## Baseball Retrograde Analysis

## Jerry Butters \& Jim Henle

The Mathematical Intelligencer
ISSN 0343-6993
Volume 40
Number 4

Math Intelligencer (2018) 40:71-76 DOI 10.1007/s00283-018-9846-y


Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to selfarchive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

# Baseball Retrograde Analysis 

Jerry Butters and Jim Henle ©

This is a column about the mathematical structures that give us pleasure. Usefulness is irrelevant. Significance, depth, even truth are optional. If something appears in this column, it's because it's intriguing, or lovely, or just fun. Moreover, it is so intended.

The box score is the catechism of baseball, ready to surrender its truth to the knowing eye. -Stanley Cohen, The Man in the Crowd

Retrograde analysis was introduced in this column a few months ago. ${ }^{1}$ In a retrograde analysis puzzle you are presented with a position in a game and asked questions about what happened earlier. The term is almost exclusively applied to chess games and chess positions, but one can devise retrograde problems in many two-player games. ${ }^{2}$ It has never been applied, however, to the eighteen-player game of baseball. Never, that is, until now.

The idea of applying logical deduction to baseball is due to the first author of this column, who began amusing himself some years ago by analyzing baseball box scores. He found he could prove surprisingly detailed facts about games from certain box scores. The second author was charmed and excited when he learned of this.

Consider the following problem: Here's the batting order of the Mudville Slugs:

1. Flynn
2. Blake
3. Casey
4. Hobbes
5. Davis
6. Shlabotnick
7. Thayer
8. Cooney
9. Barrows

Suppose I tell you that in the ninth inning, Casey came to bat for the fourth time. The bases were loaded with two out. Casey struck out, leaving the Mudville nine with another loss. How many runs did Mudville score over the course of the game?

Surprisingly, it's possible to answer this question with very little knowledge of baseball. We'll do that in a moment. The simple tool will be logic, the same tool one uses in chess retrograde analysis and in mathematics.

We realize that not all readers of the Intelligencer are baseball fans. We ourselves weren't fans of chess when we first encountered Raymond Smullyan's The Chess Mysteries of Sherlock Holmes, but we quickly fell in love with his logical puzzles. In addition to the pleasure of the problems, retrograde analysis gave us a chance to relate to chess free of all its baggage.
$>$ Jim Henle, Department of Mathematics and Statistics, Burton Hall, Smith College, Northampton, MA 01063, USA
e-mail: pleasingmath@gmail.com

[^0]Baseball has baggage too, but perhaps readers can come to enjoy its logical treats.

## The First Puzzle

Casey came to bat for the fourth time. The bases were loaded with two out. Casey struck out, leaving Mudville with another loss. How many runs did Mudville score over the course of the game?

In baseball, a batter either gets on base or is put out. Once on base, the player either scores, makes an out, or is left on base at the end of the inning. That means that every plate appearance (PA) is recorded either as an out, a run, or a player left on base. We could write this as

$$
\mathrm{PA}=\text { outs }+ \text { runs }+ \text { left on base } .
$$

We know how many plate appearances there were in this game, since the last player to bat was the third player in the lineup, Casey, and moreover, Casey had exactly four plate appearances. This means that he, Flynn, and Blake each had four appearances, and the rest of the players had three appearances, for a total of 30 .

Mudville played nine innings. At three outs per inning, that's 27 outs. Using our formula, we get

$$
30=27+\text { runs }+ \text { left on base }
$$

That leaves only three for the number of runs and the number of players left on base. But we know that in the ninth inning, the bases were loaded when Casey strode to the plate, that is, there were three players on base. When Casey struck out, all three were left on base. That leaves zero for runs. Mudville scored no runs!

## Box Scores

Before the internet, newspapers printed box scores for all major league games. The box score for the game above might have looked like this:


For the batters, "ab" is at bats, "r" is runs, " $h$ " is hits, and "bi" is runs batted in. You may notice that Mudville has only 27 at bats, though you know they had 30 plate
appearances. The difference comes from the fact that walks don't count as at bats, and Mudville had three walks-you can see that in the pitching line. The Waffletown pitcher Keefe gave up three bases on balls ("bb").

For the pitchers, "ip" is innings pitched, " $h$ " and " $r$ " are as above, "er" is earned runs-more about that later-"bb" is walks, or bases on balls, and "so" is strikeouts. The number for innings pitched is calculateed as 1 for every three outs for which the pitcher was on the mound. Later, you will see a fraction like $4 \frac{2}{3}$ under ip. That means that while that player was pitching, the opposing team made $4 \frac{2}{3} \cdot 3=14$ outs.

At the bottom, "LOB" is left on base, "DP" is double plays, "3b" and "2b" are triples and doubles.

## A Second Puzzle

There's an old saying in baseball: "a walk is as good as a hit." But with a little mathematics (and some misdirection) we can show that a walk is better than a hit, in fact, that it's better than a home run! We'll do it in a series of three puzzles.

Here's the box score of another game between Waffletown and Mudville, but all the information about runs is missing:


Three puzzles:

1. Can you figure out who won?
2. There were no home runs in the game. Suppose we keep all the numbers the same but for each of three Mudvillians-Casey, Davis, and Shlabotnick-we change one hit to a home run. How much does that help Mudville?
3. Suppose instead we change those hits of Casey, Davis, and Shlabotnick to walks (and keep the other stats the same). How much does that help Mudville?

Answers to the three puzzles:

1. We can calculate the score of the game. The sum of the numbers of at bats and walks is the number of plate appearances: ${ }^{3}$

$$
\mathrm{PA}=\mathrm{ab}+\mathrm{bb} .
$$

We also have as before that

$$
\mathrm{PA}=\text { outs }+ \text { runs }+ \text { left on base } .
$$

Putting these together, we find that Mudville scored three runs, and Waffletown scored four. Mudville lost again!
2. Now what is the effect of changing three hits to home runs?

| Waffletown |  |  |  | Mudville |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | ab | r | h bi |  |  | ab | r | h | bi |
| Patkin ss | 5 |  | 2 |  | Flynn 3b | 3 |  | 1 |  |
| Tatum 3b | 5 |  | 1 |  | Blake lf | 4 |  | 0 |  |
| Schacht lf | 4 |  | 1 |  | Casey 1b | 4 |  | 1 |  |
| Altrock rf | 4 |  | 1 |  | Hobbes cf | 4 |  | 1 |  |
| Lanigan c | 4 |  | 2 |  | Davis c | 4 |  | 1 |  |
| Abbott 1b | 4 |  | 0 |  | Shlabotnick rf | 4 |  | 3 |  |
| Costello 2b | 4 |  | 2 |  | Thayer ss | 4 |  | 2 |  |
| Lardner cf | 3 |  | 0 |  | Cooney 2b | 4 |  | 1 |  |
| Keefe p | 3 |  | 0 |  | Barrows p | 4 |  | 0 |  |
| Totals | 36 |  | 9 |  | Totals | 35 |  | 10 |  |
| Waffletown <br> Mudville |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | ip | h | $r$ er | bb | so |  |  |  |  |
| Keefe (W) |  | 10 |  | 1 |  |  |  |  |  |
| Barrows (L) | 9 | 9 |  | 2 | 5 |  |  |  |  |
| LOB--Waffletown 7, Mudville 4, DP--Waffletown 2, HR--Casey, Davis, Shlabotnick, 3b--Abbott, 2b---Thayer, Costello (2) |  |  |  |  |  |  |  |  |  |

Surprisingly, there is no change! We haven't changed ab or bb or outs or LOB (left on base). The equations remain the same. Mudville's score is still 3. Mudville still loses!

Note that if we had changed four hits to home runs, the situation would be different. We would then have an impossible box score, because we could prove that Mudville scored only three runs and we could also prove that it scored four runs (four home runs guarantees four runs). This might cause you to worry that perhaps one of the box scores above is impossible, but they are all good. It's not difficult to construct games that have the appropriate numbers.
3. And finally, what is the effect of changing hits to walks?


The change raises plate appearances, PA, by 3 , and that in turn raises the number of runs by 3. Mudville wins, 6 to 4. Home runs didn't help. Walks did!

The misdirection here is that we didn't actually trade hits for walks. If we had, that would have reduced the number of at bats (because walks don't count as at bats). Instead, we piously said we wouldn't change any numbers except the numbers of hits and walks. That meant keeping the at bats constant (which consequently increased the number of plate appearances). It's not surprising that doing so adds runs, especially when we also kept the number of players left on base constant.

## A Puzzling Problem

Hadleyburg was the home team. They won the game 6 to 4 . One of their players, Ike Farrell, scored all six runs. How many players did Hadleyburg leave on base in the seventh inning?
How on earth can the information given tell us anything about the seventh inning?!

The key is that Ike scored all the runs; no one else scored. For Ike to score a run in an inning in which no one else scores, he must be one of the first three batters, because the players appearing before him must be out before he scores. ${ }^{4}$ Furthermore, in an inning in which he scores, there can be no more than seven plate appearances.

[^1]This is just an application of the same equation, $\mathrm{PA}=$ outs + runs + LOB, but applied to a single inning. Since there are three outs in an inning and no more than three left on base, $\mathrm{PA} \leq 3+1+3$. Thus Ike can't lead off in consecutive innings.

Ike can, however, score in consecutive innings. The way that can happen (and the only way) is if he leads off in one inning and seven players come to the plate that inning. Then Ike will come up third in the next inning. He can't, however, score three innings in a row.

We're given that Hadleyburg was the home team. The visiting team always bats first, so they finish their nine innings before the home team. If after eight and a half innings the home team is ahead, the game is over, since additional runs for the home team wouldn't change the result. But after eight innings, Hadleyburg must already have at least five runs, because Ike can't score two runs in the ninth inning. So Hadleyburg must have won the game after eight and a half innings. So Hadleyburg wouldn't have batted in the ninth inning. Thus Hadleyburg scored all its runs in just eight innings.

For Ike to score six times in eight innings, the scoring must have gone like this:

## 11011011 x.

In particular, in the seventh inning, Ike must be the first batter (as described earlier), and seven players must come to the plate. That results in three players being left on base!

## A Real Game

Baseball retrograde analysis has many similarities to ordinary (chess) retrograde analysis-a range of interesting puzzles from easy to exceedingly intricate, some standard techniques for analyzing problems, and a reliance (in especially tricky problems) on some of the less well known rules of the game. But there is also an exciting difference. Retrograde analysis is seldom applied to actual chess games, because for nearly all positions appearing in actual games, there are too many ways in which the board could have developed. But baseball retrograde analysis can be applied to real games. You can learn a lot, though many questions will remain unanswered.

Consider this game, the fourth of the 1941 World Series between the New York Yankees and the Brooklyn Dodgers:

| New | York |  |  |  |  | Brooklyn |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | ab | r | h | bi | so |  |  | ab | r | h | bi | so |
| Sturm 1b | 5 | 0 | 2 | 2 | 0 |  | Reese ss | 5 | 0 | 0 | 0 | 0 |
| Rolfe 3b | 5 | 1 | 2 | 0 | 0 |  | Walker rf | 5 | 1 | 2 | 0 | 0 |
| Henrich rf | 4 | 1 | 0 | 0 | 1 |  | Reiser cf | 5 | 1 | 2 | 2 | 1 |
| DiMaggio of | 4 | 1 | 2 | 0 | 0 |  | Camilli 1b | 4 | 0 | 2 | 0 | 0 |
| Keller lf | 5 | 1 | 4 | 3 | 0 |  | Riggs 3b | 3 | 0 | 0 | 0 | 1 |
| Dickey c | 2 | 2 | 0 | 0 | 0 |  | Medwick lf | 2 | 0 | 0 | 0 | 0 |
| Gordon 2 b | 5 | 1 | 2 | 2 | 0 |  | Allen p | 0 | 0 | 0 | 0 | 0 |
| Rizzuto ss | 4 | 0 | 0 | 0 | 0 |  | Casey p | 2 | 0 | 1 | 0 | 1 |
| Donald p | 2 | 0 | 0 | 0 | 1 |  | Owen c | 2 | 1 | 0 | 0 | 0 |
| Breuer p | 1 | 0 | 0 | 0 | 0 |  | Coscarart 2b | 3 | 1 | 0 | 0 | 2 |
| Selkirk ph | h 1 | 0 | 0 | 0 | 0 |  | Higbe p | 1 | 0 | 1 | 0 | 0 |
| Murphy p | 1 | 0 | 0 | 0 | 0 |  | French p | 0 | 0 | 0 | 0 | 0 |
| Totals | 39 | 7 | 12 | 7 | 2 |  | Wasdell lf | 3 | - | 1 | 2 | 0 |
|  |  |  |  |  |  |  | Totals | 35 | 4 | 9 | 4 | 5 |
| New York Brooklyn | 10 | 0 | 2 | 0 | 00 | 4--- |  |  |  |  |  |  |
|  | 00 | 0 | 2 | 20 | 00 | 0--- |  |  |  |  |  |  |
|  | ip h |  | r er |  | bb | so |  |  |  |  |  |  |
| Donald | 4 | 6 | 4 | 4 | 3 |  |  |  |  |  |  |  |
| Breuer | 3 | 3 | 0 | 0 | 1 | 2 |  |  |  |  |  |  |
| Murphy (W) | 2 | 0 | 0 | 0 | 0 | 1 |  |  |  |  |  |  |
| Higbe 3 | $32 / 3$ | 6 | 3 | 2 | 2 | 1 |  |  |  |  |  |  |
| French | 1/3 | 0 | 0 | 0 | 0 | 0 |  |  |  |  |  |  |
| Allen | 2/3 | 1 | 0 | 0 | 1 | 0 |  |  |  |  |  |  |
| Casey (L) 4 | $41 / 3$ | 5 | 4 | 0 | 2 | 1 |  |  |  |  |  |  |
| HBP--Henrich (Allen), DP New York 1, LOB--New York 11 Brooklyn 8, PB--Owen, HR--Reiser, 2b---Keller (2), Gordon, Walker, Camilli, Wasdell |  |  |  |  |  |  |  |  |  |  |  |  |

This is a famous game. It was featured in Ken Burns's documentary Baseball. The game is famous because of a singular event that took place in the ninth inning. The Yankees were leading in the series, two games to one. At the top of the ninth inning, the Dodgers were ahead, 4 to 3 . A win for Brooklyn would tie the series at two games all.

As you can see from the box score, the Yankees scored four runs in the ninth to take the lead and eventually win the game. That gave them a three to one lead in the series. They went on to become world champions.

Let's imagine that we are baseball anthropologists centuries in the future puzzling over the box score of that game. Can we deduce what happened that enabled the Yankees to win?

The pitching line for Casey shows that he gave up four runs in the ninth but the runs were unearned (the column labelled " $r$ " is runs, the column labeled "er" is earned runs). A run is unearned if there was a defensive misplay and if, in the opinion of the official scorers, the run would not have been scored had the misplay not happened.

There were no errors (E) listed in the box score of this game, but there was a passed ball (PB). A passed ball occurs when the catcher fails to catch the pitch. That's a misplay. Since this was the only possible cause of the unearned runs in the ninth inning, the passed ball must have occurred in the ninth inning. ${ }^{5}$

When the catcher fails to catch a pitch, a player on base can run to the next base and will be safe unless the catcher recovers the ball quickly and the runner is tagged out. If one or more of the players on base score then or later, the runs might be unearned. But that wouldn't affect runs scored by later base runners. Thus, a passed ball that only moves existing runners along would yield at most three unearned runs. But there were four unearned runs in this inning.

As all baseball anthropologists know, when a catcher drops the third strike, the batter is not automatically out. He can run, and if he can get to first base before the ball does, he is safe. ${ }^{6}$

Thus, a passed ball can result in putting an additional runner on base, and that can happen only in the context of an uncaught third strike. When this happens, the passed ball gives the offense both an extra base runner and an extra out, so that any number of subsequent runs could be unearned.

Striking out and getting safely to first is a rare event. But we, as mathematical anthropologists, have just proved that it must have happened in this game.

We can, by the way, identify the player who famously struck out and scored. The only Yankees who struck out were Henrich and Donald. Donald was out of the game by the ninth inning, so Henrich must have struck out and reached first base.

## Puzzles for You

Most of the puzzles that follow avoid special situations. But if you're curious, the official rules can be found on the web. A link to the rules will appear on this column's website.

## The Special Day

$\star$ Below is the box score of a major league game. The player's names have been omitted but will be included in
the answer (on the website). In this game, player K achieved a special feat (as rare for hitters as pitching a nohitter is for pitchers). What did K achieve?


The stars in front of these puzzles are a rough measure of their difficulty. In this first puzzle, you should discover K's feat pretty easily (one star). If you additionally figure out what he did in the eleventh inning, we'll give you two stars.

It's actually possible to discover what he did in the eighth inning, but we don't have enough stars to give you for that.

## The Missing Data

$\star \star$ In the following box score, one row of data is missing. Can you fill it in?

[^2]
$\star \star$ Can you show that it's possible for a game to have that box score?

## Hitting for the Cycle

Shlabotnick, Mudville's center fielder, performed the rare feat of hitting for the cycle, that is, hitting a single, a double, a triple, and a home run-all in one game, as shown in the following box score.

| Waffletown |  |  |  |  | Mudville |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $r$ | h | bi |  |  | ab | r | h | bi |
| Patkin ss | 5 | 1 | 2 | 0 |  | Flynn 3b | 3 | 0 | 1 | 0 |
| Tatum 3b | 5 | 0 | 2 | 0 |  | Blake lf | 4 | 0 | 0 | 0 |
| Schacht lf | 3 | 0 | 0 | 0 |  | Casey 1b | 4 | 1 | 1 | 0 |
| Altrock rf | 3 | 0 | 1 | 1 |  | Hobbes cf | 4 | 1 | 2 | 0 |
| / / Gaedel rf | 2 | 0 | 0 | 0 |  | Davis c | 4 | 1 | 1 | 0 |
| Lanigan c | 5 | 0 | 1 | 0 |  | Shlabotnick rf | 4 | 2 | 4 | 4 |
| Abbott 1b | 3 | 1 | 0 | 0 |  | Thayer ss | 2 | 0 | 0 | 1 |
| Costello 2b | 4 | 0 | 1 | 0 |  | Cooney 2b | 3 | 0 | 0 | 0 |
| Lardner cf | 2 | 0 | 0 | 1 |  | Barrows p | 2 | 0 | 0 | 0 |
| Keefe p | 2 | 0 | 0 | 1 |  | Simpson p | 1 |  | 0 | 0 |
| Hamman p | 0 | 0 | 0 | 0 |  | Laloosh p | 0 | 0 | 0 | 0 |
| Veeck p | 1 | 0 | 0 | 0 |  | Totals | 31 | 5 | 9 | 5 |
| Thurber p | 0 | 0 | 0 | 0 |  |  |  |  |  |  |
| Totals | 33 | 2 | 7 | 2 |  |  |  |  |  |  |
| Waffletown ... | 1 | 1 | 0 | 00 | 00 | 0 0---2 |  |  |  |  |
| Mudville ...... | 1 | 0 | 0 | 00 | 10 | 0 3---5 |  |  |  |  |
|  | h | r | er | bb |  |  |  |  |  |  |
| Keefe | 5 | 1 | 1 | 1 | 4 |  |  |  |  |  |
| Hamman | 0 | 0 | 0 | 0 | 0 |  |  |  |  |  |
| Veeck | 1 | 1 | 1 | 0 | 0 |  |  |  |  |  |
| Thurber (L) 1 | 3 | 3 | 3 | 0 | 0 |  |  |  |  |  |
| Barrows 6 | 6 | 2 | 2 | 3 | 1 |  |  |  |  |  |
| Simpson | 0 | 0 | 0 | 0 | 1 |  |  |  |  |  |
| Laloosh (W) 1 | 0 | 0 | 0 | 3 | 3 |  |  |  |  |  |
| DP Waffletown, LOB--Waffletown 10, Mudville 4, HR--Shlabotnick, 2b--Shlabotnick, 3b--Shlabotnick, SF--Thayer, Keefe, WP--Veeck |  |  |  |  |  |  |  |  |  |  |

$\star \star$ Which of the four did Shlabotnick hit in the ninth inning?
$\star \star$ Which of the four did Shlabotnick hit in the first inning?
$\star \star \star \star \star$ Which of the four did Shlabotnick hit in the seventh inning?
$\star \star$ You should know now what Shlabotnick's remaining hit was. But in which inning did he hit it?

The Triple Play
$\star \star \star \star \star$ In this game, Mudville executed a triple play. In what inning did that happen?


## Let Us Hear from You

You can reach us at pleasingmath@gmail.com. Answers and comments (yours and ours) will be available at www. math.smith.edu/~jhenle/pleasingmath/.


[^0]:    1،'The Entertainer," Mathematical Intelligencer 40 (2018), no. 2, 76-80.
    ${ }^{2}$ See, for example, joekisenwether.wordpress.com/non-chess-retrograde-analysis/.

[^1]:    ${ }^{3}$ We're keeping things simple here. In general, there are five circumstances in which an at bat is not a plate appearance: a walk, being hit by a pitch, a sacrifice bunt, a sacrifice fly, and fielder obstruction. When any of these occur, they are recorded in the box score.
    ${ }^{4}$ Ike can't pass other runners. That's a rule.

[^2]:    ${ }^{5}$ The only other misplays besides errors and passed balls are "catcher interference" and "fielder obstruction."
    ${ }^{6}$ On an uncaught third strike, the batter is allowed to run if first base is unoccupied or if there are two outs, in which case running is always permitted. Weirdly, when this happens, the pitcher is recorded as having achieved a strikeout even though the runner isn't out!

