THE CHANGING GAME REVISITED: Major League Baseball Yesterday and Today.

BOB SAWYER

ABSTRACT

Retrosheet's expanding data base no provides data about Reaching base On Error and Innings batted back to 1916. Section Two demonstrates how this information enables deduction of base running outs for teams and leagues Section Three extends the format of The Baseball Encyclopedia's "The Changing Game" to opposition pitching and fielding . The six tables create a statistical profile of play during the 1916, 1921, 1971 and 2019 seasons. The tables identify areas in which Major League baseball changed rapidly between 1916 and 1921 and then continued changing at a more evolutionary pace for another century.

Section One: THE VOCABULARY OF TABLETOP SUCCESS:

Sports Illustrated Baseball™ is a board game in which statistics-based player-charts interact in simulation of baseball.[1] Board gamers control of batting orders and substitutions, with dice rolls determining the outcomes of their choices to bunt or take a gamble on the base paths. Winning SI Baseball is the result of brains and luck rather than ability to throw, catch or hit a real baseball. And yet the game is so well-designed that good tactics for real baseball are nearly always just as useful for an SI Baseball manager.

Fly outs, ground outs, singles, and doubles are subdivided in SI Baseball to allow differing amounts of potential advancement by base runners. The safe on error result designated by "E" has the same effect on batters and base runners as the type of single designated by "1". The results of "WP", "PB", and BK(for Balk) are likewise interchangeable in Si Baseball.

Real baseball has its own examples of these functional equivalencies. Does it matter whether the batters free pass on the fourth ball is scored as a Base on Balls(BB) or a Hit Batsmen(HB)? Scorekeepers of the 19th century sometimes erroneously credited batters with walks on these plays. The batter was is no more bruised one way than the other and he still got rewarded 1st base with the same runner advancement..

A groundout during which all runners advance one base is functionally equivalent to the results of successful sacrifice bunt (i,e, Batter is out, all runners advance one base.) And a fly ball that scores a run is still a run scoring play regardless of whether the plate appearance counts as an At Bat, a Sacrifice Hit, or a Sacrifice Fly. In real

[1] **The game designer was statistician David Neft. Time inc. sold the copyrights to Avlon Hill, a maker of table top board games and the All-time All-star version of Si Baseball was repackages at SUPERSTAR BASEBALL™ . Avalon Hill issued 48 charts for active stars in 1979 with Bob BIscontini as the chart designer . Since that time Avid player such as Elliot Warren, D.R."Randy Cox", Phil Graham, Biscontini, and this author have created un authorized updates. Historically minded readers will find Randy Cox's charts for Negro League stars of particular interest.**

baseball as in SI baseball these scorekeepers' distinctions have nothing to do with action on the field.

Gary Hardegee's incredibly accurate Base-Advance Average[2] is based entirely on functional equivalencies as judged after the fact. Hardegee's methodology ignores traditional classifications, evaluating each play according to three questions: (1) How far did the batter advance on the play or was he put out instead.? (2) How far did the current base runners advance on the play? (3) Which (if any) current base runners are removed during the play. Plays for which all three answers are the same have identical impacts upon Base-Advance Average and upon Run Expectations.

Hardegee's calculations require base runner information for the start and finish of each plate appearance. Score sheets contain this information, but player and league seasonal statistics do not. The problem is that the extended box scores on which official statistics are based summarizes events rather than putting them in context. The charts of SI Baseball provide an intermediate level of detail. Score sheets for the games are devoid of details about how each plate appearences are resolved. And so the tabletop manager only sees the larger picture of batters and runners advancing or being put out. Functional equivalencies are impossible to miss.

Even plays are not functional equivalent can serve the same purpose. For example, the purpose of a stolen base attempt is to change the base-out situation in a maner similar to the usual result of a Wild Pitch(WP), Passed Ball(PB) or Balk(Balk). This is only a partial equivalency because every Stolen Base represents one base of advancement by only one runner whereas more than one base may have been advanced when a Wild Pitch is scored. Nevertheless, advancement on stolen base attempts and advancement on WP, PB, BK and errors serves the same function within a baseball game.

Stolen Bases are officially batter information; Wild Pitches and Balks are pitching stats. Passed Balls and total Errors are found in the fielding section of Retrosheet. . In order to get a full picture of base runner advancement, we need to think holistically, leaving the usual distinctions behind. Given that all the opponents of a league used to be from that league, there was never any obstacle to grouping offense, defense and fielding together to get the full profile of events. Data for opponents pitching and

---

[2] "Base-Advance Average" by Gary Hardegee is posted in the Research Papers section of Retrosheet.org. Correlation of BAA with runs scored was .0.988 for teams compared to .874 with OPS. Hardegee treats advancing from 2nd base to Home as 2 bases out of a possible two but treats 1st base to third advancements as two out of a possible three. Batters outs are treated as 0 advancements out of 4 opportunities and any play that reduces runners or surrenders a base already advanced gets scored as a negative result. (for instance if a batter grounds into a triple play with bases loaded, that gets "scored" in Hardegee's system as negative 3 out of 10 opportunities.

fielding[3] opens the way to run estimation formula for teams that take into account a full spectrum of information. (see endnote 3)

Fans and have always understood that opponents WP and Errors help their team score runs and that lost base runners prevent runs. Unfortunately until Retrosheet made this information public, nobody outside the Elias Bureau had any way to know (how many errors were made by opponents of the 1971 Pirates. Published team information did not even include how innings in which each team batted or many batters reached base safely[4]

Analysts were also handicapped because official scorekeepers are tasked with recording some events but ignoring others which are functional equivalent There is no practical difference between a runner thrown out trying to go 1st to third on a single and a runner who halted at second base only to be thrown out trying to steal third base during the following plate appearance. Prior to 1898 the scorekeepers treated both of these as Caught Stealing[5], so this particular distinction is clearly arbitrary.

Batters thrown out trying to stretch their hits represent a less clear cut example of this enforced scorekeeper blindness. If there are no other runners to advance on the throw. how does this differ from the batter stopping on first base only to be thrown out trying to steal 2nd base on the following plate appearance? Under current practice, the failed base theft is duly noted as a Caught Stealing but the batter's misadventure is tallied only in the fielding events.

Although data created from extended box scores does not tell us when things happened, with a little algebra we can use events that are regularly tracked to tell us about many events that are not. Section Two demonstrates how this is done.

Section Two: THE ALGEBRA OF BASEBALL'S BALANCED BOOKS

There are eleven ways to reach first safely, four of which are types of Hits(H). Retrosheet's Reaching base On Error(ROE) covers Safe on fielder error, 3rd strike Wild Pitch, 3rd Strike Passed Ball, and Fielder's Choice all-safe. The remaining three ways are Base on Balls(BB), Hit By Pitch(HBP) and Defensive interference(XI). Thus the sum

---

[3] **Opposing Piching results are found on the team roster pages. For most of 2020 similar information was found about fielding events. This data was temporarily withdrawn due to coding errors that lead to incorrect totals for teams. Hopefull the data will become available again soon.**

[4] **Scoring was extremely high in the 1870s because runners reaching base on errors were more common than hits in many games, Individual players and teams likewise can be lucky or unlucky in this regard. Luck can also influence how many times a team comes to bat during a season. The 1931 Yankees have slightly inferior statistics to the 1930 Yankees. And yet the 1931 team score more runs per game. The only reason this happened is that the 1931 batted in more innings per game. Without Retrosheet's publication of Opposition innings pitched there would be no way for the public to know this,**

[5] **Total Baseball has caught stealing by the defense for early seasons for which caught stealing by offenses was unofficial. The definition of a stolen base attempt was modified so that there were substantially less attempts in 1898 than in previous seasons and less caught stealing. Ho despite the longer season. However the percentage of successful attempts went down. This is entirely consistent with current numbers which show runners advancing on hits and outs having a lower failure rate than runners trying to steal bases while the ball is in the pitchers hand or being pitched.**

H, BB, HBP, XI and ROE is the number of Plate Appearences in which a player or team or league Reached Base Safely . We will call this sum "REACHED"[i]endnote1

REACHED equals H plus BB plus HBP plus XI plus ROE

SI Baseball batters-charts with high percentages of ROE and XI are more valuable than than charts with lower rates because each safe outcome raises the ratio of REACH to the alternative, which we will be calling "Outs While Batting or "OWB". Every Plate Appearance(PA) results in either success or failure in reaching base safely.

PA equals REACHED plus OWB.

This binary analysis is made possible by how Fielder Choice outs, Double plays, and base runners hit by batted balls are treated. Neither of the first two are cases of in REACHED, so that makes them OWB. In the case of a double or triple play there are accompanying Outs by Base Runners(OBR). Counting these separately is an especially handy practice for seasons in which GIDP information is missing or incomplete. When the runner is hit by the batted ball, the runner is put out and current scorekeeping practice credits a single to the batter. Because the records are arranged this way the play will be counted as OBR in Section Three.. [6]

SI Baseball has separate charts for bunt attempts and batters "swinging away" but real life batters sometimes bunt in order to reach base. Official records make no distinction between failed sacrifice attempts and other ground outs. And data for bunt attempts in Baseball-Reference.com does not cover 1916 or 1921. So we are not positioned to remove bunt hit or bunt outs from the data. And as noted earlier, a ground out that advances runners and a Sacrifice Hit through bunting are functional equivalents. Section Three therefore treats productive outs as a subset of OWB.

The productive aspect of certain outs such as SH and Scoring Fly Balls (SF) can be acknowledged through separate analysis. This approach dispenses with At Bats— (whose definition changed in 1894, 1908, 1926, 1930, 1939, 1940 and 1954) –in favor of Plate Appearences(PA), REACHED, and Outs While Batting(OWB), which always have the same meaning.

Many sabrmetric analysts presume that AB minus Hits plus the sum of CS, GIDP, Sacrifice Hits(SH), and Sacrifice Flies(SF) accounts for 27 outs per nine innings, This is nowhere close to being true for 1916 because ROE were more common than base runner outs not accounted for by CS and GIDP. The analyst's assumption is doubly problematic because it leads to regarding ROE as massively negative events when they are never less valuable than BB or HBP and often more valuable.

---

[6] Given that the function result of this plays is the same as a Fielders Choice out, the author agrees this is a departure from ideal practice. Fortunately such assigned singles are very uncommon and thus represent a negligible portion of REACHED.

Fielders Choice events have a common element, the batter is not put out because the defense tried to get one of the base runners out. Sometimes the defenders succeed and the play is part of OWB. Sometimes they fail and Retrosheet counts such plays as ROE , making then part of REACHED. To avoid ambiguity, we shall refer to all safe Fielder Choice plays as "Fielders Wrong Choices(FWC)"

Baseball-Reference.com does NOT include FWC plays under its Reach Base on Errors columns. Therefore the number of FWC for a player, team or league is the difference between the listed ROE entries on the two websites. Baseball-Reference once used "SOE" for its Reached base on Error column, so "SOE" conveniently denotes data taken from Baseball-Reference.com.

FWC equals ROE minus SOE

Regardless of the season, runners have tagged up and scored on fly ball outs. We shall use Vin Scully's expression because such scoring fly balls have only been called "Sacrifice Fly's" since 1954. From 1908 to 1930 and in 1939 such plays were scored as Sacrifice Hits and tallied under "SH" on the official scorekeepers extended box scores. Today, only bunts that advance runners count as SH, the scoring fly balls are scored as Sacrifice Flys(SF).

Historians and Analysts have very much wanted to know how many times teams bunted to move the runners during the 1908 to 1930 seasons. Baseball Reference.com has now provided a path to an answer. Our terminology for runner advancing bunts will be "sBUNT" and the formula for counting these plays include the same sorts of plays for each of the four seasons. What we do is subtract the unofficial count of Scoring Fly balls found on split pages of Baseball-Reference com from the official count of Sacrifice Hits for the seasons where scoring fly balls were being scored as Sacrifice Hits.

For 1916 and 1921 sBUNT equals=S*H minus SF [ii](endnote2)

For 1971 and 2019 sBUNT equals SH

For the sake of consistent terminology, the polyglot Sacrifice Hits from 1939 and 1908 to 1930 are labeled "S*H" in the first equation rather than "SH" We know S*H from 1954 onward because sBUNT and SF are counted separately But even in recent seasons there are other productive outs that get scored as neither SH nor SF. Had Baseball-Reference listed these events by teams or leagues, Table Five of Section would have included this data for the four seasons of this study.

One indicator of the importance of advancement on defensive misplays is the accuracy of Bill James formula for Expected Runs Allowed(ExRA) by defenses. The B factor includes WP, PB and BK data with values more than twice that of walks. Errors appear in both the A factor and the B factor. Even though ExRA makes no provision for Stolen Bases, Caught Stealing, Doubles or Triples its standard error per team-

season is only 18.14 runs per team season.[7] ExPA multiplies Errors by 0.7 in the A factor for times reaching base, and uses 1.0 in B factor The A factor coefficient coincides with the approximate percentage of errors which are part of SOE. The B factor 95.4% of the coefficient for Hits which are not Home runs. <1/1.048= .954>

The reason the B factor coefficient is larger is that some errors prevent OWB and others only promote base runner advancement. These Errors of Advancement(Eadv) are not as valuable as hits and so the base advancement coefficient had to be larger than 0.7 but somewhat smaller than the coefficient for hits. See endnote3.[iii]

Eadv approximately equals Total opposing Errors(E)[8] minus SOE

The reason the formula is only approximately true is that not every SOE involves an error. STATS™ 1991 Baseball Scorebord tells us that this count would be off by 107 plays for Major League Baseball in 1990[9] , about 4.11 per team. .

The most important information missing from official records after ROE is the number of outs by players already on base(OBR) is for Out by Base Runners..[iv](endnote4) Caught stealing(CS) and Grounding Into Double Plays(GIDP) are available back to World War I. Yet CS and GIDP so rarely involve plays at 3rd base or Home plate that the run value of runners lost in other ways might possibly be greater for some teams.

Hardegee's BAA encompassed all forms of OBR because such plays are easily identified from a score sheet or play by play account. Without the score sheet, keeping track of OBR is stymied by the fact that the extended box scores homogenize base runner outs with other put outs and assists. Making matters worse is that the official categories such caught stealing and total double plays overlap with each other. [10] Fortunately, for any season for which REACHED is known or well estimated it is possible to deduce OBR by working from known data.

The best such method is linked to the fact that extended score sheets are balanced with respect to offensive and defensive events. It stands to reason that Runs Scored

---

[7] P13 The Bill James Handbook 2005 The full formula for Expected Runs allowed is (H + BB + HBP + 0.7 x E minus DP) times sum of [1.048 x (H-HR); HR x 4; 0.7 x (PB + WP + BK); 0.32 x (BB + HB minus IBB) divide by BFP.

[8] As I write this, data about errors and passed balls by opposing teams has been temporarily removed from Retrosheet.org. Before the removal I had copied the posted data through 2017. In lieu of this information for 2019, errors by members of each league were used instead. Therefore the entries for E per 1417.5 innings and Eadv per 1417 innings are only approximately correct for 2019. . When the data is again available I will repost this paper with corrected entries.

[99] Pp 21-22 STATS™ BASEBALL SCOREBOARD by John Dewan, Dan Zminda and STATS. Inc. Sports Teams Analysis & Tracking systems Inc, copyright 1991.

[10] OBR is a mishmash of events some of which fall into more than one category. For example, a successful hidden ball trick may be a case of a pickoff, a double play, a triple play, or even a runner taking an extra base unsuccessfully. Sometimes caught stealing is simultaneously a pickoff or a double play. And caught stealing but safe on error plays should be excluded from the category . It is therefore a practical impossibility to derive OBR through the sum of any combination of outfield assists, pickoffs, total Double Plays, Caught Stealing and Outs Advancing (from a Bill James Handbook.). Of course when evaluating players, every known OBR makes a difference of approximately 0.5 runs.

by the home team must equal Runs Allowed by the road team's pitchers, Nearly as obvious is that Batters Facing Pitcher(BFP) for the road team equal Plate appearences(PA) by the home team. In addition to these and similar pairings, there is also an intimate interconnection between Innings pitched and Runs Allowed with with by one team and their opponents PA and Left on Base (LOB).

When team P plays team Q

PA(by q) equals Runs(by q) plus LOB(by q) plus [3 times IP(by p)]

and

PA(by p). equals Runs(by p) plus LOB(by p) plus [3 times IP(by q)]

<u>WARNING!!!!</u> Endnote5 explains why this will not work for 2020.[v].

Before computer programs did this automatically, the paper form onto which game statistics were compiled for submission to the League office contained a section for "proving" the game reports. The official scorekeeper was instructed to apply the two equations to his extended box score. If either equation failed, that the scorekeeper had made some sort of prior mistake and the box score should not be submitted until corrected.

If proven box scores are correctly entered and summed by the league office then team and league statistics will also pass the above tests Thus when Retrosheet began to provide opposing innings pitched for teams it became possible for the first time to prove team season data. Previously only league totals could be proven, (and then only for seasons with complete records and no interleague play.[11]) The advantage of proven data is that even when elements of the above equations are only approximately accurate, (as is often the case for ROE and XI,) good approximation for LOB and or total Outs-by-Base-Runners(OBR) can be derived by arithmetic.

An Inning Pitched is actually a half inning of play with one team pitching and the other at bat. Hence opposition Innings Pitched and Innings Batted are one and the same. With three outs for each half inning

Total Outs equals 3 times Half Innings Played(IP)

Section Two looks at baseball from the point of view of the team batting. So from this point onward "Runs" means Runs scored. Readers are forewarned to mentally insert "opposition" in front of any statistic normally associated with pitchers or fielders.

---

[11] **Whereas normally we can prove a league by combining batting and pitching numbers, with Interleague play, only baseball as a whole will seem to have balanced books.. In order to prove a league after 1996 one must sum up the opposing pitching data for each team and use this sum rather than innings pitched by league pitchers.**

OWB are far more common than other outs and for some games OWB equaled 3 times IP. But there has never yet been any team-season with 100% OWB. Every team has a different own ratio of OWP to innings Batted. And Section Three shows that these ratios are quite different today than in 1916 and 1921.

As our goal is to describe the action with statistics rather than rating players, we can treat groundout DP and Fly Out DP the same. As mentioned above, a DP counts as one OWB with one OBR attached This practice assures (1) and (2) are simultaneously true.

(1) OWB equals PA minus REACHED,

(2) 3.0 times Innings Played equals OWB plus OBR

Combing (1) and (2) and applying grade school Algebra yields ,

OBR equals (3.0 times IP) minus (PA minus REACHED) .

A second approach to OBR is based on the fact that no matter how a runner reached base, there are only three possible fates: Either he scores or he gets Left On Base, or he gets Put Out. This remains true even for extra innings that begin with a runner on second. Let "FREE" be the number of base runners put on base to start an extra inning.

FREE plus REACHED equals R plus LOB plus OBR

It follows that when FREE equals Zero then

OBR equals REACHED minus (LOB plus RUNS). [12]

When the two formula for OBR are linked and common terms removed, a version of the proven equations is all that remains, This is our reassurance that when applied to proven data, both methods must lead to the same result. Whenever the result appears to be different—as happens with 1916 and 1921. then one or more of the inputs LOB, PA, REACHED or opposition Innings Pitched is inaccurate (see footnote12 and endnote5)

Both approaches required a value for REACHED. And ROE is essential to finding REACHED. Thus approximately accurate estimates for REACHED and OBR required at least approximate values for ROE. Thus outside of a few people working at the Elias

---

[12] **Although this is computationally simpler than the other approach, listed data for LOB leads in some cases to failure of the prove test. This is unfortunate because while Retrosheet's split pages use only proven data, neither LOB nor PA are directly listed. Baseball-Reference lists LOB and PA, but publishes only official numbers even when those numbers are collectively impossible unless Retrosheet's numbers for XI and ROE are substantially incorrect SEE ENDNOTE 3.**

Reasons for official errors include illegible extended box scores/ daily game reports, errors of transcription during the years in which the leagues maintained yearly ledgers, and errors of addition by scorekeeper or league officials. The double-counted game that led to Ty Cobb being declared winner of the controversial 1910 Batting Championship was essentially a transcription error. Entering the same game twice had the effect of entering his Hits and At Bats as 4 for 6 whereas the score sheet was correct that Cobb went 2 for 3 that day..

Bureau, no human being had more than the vaguest idea how many times runners were being put out in seasons prior to the advent of Project Score sheet.

That we have estimates back to 1916 is because Retrosheet's volunteers used official and reconstructed play by play accounts to supplement the extended box scores to include ROE Baseball Reference has recently done the same thing with SF and its own version of ROE,.

Although there is no official counterpart to OBR, there must be three Put Outs per completed inning. Testing for this is another one of the official score keeper's double-checks.

Fielder PO equals 3 times Innings Pitched (by fielders team) and

Opponents PO equals OWB plus OBR .

This applies even when an appeal play is the so-called 4th out of one or more innings . According to official scorer Stew Thornley, the correct treatment of this type of successful appeal officially negates one or more runs and also officially erases the Put Out prior to the appeal.. The scorekeeper then assigns a new Put Out for the fielder at the base where the appeal took place plus an Assist for whoever threw the ball to that base for the appeal. [13]

Section Three

Baseball Evolves from Runners Vs Fielders to Batters Vs Pitchers

Starting in 1921 the game ball is replaced whenever it becomes scuffed or discolored. This practice benefits offense in three ways. (1) It aides enforcement of the ban on trick pitches, especially the emery ball. (2) An unsoiled ball is more visible when pitched. (3) Because previously battered baseballs tend to lose residency, the replacement ball is on the average more resilient than the ball that is removed. Regardless of whether all three of these benefits were intentional, Home Runs, Batting Averages and Runs per nine innings increased in 1921 for the third year in a row.

Four questions dictated grouping of categories in the Tables One through Six below. (1) Does the statistic measure events Impacting players reaching base or does it link to failures to reach base? (2) Do the measured events advance base runners in some way?. (3) Do the measured events involve elimination of base runners that already reached first base? (4) Does the statistic tell us something about where on the field plays were happening?

---

[13]The case books with Major League Baseball's official instructions to scorekeepers are proprietary.. Thus the author could only request clarifications of the procedures from present or former official scorekeepers. Thanks are owed to SABR member Stew Thornley, whose email dated 2/7/2021 also pointed out what I had been calling "balancing the books" is more properly called "proving the box score" and that the daily game reports(aka "dailies") of seasons when the leagues used physical ledgers have been replaced by the paperless extended box scores of today.

All tables in this section are from the point of view of the team at bat. A game is a completed official game of any length. A season is 1417.5 innings. [81 x 9 plus 81 x 8.5 equals1417.5] Six thousand(6000) Plate Appearences is the roundest number within the 5300 to 6500 PA range in which the teams in the study fell.

TABLE ONE. Was 1921 a Respite in the Trend of Increasing True Outcomes.

	runs/game^	UER/game	% UER	*PA/game	TRUE PA/gm	%TRUE	BB/6000	HR/6000	SO/6000
1916 NL	6.69	1.61	24.06%	73.46	12.83	17.46%	395.23	31.31	632.61
1916 AL	7.13	1.64	22.98%	74.79	14.16	18.93%	520.74	18.42	601.84
1921 NL	8.97	1.63	18.11%	76.99	10.95	14.23%	371.18	58.75	431.72
1921 AL	10.00	1.67	16.72%	79.40	13.00	16.37%	486.69	58.88	441.76
1971 NL	7.61	0.87	11.46%	75.98	18.51	24.36%	492.81	112.16	857.43
1971 AL	7.56	0.83	10.99%	75.91	19.11	25.18%	532.82	122.08	856.70
2019 NL	9.35	0.82	8.80%	77.27	27.08	35.05%	513.46	212.23	1377.64
2019 AL	9.56	0.61	6.43%	77.36	27.20	35.16%	509.18	223.71	1377.47

Unearned runs barely changed between 1916 and 1921. Strikeouts temporarily declined by nearly a third and Walks also went down. The increase in HR from 1916 was modest in absolute terms; large in percentage terms.

For some reason AL pitchers issued a lot more BB than their National League counterparts in 1916 and 1921. Had this been a fleeting phenomena the likely explanation would be that National League pitchers were simply better at finding the plate But if this were the case then the difference would not have continued through the 1930's. Four alternative explanations are (A) betting hitting backgrounds in AL parks. B) differences in pitching philosophies (C) Differences in batting philosophies. (D) Difference in umpiring between the two leagues. A study of players who changed leagues is required to choose between explanations.

TABLE TWO: What Happened When the Ball Was Put In-Play?

	Ball in Play/game	% in play	BIP BA	H /game	intentional BB/yr	unintentional BB/yr	HBP/yr	HB or XI endnote6?	% HB&int
1916 NL	59.99	81.66%	0.277	15.82	16.66	367.04	40.57	0.65	0.88%
1916 AL	59.89	80.08%	0.279	15.75	14.32	500.36	40.20	0.74	0.99%
1921 NL	65.62	85.23%	0.308	19.55	22.00	355.64	30.41	0.42	0.54%
1921 AL	65.87	82.96%	0.312	19.88	14.17	496.49	40.67	0.53	0.67%
1971 NL	57.05	75.09%	0.282	16.61	60.11	434.72	32.26	0.42	0.55%

1971 AL	56.35	74,23%'	0.275	16.15	54.46	480.02	35.15	0.45	0.59%
2019 NL	49.30	63.79%'	0.301	16.83	31.41	492.91	68.87	0.89	1.16%
2019 AL	49.36	63.8%0'	0.301	17.06	18.08	502.45	61.55	0.80	1.04%

BIP BA is inversely related to Defensive fielding Efficiency. Thus the defenders of 1921 were notably less efficient than five years before in dealing with the increase balls in play. Perhaps the fresher balls were leaving the bat at greater velocity than in previous seasons and the fielders could not cover the same territory as in 1916. Perhaps batters adjusted their launch angles to take better advantage of livelier balls and the outfielders of the early 1920s had not yet backed far enough away from the plate to compensate. Perhaps the mix of these causes was unique to each batter. Either way, we cannot attribute this to a change in ballparks or playing surfaces because the American League parks were completely unchanged from 1916

TABLE THREE: What Was the Distribution of the Batters' Good Outcomes?

	REACHED /game	% REACHED	1b / yr	2B /yr	triple/yr	IP-HR /yr	OTF-HR/yr	TB/Hit	FWC /yr	Safe On Error/yr
1916 NL	23.21	31.60%	999.48	185.04	74.91	11.19	19.20	1.33	28.20	140.02
1916 AL	24.79	33.14%	1001.80	194.70	69.79	5.18	13.02	1.30	41.66	121.80
1921 NL	26.35	34.22%	1208.81	238.98	87.07	8.19	51.59	1.37	19.31	124.30
1921 AL	28.39	35.76%	1191.90	277.41	89.88	5.57	56.21	1.40	15.39	126.09
1971 NL	24.15	31.78%	1000.36	204.58	37.32	0.65	111.97	1.46	9.82	76.83
1971 AL	24.16	31.82%	965.16	200.19	28.96	0.91	121.55	1.48	9.32	73.86
2019 NL	24.91	32.24%	852.55	276.84	26.28	0.46	216.26	1.71	13.73	50.60
2019 AL	25.02	32.34%	853.07	283.94	25.32	0.53	228.18	1.73	13.35	51.88

Every type of Hit increased from 1916. The 1921 season saw more batters reaching first base safely(REACHED) than in recent times, but far less often through over-the-fence Home Runs(OTF-HR) than in 1971. With so many base runners and batters in motion and relatively few true outcomes, Major League Baseball become noticeably more entertaining than 1916..The author concludes that the surge in attendance of the Roaring Twenties was as much due to more action on the field than to an improved economy.

. TABLE FOUR: How Were Outs by Base Runners Distributed?

	OBR/yr	GIDP/yr	OF DP/yr	other DP	triple play	TOTAL DP	CS	Other	Pkoff
1916 NL	350.08	70.67	13.74	33.61	0.127	118.15	130.36	101.58	23.65
1916 AL	372.94	71.61	17.45	34.71	0.379	124.15	134.14	114.66	15.55
1921 NL	326.06	117.75	12.87	12.33	0.780	143.73	100.19	82.14	21.31

1921 AL	301.43	80.57	14.51	51.01	0.518	146.61	70.71	84.10	17.35
1971 NL	226.47	126.34	6.62	21.23	0.327	154.52	40.18	31.77	12.09
1971 AL	232.96	127.41	6.35	19.47	0.083	153.32	45.14	34.49	10.48
2019 NL	186.75	*114.01*	*4.67*	*15.51*	*0.000*	134.18	26.09	26.48	9.73
2019 AL	188.13	*113.63*	*4.80*	*12.56*	*0.197*	131.19	28.60	28.34	11.18

Even with the huge increase in base runners in 1921, OBR declined significantly from 1916 The drop in OBR is the same size as the drop in caught stealing for both leagues so there was very little change in other forms of base runner outs. 1921 was one of the first seasons in which managers had data about how often their players were bing thrown out trying to steal bases. Perhaps this additional information led to teams to attempt less steals. .

The "other" column is for any play not covered by columns to the left, some of which are pickoff plays which did not also get scored as caught stealing. The Bill James Handbook indicates 5.22 such pickoffs per team for 2019.[14] This means that the remaining 20 odd plays in the column represent runners out trying to advance on hits (or on outs not covered by OF DP). Since there cannot be more pickoffs which are not caught stealing than total pickoffs, the American League of counts for 1921 indicate that there were no less than 66 such outs per team, more than three times as many as in 2019. Such outs in 1971 were probably slightly more come than today. Either Major League players and coaches became progressively better at judging risks since 1921 or they are taking far less chances on being thrown out. Of course both might be true to some extent.

There are only three ways to have a DP that is neither GIDP nor outfield DP: (a) strikeouts with a pickoff or caught stealing on the final pitch;.[15] (b) runners doubled off when an infielder caught a line drive; and (c) runners unsuccessfully trying to advance two bases on ground outs. For some reason, the American league has a VERY HIGH count of these miscellaneous DP in 1921; while the corresponding NL count is very LOW. Either AL batters struck out FAR more often when the Hit and Run play had been called or AL base runners took many foolish risks that NL runners declined. Could this be the influence of Ty Cobb?

TABLE FIVE: How Often Did Teams Employ Small Ball Tactics?

[14] **Pp 453 to 458 The Billl James Handbook 2020. ACTA Sports, copyright 2019.**

[15] **One indicator of a decline in failed hit and run double plays would be a decline in number of double plays involving the catcher. Unfortunately this author neglected to compile and include this data in the study, which was originally intended to compare and contrast play in 1921 as part of the SABR Centennial Commemoration.**

	sBUNT /yr[16]	SF/yr[17]	CS/y	SB %	E adv /Yr	WP/yr	PB/yr	BK/yr	SB/yr	total ADV	little ball outs
1916 NL	120.6	43.9	130.4	56.73%	106.8	31.2	20.2	2.4	170.9	496.0	250.9
1916 AL	146.6	47.6	134.1	57.46%	110.8	32.2	17.3	3.9	181.2	539.7	280.7
1921 NL	131.3	56.0	100.2	51.30%	91.7	23.5	11.8	2.3	105.5	422.1	231.4
1921 AL	139.4	60.7	70.7	56.67%	100.7	23.2	12.0	4.1	92.5	432.6	210.1
1971 NL	75.0	39.5	40.2	64.66%	54.9	48.1	17.6	4.2	73.5	312.8	115.2
1971 AL	72.9	41.5	45.1	61.26%	50.8	43.4	16.8	3.7	71.4	300.5	118.1
2019 NL	34.9	36.5	26.1	73.71%	44.2	51.5	11.1	4.7	73.1	255.9	61.0
2019 AL	16.1	39.1	28.6	72.33%	43.8	66.1	11.8	5.4	74.8	257.1	44.7

Even with the much higher OPS and Batting Averages of 1921, the average team expended over 200 outs per season to one-run-tactics. Advancement on Wild pitches and Passed Balls was much less frequent in 1921 than in 1971 or 2019, but additional Eadv plays meant more total advancement in 1921 Despite the much lower success rate of 1921, Stolen bases were more common in 1921 than in 1971 Once again we are seeing evidence that Major League Baseball was more of a base runner's game in 1921 than in 1971. Table Five shows that the game of today is even less of base runners' game than baseball of 1971

TABLE SIX Which Defenders Were Responsible for Batters Outs?

	Total outs / team	Strikeouts	Catcher PO w/o SO	Outfield PO	Infield Assist	Other Infield Put Outs
1916 NL	4208.37	14.34%	3.66%	22.45%	41.56%	17.99%
1916 AL	4233.38	13.89%	4.16%	22.05%	42.11%	17.78%
1921 NL	4118.50	10.26%	3.81%	25.25%	43.81%	16.88%
1921 Al	4132.50	10.83%	3.77%	24.66%	42.42%	18.32%
1971 NL	4368.92	20.11%	1.67%	23.63%	38.71%	15.89%
1971 AL	4323.83	20.07%	1.66%	24.14%	34.02%	20.11%
2019 NL	4343.87	32.86%	0.29%	22.06%	32.13%	12.67%
2019 AL	4340.87	32.89%	0.20%	23.12%	30.75%	13.04%

Table Six counts pitchers as infielders. Therefore "other infield outs" can only be either unassisted infielder Put Outs or Put Outs with Assists from either the catcher or outfielders. Total outs are not prorated to 1417 innings but instead show the

[16] **sBUNT for 1916 and 1921 are estimated by subtracting estimated SF from the know official SH. .**

[17] **A scoring fly ball was scored as S*H in 1921 and 1916. Baseball Reference had incomplete data based on incomplete play by play records. Partial counts were extrapolated to each teams' actual PA to get team estimated SF, which were then totaled to get a league estimate for use in this column**

actual volume divided by the number of teams so that readers can better measure the disappearance of catcher independent put outs Catchers used to get about 22% of their putouts on plays other than strikeouts.

The percentage of Outfield Putouts relative to all other outs is much the same today as in 1921. However the ratio of such PO to infield Assists is now far higher. Bill James noted that team assist counts are a proxy for ground balls hit by opposing batters. [18] If so then, then Table Six depicts a gradual trend to lower and lower ground ball to air ball ratios for Baseball as a whole. The AL was more of a Fly ball League in 1916 but the NL mostly caught up by 1921. Adaptation to more of a Fly ball approach to hitting continued for a century at a lesser pace after 1921.

The two best explanations for this trend do not exclude the other. Table Six could be showing that many existing hitters made immediate efforts to change to their approach and take fuller advantage of more resilient baseball, The alternative hypothesis, which is supported by Major League Baseball's continuing trend, is the tables show the effect of one "generation" of hitters giving way to a new one with a different approach.

Why might one baseball generation of hitters be different from the next? This would happen if the new generation is taught different techniques by different coaches. Or it could happen because hitters learned to hit under different conditions. Hitting is mostly a matter of reflex and reflexes become harder to unlearn the older a person becomes. The cushioned cork center ball introduced when Babe Ruth and Rogers Hornsby were teenagers was very different from the baseballs of the youth of every previous generation. This put Ruth and Hornsby at a temporary disadvantage as rookies because from 1915 to 1918 baseballs were made from inferior wool.

But then the ball became fully resilient again and by the end of 1920 the Babe and the Rajah had home parks friendly for hitters of deep fly balls. Such advantages would have been fleeting in other eras, but in this case the slightly younger players had come of age trying to hit the mushy balls of World War I. There were effective at hitting for average. During the 1920s starting flank fielders other than Ruth hit .321. [19] yet it was 1925 before a successor generation of power hitters began becoming regulars

We know today that some launch angles are more optimal for hitting for distiance than others. And that many players have started using slow motion to fine tune their swings, which would be unnecessary had they started with ideal mechanics. Does it not follow that the smaller the percentage of batters with optimal swings then the

---

[18] page 114 Win Shares(Stats Inc 2002) Bill James points out that ground balls cannot help but correlate very strongly with Assists, and this is especially true if Assists by Catchers and Outfielders are removed. Likewise outfield Put outs are 100% fly balls. While the ratio of infield assists to outfield put outs need not correlate perfectly with the ground ball/Fly ball ratio. It cannot possibly not correlate strongly with the ratio of ground outs to fly outs.

[19] P 396 The Diamond Appraised, Craig Wright and Tom House Fireside Press div of Simon and Shuster 1989.

bigger the comparative advantage to whoever can regularly produce the best launch angle?

Thus perhaps the reason for Ruth and Hornsby's sustained dominance of their leagues was that they were the first to figure out the best way to hit. If so, then their positions near top of the all-time leader boards for relative slugging average, relative Isolated power and relative OPS[20] should not be such a source of awe.

And yet we are still waiting for a any batter to exceed Ruth's 1921 production of Extra Base Hits, Total Bases or Runs Scored. And no right-handed hitter has bested Hornsby's highs for Total Bases(1922) or Slugging Average(1925) So maybe there was never a better time to be a hitter, but the other stars failed to take full advantage.

CONCLUSIONS

(1) Inside the park home runs were a very significant percentage of all home runs in both 1921 and 1916 but represent a negligible percentage in 1971 and 2019. The most probable explanation is park dimensions. The dimensions for of 8 of the 14 stadiums in 1921 made out of the park Home Runs prohibitively difficult for most hitters but made triples considerably more common than today. Ballparks of today's make triples unlikely and inside the parks all but impossible

(2) With BB, Strikeouts, Hit by Pitches, and OTF Home Runs being less common than in other years of the study, a much higher percentage of runners and batters in 1921 were exposed to risk of being put out while the ball was in play. Although Stolen Base attempts were down from previous seasons, Major League baseball in 1921 remained very much a base runner's game. For every 1417 .5 innings in 1921 a typical team lost more than 210 base runners to causes other than GIDP and expended more than 130 of its outs to advance runners with bunts.

(3) The distribution of outs shows that batters of 1921 were willing to put thee ball on the ground , which would be self-sabotage today but worked well enough during 1920-1922 that eight batting title qualifiers hit better than .385 even with none of them hitting more than 21 home runs. Given that the rate of SOE was so much lower in 1971, it is probable that bad hops on infield ground balls were so common that they formed the basis of many players' offensive philosophy.

(4) Prior to this study that the author believed that better grounds keeping and bigger gloves entirely explained the decrease in unearned runs since 1921. But had these been the causes, Eadv should not have declined in such perfect tandem with SOE. An explanation that better fits the data is that base-path aggressiveness (which generates Eadv) is linked to the type and number of balls in play to nearly the same extent that SOE are.

Less ground balls and increased strikeouts would lead to fewer runners in motion at the instant of contact. In the absence of runners in motion, outfielders have

---

[20] **Pp 268-271 and 288-290 The Hidden Game of Baseball reprinting with foreword by Keith Law. John Thorn and Pete Palmer with David Reuther authors. University of Chicago Press 1984**

additional fractions of a second to get the ball to the infield in time to inhibit base runner advancement. With both fewer runners over all and less of them in motion, each season would have many fewer plays which put outfielders at risk of being charged with an error of advancement if they failed to pick up the ball cleanly or failed to make a strong throw. Meanwhile, even without bigger gloves or consistent playing surfaces, decreasing groundballs will lead to fewer ROE via infield errors..[vii](endnote7))

Under these new conditions and assisted by bigger gloves and , we would expect to find a drastic decline in both types of errors between 1921 and 1971 with its artificial surfaces. The continued rise in true outcome plate appearences after 1971 further reduced errors of all types even though every turf team of 1971 subsequently switched to grass surfaces by 2019..

END NOTES.

---

[i] **Note 1.**

To the maximum extent practical, I have adopted the terminology of Richard Schell's "A Model for Estimating Run Creation" By the Numbers Vol 19 Number 1, Feb 2009 Schell created a formula that he believed was more accurate than BaseRuns,. Unfortunately the full formula is not included in the published article. The author created a spreadsheet to link the elements of the formula together and used Retrosheets ROE where Schell was using SOE. The resulting formula did not replicate the small standard error announced in the article. When contacted by the author, Schell was unable to provide show a different way of assembling the pieces. Nonetheless this author remains grateful for Schell's conceptual framework of REACHED, Outs While Batting, and Outs by Base Runners,

[ii] **Note 2.**

The changing definition of Sacrifice Hits causes a changing portion of Plate Appearences to count as At Bats. This caused programming and display problems for Retrosheet. If SF are listed when they are unofficial then Batters Faced pitcher will not sum up to opponents AB + BB + HBP + XI + official SH + SF and equal BFP because SF will be double counted. Using sBUNT removes the dilemma . PA appearences will always equals AB + BB + HBP + XI + sBUNT + SF. Scoring Fly data is incomplete for 1916 to 1930 and currently completely unavailable for 1931 to 1938 and 1940 to 1953.

[iii] **Note 3**

The structure of Expected Runs Allowed is A time B divided by C. IF the A factor is 5% too small but the B factor is 5% too big then on average the formula will still estimate the right number of runs for a league. James' A factor uses 70% of errors, but there are ROE being missed by using 70% of errors committed (I,e FWC plays and also some failed strikeouts involving PB or WP.)

Using Baseball Scoreboard data we learn that doubles and triples are somewhat more common relative to Singles than two-base and three-base errors are to one-base SOE. Using 1989-91 and 1997 seasons as our guide, we discover that SOE plays have n average only 92.88% as many total bases as an equivalent number of Hits.. That would make 0.973 a better coefficient than 1.0 for these 70% of errors. As Eadv would be expected to have importance roughly similar to WP and

PB, the right overall coefficient for E in the B factor seems to be about .0.891. <.973 x 0.7 + 0.7 x (1-0.7) = .891>

But " 0.973" as value SOE Errors was based upon an unspoken presumption that 1.048 is the right coefficient for (H –HR). And in the very same article that introduced Ex RA James had spent the two previous pages explaining why he was changing Technical Runs Created formula by reducing the coefficient for HR and increasing the coefficient for singles so that the B value of a singels is closer to one third of the B value of a HR and than to the previous ratio of o one fourth. As HR have a coefficient of 4.0 and the ExRA formula assigns B value of 1.048 even to 2B and 3B, clearly the non-home runs are undervalued relative to Home runs. So perhaps ExRA undervalues E relative to HR yet over values E relative to singles with the fortuitous final result that James' errors canceling each other out.

The degree of accuracy of the two separate formula suggests that a even more accurate formula is possible which would employ all the elements of both plus additional information from Retrosheet or Baseball-Reference.com The A factor could be Reached minus OBR. The C factor could be either PA or opponents BFP. And the B factor could be the SUM of [c1 x Singles c2 x Doubles c3 x Triples, c4 x HR, c5 x (BB+HB minus IBB); c6 x SH; c7 xSF; c8 x SB; c9 x (PB + WP) c10 * BK; c11 x XI; c12 x FWC; c12 x SOE, c13 x Edv; and c14 x Strikeouts] where c1 through c14 are coefficients whose best values have yet to be determined with c1 through c5 being in the same proportions to each other as the coefficients in the 2004 Technical Runs Created formula..

[iv] Note4:

For many seasons prior to 1950 data on times caught stealing and grounding into double plays was unofficial and consequently went unmentioned in the daily game reports to the league offices. In Hidden Game of Baseball, with forward by Keith Law, University of Chicago Press Pete Palmer utilized something he called "OOB" when applying Linear Weights to teams. OOB equals Hits plus Walks plus, Hit Batsmen minus the sum of Runs, Left On Base, and Caught Stealing. Thus OOB always includes hitting into double plays such as the National League tallied from 1933 to 1938 and encompassed CS when Palmer was missing this data. . We see on page 66 that Palmer is subtracting 0.5 runs for each OOB in the linear weights calculation of how many runs a team would be expected to score.

Palmer never shows the accuracy lost from Linear weights by eliminating this correction, but on page 59 he shows that OPS, Runs Created, Total Average, Run Productivity Average and Earnshaw Cook's Scoring Index all become better correlated with team scoring when outs on base are taken into account

Note 5

For the 2020 and 2021 seasons the Major League adopted the practice of beginning each extra half inning by placing the first player up on 2nd base without benefit of a plate appearance. As this runner will then subsequently score, make an out, or be Lett on base at the end of the inning, the count of Plate Appearance will come up 1 short for each side if the game ends after ten innings and 2 short if the game ends after 10 innings.

Thus the new proving formula is FREE plus PA = Runs + LOB + 3 x Opposing innings. PA continues to Equal BFR and PA continues to equal REACH + OWB. OWB Plus OBR still account for 100% of outs. The problem is that that extra runner can lead to one more run, or one more LOB than can be accounted for by the number of PA a

[vi]Note6

Although Retrosheet's tables indicate almost no cases of defensive interference prior to 1916 the numbers for6 BFP, plate appearences and LOB for 1916 and 1921 are very much out of balance with other data unless defensive interference was being regularly called. The numbers for LOB and PA from Baseball Reference are out of balance with the counts of Innings pitched, and Runs. Furthermore PA and BFP did not agree. Retrosheet's counts of AB, BB SH, HBP and estimated XI were out of line with Baseball-References numbers for PA as well being slightly different from the corresponding Baseball Reference counts.

The author's resolution of the discrepancies "fudges" the numbers. I found the average of the conflicting league totals and then adjusted by those averages by small increments until the two methods of calculating OBR produced approximately the same result. The best balanced numbers hypothesizes more XI plays than one gets by extrapolating the rate of known plays over the plate appearences for which play by play data was unavailable to Retrosheet and Baseball Reference.com

[vii]Note 7

Studies of SOE rates for players show that ground ball hitters have substantially higher rates of SOE. If ground balls are more difficult to field cleanly as this data suggests, then the result of batters striving to get the ball in the air to the outfield more often would be fewer SOE. Given the current emphasis on launch angle, SOE will probably continue diminishing.