of runs resulting from these various situations.

The object of all choices is, presumably, to maximize the probability \( \Omega \) of eventually winning the game.

Suppose that, if a team in a situation \( S \) decides on strategy \( A \), the probability that \( r \) runs will be scored before the completion of the half-inning is \( A(r|S) \). But, if they choose strategy \( B \), the probability that \( r \) runs will be scored is \( B(r|S) \).

Suppose, further, that the situation \( S \) arises during the Visitors' half of the \( i \)th inning, with the Visitors \( l \) runs ahead, and that the probability that a visiting team \( l \) runs ahead at the end of its half of the \( i \)th inning will win the game is \( V_i W(l) \). In these formulas \( l \) may be negative, signifying that the team is behind.

The probability that the Visiting Team (hereafter called \( V \); the Home Team will be called \( H \) ), will win the game if strategy \( A \) is selected is

\[
\Omega(V,i,l,S,A) = A(0|S) V_i W(l) + A(1|S) V_i W(l+1) + \cdots,
\]

the terms of which correspond to the possibilities that strategy \( A \) will result in the scoring of \( 0, 1, 2, \cdots \) runs by the end of the half-inning \( V_i \).

In the stated circumstances, \( A \) is a better strategy than \( B \) if

\[
\Omega(V,i,l,S,A) = \sum_{r=0}^{\infty} A(r|S) V_i W(l+r) > \Omega(V,i,l,S,B) = \sum_{r=0}^{\infty} B(r|S) V_i W(l+r).
\]

The comparison depends not only on the probabilities \( A(r|S) \) and \( B(r|S) \) of scoring various numbers of runs, but also on the state of the game, as determined by the inning \( (V_i) \) and the difference in the scores \( (l) \).

It may seem surprising to the reader that the value of a strategy cannot simply be determined by the expected number of runs \( E(S) = \sum_{r=0}^{\infty} r A(r|S) \). However, the significance of the state of the game may be illustrated by a situation arising in the last half of the ninth inning \( (H9) \) with the Home Team \( H \) two runs behind \( (l = -2) \). Under these circumstances, \( H \) certainly lose the game if they score \( r = 0 \) or 1, retain a
probability of winning of ½ if \( r = 2 \) (i.e., extra innings will be played), and certainly win if \( r \geq 3 \).

\[
m_iW(l) = \begin{cases} 
0, & (l < 0) \\
\frac{1}{2}, & (l = 0) \\
1, & (l > 0)
\end{cases}
\]

The measure of value of strategy \( A \) in these circumstances is

\[
\Omega(H, 9, -2, S, A) = \frac{1}{2} A(2|S) + \sum_{r=3}^{\infty} A(r|S).
\]

Thus, a strategy producing a very high probability \( A(1|S) \) of scoring one run only would be a poor choice.

If, on the other hand, the score were tied in \( H9 \), then one run would win the game for \( H \), and the strategy just mentioned would be a good choice.

Recent examinations of the relative values of alternative strategies have taken the form of computer studies using Monte Carlo simulations\(^{[1,2]}\) or a mathematical model,\(^{[3]}\) based on average performance for batters. HOWARD\(^{[8]}\) showed that the strategy maximizing the expected number of runs in the half-inning was not the same as the one maximizing the expected number of runs immediately following the next play.

The present study follows a somewhat different line in two respects, in that it derives data from a large number of actual games, and that it attempts to relate the value of a strategy to the ultimate end—the winning of the game—rather than to the intermediate means—the scoring of runs.

### THE PROBABILITY OF WINNING THE GAME

In a previous paper\(^{[4]}\) the inning-by-inning progress of the score of a large number of baseball games was analyzed, and a table of probabilities obtained for the function \( H_iW(l) \), the probability that a team with a lead of \( l \) runs at the end of the home half of the \( i \)th inning will eventually win the game.

Using the data from reference 4, it is also possible to compute

\[
v_iW(l) = \sum_{r=0}^{\infty} f_i(r) H_iW(l-r),
\]

the probability that the visiting team \( V \) will win the game if they hold a lead \( l \) at the end of their half of the \( i \)th inning. Here \( f_i(r) \) is the probability that a team will score exactly \( r \) runs in the \( i \)th inning.

These two functions \( v_iW(l) \) and \( H_iW(l) \) are shown on Figs. 1 and 2. \( v_iW(l) \) is not symmetrical about \( l = 0 \), and the probability that the Home Team \( H \) will win if they have a lead \( l \) at the end of \( Vi \) is \( 1 - v_iW(-l) \).

\( H_iW(l) \) is symmetrical about \( l = 0 \), and the same function can be applied to either team.
Fig. 1. The probability that the visiting team will win the game, if they are \( l \) runs ahead at the end of their half of the \( i \)th inning.

The reason for the difference is, of course, that at the end of \( Vi \) the Home Team have one more half-inning in which to try to score than do the Visitors.

It is seen from both figures that the value of a lead of a given number of runs increases as the game progresses, and the value of one additional run, \( W(l+1) - W(l) \), is greatest for \( l = -1 \) and 0, and decreases for successively larger absolute values of \( l \). In other words, the most valuable runs are those that tie the score or put the team one run ahead.
The point is illustrated on Fig. 3, which shows differences $v_i W(l+1) - v_i W(l)$ for the Visitors' half-innings $i = 1, 4, 6, 8,$ and $9$, plotted against the lead $l$. As the game progresses, the tying and lead runs ($l = 0, 1$) become relatively more and more important. For example, the lead run, which changes $l$ from 0 to 1, increases the probability that $V$ will eventually win the game by 0.105 in the first inning, but by 0.440 in the ninth inning.

In an ordinary baseball game, $V$ cannot fall behind ($l < 0$) in the first half of the first inning ($V1$), so that $v_1 W(l)$ is useful only for $l \geq 0$. 

**Fig. 2.** The probability that a team will win the game if they are $l$ runs ahead at the end of the home half of the $i$th inning.
A diagram for $nW(l+1) - nW(l)$ would be symmetrical about $l=0$, but otherwise generally similar to Fig. 3.

These facts are recognized, at least in a qualitative way, by practicing managers, and to illustrate that baseball strategy does alter with the state of the game some quotations are given from Ethan Allen's book *Baseball Play and Strategy*, with the kind permission of the publishers:

Try for a big inning early in the game, and sacrifice in the late innings (p. 324).

It is a general practice to play for one run in the late innings and to hit straight away in the early part of the game (p. 336).
Make the safe play early in the game (referring to fielding strategy, p. 59).

The infield is played back or in when third base is occupied with less than two out, according to the score and the stage of the game. As a general rule the deep position is favoured during the first five innings unless the opposing team is ahead more than one run (p. 319).

The position of the infield is especially important late in the game because there may be little or no opportunity to overcome runs scored by the opposing team (p. 320).

Generally speaking, the use of the intentional pass is confined to the latter part of the game. Runs resulting from such strategy in the early innings are actually gifts. Furthermore, such runs often build an insurmountable lead (p. 319).

**PROBABILITY DISTRIBUTION OF RUNS FOLLOWING VARIOUS SITUATIONS**

If one were to specify all the details of each possible situation in which a choice of strategy is available, the number of variables would be large and the number of possible combinations enormous.

To reduce the study to more manageable proportions, attention has been confined to the situations occurring as a new batter comes to the plate. Thus, the ball and strike count is always 0 and 0. Moreover, the collection of data has been made in two separate sets: one pertaining to the relation between the score and the inning and the probability of winning the game, which was described in the preceding section, and the other, to be described in this section, which examines only the number of runs scored between the occurrence of a particular situation $S$ and the end of the half-inning in which $S$ occurred.

For this latter purpose the situation $S$ is completely described by two variables, $T$ and $B$. $T$ is the number of men out (0, 1, or 2) and $B$ is the state of occupation of the bases (0, 1, 2, 3, 12, 13, 23, or F). There are twenty-four basic situations, which will be designated by the symbol $\{T, B\}$. The desired distributions are $P(r | T, B)$, the probability that, between the time that a batter comes to the plate with $T$ men out and the bases in state $B$, and the end of the half-inning, the team will score exactly $r$ runs.

To determine these distributions from actual baseball experience, a large number of games were observed (by radio, television, or personal attendance) and data were recorded in such a way that the number of runs scored subsequent to the occurrence of each new situation as defined above could be tabulated. If the situation changed while the batter was still at the plate (e.g., by a stolen base, wild pitch, or runner being picked off) only the situation pertaining when he first came up was recorded. If a half-inning was not completed (e.g., in a game won in the last half of the
ninth or an extra inning), nothing was recorded. Data taken from all completed half-innings was pooled. Exclusion of incomplete half-innings will introduce a bias into the data, since, with the rare exception of games abandoned because of weather, a half-inning is incomplete because the winning run has been scored, and the average score in an incomplete half-inning is greater than in a complete half-inning. However, the proportion of all half-innings that are incomplete is less than 0.7 per cent, and the proportion of runs scored in incomplete half-innings is less than 2.2 per cent, so that the bias will be very small.

The majority of the recording and tabulation was contributed by Col. Charles Lindsey. Most of it came from 3033 half-innings in 1959 (from all or parts of 176 major league games) and 3366 half-innings in 1960 (from all or parts of 197 major league games). For some purposes use was also made of data taken in 1952, 56, 57, and 58 concerning only situations with the bases full.

A summary of the results is given in Table I. For the subsequent calculations a full table was used, including the larger values of $r$.

The column headed $N$ gives the number of cases recorded. It will be seen that there is a very wide variation. Over half of the 27,027 cases are for the bases empty. Some situations ($\{0,3\}, \{0,23\}, \{0,F\}$) occur only once in about 300 cases.

In addition to values of $P(r|T,B)$, the probability that $r$ runs will be scored, Table I also shows the expected number of runs

$$E(T,B) = \sum_{r=0}^{T} rP(r|T,B)$$

and $\sigma/\sqrt{N}$, the standard deviation of $E(T,B)$ as computed from these statistics.

As would be expected, the values of $E(T,B)$ show a sharp reduction as $T$ rises from 0 to 1 to 2 out.

**THE ADVISABILITY OF THE INTENTIONAL BASE ON BALLS**

For most of the strategies to be discussed, it is never certain that the intended play will or even can be carried out as intended. However, the issuing of an intentional base on balls is completely under the control of the fielding side, barring such improbable incidents as a wild pitch or an attempt by the batter to hit.

There are two usual purposes behind the decision to issue an intentional walk. One is to avoid pitching to a particularly dangerous batter who is followed in the order by a much weaker batter. The other is in the situations $\{T,2\}$ or $\{T,23\}$, when filling of first base will permit a force play or a double play on a subsequent ground ball.
Strategies in Baseball

| Data  | B | T  | N(T,B) | P(0|T,B) | P(1|T,B) | P(2|T,B) | P(>2|T,B) | E(T,B) | \(a/\sqrt{N}\) |
|-------|---|----|--------|---------|---------|---------|----------|--------|----------|
| 59/60 | 0 | 0  | 6561   | .747    | .136    | .068    | .049     | .461   | .012     |
| 59/60 | 0 | 1  | 4664   | .855    | .085    | .039    | .021     | .243   | .011     |
| 59/60 | 0 | 2  | 3710   | .933    | .042    | .018    | .007     | .102   | .008     |
| 59/60 | 1 | 0  | 1728   | .604    | .166    | .127    | .103     | .813   | .031     |
| 59/60 | 1 | 1  | 2063   | .734    | .124    | .092    | .050     | .498   | .022     |
| 59/60 | 1 | 2  | 2119   | .886    | .045    | .048    | .021     | .219   | .016     |
| 59/60 | 2 | 0  | 294    | .381    | .344    | .129    | .146     | 1.194  | .083     |
| 59/60 | 2 | 1  | 657    | .610    | .224    | .104    | .062     | .671   | .043     |
| 59/60 | 2 | 2  | 779    | .788    | .158    | .038    | .016     | .297   | .024     |
| 59/60 | 3 | 0  | 67     | .12     | .64     | .11     | .13      | 1.39   | .09      |
| 59/60 | 3 | 1  | 202    | .307    | .529    | .104    | .060     | .980   | .072     |
| 59/60 | 3 | 2  | 327    | .738    | .208    | .030    | .024     | .355   | .040     |
| 59/60 | 12 | 0 | 367    | .395    | .220    | .131    | .254     | 1.471  | .087     |
| 59/60 | 12 | 1 | 700    | .571    | .163    | .119    | .147     | .939   | .051     |
| 59/60 | 12 | 2 | 896    | .791    | .100    | .061    | .048     | .403   | .032     |
| 59/60 | 13 | 0 | 119    | .13     | .41     | .18     | .28      | 1.94   | .15      |
| 59/60 | 13 | 1 | 305    | .367    | .400    | .105    | .128     | 1.115  | .077     |
| 59/60 | 13 | 2 | 419    | .717    | .167    | .045    | .071     | .532   | .054     |
| 59/60 | 23 | 0 | 73     | .18     | .25     | .26     | .31      | 1.98   | .18      |
| 59/60 | 23 | 1 | 176    | .27     | .24     | .28     | .21      | 1.56   | .10      |
| 59/60 | 23 | 2 | 211    | .668    | .095    | .170    | .067     | .687   | .080     |
| 59/60 | F | 0  | 92     | .18     | .26     | .21     | .35      | 2.22   | .20      |
| 59/60 | F | 1  | 315    | .303    | .242    | .172    | .283     | 1.642  | .105     |
| 59/60 | F | 2  | 283    | .671    | .092    | .102    | .135     | .823   | .085     |

\[
\sum N = 27027
\]

| 52/60 | F | 0  | 172    | .17     | .27     | .17     | .39      | 2.254  | .145     |
| 52/60 | F | 1  | 419    | .310    | .242    | .186    | .262     | 1.632  | .080     |
| 52/60 | F | 2  | 527    | .645    | .114    | .110    | .131     | .861   | .06      |

The advisability of the strategy (aside from the factor associated with the individual skills of the batters) can be examined by using the data from Table I to compare probable scores following the situation \(\{T,B\}\) before the walk with \(\{T,B'\}\) afterwards. This is done in Table II.
There are several cases for which
\[ \Delta = P(> r | T, B') - P(> r | T, B) < 0, \]
indicating that the probability of scoring more than \( r \) runs has been reduced by issuing the walk. However, in order to determine whether the difference \( \Delta \) is statistically significant, it must be compared to \( \sigma_\Delta \), the standard error of \( \Delta \). This is computed from the formula
\[ \sigma_\Delta^2 = P(r | T, B) [1 - P(r | T, B)] / N(T, B) + P(r | T, B') [1 - P(r | T, B')] / N(T, B') \]
derived in the Appendix.

**TABLE II**

**The Circumstances under Which the Issuing of a Base on Balls Increases the Probability of Scoring**

| Before | After | \( P(> r | T, B) \) | \( P(> r | T, B') \) | \( \Delta \) | \( E(\Delta | T, B) \) | \( E(\Delta | T, B') \) | \( \Delta \) |
|--------|-------|--------------------|--------------------|----------|----------------|----------------|--------|
| 2      | 0     | 12                 | .619               | .605     | .275           | .254           | .671   |
| 2      | 1     | 12                 | .390               | .429     | .166           | .147           | .671   |
| 2      | 2     | 12                 | .390               | .429     | .054           | .048           | .671   |
| 23     | 0     | \( F \)            | .82                | .83      | .57            | .31            | .939   |
| 23     | 1     | \( F \)            | .73                | .69      | .49            | .21            | .939   |
| 23     | 2     | \( F \)            | .332               | .355     | .237           | .131           | .939   |

We will adopt the weak criterion \( \Delta > \sigma_\Delta \) to determine whether a difference is to be regarded as significant.

In Table II, none of the differences \( \Delta \) that are negative are as large (in absolute value) as \( \sigma_\Delta \), so that there is no clear statistical evidence that an intentional walk given to a batter in a team consisting only of average batters ever reduces the probability of scoring.

In a number of cases \( \Delta \) is positive and exceeds \( \sigma_\Delta \). These are indicated by the word ‘Yes’ in the column headed ‘\( \Delta \)’, and represent cases in which a walk is detrimental to the chances of the fielding team.

A blank in the \( \Delta \) column indicates that the difference is not significant according to the stated criterion. It will be observed that there are more blanks than Yesses, so that in the majority of these circumstances it cannot be determined whether the issuing of a base on balls to a member of a team of average batters is a good or a bad strategy. The wisdom of the strategy will turn mainly on the above- or below-average skill of the batter and of those who follow him in the order.
As explained previously, the decision regarding the advisability of a particular strategy must be made with regard to the state of the game.

Suppose that the Home Team $H$ are batting in the $i$th inning, and are 1 run ahead, with the situation $\{T, B\}$. The fielding team $V$ can convert this situation into $\{T, B'\}$ by giving up an intentional base on balls. This move will increase their probability of winning the game if

$$\Omega(H, i, l, T, B) = \sum_{r=0}^{\infty} P(r | T, B) W(l+r)$$

$$\Omega(H, i, l, T, B') = \sum_{r=0}^{\infty} P(r | T, B') W(l+r)$$

i.e., $\Delta = \Omega(H, i, l, T, B') - \Omega(H, i, l, T, B) < 0$.

However, to be statistically significant, it is necessary that

$$\Delta^2 > \sigma^2 = \text{Var} \Omega(H, i, l, T, B') + \text{Var} \Omega(H, i, l, T, B).$$

### Table III

<table>
<thead>
<tr>
<th>Situation before</th>
<th>$B$</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number out</td>
<td>$T$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>$H$ ahead by</td>
<td>$l$</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BB$ helps $V$?</td>
<td>$\Delta$</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### Intentional Walk in Last Half of Ninth Inning

Table III shows the results of such calculations for the last half of the ninth inning.

In Table III, in the row labelled ‘$\Delta$’, the word ‘Yes’ signifies that $\Delta < -\sigma$, and occurs for the case with men on second and third, one out, and the Home Team one run behind. As a matter of fact, $\Delta = -0.041$, and $\sigma = 0.038$, and a difference of this sign and magnitude would be expected to be observed on 14 per cent of occasions between two random variables drawn from the same population.

There are a number of cases in Table III, indicated by blanks, for which $|\Delta| < \sigma$, and a number marked by ‘No’ for which $\Delta > \sigma$, indicating that the base on balls is inadvisable.
**Intentional Walk in First Half of Ninth Inning**

Similar calculations for the first half of the ninth inning show a large number of cases for which the change in probability of winning is less than \( \sigma_\Delta \), but only one for which the fielding team \( H \) can produce a significant increase in their probability of winning, by issuing a base on balls. This instance occurs when the score is tied and \( V \) have runners on second and third with one out. \( \Delta = -1.33\sigma \), which sign and magnitude would occur on 9 per cent of occasions if there were no true difference.

In most cases where \( H \) are ahead, they are distinctly ill-advised to give an intentional walk.

**Intentional Walk in Other Innings**

Tabulation of all the possible circumstances would be quite lengthy, and will not be attempted here.

In general, it may be concluded that, aside from the personal abilities of the batters about to come up in the actual situations, there are not many circumstances in which an intentional walk improves the chances that the fielding team will win the game. The circumstances are most favourable for the strategy when the fielding team is ahead by one run and the situation is \( \{1,23\} \).

**THE CIRCUMSTANCES IN WHICH A DOUBLE PLAY SHOULD BE ATTEMPTED**

In most situations a double play will be very much to the advantage of the fielding team. But, with two or three men on base and none out, a double play extinguishing runners at second and first base will allow the other(s) to advance. In particular, it is easy to see that the fielding team would not want to try a second-to-first double play that allowed the winning run to score from third base.

Some information regarding this situation was derived in reference 6, but it will now be possible to analyze the situation more fully.

The fielding team will place their infield differently, according as they wish to make their play at home plate or second-to-first. For the play at home the infield will be drawn in; for the double play it will be set deep. Thus, there is a strategic decision to be taken before the ball is hit. It is, of course, very probable that neither play will be possible, but the calculations will compare the effects that the two strategies would have on the probability of winning the game if one or the other were in fact carried out as planned.

For this problem we have to compare only the alternate situations after the execution of the selected strategies. The probabilities of scoring more
than 0, 1, and 2 runs are shown in Table IV. In the column headed \('\Delta\)', either \('DP\)' or \('H\)' is entered according as the double play or the play to home reduces the probability of scoring most. The symbol \('H\)' in the column headed \('B'\) means that a run has come home. In all cases the differences are statistically significant.

With men on first and second, either kind of double play (third-to-first or second-to-first) reduces the probability of scoring, and also the expected score, by more than does a force on the runner at third.

With men on first and third, the second-to-first double play produces a

<table>
<thead>
<tr>
<th>Situation before</th>
<th>Play</th>
<th>Situation after</th>
<th>(P(\Delta^*,B'))</th>
<th>(\Delta)</th>
<th>(P(\Delta^*,B'))</th>
<th>(\Delta)</th>
<th>(P(\Delta^*,B'))</th>
<th>(\Delta)</th>
<th>(P(\Delta^*,B'))</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>DP</td>
<td>3</td>
<td>.262</td>
<td>DP</td>
<td>.054</td>
<td>DP</td>
<td>.024</td>
<td>DP</td>
<td>.355</td>
</tr>
<tr>
<td>DP</td>
<td>2</td>
<td>2</td>
<td>.212</td>
<td>DP</td>
<td>.054</td>
<td>DP</td>
<td>.016</td>
<td>DP</td>
<td>.297</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>1</td>
<td>.429</td>
<td>DP</td>
<td>.266</td>
<td>DP</td>
<td>.147</td>
<td>DP</td>
<td>.939</td>
</tr>
<tr>
<td>13</td>
<td>DP</td>
<td>H,0</td>
<td>2</td>
<td>H</td>
<td>.067</td>
<td>DP</td>
<td>.025</td>
<td>DP</td>
<td>1.102</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>1</td>
<td>.429</td>
<td>H</td>
<td>.266</td>
<td>DP</td>
<td>.147</td>
<td>DP</td>
<td>0.939</td>
</tr>
<tr>
<td>F</td>
<td>DP</td>
<td>H,3</td>
<td>2</td>
<td>(DP)</td>
<td>.332</td>
<td>DP</td>
<td>.054</td>
<td>DP</td>
<td>1.355</td>
</tr>
<tr>
<td>DP</td>
<td>23</td>
<td>2</td>
<td>.332</td>
<td>(DP)</td>
<td>.237</td>
<td>DP</td>
<td>.067</td>
<td>DP</td>
<td>0.687</td>
</tr>
<tr>
<td>H</td>
<td>F</td>
<td>1</td>
<td>.690</td>
<td>H</td>
<td>.448</td>
<td>H</td>
<td>.262</td>
<td>H</td>
<td>1.632</td>
</tr>
</tbody>
</table>

very much smaller probability of scoring more than one or more than two runs. However, the expected score is smaller after the play to home. Naturally, the probability of scoring one or more runs, which is certainty for the double play, is smaller after the play to home.

With the bases full, a home-to-first double play is the best for the fielding side, but in practice there is seldom time to execute this, and a play to home plate is likely to produce one out only. Naturally, the force at home produces the smaller probability of scoring one or more, but the second-to-first double play produces a substantially smaller expected score.

**Fielding Strategy in the Last of the Ninth**

Comparison of the values of

\[
\Omega(H,9,l,T',B') = \sum_{r=0}^{\infty} P(r|T',B') H_0 W(l+r)
\]

shows that with men at first and third, or first, second, and third, and none out, the fielding team \(V\) should play to home plate if the score is tied or
if they lead by one run. In all other circumstances the double play is more advantageous. All of the differences are statistically significant.

**Fielding Strategy in the First of the Ninth**

In the first of the ninth, if they are ahead by more than one run the fielding side should always attempt the double play. If they lead by one run, and the bases are full with none out, there is no significant difference between the two strategies. The play to home plate is the better strategy for $\{0,F\}$ when the score is tied or the Visitors lead, and for $\{0,13\}$ when $H$ lead by one, the score is tied, or $V$ lead.

**THE CIRCUMSTANCES UNDER WHICH A SACRIFICE IS ADVISABLE**

By the use of a skillful sacrifice bunt it is usually possible to advance all runners one base while the batter is put out at first. The laying down of a successful sacrifice bunt is not an easy matter, especially when it has been anticipated by the fielding side, and several undesired results may follow, such as a double play, or the extinction of a runner instead of the batter. It is also possible that everyone will be safe. However, for this study we will assume that the sacrifice will produce the intended result, and we will examine the desirability of this result in various circumstances.

**Runner on First Base, with None or One out**

The two usual reasons given for sacrificing a runner from first to second base are that he can probably score directly from second if a succeeding batter hits a single, and that the opportunity for the fielding side to make a double play is reduced.

A successful sacrifice converts the situation from $\{T,1\}$ to $\{T+1,2\}$. Comparison of $P(r|T,1)$ with $P(r|T+1,2)$ in Table I shows that the probability of scoring exactly one run is increased, but the probabilities of scoring at all, of scoring two, and of scoring more than two are all decreased. The expected score is decreased substantially. These statements are true both for $T=0$ and $T=1$, and all are statistically significant except for the probability of scoring at all with a man on first and none out.

To decide between the two alternative strategies of trying to hit and trying to sacrifice it would really be better to employ data obtained from situations $\{T,1\}$ when the team decided to try to hit. In fact, the data from Table I includes a number of cases in which a sacrifice was attempted (not always with the expected result). However, the inclusion of this data will not alter the sign of $P(r|T+1,B') - P(r|T,B)$, for which a positive sign is taken to signify that a sacrifice increases the probability of scoring $r$, nor will it reverse the sign of the inequalities used to determine whether
the execution of a sacrifice will increase the probability of winning the game in a particular circumstance.

For the last half of the ninth inning it is found that

\[ \Omega(H,9,l,T+1,2) = \sum_{r=0}^{\infty} P(r|T+1,2)_{H9} W(l+r) \]

\[ <\Omega(H,9,l,T,1) = \sum_{r=0}^{\infty} P(r|T,1)_{H9} W(l+r) \]

for \( T=0,1 \), and all values of \( l \leq 0 \). The difference is less than \( \sigma_\Delta \) for \( T=l=0 \), but is significant for the other situations.

This indicates that, with average batters coming up, the Home Team is not well advised to sacrifice with a runner on first base in the last of the ninth, except possibly with none out and the score tied.

In the first half of the ninth, there is no case for which a sacrifice by the Visiting Team with a runner on first base (and average batters coming up) makes a significant increase to their probability of winning the game. The sacrifice is a bad strategy when they are behind, or the score is tied with one out. When they are ahead, or the score is tied with none out, the difference is not significant.

Other Situations on the Bases, with None or One out

Table V shows all of the situations in which a sacrifice could conceivably help the batting team. The entry ‘H’ in the column ‘B’ indicates that a runner has scored. The entry ‘Yes’ in the table signifies that the probability of scoring more than \( r \), or the expected score \( E \), would be increased by more than \( \sigma_\Delta \) by the execution of a sacrifice.

Most of the ‘Yes’ items occur in the column giving the probability of scoring one or more runs. Also, most of them are found for situations with a runner on third base.

The placing of a bunt to bring a runner home from third base is the famous ‘squeeze play,’ and is difficult to execute. If the batter misses the ball, the runner is very likely to be trapped between third and home. Thus, the possibility that the attempted sacrifice will produce an unsought result, which is neglected in these calculations, is particularly important in the cases with runners on third base.

BASE STEALING

In the three previous sections it was assumed that the chosen strategy could be executed, but it was not certain, and required calculation, to determine whether the carrying out of the strategy increased the probability of winning the game.

Consideration of the advisability of attempting to have a runner steal a base is different. In this case there is no doubt that the advancement
of the runner increases the probability of winning. But it certainly cannot be assumed that the attempt to steal will be successful.

The combined records of both major leagues for the years 1954, 55, 58, and 59 show 2,959 stolen bases and 2,057 instances when runners were caught stealing, so that the proportion of successes was 0.588. However, the success of certain players was very much better than this. Table VI shows the results during 1959 of those players who attempted to steal on twenty or more occasions.

These figures include attempts to steal second, third, and home base, and double steals. However, the vast majority are undoubtedly single attempts to steal second base. This is attempted much more frequently than the others, because of the importance of getting the runner to second base (from where he can score on a clean single, and is not subject to a second-to-first double play), and because the catcher must make a long throw across the diagonal of the diamond.

When a runner attempts to steal a base, he may succeed, he may be caught stealing, or another play may occur (for example, if the batter hits
Strategies in Baseball

the ball). In this analysis we will consider only the first two possibilities, and compare the probabilities of scoring and of winning the game in the two alternatives that

(a) the attempt to steal succeeds, and converts the situation

\[ \{T, B\} \text{ into } \{T', B'\}, \]

(b) the runner is caught stealing, converting the situation

\[ \{T, B\} \text{ into } \{T+1, B''\}. \]

Let \( S \) be the probability that the attempted steal will succeed. Then the probability of scoring \( r \) runs by the end of the half-inning will be converted by the decision to attempt a steal from \( P(r|T,B) \) to \( SP(r|T,B') + (1-S)P(r|T+1,B'') \). This will be an increase if

\[ S > \frac{P(r|T,B) - P(r|T+1,B'')}{P(r|T,B') - P(r|T+1,B'')} \]

where we define \( S_c \) as the critical probability of success above which the probability of scoring will be increased.

---

**TABLE VI**

Success of Base Stealing Attempts by Individual Players

<table>
<thead>
<tr>
<th>League, player</th>
<th>Stolen bases</th>
<th>Caught stealing</th>
<th>Probability of success</th>
</tr>
</thead>
<tbody>
<tr>
<td>National 1959</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mays............</td>
<td>27</td>
<td>4</td>
<td>.87</td>
</tr>
<tr>
<td>Pinson..........</td>
<td>21</td>
<td>6</td>
<td>.78</td>
</tr>
<tr>
<td>Neal............</td>
<td>17</td>
<td>6</td>
<td>.74</td>
</tr>
<tr>
<td>Cepeda..........</td>
<td>23</td>
<td>9</td>
<td>.72</td>
</tr>
<tr>
<td>Taylor..........</td>
<td>23</td>
<td>9</td>
<td>.72</td>
</tr>
<tr>
<td>Moon............</td>
<td>15</td>
<td>6</td>
<td>.72</td>
</tr>
<tr>
<td>Gilliam.........</td>
<td>23</td>
<td>10</td>
<td>.70</td>
</tr>
<tr>
<td>Robinson........</td>
<td>18</td>
<td>8</td>
<td>.69</td>
</tr>
<tr>
<td>White...........</td>
<td>15</td>
<td>10</td>
<td>.60</td>
</tr>
<tr>
<td>Blasinghame....</td>
<td>15</td>
<td>15</td>
<td>.50</td>
</tr>
<tr>
<td>Ashburn.........</td>
<td>9</td>
<td>11</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>206</td>
<td>94</td>
<td>.69</td>
</tr>
</tbody>
</table>

| American 1959  |              |                 |                        |
| Mante...........| 21           | 3               | .87                    |
| Aparicio.........| 56           | 13              | .81                    |
| Jensen..........| 20           | 5               | .80                    |
| Landis..........| 20           | 9               | .69                    |
| Allison.........| 13           | 8               | .62                    |
| Power...........| 9            | 13              | .41                    |
|                | 139          | 51              | .73                    |
Table VII shows $S_e$ for a number of possible circumstances. The column headed ‘Contested Steal’ indicates the base at which the fielding team attempts to make the putout, and in the attempt for which $S_e$ is calculated. ‘Other Advance’ signifies that it is an attempted double steal, in which the advance indicated in this column is assumed to succeed, whether the contested steal succeeds or fails.

**TABLE VII**

**Critical Probabilities of Success if an Attempted Steal Is to Increase the Probability of Scoring**

<table>
<thead>
<tr>
<th>Situation before</th>
<th>Contested steal</th>
<th>Other advance</th>
<th>Critical probability of success $S_e$</th>
<th>$P(&gt;0)$</th>
<th>$P(&gt;1)$</th>
<th>$P(&gt;2)$</th>
<th>$E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2nd</td>
<td>.53</td>
<td>.79</td>
<td>.66</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.62</td>
<td>.83</td>
<td>.78</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.54</td>
<td>.73</td>
<td>.88</td>
<td>.96</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>2nd</td>
<td>.79</td>
<td>.41</td>
<td>.56</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.85</td>
<td>.49</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.00</td>
<td>.67</td>
<td>.73</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>2nd $3\rightarrow$H</td>
<td>.00</td>
<td>.51</td>
<td>.73</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.28</td>
<td>.55</td>
<td>.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.64</td>
<td>.83</td>
<td>.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3rd</td>
<td>.52</td>
<td>.81</td>
<td>1.00</td>
<td>.67</td>
<td>.84</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.50</td>
<td>.54</td>
<td>.77</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.42</td>
<td>.49</td>
<td>.68</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>3rd $1\rightarrow$2</td>
<td>.63</td>
<td>.46</td>
<td>.72</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.52</td>
<td>.67</td>
<td>1.00</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.26</td>
<td>.81</td>
<td>.96</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>$H$</td>
<td>.79</td>
<td>.65</td>
<td>.75</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.53</td>
<td>.53</td>
<td>.77</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.28</td>
<td>.55</td>
<td>.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>$H$ $1\rightarrow$2</td>
<td>.41</td>
<td>.57</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.63</td>
<td>.68</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.33</td>
<td>.90</td>
<td>.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>$H$ $2\rightarrow$3</td>
<td>.37</td>
<td>.21</td>
<td>.50</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.54</td>
<td>.43</td>
<td>.46</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.36</td>
<td>.73</td>
<td>.55</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>$H$ $2\rightarrow$3</td>
<td>.70</td>
<td>.87</td>
<td>.86</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>.61</td>
<td>.98</td>
<td>.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td>.36</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A blank in a column headed $P(>r)$ signifies that, according to the data in Table I, completion of a successful steal will not increase the probability of scoring more than $r$ runs. It seems illogical that this should really be the case. However, out of the 21 cases in Table VII for which
\[ \Delta = P(>r|T,B') - P(>r|T,B) < 0 \]
we have $|\Delta| < \sigma_\Delta$ 17 times and $|\Delta| < 2\sigma_\Delta$ every time, so that these results can be attributed to sampling error.

In examining Table VII we should remember that the probability of success in an attempted steal of second by an average runner is about 0.59, but that the best runners achieve a much higher probability (0.7 to 0.8 or higher). Moreover, the characteristics of the pitcher’s motion and the catcher’s arm are very important in determining the probability of success in any given circumstance.

The situation in which a steal is most frequently attempted is \{T,1\}, i.e., a man on first base only, especially if he is a better than average runner. The first three lines of Table VII indicate that this strategy is marginal if one run is badly needed (i.e., to tie the score or go one ahead, especially near the end of the game) and the runner is an average one, but is not generally advisable if two or more runs are required.

With men on first and third, it would usually be advisable to attempt the steal of second if the runner on third could come in to score, but if this cannot be assumed (as it cannot be for baseball of a high standard) the attempted steal is attractive only in the cases with one or two out and an important requirement for two or more runs.

In default of knowledge regarding the probability of success of stealing third or home, it is more difficult to draw conclusions from the lower parts of the table. If the probability of a successful theft of third base is higher than 0.5, then the double steal with men on first and second is attractive, especially with one out. The advisability of an attempt to steal home when one run is badly needed increases with the number of men out. The critical probability of success with two out is only about 1/4 to 1/3.

Calculations of $S_c$ for the ninth inning were made for the case \{T,1\} of a runner on first base only. For the last half of the ninth, the only situations for which $S_c < 0.60$ are with the score tied and less than two out, but $S_c < 0.70$ when the Home Team is one run behind with any number out. For the first half of the ninth $S_c < 0.60$ with none out and the score tied, or the Visitors two or three ahead, and with one out and the Visitors one run behind. $S_c \leq 0.72$ with none out, and $S_c > 0.72$ with two out. In general, the attempt to steal second near the end of the game appears more advisable with none out than with one out, and distinctly inadvisable with two out, except when the batting team is one run behind.
These data for stolen bases can also be used to study the advisability of attempting to extend or 'stretch' a hit to one more base, or for base runners to take chances as they advance after a hit. The base runner or coach must make a rapid judgment of the probability of success, but Table VII indicates what it must be in order to warrant the attempt. It is, for example, much more desirable to try for home if there are two out.

A MEASURE OF BATTING EFFECTIVENESS

The measure of batting effectiveness most commonly quoted, for an individual and for a team, is the simple batting average, the proportion of official times at bat \( (AB) \) which resulted in safe hits \( (H) \). An official time at bat is supposed to represent an occasion on which an attempt was made to hit, so that actions other than a safe hit that helped the team (such as a base on balls or a sacrifice) do not reduce the batting average. However, for this purpose all safe hits are deemed equal, so that a lucky single with the bases empty and two out counts as much as a home run with the bases full.

Another measure less frequently used is the 'slugging percentage.' To obtain this measure the total bases \( (TB) \) gained by hitting are added, with each single \( (1B) \), double \( (2B) \), triple \( (3B) \), and home run \( (HR) \) contributing one, two, three, and four total bases respectively. The slugging percentage is the ratio of total bases to official times at bat.

A third index of batting effectiveness is 'runs batted in' \( (RBI) \). However, to obtain an \( RBI \) it is necessary that a teammate be on base (unless a home run is hit), so that the \( RBI \) total represents exploits of others as well as the individual batter. Also, certain actions such as being hit by pitcher or receiving a base on balls with the bases full, or being thrown out at first while a runner scores from third, count as runs batted in, although they may not reflect intentional skillful acts on the part of the batter.

It is interesting to compare the value of two batters when one has the higher batting average but the other has a larger proportion of long hits. For example, at the beginning of the 1960 season Kuenn of Detroit, who had the highest batting average in the league for the previous season, was traded for Colavito of Cleveland, who had tied with Killebrew of Washington for the most home runs. Some of the details of the batting records of these three players and for Kaline of Detroit, who finished second in the batting averages, are given in Table VIII for the 1959 season.

Note that Colavito's 151 hits and Killebrew's 132 hits each produced more total bases (and more runs batted in) than did Kuenn's 198 hits.

A new approach to the assessment of batting effectiveness could be based on three assumptions:

(a) that the ultimate purpose of the batter is to cause runs to be scored
(b) that the measure of the batting effectiveness of an individual should not depend on the situations that faced him when he came to the plate (since they were not brought about by his own actions), and

(c) that the probability of the batter making different kinds of hits is independent of the situation on the bases.

It is generally believed that the third assumption is not true, but that there are so-called ‘clutch hitters’ who are particularly successful in critical situations. Evidence on this point is difficult to secure.

The value of a hit toward the scoring of runs can be estimated from the data in Table I, since a hit which converts situation \{T,B\} into \{T,B'\} increases the expected number of runs by \(E(T,B') - E(T,B)\). The relative frequency with which a batter is faced with each of the 24 possible situations \{T,B\} can be estimated from the values of \(N(T,B)\) in the rows of Table I labelled ‘59/60.’ For the calculation it is assumed that runners always score from second or third base on any safe hit, score from first on a triple, go from first to third on 50 per cent of doubles, and score from first on the other 50 per cent of doubles.

For example, in computing the value of a double, one of the 24 situations is \{2,12\}. The batter will find this awaiting him on \(896/27027 = 0.033\) of the occasions that he comes to bat. The expected score is \(E(2,12) = 0.403\). After the double the situation may be \{2,23\} with one run in, or \{2,2\} with two runs in. Weighting equally, the new expected score is

\[
1/2 \left[ 1 + E(2,23) \right] + 1/2 \left[ 2 + E(2,2) \right] = 1.992.
\]

The value of the double was 1.589 runs in this situation, but to obtain the value for an ‘average’ occasion we must add 0.033 \(\times 1.589\) to the terms for the 23 other situations. The resulting sum is 0.817.

The result of these calculations is shown in Table IX.

Thus, a home run increases the expected score, on the average, by 1.42 runs, about 3\(\frac{1}{2}\) times as much as a single.

The weighting of 1:2:3:4 used to compute the slugging percentage is not very different, but overvalues the triple and the home run.
Using these values for the performances of the four batters of Table VIII, we obtain in Table X their values in equivalent runs.

The 42 home runs of Colavito and Killebrew slightly exceed the value of Kuenn's 140 singles. However, when based on official times at bat, the most effective of the four was Kaline, who produced 0.206 equivalent runs per official time at bat.

D'Esopo and Lefkowitz used a simplified mathematical model to relate batting records to run production, including bases on balls and hit by pitcher. They found that the individual rankings by their method correspond more closely with the ranking by slugging percentage than by the batting average.

**TABLE IX**

<table>
<thead>
<tr>
<th>Type of hit</th>
<th>Single</th>
<th>Double</th>
<th>Triple</th>
<th>Home Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.41</td>
<td>0.82</td>
<td>1.06</td>
<td>1.42</td>
</tr>
<tr>
<td>Ratio</td>
<td>1.00</td>
<td>1.97</td>
<td>2.56</td>
<td>3.42</td>
</tr>
</tbody>
</table>

**TABLE X**

<table>
<thead>
<tr>
<th>Player</th>
<th>AB</th>
<th>1B</th>
<th>2B</th>
<th>3B</th>
<th>HR</th>
<th>Total</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kuenn</td>
<td>561</td>
<td>58.1</td>
<td>34.3</td>
<td>7.4</td>
<td>12.8</td>
<td>112.6</td>
<td>.291</td>
</tr>
<tr>
<td>Kaline</td>
<td>511</td>
<td>49.4</td>
<td>15.5</td>
<td>2.1</td>
<td>38.3</td>
<td>105.3</td>
<td>.206</td>
</tr>
<tr>
<td>Colavito</td>
<td>588</td>
<td>35.3</td>
<td>19.6</td>
<td>0.0</td>
<td>59.6</td>
<td>114.5</td>
<td>.195</td>
</tr>
<tr>
<td>Killebrew</td>
<td>546</td>
<td>28.2</td>
<td>16.3</td>
<td>2.1</td>
<td>39.6</td>
<td>106.2</td>
<td>.195</td>
</tr>
</tbody>
</table>

**REMARKS**

It must be reiterated that these calculations pertain to the mythical situation in which all players are 'average.' The allowance for the deviation from average performance of the batter at the plate, and those expected to follow him, or of runners on the bases, can be made by a shrewd manager who knows his players.

If it were desired to provide a manager with a guide to the advisability of attempting various strategies in different situations, it would be possible to complete calculations of the type outlined here for all stages of the game and scores, pertaining to average players. It would also be possible, although onerous, to compute tables for nonaverage statistics, perhaps based on the past records of the individual players on the team.
The concept of the electronic computer in the dugout is a distasteful one, but, if progress demands it, this is the type of calculation for which it could be programmed.

SUMMARY AND CONCLUSIONS

1. In order to judge the advisability of any particular strategy, it is necessary to take into account not only the immediate situation on the bases, but also the inning and the score.

2. The most valuable run for increasing the probability of winning the game is the run that would put the team one ahead. The run that would tie the score is nearly as valuable. All others are slightly less valuable early in the game, and become increasingly less valuable as the game proceeds.

3. The statistics summarized in this paper allow an estimate to be made of the probability distribution of the number of runs that will be scored between the arrival of a new batter in any situation and the end of the half-inning.

4. Leaving aside the particular skills of individual batters, there are not many circumstances in which the issuing of an intentional base on balls increases the probability that the fielding team will win the game. The most favourable case for this strategy is with runners on second and third, one out, and the fielding team one run ahead.

5. When there is a runner at first base and none out, a double play represents the best strategy for the fielding team except when it would permit another runner to score the tying or lead run from third base. In this case it is better to make the single putout at home plate.

6. With an average batter (followed in the order by other average batters), the strategy of sacrificing with a runner on first base only does not appear to be a wise one. The ‘squeeze play,’ in which a runner scores from third base on a sacrifice, is a good strategy if one run is badly needed, and if it can be accomplished.

7. The success of individuals in stealing bases varies greatly, and the advisability of attempting to steal a base is very much dependent on the probability that the individual runner can complete the theft successfully. An estimate of this probability should take into account the idiosyncracies of the pitcher and catcher as well. With an average runner on first only, an attempt to steal second may be justified if one run is needed to tie or to go one ahead. In the late innings, the attempt to steal second is more advisable with 0 out than 1 out, and is unfavorable with two out except when the batting team is one run behind. The advisability of an attempt to steal home is highest with two out, especially when one run is badly needed, but should not be attempted unless the probability of success is greater than 1/4.

8. The ‘slugging percentage,’ which measures the average number of bases gained by the batter by his own hitting, is also quite a good indication of the contribution that the batter’s performance would have made towards the expected number of runs scored by his team, if he had been faced by an average set of situations. A more accurate ratio than 1:2:3:4 for the relative value of singles, doubles, triples, and home runs is 1:2:2\frac{1}{2}:3\frac{1}{2}. A value index based on these ratios shows
George R. Lindsey

some players fairly far down in the listing of batting averages to be more valuable than some near the top.

APPENDIX

ESTIMATION OF SAMPLING ERRORS

Expected Score

The variance \( \sigma^2 = \text{Var} \ P(r|T,B) \) can be calculated directly from the observed data. The standard error of \( E(T,B) \), the expected score, is simply \( \sigma / \sqrt{N(T,B)} \), and is listed in the last column of Table I.

Difference between Two Probabilities of Scoring

When we compare \( P(>r|T,B) \) with \( P(>r|T,B') \), the two terms come from different distributions. To estimate the standard error of \( P(>r|T,B) \) within its own distribution, note that it is computed by making \( N(T,B) \) observations, out of which \( n_r \) produced more than \( r \) runs, and \( N(T,B) - n_r \) produced \( r \) or less. \( P(>r|T,B) \) is estimated by \( n_r / N(T,B) \), and its variance will be that for a binomial distribution, i.e.,

\[
\text{Var} \ P(>r|T,B) = P(>r|T,B)(1 - P(>r|T,B))/N(T,B).
\]

A similar formula applies for \( \text{Var} \ P(>r|T,B') \), and since they are independent the variance of the difference \( \Delta \) is simply the sum of the two variances.

Difference between the Two Probabilities of Winning

If we set \( \mu_W(l+r) = \lambda_{l+r} \) and consider these to be numerical coefficients with no statistical variance of their own, we then wish to estimate the variance of the function

\[
\Omega = \sum_{r=0}^{m} \lambda_{l+r} P(r|T,B).
\]

The terms \( P(r|T,B) \) have been estimated by \( n_r / N \), where \( N = N(T,B) \) observations of the situation \( \{T,B\} \) produced \( n_0 \) cases for which \( r = 0 \), \( n_1 \) for \( r = 1 \), \ldots, etc.

If the true probabilities of scoring \( r \) runs, in an infinitely large population from which we took our sample of \( N \) observations, are \( p_0, p_1, \ldots, p_r, \ldots \), then the probabilities of obtaining \( n_0, n_1, \ldots \) occurrences of the different values of \( r \) are generated by the multinomial distribution

\[
(N! / n_0! n_1! \cdots) p_0^{n_0} p_1^{n_1} \cdots,
\]

for which

\[
\text{Var}(n_r) = N p_r (1 - p_r),
\]

and

\[
\text{Cov}(n_r, n_t) = -N p_r p_t.
\]

Using the formula\([b]\) for the variance of the sum \( S_m \) of correlated random variables \( Z_0, \ldots, Z_m \),

\[
\text{Var}(S_m) = \sum_{r=0}^{m} \sigma_r^2 + 2 \sum_{r<t} \text{Cov}(Z_r, Z_t),
\]

and, putting

\[
Z_r = \lambda_{l+r} P(r|T,B) = \lambda_{l+r} n_r / N,
\]
we obtain
\[ \sigma^2 = \text{Var}(Z_r) = \lambda^2_{t+r} p_r (1 - p_r)/N, \]
and, finally,
\[ \text{Var} \Omega = \sum_{r=0}^{\infty} \lambda^2_{t+r} p_r (1 - p_r)/N - 2 \sum_{r<t} \lambda^2_{t+r} \lambda_{t+t} p_r p_t/N \]
in which we use \( P(r|T,B) \) as our best estimate of \( p_r \).

In the last half of the ninth inning the calculation of \( \text{Var} \Omega \) can be considerably simplified, since
\[ \lambda_{t+t} = H_{9W}(l+r) = \begin{cases} 0 & \text{for } r < -l \\ \frac{1}{2} & \text{for } r = -l \\ 1 & \text{for } r > -l \end{cases} \]
\[ \Omega(H,9,l,T,B) = \frac{1}{2} P(-l|T,B) + \sum_{r=-l-1}^{\infty} P(r|T,B) \]
\[ = \frac{1}{2} P(-l|T,B) + P(> -l|T,B) \quad \text{for } l = 0, -1, -2, \ldots \]

\[ \text{Var} \Omega = \frac{1}{4} P(-l|T,B)[1 - P(-l|T,B)]/N(T,B) \]
\[ + P(> -l|T,B)[1 - P(> -l|T,B)]/N(T,B) \]
\[ - P(-l|T,B)P(> -l|T,B)/N(T,B), \]
and, for the case of \( l = 0 \), for which \( P(0|T,B) + P(> 0|T,B) = 1 \),
\[ \text{Var} \Omega = \frac{1}{4} P(0|T,B)[1 - P(0|T,B)]/N(T,B). \]

REFERENCES