BASEBALL RETROGRADE ANALYSIS

JERRY BUTTERS AND JIM HENLE

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.

¹ "The Entertainer," The Mathematical Intelligencer (40) 2, 2018.

 $^{^2 \}rm See,$ for example, https://joekisenwether.wordpress.com/non-chess-retrograde-analysis/

Consider the following problem:

Here's the batting order of the Mudville Hamsters:

- 1. Flynn
- 2. Blake
- 3. Casey
- 4. Hobbes
- 5. Davis
- 6. Shlabotnik
- 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 Mudville 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.

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 is recorded either as an out, a run, or a player left on base. We could write this as

PA = Outs + Runs + 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 line up, Casey. Moreover, Casey had exactly four plate appearances. That 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 equation,

30 = 27 + Runs + Left on base.

That leaves only 3 for runs and left on base. But we know that in the ninth inning the bases were loaded for Casey, that is, there were three players on base. When Casey struck out all three were left on base. That leaves 0 for runs. Mudville scored no runs!

Box Scores

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

	ab	r	h	b	i					ab	r	h	bi
Patkin ss	4	1	1	C)		Flyr	ın 3b		3	0	0	0
Tatum 3b	3	1	1	C)		Blak	e lf		3	0	1	0
Schacht lf	4	1	1	C)		Case	ey 1b		4	0	0	0
Altrock rf	4	2	2	З	3		Hobb	es cf		3	0	0	0
Lanigan c	4	0	1	2	2		Davi	s c		3	0	0	0
Abbott 1b	3	0	0	C)		Shla	botnik	rf	3	0	0	0
Costello 2b	3	0	0	C)		Thay	ver ss		3	0	1	0
Lardner cf	3	0	1	C)		Coor	ley 2b		2	0	1	0
Keefe p	3	0	0	C)		Barr	ows p		3	0	0	0
Totals	31	5	7	5	;		Tota	ıls		27	0	3	0
Waffletown ·		0 0	0	3	0	2	0 0	05					
Mudville		0 0	0	0	0	0	0 0	00					
	ip	h	r	er	1	bb	so						
Keefe (W)	9	3	0	0		3	5						
Barrows (L)	9	7	5	5		4	5						
LOBWaffletc HRAltrock 2								affleto	wn 3	, Mu	dvi	lle	2,

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 when 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 3 base 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 base on balls, and "so" is strike-outs. The number for ip counts 1 for every three outs while the pitcher was on the mound. Later you will see a fraction like 4 2/3 under ip. That means that while the 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:

Waffl	et	οw	n			Mudvi	110	Э		
	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			Shlabotnik 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 ·										
Mudville										
	ip	h	r	er	bb	so				
Keefe (W)	9	10			1	5				
Barrows (L)	9	9			2	5				
LOBWaffleto 2bThayer,		-			6, I	PWaffletown 2	2, 3b	A1	obott	;,

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 Shlabotnik, change one hit to a home run. How much does that help Mudville? 3. Suppose instead we change those hits of Casey, Davis, and Shlabotnik 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

$$PA = ab + bb$$

We also have as before that

PA = Outs + Runs + Left on base.

Putting these together we find that Mudville scored 3 runs and Waffletown scored 4. Mudville lost again!

2. Now what is the effect of changing three hits to home runs?

³We're keeping things simple here. In general, there are five circumstances where 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.

Waffl	et	OW	n			Mud	/i	110	Э		
	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			Shlabotnik	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 ·											
Mudville											
	ip	h	r	er	bb	so					
Keefe (W)	9	10			1	5					
Barrows (L)	9	9			2	5					
LOBWaffletc Davis, Shlabo		-									,

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 one of the box scores above is impossible, but they are all good. It's not difficult to construct games which have the appropriate numbers.

3. And finally, what is the effect of changing hits to walks?

Waffl	et	٥ĩ	n		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		0				
Altrock rf	4		1			Hobbes cf	4		1				
Lanigan c	4		2			Davis c	4		0				
Abbott 1b	4		0			Shlabotnik rf	4		2				
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		7				
Waffletown ·													
Mudville	• •												
	ip	h	r	er	bb	so							
Keefe (W)	9	9			4	-							
Barrows (L)	9	7			2	5							
LOBWaffleto 2bThayer,					6,	DPWaffletown 2	9, 3b	A	bbot	t,			

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 town team. They won the game 6-4. One of their players, Ike Farrell, scored all six runs. How many players were left on base for Hadleyburg 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. This is just an application of the same equation, 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, 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 town 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.

 $^{^4\}mathrm{Ike}$ can't pass other runners. That's a rule.

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

110 110 11x.

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. 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 game of the 1941 World Series between the New York Yankees and the Brooklyn Dodgers:

New	Yo	rl	K				Broo	kl	yr	ı		
	ab	r	h	bi	so			ab	r	h	bi	so
Sturm 1b	5	0	2	2	0	F	eese ss	5	0	0	0	0
Rolfe 3b	5	1	2	0	0	W	alker rf	5	1	2	0	0
Henrich rf	4	1	0	0	1	F	eiser cf	5	1	2	2	1
DiMaggio cf	4	1	2	0	0	C	amilli 1b	4	0	2	0	0
Keller lf	5	1	4	3	0	F	iggs 3b	3	0	0	0	1
Dickey c	2	2	0	0	0	Μ	ledwick lf	2	0	0	0	0
Gordon 2b	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	C	wen c	2	1	0	0	0
Breuer p	1	0	0	0	0	C	loscarart 2b	3	1	0	0	2
Selkirk ph	ı 1	0	0	0	0	Н	ligbe 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	0	1	2	0
						ī	otals	35	4	9	4	5
New York ···	10	0	2 0	0	0 0	4	7					
Brooklyn	0 0	0	22	0	0 0	0	4					
	ip	h	r	er	bb	so						
Donald	4	6	4	4	3	2						
Breuer	3	3	0	0	1	2						
Murphy (W)	2	0	0	0	0	1						
Higbe 3	2/3	6	3	2	2	1						
French	1/3	0	0	0	0	0						
Allen	2/3	1	0	0	1	0						
Casey (L) 4	1/3	5	4	0	2	1						
HBPHenrich PBOwen, HR 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, 2-1. At the start of the ninth, The Dodgers were ahead, 4-3. A win for Brooklyn would tie the series at 2-2.

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 lead in the series 3-1. 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 the 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 labelled ''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 is 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 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 to first and if he can get there before the ball does, he is safe at first.⁶

⁵The only other misplays besides errors and passed balls are "catcher interference" and "fielder obstruction."

⁶For this to be allowed either first base must be open (unoccupied) or else there are two outs. Weirdly, when this happens, the pitcher records a strike out even though

Thus, a passed ball can result in putting an additional runner on base. Indeed, that is the only way a passed ball can do that. 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.

Puzzles for You

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

The Special Day

★ 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 the game, player K achieved a special feat (as rare for hitters as pitching a no-hitter is for pitchers). What did K achieve?

the runner isn't out!

Vis	sit	0	rs				Hom	е	Τe	ear	n	
	ab	r	h	bi	so			ab	r	h	bi	so
A 2b	5	1	1	0	0		J 3b	7	2	2	1	0
B lf	6	1	2	0	0		K rf	7	4	4	3	1
C 1b	6	1	2	3	1		L lf	6	1	4	1	0
D rf	6	2	3	1	2		M cf	5	0	1	1	3
E cf	5	1	2	0	2		N dh	5	1	2	1	0
F1 dh	1	0	0	0	0		0 1b	4	0	1	2	0
F2 ph	0	0	0	0	0		P1 c	3	0	1	0	0
G1 3b	3	0	0	0	2		P2 c	3	0	0	0	0
G2 3b	1	0	0	0	1		Q1 2b	3	0	1	0	0
G3 lf	0	0	0	1	0		Q2 ph	1	0	0	0	0
Нc	4	0	0	1	0		Q3 2b	1	0	0	0	0
I ss	3	0	0	0	0		R1 ss	3	0	0	0	1
Totals	40	6	10	6	8		R2 ph	0	0	0	0	0
							R3 ss	1	1	0	0	1
							Totals	49	9	16	9	6
Visitors		0	1 0	0	0 0	030	2 06					
Home Team	ı	1	0 1	0	0 0	0 1 1	2 39					
	ip	h	r	er	bb	so						
S	6	8	2	2	1	1						
Т	1	1	1	1	0	1						
U	2	5	3	3	3	3						
V	2/3	0	0	0	1	1						
W (L)	1	1	2	0	1	0						
Х	0	1	1	1	0	0						
Y 7	2/3	7	4	4	2	6						
Z 1	1/3	0	0	0	3	2						
ZZ (W)	2	3	2	2	2	0						
	2ъ						am 1, LOB– O 3bK, F					

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?

Waffle	et	OW	n		Mudville							
	ab	r	h	bi					ab	r	h	bi
Patkin ss	4	2	2	0		F	.ynn 3b		3	0	0	0
Tatum 3b						BI	ake lf		4	0	2	0
Schacht lf	4	0	0	0		Ca	asey 1b		3	0	0	0
Altrock rf	4	1	1	2		Ho	bbes cf		3	0	1	0
Lanigan c	3	0	0	0		Da	avis c		4	2	1	0
Abbott 1b	3	0	0	0		Sł	labotni	k rf	4	2	2	1
Costello 2b	3	0	0	0		Tł	hayer ss		4	1	1	3
Lardner cf	3	0	0	0		Co	oney 2b		3	1	1	2
Keefe p	2	0	0	0		Ba	arrows p		1	0	0	0
Veeck p	0	0	0	0			Hardy p	h	1	0	0	0
Gaedel ph	1	0	0	0			Simpson	р	2	0	0	0
Totals							Laloosh	р	0	0	0	0
Waffletown ··· Mudville		30 00 ip	0	3 0		0 0	25 x6		32	6	8	6
Keefe (L. 0-3)	7				1	4					
Veeck	·	1	2	1	1	0	1					
Thurber		0	0	0	0	0	0					
Barrows (W, 4	-2)	4	2	3	3	2	3					
Simpson		4	0	0	0	1	3					
Laloosh		1	2	2	2	0	1					
LOBWaffletown 3, Mudville 6, DP Waffletown 1, Mudville 1, CSTatum, HRHobbes, Altrock, Tatum, Davis, 2bShlabotnik, EAbbott												

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

Hitting for the Cycle

Shlabotnik, 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:

		ab	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	Shlabotnik 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	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	01	0	0 0	0 0 02				
Mudville		1	0 0	0	00	1 0 35				
	ip	h	r	er	bb	SO				
Keefe	5	5	1	1	1	4				
Hamman	1	0	0	0	0	0				
Veeck	1	1	1	1	0	0				
Thurber (L)	1	3	3	3	0	0				
Barrows	6	6	2	2	3	1				
	1	0	0	0	0	1				
Simpson	1	0	0	0	3	3				

 $\star\star$ Which of the four did Slabotnick hit in the ninth inning?

 $\star\star$ Which of the four did Slabotnick hit in the first inning?

 $\star \star \star \star \star$ Which of the four did Slabotnick hit in the seventh inning?

 $\star\star$ You should know now what Slabotnick'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?

Waffleto	wn				Mudville						
	ab	r	h	bi		ab	r	h	bi		
Patkin ss	3	3	3	1	Flynn 3b	5	0	1	0		
Phanatic ss	2	0	0	0	Blake lf	3	0	1	1		
Tatum 3b	5	2	2	1	Casey 1b	4	0	1	0		
Schacht lf	5	1	1	2	Hobbes cf	4	0	1	0		
Altrock rf	4	0	1	1	Davis c	4	0	1	0		
Lanigan c	3	0	1	0	Shlabotnick rf	4	0	1	0		
Abbott 1b	2	0	0	0	Thayer ss	4	0	1	0		
Du Monville 1b	0	0	0	1	Cooney 2b	4	1	1	0		
Costello 2b	3	0	0	0	Barrows p	0	0	0	0		
Lardner cf	4	0	0	0	Simpson p	2	0	1	0		
Keefe p	4	0	0	0	Laloosh p	2	0	0	0		
Totals	36	6	6	6	Totals	36	1	9	1		
Waffletown ··· 2	2 1	0	0 0	1	0 06						
Mudville 0	0 0	0	1 0		0 01						
	h	r	er	bb	50						
ip Keefe (W) 9	9	1	1	1	9						
veele (w) 9	9	1	1	1	9						
Barrows (L) 2 2/3	6	5	5	0	0						
Simpson 3 1/3	0	0	0	0	0						
Laloosh 3	0	1	0	0	0						
LOBWaffletown 6,	Mudv	ill	e 9,	HR-	-Patkin, SFDu Monv	ille	, E-	-F1	ynn		

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