

Downey, Jim, and Joseph P. McGarrity (2015). Pick off throws, stolen bases, and southpaws: A comparative static analysis of a mixed strategy game. *Atlantic Economics Journal*, Vol. 43 No. 3, pages 319-335.

Downey and McGarrity (2015) looked at the issue at hand as a cat and mouse game between baserunners on first thinking about stealing and pitchers trying to keep them from doing so. Their Retrosheet data set consisted of all pitches (and attempted pickoffs) between June 9th and 13th, 2010 with a runner on first base during games in American League parks, i.e. with the DH, purposely chosen to sidestep the complexities involved with pitcher at bats, in the middle of the season, and 5 games to include each member of the standard starting rotation once. Several models imply that there are more pickoff attempts with righty pitchers than lefties, which the authors attribute to the idea that lefties have more success when they do try a pickoff, resulting in baserunners taking shorter leads and attempting fewer steals. There were also more throws to first base with lower OPS batters (allowing pitchers to concentrate more on the baserunner), a catcher less successful at throwing out runners (giving the pitcher a greater incentive to throw over), a closer game score (increasing the baserunner's incentive to steal), better base stealers on first, and fewer balls and more strikes to the batter. As for steal attempts, they increased with better base stealers, higher pitcher ERAs (more baserunners), a closer game score (as before), and right-handed pitching (again, less success keeping batters from stealing).

Downey, Jim and Joseph McGarrity (2019). Pressure and the ability to randomize decision-making: The case of the pickoff play in major league baseball. *Atlantic Economic Journal*, Vol. 47 No. 3, pages 261-274.

The authors build on their previous work, in which they described when pickoff attempts were more versus less likely, with a study of the sequence of pickoff throws as an alternative to pitches. The data set (from Retrosheet) was the same as the previous study. In summary, pitchers were pretty good at randomizing their alternation between throws to plate and to first, with the exception of righty pitchers against good base stealers (those in the upper third of a measure of proficiency; stolen bases divided by times on first) in relatively close games (2 runs or less score difference). In this case they tend to alternate between pitches to plate and throws to first in a predictable pattern. In addition, the authors hypothesize that when it is more likely for batters to be successful, there is less an incentive for a baserunner to try to steal and so less reason to throw to first. As a consequence, there were more throws to first with an increased number of strikes and fewer throws to first with a three-ball count compared to fewer.

Eisen, Michael (2024). The first run isn't the most important.

<https://www.baseballprospectus.com/news/article/88683/prospectus-feature-the-first-run-isnt-the-most-important/>

The first run of the game is not the most important, in the sense that the team that scores it (between 2000 and 2023, winning average .665) does not eventually win the game as much as often as the team that scores the third run (winning average .722); and in fact runs two through 11 all signaled the winner more often than the first. It is, however, the most "decisive" run in a game, in the sense that the team that scores is more likely to never give up

the lead than any other run. The decisive run must be an odd number, because the score must be tied for the next run to be decisive, and the number of runs in a tied game must be either 0 or even. Anyway, the first run is decisive more than 40 percent of the time, the third run less than 20 percent, and subsequent odd-numbered runs ever-smaller percentages.

Florez, Mauro, Michele Guindani, and Marina Vannucci (2025). Bayesian bivariate Conway-Maxwell-Poisson regression model for correlated count data in sports. *Journal of Quantitative Analysis in Sports*, Vol. 21 No. 1, pages 51-71.

Based on Retrosheet data, the authors proposed a mathematical technique that they used to model game-by-game home and away team run scoring using Retrosheet data for 2019, 2020 (they noted that the COVID season resulted in fewer runs scored than the other two), and 2021. Home teams averaged only nine more runs per game (4.72) than away teams (4.63), and the two were uncorrelated (0.004). Their substantive findings were trite, such as the fact that the Rockies “have one of the weakest attacks when playing away, they possess the strongest attack when playing at home”; yes, they knew why.

Fox, Dan aka Dan Agonistes (2004). Swing away.

<http://danagonistes.blogspot.com/search?updated-max=2004-10-01T15:02:00-06:00&max-results=20&reverse-paginate=true>

These data (2003 and 2004 up to Sept 14th) were compiled by Dave Smith, further analyzed by Bruce Cowgill, and reported here. They report outcomes by the end of a plate appearance when the batter swung on the first pitch.

	BA	OBA	SA	OPS
Swing	0.282	0.302	0.457	0.759
NoSwing	0.259	0.346	0.413	0.759
Total	0.266	0.334	0.426	0.760

Note that BA and SA were better and OBA worse for swing, canceling one another out in OPS. Simply, swinging meant more hits and extra bases but fewer walks.

Fox, Dan aka Dan Agonistes (2004). Triples galore.

<http://danagonistes.blogspot.com/search?updated-max=2004-10-18T21:42:00-06:00&max-results=20&reverse-paginate=true>

1992 Retrosheet data on the relationship between triples and run scoring.

Fox, Dan aka Dan Agonistes (2004). Defensive indifference.

<http://danagonistes.blogspot.com/search?updated-max=2004-10-18T21:42:00-06:00&max-results=20&reverse-paginate=true>

Posting data compiled by Dave Smith and posted on SABR-L on defensive indifference between 1990 and 2004.

Defensive Indifference by year:

2004 247

2003 219
 2002 201
 2001 213
 2000 199
 1999 166
 1998 54
 1997 122
 1996 124
 1995 88
 1994 82
 1993 85
 1992 85
 1991 78
 1990 42

Defensive Indifference by base:

2nd base 1940
 3rd base 65

Defensive Indifference by inning:

1st 1
 2nd 1
 3rd 1
 4th 3
 5th 12
 6th 36
 7th 69
 8th 212
 9th 1498
 extra 172

What is interesting here is the increase over time.

Fox, Dan aka Dan Agonistes (2004). Measuring baserunning: Setting a baseline.

<http://danagonistes.blogspot.com/search?updated-max=2004-11-11T12:29:00-07:00&max-results=20&reverse-paginate=true>

Based on play-by-play data for 2003, almost certainly from Retrosheet, Dan Fox posted the probabilities for various baserunner results from singles and doubles: going one base on a single or two on a double (labeled +1), going an extra base on each (+2), thrown out in attempted advancement (OA), and how often the base directly in front was already occupied. Here it is:

Outs	To	Typ	+1	+2	OA	Next	Base	Occ
All	All	70.5%	27.2%	0.9%	1.4%	29.2%		
	0	All	73.4%	25.0%	0.5%	1.2%	21.1%	
	7		84.5%	14.1%	0.6%	0.7%	21.6%	
	8		68.6%	30.1%	0.3%	1.1%	25.0%	

	9	59.7%	38.3%	0.6%	1.4%	16.5%
1	All	72.4%	25.5%	0.7%	1.3%	30.6%
	7	84.7%	13.4%	1.0%	0.9%	31.5%
	8	70.3%	28.6%	0.4%	0.7%	34.0%
	9	58.1%	39.0%	0.9%	2.0%	27.9%
2	All	66.3%	30.7%	1.4%	1.6%	33.9%
	7	81.1%	15.8%	1.4%	1.6%	33.8%
	8	60.0%	35.9%	1.8%	2.3%	33.1%
	9	48.3%	49.7%	1.0%	1.1%	31.1%
All	7	83.4%	14.4%	1.0%	1.1%	29.6%
All	8	65.8%	31.9%	0.9%	1.4%	31.5%
All	9	55.3%	42.3%	0.8%	1.5%	25.8%

7, 8, and 9 stand for balls fielded by the left, center, and right fielder respectively. I do not find anything in the fielder/outs breakdown of note.

Fox, Dan aka Dan Agonistes (2004). The effect of pitchers.

<http://danagonistes.blogspot.com/search?updated-max=2004-08-31T13:58:00-06:00&max-results=20&reverse-paginate=true>

Based on 1999-2002 (most likely Retrosheet) data, difference between average non-pitcher and average pitcher (quoted)

“With a runner on first and nobody out the Pocket Manager says that an average hitter should never bunt. With the pitcher up the strategy makes sense if the goal is to score one run and if the pitcher's odds of laying down the bunt are 83.1% or better. In another example, with runners on first and second and nobody out an average hitter should not sacrifice if the goal is maximize runs and must be successful 79.9% of the time if the goal is to score a single run. With the pitcher up these odds change dramatically as it makes sense to sacrifice with a break-even percentage of just 35.9% to score one run and 62.1% to maximize runs.”

Fox, Dan aka Dan Agonistes (2004). Full count pitch outcomes.

<http://danagonistes.blogspot.com/search?updated-max=2004-06-10T10:36:00-06:00&max-results=20&reverse-paginate=true>

The following is from data supplied by Dave Smith, for 2004 games through June 2

Pitch	Not 3-2		3-2 count	
	No.	%	No.	%
Called strike	8548	17.2	463	4.5
Foul ball	34908	16.6	2875	28.2
Batted ball	41248	19.7	3414	33.5
Called ball	83681	37.8	2270	22.3
Swinging strike	18619	8.7	1157	11.4
	217,004		10,179	

Note that batters swing more at full counts. [But they swing more at any two strike count.]

Fox, Dan aka Dan Agonistes (2005). When to steal?

<http://danagonistes.blogspot.com/search?updated-max=2005-03-02T07:48:00-07:00&max-results=20&reverse-paginate=true>

The following I assume is from Retrosheet, 2004 data: Stolen bases by inning:

Inning	PerInn	SB2	CS2	PCT	SB3	CS3	PCT	SB4	CS4	PCT	
1	4857	0.129	400	151	0.726	60	15	0.800	2	0	1.000
2	4858	0.074	200	118	0.629	26	11	0.703	1	5	0.167
3	4856	0.098	280	133	0.678	38	22	0.633	0	5	0.000
4	4857	0.081	243	101	0.706	27	16	0.628	2	4	0.333
5	4856	0.086	237	119	0.666	42	14	0.750	1	6	0.143
6	4853	0.078	221	106	0.676	38	10	0.792	0	3	0.000
7	4851	0.081	255	80	0.761	43	10	0.811	1	2	0.333
8	4850	0.070	219	79	0.735	26	13	0.667	1	3	0.250
9	3771	0.055	130	50	0.722	24	3	0.889	0	2	0.000
10+	946	0.096	65	17	0.793	7	1	0.875	0	1	0.000

Next, stolen bases by score differential:

Diff	PA	Att/PA
0	48766	0.024
1	43536	0.022
2	32214	0.020
3	21834	0.020
4	14750	0.019
5	9807	0.011
6	6573	0.007
7	3996	0.001
8	2711	0.001
9	1890	0.001
10	1071	0.001
11	672	0.000
12	292	0.000
13	162	0.000
14	82	0.000
15	92	0.000
16	48	0.000
17	2	0.000
18	1	0.000
19	4	0.000
20	5	0.000
21	23	0.000
22	8	0.000

I assume that the data for the 21st inning is a typo

Fox, Dan aka Dan Agonistes (2004). Homerun distribution.

<http://danagonistes.blogspot.com/search?updated-max=2004-09-14T16:32:00-06:00&max-results=20&reverse-paginate=true>

Fox, Dan aka Dan Agonistes (2004). Homerun distribution amended.

<http://danagonistes.blogspot.com/search?updated-max=2004-09-14T16:32:00-06:00&max-results=20&reverse-paginate=true>

After making errors in his 2004k, in 2004d Dan Fox under his pseudonym Dan Agonistes presented accurate percentages of home runs to each field by batter handedness from 1992 Retrosheet data.

	L	Pct	R	Pct
LF	40	4%	1076	56%
LF-CF	41	4%	498	26%
CF	117	10%	191	10%
CF-RF	331	30%	84	4%
RF	592	53%	60	3%
	1121		1909	

Fox, Dan aka Dan Agonistes (2004). The impact of count.

<http://danagonistes.blogspot.com/search?updated-max=2004-08-31T13:58:00-06:00&max-results=20&reverse-paginate=true&start=20&by-date=false>

Fox, Dan aka Dan Agonistes (2005). Counts for 2004.

<http://danagonistes.blogspot.com/search?updated-max=2005-09-16T12:12:00-06:00&max-results=20&reverse-paginate=true>

Lost the material I had written, but 2005 is a redo of 2004 with 2004 data.

Fox, Dan (2005). Tony LaRussa and the search for significance.

<https://tth.fangraphs.com/tony-larussa-and-the-search-for-significance/>

Fox, Dan (2005). A short digression into log5. <https://tth.fangraphs.com/a-short-digression-into-log5/>

In the first of these, Dan used the Dallas Adams batter/pitcher matchup version of log5 to compute the number of batter/pitcher matchups with outcomes significantly different from chance given their and league average BA. The data was 2003-2005 play-by-play (I'm guessing from Retrosheet) for batters with at least 50 PA and at least 5 matchups; N = 30,481. Of these, only 956 (3.1%) led to more hits than chance would allow for given batter's overall BA. The test is problematic because five matchups are too small a sample size, but further tests validated the reasonableness of this analysis. Given league average, the actual number of 5 PA matchups with either 4 or 5 hits (150) was close to the chance expectation (144). This work provides evidence that (1) the outcome of batter/pitcher matchups is random over a large sample size and (2) log5 works well in this context.

The second of these webposts provides further evidence that log5 works well using the same data.

Fox, Dan aka Dan Agonistes (2005d). The luck of the single.

<http://danagonistes.blogspot.com/search?updated-max=2005-04-08T12:14:00-06:00&max-results=20&reverse-paginate=true>

The following argument for why BA is less reliable than ISO and SLG comes from Dan Fox aka Dan Agonistes (2005d), based on what is certainly Retrosheet data from 2004.

	BIP	Hits	Pct	Single	Double	Triple	Homerun
Ground	60234	14801	33%	45%	14%	42%	0%
Line	25654	18942	43%	46%	50%	28%	14%
Fly	47693	10507	24%	8%	35%	30%	86%
Pop	11010	226	.5%	1%	0%	0%	0%
	144591	44476					

Fox, Dan aka Dan Agonistes (2005). When to steal?

<http://danagonistes.blogspot.com/search?updated-max=2005-03-02T07:48:00-07:00&max-results=20&reverse-paginate=true>

The following I assume is from 2004 Retrosheet data: Sacrifice bunts by inning:

Inning	SH	Per Inning
1	129	0.027
2	187	0.038
3	225	0.046
4	170	0.035
5	223	0.046
6	172	0.035
7	178	0.037
8	185	0.038
9	139	0.037
10	123	0.130

Fox, Dan aka Dan Agonistes (2005). Pitch outs.

<http://danagonistes.blogspot.com/search?updated-max=2005-06-23T21:36:00-06:00&max-results=20&reverse-paginate=true>

This is based on 2003 and 2004 most certainly Retrosheet data; 232 occurrences with pitch outs on steal attempts. The results: a 43 percent successful steals, with 53 percent thrown out, the rest passed ball/wild pitch with one pickoff (Dan asked whether the runner stopped despite a steal attempt and went back to first?).

Fox, Dan aka Dan Agonistes (2005). Looking at DIPS for 2005.

<http://danagonistes.blogspot.com/search?updated-max=2005-03-02T07:48:00-07:00&max-results=20&reverse-paginate=true>

Following are year-to-year correlations for 86 pitchers with at least 120 IP in both 2003 and 2004:

BABIP	0.09	Defense-independent ERA	0.32
ERA	0.19	WHIP	0.41
Component ERA	0.23	Strikeouts/Innings Pitching	0.72
Homers/Innings Pitched	0.31	Walks/Innings Pitched	0.73

This table gives an indication of which measures are reflecting pitching skill and which are not.

Fox, Dan aka Dan Agonistes (2005). Walks followed by homeruns.

<http://danagonistes.blogspot.com/search?updated-max=2005-06-23T21:36:00-06:00&max-results=20&reverse-paginate=true>

2004 certainly Retrosheet data – how often homers occur directly after each of these events. The sample sizes for triples, intentional walks, and hit by pitches are too small to trust. In addition, batting order position is a confound. Hitter type and quality influences who bats next to whom, and so home run hitters tend to be grouped with home run hitters, who walk and strike out more than others, placing those events next, and out makers with non-homer-hitting out makers.

Event	HR	Percentage
Homerun	197	3.6%
Walk	430	2.9%
K	920	2.9%
Double	240	2.7%
Single	751	2.6%
Triple	22	2.5%
I-Walk	30	2.2%
HBP	39	2.1%
Out	2584	2.1%

Fox, Dan aka Dan Agonistes (2005). Love to bunt.

<http://danagonistes.blogspot.com/search?updated-max=2006-01-10T19:30:00-07:00&max-results=20&reverse-paginate=true>

Fox, Dan (2006). Sacrificing in 2005 Redux. <https://tft.fangraphs.com/sacrificing-2005-redux/>

Fox, Dan aka Dan Agonistes (2006). Last time, I promise.

<http://danagonistes.blogspot.com/search?updated-max=2006-02-06T12:51:00-07:00&max-results=20&reverse-paginate=true>

All three of these are 2003-2005 Retrosheet data about sacrifice hit attempts, from the 2005 webposts. The 2006 webpost includes only 2005 data for the same breakdowns. First, by position; 10 is DH, 11 is PH.

Position	Att	Succ	Pct
7	356	304	0.854
9	267	223	0.835
4	861	719	0.835
3	137	114	0.832
6	1006	836	0.831
8	693	571	0.824
2	549	451	0.821
10	84	69	0.821
5	378	292	0.772
11	217	161	0.742
1	2704	1810	0.669

Note that many of the top success rates are for positions not usually associated with bunting; as Dan noted, this is likely due to the surprise element. By inning:

Inning	Att	Succ	Pct
1	512	448	0.875
2	783	562	0.718
3	998	714	0.715
4	704	527	0.749
5	961	705	0.734
6	683	551	0.807

7	822	647	0.787
8	819	638	0.779
9	588	450	0.765
10	159	138	0.868
11	105	73	0.695

Dan included innings 12 to 19, but I deleted them due to small sample sizes. Fewer in the first inning makes sense, and the higher success rate perhaps again indicates defensive team expectations being violated. My guess is that fewer in the ninth inning is due to the home team often not batting. By score differential:

Diff	Runners On	SacAtt	PctAtt	SuccPct
<=-5	5591	21	0.004	0.909
-4	2946	23	0.008	0.739
-3	4426	54	0.012	0.722
-2	6653	138	0.021	0.725
-1	9763	377	0.039	0.769
0	20293	809	0.040	0.782
1	10567	352	0.033	0.730
2	7677	284	0.037	0.771
3	5217	158	0.030	0.747
4	3465	80	0.023	0.800
>=5	6527	59	0.009	0.750

No surprises here; the closer the score, the more bunting.

Fox, Dan (2006). The irreducible essence of platoon splits.

<https://www.baseballprospectus.com/news/article/4970/schrodingers-bat-the-irreducible-essence-of-platoon-splits/>

Using Retrosheet data from 1970 through 1992, Dan Fox (2006) discovered the usual batter advantages when facing opposite side hitters, again as usual more extreme for lefty hitters than righty. More interestingly, based on 505 batters with at least 2000 plate appearances during that time, the platoon differentials for batting, on-base, and slugging averages and for walk and strikeout rates were approximately normally distributed, and the correlations between odd and even years for the first three of these were all less than +0.2, although somewhat higher for the last two. These figures imply that individual differences may be random fluctuation such that batters are not consistently more or less susceptible than one another. This in no way disconfirms the existence of the general tendency.

Fox, Dan aka Dan Agonistes (2006). Pitch outcomes.

<http://danagonistes.blogspot.com/search?updated-max=2006-02-06T12:51:00-07:00&max-results=20&reverse-paginate=true>

Here are outcomes for specific pitches for two seasons. Note their similarity.

Year	Ball	Called Strike	Swinging Strike	Foul Ball	Foul Tip	Batted Ball in Play
2004	37.1%	17.2%	8.6%	17.1%	0.2%	19.8%
2005	36.6%	17.4%	8.4%	17.0%	0.5%	20.2%

Dan Fox aka Dan Agonistes (2007). Double steals and contentment.

<http://danagonistes.blogspot.com/search?updated-max=2007-04-24T06:38:00-06:00&max-results=20&reverse-paginate=true>

From Retrosheet 1970-2006 except 1999. These are double steal attempts, success rate, and breakeven rate per base-out situation assuming that the worst case is the lead runner thrown out but the trailing runner safe.

Base	Outs	Succ	Att	Percent	Avg BE
12x	0	643	1127	57.1	0.587
12x	1	1595	2258	70.6	0.667
1x3	0	21	70	30.0	0.728
1x3	1	147	478	30.8	0.590
x23	0	2	4	50.0	0.717
x23	1	2	51	3.9	0.633
123	0	0	5	0.0	0.544
123	1	5	39	12.8	0.524
Total		2415	4032	59.9	0.635

Overall, looks like a losing strategy.

Fox, Dan aka Dan Agonistes (2007). The 100 RBI men.

<http://danagonistes.blogspot.com/search?updated-max=2007-05-31T12:59:00-06:00&max-results=20&reverse-paginate=true>

As read off a graph, the following are the average runners on base per plate appearance during 2006, which I assume are based on Retrosheet data. Dan Fox's point in compiling this is to demonstrate the position bias in RBI opportunity. Read off graph.

Inning	1	2	3	4	5	6	7	8	9
Runners	0.46	0.55	0.65	0.7	0.7	0.67	0.66	0.65	0.65

Fox, Dan aka Dan Agonistes (2007). Of crowds and splits.

<http://danagonistes.blogspot.com/2007/11/of-crowds-and-splits.html>

This a repost of a 2006 Baseball Prospectus post) presenting the following slash metrics based on 1970 through 1992 for all 505 players with at least 2000 plate appearances during that time.

			-----Right-----				-----Left-----				----Platoon Split----				
ISO	Count	Bats	AVG	OBP	SLG	OPS	AVG	OBP	SLG	OPS	AVG	OBP	SLG	OPS	
<.1	122	R	.253	.300	.325	626	.269	.317	.358	675	.016	.017	.033	49	
		L	.276	.328	.364	692	.250	.302	.313	615	.026	.025	.052	77	
.1-.15	183	R	.262	.312	.381	693	.280	.332	.423	755	.019	.020	.042	62	
		L	.283	.340	.417	757	.258	.312	.354	665	.025	.029	.063	92	
>.15	200	R	.261	.326	.437	762	.278	.348	.479	827	.017	.023	.042	65	
		L	.277	.351	.470	822	.251	.319	.396	715	.026	.032	.074	106	

See that it is bigger for higher ISOs. Dan also included graphs showing 104 of these players with at least 15 consecutive seasons, which showed basically no change in handedness

breakdowns over time for BA, OBA, and SLG.

Fox, Dan aka Dan Agonistes (2007). Double steals and contentment.

<http://danagonistes.blogspot.com/search?updated-max=2007-04-24T06:38:00-06:00&max-results=20&reverse-paginate=true>

Fox, Dan (2007). Double steals and more.

<https://www.baseballprospectus.com/news/article/6003/schrodingers-bat-double-steals-and-more/>

Dan Fox examined the history of double steals based on 1970-2006 Retrosheet data in two webposts. I begin with basic figures classified by base-out situation, from the first webpost:

Base	Outs	Succ	Att	Percent	Avg BE
12x	0	643	1127	57.1	0.587
12x	1	1595	2258	70.6	0.667
1x3	0	21	70	30.0	0.728
1x3	1	147	478	30.8	0.590
x23	0	2	4	50.0	0.717
x23	1	2	51	3.9	0.633
123	0	0	5	0.0	0.544
123	1	5	39	12.8	0.524
Total		2415	4032	59.9	0.635

The rest of this report is from the second webpost. Per 162 games for both teams combined double steal attempts per 162 games, these remained in the 8 to 12 range from the early 1970s through around 1986. At that point, the numbers rose substantially, reaching a high of over 17 in 1995 before trending downward once more to where that number has hovered between 7.5 and 8.9 since 2003. These figures did not approximate the general historical stolen-base attempt pattern during those years, lagging behind. Dan viewed this pattern as perhaps a late managerial response to increasing success on the basepaths, or simply a kind of fad that soon began to wane.

Strategically, here are breakeven numbers for double steals with runners on first and second (79% of the total between 1970 and 2006) between 1999 and 2002: .639 with 0 out, .558 with 1 out, .735 with 2 out. In leaner offensive times like those that persisted during much of the rest of the period since 1970, the breakeven percentages would be lower with less than two outs, since making outs on the bases would not have been as costly. For example, in 1980 the breakeven percentages fall to .600 and .530 with zero and one out, and raises slightly to .778 with two outs. Success rates measured by at least two runners successful and no outs occurring: .566 with 0 out, .703 with 1 out, .991 with 2 outs.

Nineteen percent of attempts occurred with runners on first and third. Success rates were .286 with 0 out, .305 with 1 out, .764 with 2 outs.

Here are delayed double steals with first and third, with success meaning the runner on third scored no matter whether or not the runner on first got thrown out, leading to higher success rates: higher: .443 with 0 outs, .408 with one out, .895 with two outs. The reason why the two out figure is higher is that with 2 outs any stolen base is only noted when the third out does not occur.

Fox, Dan aka Dan Agonistes (2007). The daily double.

<http://danagonistes.blogspot.com/search?updated-max=2008-01-03T06:01:00->

[07:00&max-results=20&reverse-paginate=true](#)

Dan Fox aka Dan Agonistes (2007) compiled stolen base success rate for seasons with then-uploaded Retrosheet data. At about 55 percent in 1914-1915, it started a generally linear increase to over 70 percent in 2006. There were however eras in which it was considerably above any regression line (none provided here); in the 1930s it jumped well over 60 percent but well below that in the 1950s.

Fox, Dan (2007). Dropping one down.

<https://www.baseballprospectus.com/news/article/6446/schrodingers-bat-dropping-one-down/>

Fox, Dan (2007). Dropping one down, part 2.

<https://www.baseballprospectus.com/news/article/6475/schrodingers-bat-dropping-one-down-part-two/>

NOT IN BIBLIOGRAPHY

Dan Fox (2007) supplied a detailed examination of attempted bunt hits between 1970 and 2006, defined as events in which batters bunt and not charged with a sacrifice. Unfortunately this includes sacrifice attempts with a lead runner forced out, and on the other side sacrifice attempts in which the batter beats it out. Anyway, being closer to first it makes sense that lefty batters were more successful (43.8 percent of the time) than righties (37.4%). These attempts occurred most often with no outs (59.9%) and were least likely with two outs (12.2%) with one out intermediate (27.9%). This makes sense, as a runner on first is more valuable the fewer outs there are. And probably for this reason fielders not ready for one with two outs, so success rate (48.8%) was higher than with one (39.8%) or no (39.1%) outs. Attempts were most frequent with bases empty (49.8%), a runner on first (26.7%), or runners on first and second (10.9%). Among these three, success rate was higher with bases empty (45.0%) than first (32.2%) or first and second (32.3%), as latter two allow for forces on base runners. Attempts occurred much more often on the first pitch of a plate appearance (69.4%, with a success rate 42.2%), with no other count as high as 10%. Success rate was much lower with two strikes (between 13.6% and 9% depending in the number of balls, with 58% of these ending in strikeouts), and over 50 percent for 2-0, 3-0, and 3-1 counts. Dan's follow-up (2007) includes break-even figures based on run expectancies at different base-out situations, which could be discouragingly low (.021 with runners on first and second and no outs) and encouragingly high (.690 with runner on second and two outs).

Fox, Dan (2007). Defense and alphabet soup.

<https://www.baseballprospectus.com/news/article/6976/schrodingers-bat-defense-and-alphabet-soup/>

Fox, Dan (2007). Inching along.

<https://www.baseballprospectus.com/news/article/6990/schrodingers-bat-inching-along/>

Fox, Dan (2007). The issue of the day, and ranging into the outfield.

<https://www.baseballprospectus.com/news/article/7006/schrodingers-bat-the-issue-of-the-day-and-ranging-into-the-outfield/>

Fox, Dan (2007). And even more refinements in SFR.

<http://danagonistes.blogspot.com/search?updated-max=2008-01-03T06:01:00-07:00&max-results=20&reverse-paginate=true>

Dan Fox's (four 2007 webposts summarized in 2008) Simple Fielding Runs (SFR) works well with Retrosheet data. The idea is to take the proportion of batted balls of each type (grounders, liners, pop-ups) that an infielder makes plays on relative to the average player at the same position, and adjust that proportion for batter handedness, number of bunts, whether there is a runner on first (all affecting positioning), and turning that adjusted proportion into a run figure. Fox was not as clear about balls hit in the first/second and shortstop/third holes – it appears that they are split according the proportions fielded on average by each – as he was about the second/shortstop hole, which is divided 50/50. First base was an outlier of sorts, correlating at .68 with UZR. In the third of the 2007 webposts, Dan expanded the concept to outfielders, with a fourth hit type (fly balls) included.

Fox, Dan aka Dan Agonistes (2007). The hook part II.

<http://danagonistes.blogspot.com/search?updated-max=2007-02-08T12:55:00-07:00&max-results=20&reverse-paginate=true>

I believe Retrosheet 2006 data. First, the proportion of pitching changes intended to get a platoon advantage given different score differentials.

Tied	65.5%
1 Run	67.0%
2 Run	65.9%
3 Run	64.6%
4 Run	64.0%
5 Run	62.1%
6 Run	55.3%
>6	53.3%

The more that the outcome of the game seems certain, the less managers care about platoon differentials. Second, for games with run differential of 3 or less, given different innings

Inning 4	69.4%
Inning 5	69.0%
Inning 6	69.4%
Inning 7	67.9%
Inning 8	66.6%
Inning 9	58.4%
Inning 10+	58.1%

Thanks to the myth of the proving closer, managers trust their #1 guy against both handedness hitters in the 9th and beyond.

Fox, Dan (2007). Beautiful theories and ugly facts.

<https://www.baseballprospectus.com/news/article/5040/schrodingers-bat-beautiful-theories-and-ugly-facts/>

Fox, Dan (2007). Strike zones, trilobites, and a vicious cycle.

<https://www.baseballprospectus.com/news/article/5069/schrodingers-bat-strike-zones-trilobites-and-a-vicious-cycle/>

Fox, Dan (2007). The moral hazards of the hit batsmen.

<https://www.baseballprospectus.com/news/article/5093/schrodingers-bat-the-moral-hazards-of-the-hit-batsmen/>

In these webposts, Dan Fox considered several theories for the changes in HBP rate over time and, while seeing virtue in many of them, rejected all of them as complete explanations using Retrosheet data to show when each is and is not consistent with what occurred. As for the moral hazards argument made by Bradbury and Drinen, he believed that about a third of the A.L. HBP surplus was due to having more “real” aka non-pitcher hitters in the lineup that pitchers believe need to be pitched to inside, but that the theory as a whole is valid.

Fox, Dan aka Dan Agonistes (2007). Where they ain't redux.

<http://danagonistes.blogspot.com/search?updated-max=2007-06-30T12:13:00-06:00&max-results=20&reverse-paginate=true>

The following is the second version of data concerning outcomes from batted ball types. The third from the last column lists the percentage of non-homer hits, the last column the same but with HR subtracted from the denominator as suggested by Tom Tango.

Year	Type	BIP	H	Non-HR	TB	%H	%Non-HR	SLUG	%Non-HR2
2003	Fly	36744	9898	5390	27314	26.9%	14.7%	0.743	16.7%
2004	Fly	37052	10494	5786	28619	28.3%	15.6%	0.772	17.9%
2005	Fly	37268	10442	5913	28207	28.0%	15.9%	0.757	18.1%
2006	Fly	37712	10863	6034	29557	28.8%	16.0%	0.784	18.3%

2003	Ground	60783	14355	14353	15687	23.6%	23.6%	0.258	23.6%
2004	Ground	60212	14267	14267	15623	23.7%	23.7%	0.259	23.7%
2005	Ground	60373	14092	14092	15388	23.3%	23.3%	0.255	23.3%
2006	Ground	59912	14367	14367	15690	24.0%	24.0%	0.262	24.0%
2003	Line	25846	18985	18289	26505	73.5%	70.8%	1.025	72.7%
2004	Line	25663	18951	18208	26495	73.8%	71.0%	1.032	73.1%
2005	Line	25425	18649	18162	25240	73.3%	71.4%	0.993	72.8%
2006	Line	25902	19012	18456	26072	73.4%	71.3%	1.007	72.8%
2003	Pop	10853	168	168	207	1.5%	1.5%	0.019	1.5%
2004	Pop	11007	226	226	268	2.1%	2.1%	0.024	2.1%
2005	Pop	11123	223	223	258	2.0%	2.0%	0.023	2.0%
2006	Pop	10656	238	238	309	2.2%	2.2%	0.029	2.2%
2003		30639	15	14	18	0.0%	0.0%	0.001	0.0%
2004		31657	0	0	0	0.0%	0.0%	0.000	0.0%
2005		30463	1	0	4	0.0%	0.0%	0.000	0.0%
2006		31558	47	47	93	0.1%	0.1%	0.003	0.1%
Totals		660848	175293	154233	281554	26.5%	23.3%	0.426	24.1%

Fox, Dan (2008). Clearing the decks.

<https://www.baseballprospectus.com/news/article/7252/schrodingers-bat-clearing-the-decks/>

Between 1959 and 2007, although the number of attempted bunt hits declined from age 21 through 37, success rate increased to its peak at age 31 before declining quickly thereafter.

Fox, Dan and Neal Williams (2007). Quantifying coaches, part I.

http://baseballanalysts.com/archives/2007/03/quantifying_coa.php

Fox, Dan and Neal Williams (2007). Quantifying coaches, part II.

http://baseballanalysts.com/archives/2007/03/quantifying_coa_1.php

Fox, Dan and Neal Williams (2007). The traffic directors. *Baseball Research Journal*, No. 36, pages 19-26.

Fox, Dan (2008). The traffic directors addendum.

<http://danagonistes.blogspot.com/search?updated-max=2008-04-03T09:11:00-06:00&max-results=20&reverse-paginate=true>

In a two-part posting, Dan Fox and Neal Williams (2007), see also the *BRJ* summary) tried to evaluate third base coaches in terms of decisions concerning taking extra bases on outfield hits. They judged the third base coach, as opposed to the baserunner, as relevant to these decisions with a runner on first and the batter either singles or doubles or with a runner on second and the batter singles, assuming that the ball is fielded by the right fielder. Basically, Dan and Neal first computed the ratio between the number of runs gained or lost through runner attempts at advancement in these situations, as measured by Dan's EqHAR measure and the number of opportunities

meeting the relevant situations. Second, they performed the same calculation for situations in which the ball is fielded by the left or center fielder, under the assumption that in these cases the baserunner is responsible for the decision. Third and last, they divided the first figure by the second, providing a metric measuring whether EqHAR is higher or lower when the coach is responsible rather than the baserunner. These final numbers range from 1.44 to 0.77, with the higher ones indicating positive EqHARs when the coach was responsible and negative ones when the runner was, and the lower ones the opposite. However, Dan and Neal correlated the final figures in consecutive seasons for those third base coaches who retained their positions over the winter for 2000-2001 through 2005-2006, a sample size of 2004 season-pairs, and overall the correlation across seasons was a nonexistent 0.04. In short, one can calculate a third base coaches' single season performance, but there is no evidence that these judgments indicate an actual baserunner-sending skill. A correlation between odd and even years for the 35 coaches with the most experience from 1993 to 2007 was just as small (0.03; Fox, 2008).

Fritz, Kevin and Bruce Bukiet (2010). Objective method for determining the Most Valuable Player in major league baseball. *International Journal of Performance Analysis in Sport*, Vol. 10, pages 152-169.

Fritz and Bukiet (2010) developed a Markovian method for determining the “best” candidate for MVP awards. The authors applied Retrosheet data to determine actual advancement probabilities, in so doing halving the error in runs prediction from 4 percent in previous work but Bukiet to 2 percent here. They then used a standard lineup (e.g., shortstop leads off, outfield second and third, first base cleanup, etc.) and average offensive performance for a given position (e.g., mean shortstop in the first position etc.) to provide a baseline run distribution, substituted a given MVP candidate’s performance for the average in their position, and compared the two to provide a runs-greater-than-average figure for that candidate. Excluding MVP winners who were pitchers and so irrelevant to the model, the sportswriters’ choice and their “best player” were the same 45 percent of the time. and the winner was among their three “best players” 65 percent of the time, between 1988 and 2007.

Fuld, Elan (n.d.). Clutch and choke hitters in major league baseball: Romantic myth or empirical fact? Unpublished paper.

A well-publicized paper by a University of Pennsylvania student named Elan Fuld that unpublished but easy to access online (search for “Elan Fuld clutch”) claims that clutch hitters really do exist. Fuld defined the importance of clutch situations according to his computation of their leverage, and then compared through regression analysis the batter’s performance in terms of bases gained per plate appearance (0 to 4) on the plate appearance’s specific leverage. If a player did substantially better (worse) in high leverage situations than in low during a given season, then Fuld labeled the player as clutch (choke) in that season. The real issue was whether a player was consistently clutch or choke across their entire career. He used Retrosheet data for 1974 through

1992 for 1075 player with at least two seasons with 100 PAs, including each season reaching that threshold of play (6784 player-seasons in all). He then computed a measure of clutch tendencies across seasons with a threshold defined such that only 1 percent (11 of 1075) of players would be considered clutch and another 1 percent (another 11) choke by chance. When Fuld treated sacrifice flies under the very strange assumption that they are analogous in value to walks, as many as 24 players met the criteria of consistent clutchness across seasons, although never more than 7 reached that for chokeness. As Phil Birnbaum noted (2005c), this assumption inflates the value of a fly ball with a runner on third over fly balls in other situations, as SFs are more likely to occur in clutch situations than the average base/out configuration, while at the same time treating them as walks credits the batter an extra base they did not really earn, artificially inflating their bases gained in clutch situations. When Fuld excluded SFs from the data set, no more than 8 hitters met his criteria for clutchness. Therefore, despite a U. Penn press release claiming that the existence of clutch hitters had been proven along with the media sources that accepted that claim, Fuld's study failed to find the existence of clutch hitters.

Gantner, Ryan (2016). Never make the first or third out at third base...perhaps.
Baseball Research Journal, Vol. 45 No. 1, pages 17-24.

Ryan Gantner has contributed a computation of breakeven points for advancing while on base. The basis of Gantner's work was an examination of the wisdom of, in his words, "the familiar adage Never make the first or last out at third base" (page 17). Beginning with the relevant baserunner on second and assuming no one else on base (Gantner also looked at an additional runner on first, with similar findings) and using 2013 Baseball Prospectus run expectancy tables, the breakeven points are success rates of 76.4% for no out, 67.1% for one out, and 87.6% for two outs; a replication for 2014 provided almost the same figures. This data appears to corroborate the adage; higher break-evens for no and two outs than for one. However, now including the impact of subsequent possible batters, in 2014 the expected number of runs forfeited by unsuccessful attempts was highest for no outs (.7999), intermediate for one out (.5373), and lowest for two outs (.2901), which stand to reason given the impact of number of outs on scoring. This implied that making the second out is worse than making the third. Further, using Retrosheet play-by-play data, Gantner noted that break-evens are way lower (.651 for no outs, .540 for one out, .806 for two outs) when only one run is needed than for higher numbers of needed runs, implying that when the score is tied in the ninth the runner should more often go for it. Gantner went on to study the impact of baserunning outs at second (overall break-evens about .70 no matter the outs, but about .60 if only one run needed) and home plate (very dependent on number of outs and again lower if only one run needed). He concluded with the following revised adage:

Never make the last out at third base. Never make the first out at home plate.
And never make any out at home plate if more than one run is needed in the inning.

Goldschmied, Nadav, Michael Harris, Damien Vira, and Jason Kowalczyk (2014). Drive

theory and home run milestones in baseball: An historical analysis. *Perceptual and Motor Skills: Exercise and Sport*, Vol. 118 No. 1, pages 1-11.

In an attempt to relate drive theory to baseball, these authors examined the 24 players who had reached 505 home runs before the publication date (Albert Pujols got there too late to be included), comparing how many at bats it took for them to hit the last five home runs before their last milestone (either 500, 600, 700, 715 in the case of Henry Aaron and 756 in the case of Barry Bonds) with the first five homers after it. On average, the five leading up took 117.7 at bats and the five afterward 77.5 at bats, consistent with the authors' hypothesis that stress before the milestone restricted performance. Data came from baseball-reference.com and Retrosheet.

Green, Bret and Jeffrey Zwiebel (n.d.). The hot hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from baseball. Downloaded from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2358747

Lichtman, Mitchel (2016). Revisiting the hot hand. In Paul Swydan (Prod.), *The Hardball Times Baseball Annual 2016*, pages 213-227. FanGraphs. Green, Bret and Jeffrey Zwiebel (2018). The hot hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from baseball. *Management Science*, Vol. 64 No. 11, pages 5315-5348.

Another reported demonstration that received a good bit of publicity was an unpublished study by Green and Zwiebel, based on Retrosheet data from 2000 through 2011. In essence using the second, conditional probability method, Green and Zwiebel wanted to see if the outcome of a particular plate appearance for both batters and pitchers could be predicted more accurately using the outcomes of the previous 25 at bats than overall performance for the given season, minus a 50 at bat window around the plate appearance under question. They provided various operational definitions for hot and cold streaks. Some of these definitions seem to bias the study in favor of finding streakiness; these established criteria based on the assumption that the average player is hot five percent and cold five percent of the time, which strikes me as out of bounds given that it presumes streakiness exists. A more defensible definition required the batter to be hot or cold if in the upper or lower five percent of a distribution based on his own performance. Their equations also controlled for handedness and strength of opposing pitchers and ballpark effects, but not, as Mitchel Lichtman (2016) pointed out, for umpire and weather. Unfortunately, ballpark effect was poorly conceived, as it was based solely on raw performance figures and did not control for relative strength of the home team (i.e., a really good/bad hitting home team would lead to the measure indicating a better/worse hitting environment than the ballpark is in truth). The authors' results indicated the existence of hot/cold streaks for all examined measures: hits, walks, home runs, strikeouts, and times on base for both batters and pitchers. Interestingly, after noting improved performance after the plate appearance under question than before, the authors attributed half of the reported increase in that PA to a "learning effect," in essence true improvement in hitting. As Mitchel Lichtman (2016) pointed out, if so, then it should not be considered evidence for the existence of

streakiness.

Green and Zwiebel's work elicited a lot of critical comment. Along with the ballpark problem, which Zwiebel acknowledged in email correspondence with Mitchel Lichtman, one comment was that subtracting the 50 at bat window biased the study in favor of finding streaks. Here's an example showing why: let us assume that a player is a .270 hitter. If a player happens to be hitting .300 or .240 during that window, then the rest of the season he must be hitting say .260 or .280 to end up at that .270. In this case, the .300 and .240 are being compared to averages unusually low and high rather than the player's norm. But it strikes me this would only be a problem if hot and cold streaks actually existed – if not, it would be .270 all the way. It is the case that subtracting the 50 at bat window lowers the sample size of comparison at bats, increasing random fluctuation and again adding a bias in favor of finding streakiness. Whether this loss of 50 at bats is catastrophic during a 500 at bat season for a regular player is a matter for debate. In any case, Lichtman (2016) performed his own study using 2000-2014 Retrosheet data, but in this case used the sixth PA after the 25 window, in order to insure it occurred in a different game in most cases. He also used a normal projection method (i.e. three years of past performance with more recent weighted over less) rather than a within-season window. The results were a small hot and slightly larger cold hand effects for BB/PA, OBA, wOBA, and HR/PA, and almost none for BA. Mitchel speculated that changes in both batting (such as swinging for homers after hitting a few) and pitching (such as pitching more carefully to the hot batter and less so to the cold) strategies might be at least partly responsible, along with cold batters playing with an injury.

Green and Zwiebel were finally able to publish their work in 2018, basically unchanged with an additional section in which they claimed to show that the opposition responds to hot streaks by walking the batter in question more often than the batter is normally. They also included a criticism of the Tango, Lichtman and Dolphin analysis of streaky batting described below, based on perceived problems with TMA's use of a batter's average performance as a baseline for identifying streaks. As before, I believe this criticism is flawed by the continued implicit presumption that streaks and slumps exist inherent in Green and Zwiebel's work.

Gross, Alexander and Charles Link (2017). Does option theory hold for major league baseball contracts. *Economic Inquiry*, Vol. 55 No. 1, pages 425-433.

Gross and Link (2017) likely began a new area of study in examining the factors that motivate teams to seek team options for seasons included in free agent contracts. They restricted their sample to 109 circumstances in which position players eligible for free agency signed new contracts between 2003 and 2011, with those contracts either including team options or performance standards that needed to be reached for additional years to vest. Using performance data from Retrosheet, the authors discerned that team options/performance standards were more likely to be included to the extent that player OPS had been variant over the past three seasons, which makes sense as such players were could be thought more likely to perform poorly than more consistent players.

Guérette, Joel, Caroline Blais, and Daniel Fiset (2024). Verbal aggressions against major league baseball umpires affect their decision making. *Psychological Science*, Vol. 35 No. 3, pages 288-303.

Since pitch location data became available, we have found a number of umpire biases on pitch calls. The most prominent of these I call the count compensation bias additional balls as the number of strikes increase and additional strikes in counts with no balls; two others that tendencies that perhaps affected game outcomes are a home field advantage, and different strike zones for left-handed versus right-handed batters. There are quite a few others that are very weak and probably inconsequential Guérette, Blais, and Fiset (2024) may have found a new bias; how significant it is on game outcomes is unclear. Using 2010-2019 data (sample size of 153255 pitches with no swing) including a list of ejections from Retrosheet plus data from Statcast, Baseball Reference, and FanGraphs, they examined results from a series of models with a large and varying set of controls (including pitch location) and concluded the following: The odds of a strike call after a batter or manager/coach was ejected for arguing a strike call were lower for the ejected player's or manager/coach's team, and higher for other team but only for a manager/coach ejection. Before the ejection, the ejected player/manager/coach's team was getting more strike calls than the opposition; after the ejection, it received got fewer. Player status (All-Star) and fWAR had no significant effect, and ejection for other reasons had no comparable impact.

Haechrel, Matt (2014). Matchup probabilities in major league baseball. *Baseball Research Journal*, Vol. 43 No. 2, pages 118-123.

Back in the 1983 *Baseball Analyst*, Bill James presented a formula for the prediction of batting averages in specific batter/pitcher matchups proposed by Dallas Adams which was a spin-off on James's log5 method for predicting two-team matchups. This formula only works for two-event situations; hits versus outs. Matt Haechrel (2014) proposed and mathematically justified a generalization allowing for probability predictions for multiple events (outs, singles, doubles, triples, homeruns, walks, hit by pitches), and using Retrosheet event data showed that the generalization does a good job of predicting the actual proportion of these events for the 2012 season.

Hamrick, Jeff and John Rasp (2015). The connection between race and called strikes and balls. *Journal of Sports Economics*, Vol. 16 No. 7, pages 714-734.

Hamrick and Rasp (2015) took on the issue of racial bias in umpiring, using 1989-2010 data from Retrosheet. They discovered slight increases (.004) in the probability of a strike if the umpire and batter were of different races, which accounts for perhaps a pitch every two games. That increase was greater (.005) with three-ball counts and smaller (.003) with two-ball counts. They also noted some slight differences among races in for umpires (relatively speaking, Hispanic umps favored hitters and Black umps favored pitchers in three-ball counts), pitchers (with three balls, Latin pitchers got more strikes and Black pitchers got fewer; with two strikes, this tendency was reversed), and batters (with two strikes, Black hitters got more strikes and Latin hitters fewer; both were

disadvantaged relative to White hitters with three balls). However, there were no significant three-way interactions between the races of umpires, pitchers, and batters; in other words, no evidence for discrimination based on similarity of race. They also noted significant but tiny increases in the probability that a pitch would be called a strike if the hitter were on the visiting team or on the team with the worse record, if the pitcher were on the better team, if either the hitter or pitcher was relatively inexperienced, if the score difference was greater, if there were more balls or less strikes, and if QuesTec or PITCHf/x were in use.

Harrison, Willie K. and John L. Salmon (2019). Leveraging pitcher/batter matchups for optimal game strategy (2019). *2019 MIT Sloan Sports Analytics Conference*.

Using 2000 to 2018 Retrosheet data, Harrison and Salmon (2019) uncovered 5170 pitcher/batter matchups with at least 35 PA (they say AB, but they include walks) and used data from that as the basis for simulating 500,000 innings in which they randomized the matchups in order to find the best sequence of pitchers for facing each simulated "lineup" of players. This provided them with 15 clusters of matchup types, with each cluster maximizing certain outcomes and minimizing others. For example, Cluster 12 (the numbers serve only as labels) maximized strikeouts and homers but minimized doubles/triples whereas Cluster 8 maximized flyouts and groundouts. They used those clusters to compare what actually occurred in two innings during the 2018 playoffs with what their simulations would predict were the best matchups from the pitcher's team's point of view.

Healey, Glenn (2015). Modeling the probability of a strikeout for a batter/pitcher matchup. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27 No. 9, pages 2415-2423.

Healey (2015) proposed models based on Dallas Adams's and Bill James's log 5 method for predicting the general outcome (strikeout versus ground ball) in specific batter/pitcher matchups. Basically, his models establish overall parameters for four categories (lefty and righty pitchers paired with lefty and righty batters), which can then be used for predicting the strikeout and ground ball tendencies for specific batter/pitcher matchups. Healey used Retrosheet plate appearance data for 2003 through 2013, and included every player with at least 150 PAs against both righty and lefty opponents. One interesting overall finding emerged; the closer the ground ball rate of the batter and pitcher in a matchup, the greater the odds of a strikeout. His explanation rings true; ground ball pitchers tend to pitch under bats and ground ball hitters tend to swing over pitches, leading to more strikes. Analogously, fly ball pitchers tend to pitch over bats and fly ball hitters tend to miss under pitches, leading to more strikes.

Healey, Glenn (2017). Matchup models for the probability of a ground ball and a ground ball hit. *Journal of Sports Analytics*, Vol. 3 No. 1, pages 23-35.

Healey's study, based on Retrosheet data from 2003 to 2014, was intended to examine a model for predicting groundball rates and batting averages on ground balls in

specific matchups. It included as predictors fairly obvious individual indices; individual pitcher and batter strikeout rate, pitcher groundball rate (although not batter, but instead overall league BA on grounders), batter speed, and pitcher's team's fielding range. Healey claimed that his model allowed for smaller sample sizes than an alternative based log5 for the same accuracy rate. However, extreme cases were poorly predicted. As a byproduct of this work, Healey also gained some insight into the standard platoon advantage/handedness issue. The data revealed that same handed matchups have tended to result in more strikeouts and groundballs than have opposite handed matchups. This leads in turn has led to a lower batting average on grounders but a higher batting average on flies, perhaps due to the tendency for same handed hitters to hit pitches higher in the strike zone than opposite handed. BA on grounders was higher from righthanded hitters than lefties overall, probably due to the preponderance of balls hit to the left side of the infield and thus the longer throw needed to erase the hitter.

Healey, Glenn and Shiyuan Zhao (2020). Learning and applying a function over distributions. *IEEE Access*, Vol. 8, pages 172196-172203.

Using PITCHf/x and Retrosheet data for all 149 pitchers throwing at least 1500 pitches in 2016, Healey and Zhao (2010) proposed a method for modeling the odds of strikeouts based on the variation in pitch location and speed.

Heipp, Brandon aka Patriot (2010). Relief Run Average.
<http://walksaber.blogspot.com/2010/09/>

This is an extension of Sky Andrecheck's RRA, which was based only on inherited runners, to include bequeathed runners. Brandon called it Relief Run Average (RRA). The following is a simplified form of it.

Step 1 – Compute the league average proportion of inherited runners that score in a given year. In 2009 it was .337 in the A.L. and .303 in the N.L., so I imagine it is some figure around .3 or .4 annually.

Step 2 – Multiply the result of Step 1 by the relevant pitcher's number of bequeathed runners. This gives you an expected figure.

Step 3 – Subtract that from the pitcher's number of bequeathed runners that scored. The larger(smaller) the figure, the more(fewer) of these runs relative to average relief pitchers allowed to score.

Step 4 – Subtract that figure from the number of runs the pitcher gave up that year.

Step 5 – Multiply that by 9 and divide by innings pitched to give you a run average.

Step 6 – Multiply the result of Step 1 by the relevant pitcher's number of inherited runners. This gives you an expected figure.

Step 7 – Subtract from that (the opposite of Step 3) the pitcher's number of inherited runners that scored. The larger(smaller) the figure, the more(fewer) of these runners scored with that pitcher on the mound.

Step 8 – Same as Step 4 with the result of Step 7.

Step 9 – Same as Step 5 with the result of Step 8.

Step 10 – Sum the results of steps 5 and 8, giving you a total of runs saved.

Step 11 – Subtract the sum from the number of runs allowed.

Step 12 – Same as Steps 5 and 9 with the result of Step 11, giving you Relief Run Average.

Brandon's method also included a park factor, complicating the process; these are the (copied and pasted) actual formulas, with "i" the result of Step 1, BRSV Step 5, and ISRV Step 9.

$$\text{BRSV} = \text{BRS} - \text{BR} * i * \text{sqrt}(\text{PF})$$

$$\text{IRSV} = \text{IR} * i * \text{sqrt}(\text{PF}) - \text{IRS}$$

$$\text{RRA} = ((\text{R} - (\text{BRSV} + \text{IRSV})) * 9 / \text{IP}) / \text{PF}$$

Hersch, Philip L. and Jodi E. Pelkowski (2014). Does general manager networking affect choice of trade partners in major league baseball? *Journal of Sports Economics*, Vol. 15 No. 6, pages 601-616.

Hersch and Pelkowski (2014), examining data from 1985 through 2011 mostly gathered from Retrosheet, were on the lookout for tendencies for general managers with connections of one type or another to another team to carry out more transactions with that other team than with others. They uncovered a small tendency for general managers who had previously worked together on the same team, and a stronger tendency for two general managers who were related to either one another or to someone else in the other's organization, to trade more often than the average two-team pairing. General managers who had previously worked for another team were otherwise not more likely to do business with the other team. Other tendencies Hersch and Pelkowski discovered were teams being relatively unlikely to transact with teams in their division but more likely to work with teams in other divisions in their league.

Howard, Jeffrey N. (2018). Hit probability as a function of foul-ball accumulation. *Baseball Research Journal*, Vol. 47 No. 1, pages 60-64.

Based on Retrosheet data 1945-2015, Jeffrey Howard (2018) noted a big difference associated with batters hitting foul balls between when two of them both count as strikes one and two and when they don't (fouls after strike two, which means swing and misses for strikes). With two strikes on them, batters have hit much better in the former circumstance than in the latter; .335 versus .124 with three non-strike fouls and .413 versus .079 with four non-strike fouls (keep in mind that this means five and six foul balls total respectively for the former situation).

Huckabay, Gary and Nate Silver (2003). Looking for advantages on the ground.
<https://www.baseballprospectus.com/news/article/1928/6-4-3-looking-for-advantages-on-the-ground/>

This is a study of batter/pitcher matchups based on fly ball/ground ball tendencies, based on 1978-2000 Retrosheet data for, batters and pitchers with at least 300 PA for or against. Players have been divided into quartiles, with 1 standing for the quartile most biased toward hitting/giving up fly balls and 4 meaning most likely to hit/give up grounders. The three data lines are for BA/OBA/SLG:

Pitchers	Hitter	Hitter	Hitter	Hitter	
Overall	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Pitchers
Pitcher					
Quartile 1	0.259	0.267	0.271	0.269	0.266
	0.338	0.335	0.335	0.331	0.335
	0.462	0.442	0.426	0.384	0.430
Pitcher					
Quartile 2	0.267	0.270	0.272	0.271	0.270
	0.343	0.337	0.335	0.330	0.336
	0.459	0.429	0.415	0.382	0.426
Pitcher					
Quartile 3	0.272	0.274	0.271	0.275	0.273
	0.346	0.340	0.335	0.333	0.339
	0.454	0.427	0.401	0.378	0.415
Pitcher					
Quartile 4	0.279	0.276	0.273	0.268	0.274
	0.351	0.340	0.336	0.326	0.338
	0.447	0.416	0.391	0.358	0.403
Aggregate:					
Hitters	0.269	0.272	0.272	0.271	
	0.344	0.338	0.335	0.330	
	0.456	0.429	0.408	0.376	

First, the overall tendencies, which are no surprise. Fly ball hitters had about the same BA but higher OBA (and so more walks) and a greater proportion of extra base hits than ground ball hitters, and fly ball pitchers were responsible for about the same BA and OBA (so the same walks) but a greater proportion of extra base hits than ground ball pitchers. In addition, the most extreme fly ball hitters had higher BA against ground ball pitchers (unlike the other three categories), and analogously the most extreme ground ball pitchers BA went up for fly ball hitters.

Humphreys, Michael A. (2011). *Wizardry*. New York: Oxford University Press.

This in my opinion is the best effort to date to evaluate defensive skill based on conventional data, i.e., not through zone-rating analysis of actual gameplay. There are

actually two procedures, both titled Defensive Regression Analysis (DRA), one using Retrosheet data and the other based on conventionally available fielding indices. I will describe procedures non-technically; those interested in the details should consult the book. The goal of the effort was to rid the available data of bias in every practical case, particularly in terms of pitching staff tendencies (i.e., strikeouts versus outs on balls in play, ground ball versus fly ball, lefthanded versus righthanded innings). These tendencies are assumed independent of one another, such that for example lefties and righties on a team are presumed to have the same ground ball/fly ball tendencies. This of course is not true, and, when available, using the Retrosheet data allowed Michael to overcome these problems also. For each position, and starting with a large set of indices, Michael transformed each *relevant index (for example, strikeouts per batters faced, assists per number of balls in play) so as to make each as uncorrelated with one another as possible. The indices for different positions were of course specific to each. For the same reasons I did, and contrary to Bill James's veiled criticisms of my work, Michael only used assists for evaluating most infielders and also catchers, and made what in my probably-biased opinion provided a very persuasive argument for that decision. For analogous reasons, first basemen are only evaluated on their ground ball putouts, although this leaves one with a bias caused by the individual player's tendencies to make the play unassisted versus tossing to covering pitchers. Outfielders are of course rated by putouts.

After that, Michael associated these transformed indices with runs-allowed data, allowing the determination of the average number of runs for each event. These numbers corresponded well with past efforts (e.g., walks worth .34 runs, home runs 2.44 runs), adding a degree of credence to the calculations. Humphrey had to make some potentially controversial decisions along the way; for example, crediting responsibility for infield popups to the pitcher under the assumption that the batter was overpowered, despite his general acceptance of the DIPS principle that the result of batted balls in play are not due to the pitcher. Michael's resulting ratings correlate at about .7 with two zone-rating-type measures, Mitchell Lichtman's Ultimate Zone Rating and Tom Tippett's, and leads to analogous findings. The best fielders save about 20 runs a year, whereas the worse cost 20 runs, when compared to the average.

Hyland, David C. (2022). Rounding second: A probabilistic investigation of the major league baseball modified extra innings rule. *Baseball Research Journal*, Vol. 51 No. 2, pages 61-65.

According to Retrosheet data from 2019 and 2021, the expected probability of scoring at least one run was .29 with no out no baserunners but .61 with no out runner on second. Here are probabilities of a team scoring a given number of runs, and of both teams scoring that number of runs, in an inning under old and new rules:

	Runs	0	1	2	3	4	5-7
Old Rules	One Team	71	15	7	4	2	1
	Both Teams	51	2	1	0	0	0

New Rules	One Team	39	32	14	8	4	4
	Both Teams	16	10	2	1	0	0

Given this, the following displays the length of extra inning games occurring in 2019 2021 and their authors' prediction for 2021 using a Markov model and given the data above.

Innings	10	11	12	13	14 and up
Actual 2019	91	58	22	14	23
Predicted 2021	155	33	22	5	1
Actual 2021	154	46	13	3	2

In their table, they also included predictions for 2019 using both the old and new rules; the former was reasonably close to the 2019 actual, and the latter was closer to the 2021 actual than was their predictions.

Hyman, Barry (2021). Overall Offensive Performance (OOP). *Baseball Research Journal*, Vol. 50 No. 2, pages 130-139.

Barry Hyman (2021) proposed what he called Overall Offensive Performance (OOP), in which players receive credit for the bases gained by their own production, including getting on base due to errors, bases gained by base runners not due to “extra effort” - one base on singles, two bases on doubles, etc. - and bases gained by “extra effort” when baserunners, such as extra bases on hits, steals, and the like. They are charged for outs made, both at bat and on the basepath. This metric clearly is biased toward batters who get to the plate with a lot of baserunners aboard and away from those usually batting with bases empty. Using Retrosheet as his data source, Barry concluded that the average batter's OOP would be in the range of 1 to 1.3.

James, Bill (2006). Relative range factors. In John Dewan, *The Fielding Bible* (pages 199-209). Skokie, IL: Acta Sports.

Bill proposed a new version of range factor in order to correct for various biases in the original measure. One of these biases was the use of games played as a denominator, because it short-changed fielders who did not play full games with some regularity. Bill used Retrosheet data to compute the actual number of innings these fielders played.

James, Bill (2008). *The Bill James Gold Mine 2008*. Skokie, IL: Acta Press.

Batting performance tends to tail off between the middle and end of seasons, but Bill James (2008, pages 310-311) uncovered evidence that player size interacts with this general tendency. Among the 1000 position players with the most plate appearances

between 1957 and 2006, Bill compared the size, as measured by an undescribed combination of height and weight, the fifty largest lost 32 OPS points between June and September (from .834 to .802) whereas the fifty smallest lost 11 (from .699 to .688). Although Bill does not say so, I suspect he used Retrosheet data here.

James, Bill (2010). *The Bill James Gold Mine 2010*. Skokie, IL: Acta Sports.

At least during the 2000-2009 decade, it was not true that teams made an effort to match up their number one starters against one another; if anything, it was the opposite. Using his Season Score metric, which works well enough for this sort of analysis, here are Season Score categories for starters and their average opposition:

Pitcher Season Score	Number of Pitchers	Number of Starts	Opposition Pitcher Season Score	Pitcher Season Score	Number of Pitchers	Number of Starts	Opposition Pitcher Season Score
>299	11	366	68.88	50-99	451	10151	77.89
200-299	136	4093	77.67	0-49	980	11614	79.40
150-199	152	4660	80.13	<0	963	8711	81.63
100-149	316	8987	78.01				

The lowest (highest) average opposition starter Season Score was for the starters with the highest (lowest) Season Score. As Bill mentioned, it looks like there was a slight tendency for teams facing the absolute best starting pitchers to sacrifice the game and start their weakest.

Jane, Wen-Jhan (2022). Choking or excelling under pressure: Evidence of the causal effect of audience size on performance. *Bulletin of Economic Research*, Vol. 74 No. 1, pages 329-357.

Using 2015 to 2018 Retrosheet performance data and attendance figures from mlb.com, along with various control variables, Wen-Jhan Jane (2022) examined the influence of the latter on the former. Overall, using a metric that I believe is hits divided by plate appearances, the average performance for both home and away teams were an inverted U function across five attendance categories (less than 10K, 10K to 20K, 20K to 30K, 30K to 40 K, and more than 40K). Home team players peaked in the 30K to 40K range whereas away team players did so between 20K and 30K. Although present in every inning, the effect for the away team effect players was stronger yet in the 9th and later innings, with the peak now between 10K and 20K. However, there was evidence that “star” players, defined as those who had been All-Stars the previous season, actually improved as attendance rose. Jane's study also revealed more support for home field advantage by means of higher figures on the H/PA metric.

Jarvis, John F. (1999). An analysis of the intentional base on balls. Presented at the 1999 SABR convention and retrieved from <http://knology.net/johnfjarvis/IBBanalysis.html>

Jarvis, John F. (2002). Career summaries and projections. Presented at the 2002 SABR convention and retrieved from <http://knology.net/johnfjarvis/cftn.html>

John Jarvis (1999), using the data then available from Retrosheet (1980 through 1996 with the exception of 1991), performed simulations that actually found support for the defensive use of the intentional walk, suggesting that it decreased the number of one- and two-run innings and, although it increased the number of innings with three or more runs, the former impact outweighed the latter. However, by 2002 Jarvis was changing his tune, calculating with 1969 and 1972 to 2002 data that intentional walks only helped the defense when the batter's slugging percentage was greater than .600, which occurred in only four percent of the at bats over those years.

Jarvis, John F. (2000). Mark McGwire's 162 bases on balls: More than one record in 1998. *Baseball Research Journal*, No. 29, pages 107-112.

Adding a wrinkle to research regarding the value of the intentional walk as a strategic tool, we have the unofficial "intentional" walk, when an opposing team does not signal the IBB but the pitcher does not throw anywhere near the center of the plate. John Jarvis (2000) wanted to figure out the circumstances that most often distinguish official IBBs from other walks, so that we can at least speculate the situations when walks not classified as intentional to all extents and purposes are. Based on neural net training and a regression analysis for validation, and again using Retrosheet data, John determined that a walk is most likely intentional if, in order of importance, there is a runner on second, there is a runner on third, there is not a runner on first, the relative score between opposing and batting teams, the inning is later, and there are more outs in the inning (relative score was behind in inning and outs in the regression). The slugging average of the batter and (negatively) the next batter also had impact but, surprisingly, far less than the previous list. I would speculate that this is because IBBs often happen at the bottom of the lineup and not only when the best opposing hitter is at the plate.

At some point, John Jarvis did an unpublished study using 17 different seasons for which there was then available Project Scoresheet or Retrosheet data and demonstrated that attempted steals result in worse performance by batters. He also learned that over an entire league the stolen base led to an average of only 2.7 wins per season (with a range of 7 to -2.5).

Jordan, Douglas and David Macias (2019). Team batting average: A comprehensive analysis. *Baseball Research Journal*, Vol 48 No. 1, pages 64-69.

Based on Retrosheet data from 2017, team batting averages pretty much stabilized at by about game 70 and remained the same until the end of the season. They also began to have predictive value in terms of teams better or worse than league average by game 16. Overall between 2003 and 2017, team BA tends to increase until that 70th or so game, I imagine that this is a consequence of warmer and more humid weather encouraging batted ball flight.

Judge, Jonathan and Sean O'Rourke (2020). Measuring defensive accuracy in baseball. <https://www.baseballprospectus.com/news/article/58243/measuring-defensive-accuracy-in-baseball/>

Jonathan Judge and Sean O'Rourke (2020) used Retrosheet data to compare 2019 fielding performance with evaluations for the then-current version of FRAA (as always, details unknown) with the following set of "competitors": Sports Info Solutions' then current version of Defensive Runs Saved, Mitchel Lichtman's Ultimate Zone Rating, Chris Dial's Runs Effectively Defended, and MLB's Outs Above Average. Ignoring the details, FRAA was the most accurate for outfielders and the least accurate for infielders, OAA was the opposite, RED and DRS did okay across the board, and UZR performed relatively poorly. They speculated that fielder positioning and movement might be significant for infield defense but not for outfielders. If so, then OAA's reliance on it, as described in Tom Tango's essay, could be crucial for infielder evaluation but only add random error for outfielders. It is however important to note that to even the playing field they purposely added no controls for batter, pitcher, ballpark or overall team defense. While defensible in this case, they would need to do so if comparing FRAA to its actual closest "competitor," Michael Humphreys' Defensive Regression Analysis. For what it's worth, I challenge them to do so.

Judge, Jonathan, Harry Pavlidis, and Dan Brooks (2015). Moving beyond WOWY: A mixed approach to measuring catcher framing. <https://www.baseballprospectus.com/news/articles/25514/moving-beyond-wowy-a-mixed-approach-to-measuring-pitch-framing>

As part of a project designed to measure catcher framing, Judge et al., used 1988 to 2007 Retrosheet ball-strike data to estimate catcher framing abilities, resulting in a model that correlated at .7 with a model based on PITCHf/x data when applied to subsequent seasons. According to their method, the best framers saved about 20 runs in a season over average, comparable to what PITCHf/x data implies. In addition, the researchers calculated the the proportion of taken pitches that were called strikes during that period and on to 2014. The figure was around 29 percent at the beginning, eased up to about 30 percent in 2000, and then jumped to 31.5 percent the next year, perhaps as a product of umpires first answering to MLB as a whole rather than the leagues separately. At about 32 percent in 2008, it went up almost full percentage point in two years when PITCHf/x replaced Questec, and had gotten over 33 percent by 2014.

Kalist, David E. and Stephen J. Spurr (2006). Baseball errors. *Journal of Quantitative Analysis in Sports*, Vol. 2 Issue 4, Article 3.

Using Retrosheet data from 1969 through 2005, Kalist and Spurr discovered that errors tend to be higher for first-year expansion teams, in April than in later months, in day games rather than night (more variable lighting conditions?), in grass rather than artificial turf (again, more variation?), and against faster opposition, as measured by steals per game. Finally, there was a consistent bias in favor of the home team, but it decreased substantially over the period, possibly due to the replacement of active

sportswriters with others with perhaps less incentive to ingratiate themselves with home-team players.

Kim, Jerry W. and Brayden G. King (2014). Seeing stars: Matthew effects and status bias in major league baseball umpiring. *Management Science*, Vol. 60 No. 11, pages 2619-2644.

This is probably the best analysis of umpire bias to date. The basic argument is that umpires are predisposed toward favoring “high-status” pitchers; more likely calling “real” balls as strikes (“overrecognition” in the authors’ terminology) and less likely “real” strikes as balls (“underrecognition”) the higher the pitcher’s status, with the bias accentuated for pitchers known to have good control. To examine the argument’s validity, all 2008 and 2009 pitches without batter swings were categorized via f/x pitch data, with a long list of control measures gathered from various sources including Retrosheet. Status was based on number of All-Star appearances, which strikes me as a good index; pitcher control via walks per plate appearance. The results were as follows: In total, overrecognition occurred on 18.8% of real balls and underrecognition on 12.9% of real strikes. Both over- and underrecognition were more likely for the home team, counts favoring the batter, later innings, high leverage plate appearances, more experienced pitchers, and as hypothesized pitchers with more All-Star appearances and better control. The status effects were still apparent for pitches by high and low status pitchers matched for pitch location and type, specific umpire, and count; All-Stars received a relative 6.7% reward in overrecognition and 5.7% bonus in underrecognition. Overrecognition also occurred for lefty batters and games with higher attendance. In my view, the authors’ argument seems to generalize to more experienced pitchers, who would have status for that reason alone. In addition, the results for attendance and home team are consistent with the most strongly supported explanation for the home-field advantage; crowd noise.

In addition, analogous biases were uncovered in favor of batters with high status (again All-Star appearance) and demonstrated batting eyes (walks per plate appearance). Variance depending on catcher revealed different skill levels in pitch framing ability, which was not associated with All-Star catcher appearances; skill in pitch framing does appear less appreciated than it deserves. Finally, overcoming a problem in past umpire bias research, an on-line unpublished version of the paper included individual differences among umps in both over- and underrecognition. The authors concluded that 80% of umps are guilty of the former and 64% of the latter. Interestingly, the two biases were largely independent, correlating at only -.16.

Koch, Brandon Lee D. and Anna K. Panorska (2013). The impact of temperature on major league baseball. *Weather, Climate, and Society*, Vol. 5, pages 359-366.

Retrosheet data from 2000 through 2011 combined with data from the National Climate Data Center revealed that most offensive measures (runs scored, home runs, batting, on-base, and slugging averages) increased as game weather got hotter, with the exception of walks. Koch and Panorska also noted the impact of heat on hit batsmen; see Larrick below.

Larrick, Richard P., Thomas A. Timmerman, Andrew M. Carton, and Jason Abrevaya (2011). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science*, Vol. 22 No. 4, pages 423-428.

Krenzer, William L. D., and Eric D. Splan (2018). Evaluating the heat-aggression hypothesis: The role of temporal and social factors in predicting baseball related aggression. *Aggressive Behavior*, Vol. 44 No. 1, pages 83-88.

It has become clear that as the weather gets warmer, the number of hit batsmen goes up, and this has been explained as a consequence of discomfort resulting in increased aggressiveness. Larrick, Timmerman, Carton and Abrevaya (2011), using all games with Retrosheet data from 1952 through 2009 which included game temperature and controlling for pitcher control, discerned that the odds of a hit batsman increased as an interactive function of temperature and the number of teammates hit by the opposing team, such that more hit teammates resulted in more plunking of the opposing team, with this effect accentuated by hotter weather. Krenzer and Splan, using 2000-2015 Retrosheet data, noted both temperature and, more importantly, pitcher wildness as predictors HBPs. Further addressing the question, after dividing the season into fifths based on games played, they observed this correlation only occurring during the middle three-fifths, in other words the warmer months, implying a probable threshold temperature effect before aggression steps in. In addition, HBPs were greater against division rivals than otherwise (where the best rivalries lie), in blow-out games rather than one-runners (unfortunately they did not analyze winner versus losers separately; is this frustrating only for the blown-out team?), and for some reason against visiting teams with better records (why, and why not home teams also?).

Lei, Xinrong and Brad R. Humphreys (2013). Game Importance as a dimension of uncertainty of outcome. *Journal of Quantitative Analysis in Sports*, Vol. 9 No. 1, pages 25-36.

Of the several reasons proposed for the home field advantage in baseball, which is consistently measured at 53 or 54 percent, the most strongly backed by research is the presence of fan support, as home field advantage increases with rising attendance. Indirect corroboration comes from work by Lei and Humphreys (2013). They proposed a measure of game importance (GI), based on either how far a team leading a divisional or wild-card race is ahead of the second place team or how far a team not leading is behind the team that is. Smaller differences imply higher GI scores. Unfortunately, as the authors note, their measure is not weighted by how far in the season a game occurs, so that GI will be the same for a team one game ahead or behind after the 1st as the 161st game. Anyway, in Retrosheet data from 1994 through 2010, GI was positively related with both attendance and home team winning percentage, with the latter implying that home field advantage rises as games become more important. The authors did not know to relate all three, but we can conjecture that game importance raises attendance which increases home field advantage in turn.

Levitt, Dan (1999). Hits and baserunner advancement. *By the Numbers*, Vol. 9 No. 3,

pages 20-21.

Dan Levitt (1999) has provided us with estimates of the odds of baserunner advancement on hits based on four years of Retrosheet data (1980-1983). The following is what I believe to be the most interesting of Levitt's findings. The three right-most columns display hit locations when known.

Occurrence	Result	Sample Size	Total	Left Field	Center Field	Right Field
Single with runner on first	Runner to third	31132	31.3%	19.1%	34.6%	49.4%
Single with runner on second	Runner scores	18399	65.3%	68.4%	82.6%	71.7%
Double with runner on first	Runner scores	6997	53.6%	40.5%	58.6%	37.7%

Most of the results can be explained through considering the throwing distance from the outfielder to the relevant base. As home plate is generally farther from the outfield than third base, runners successfully take extra bases to score more often than to get to third. Baserunner advancement for first-to-third after a single is more likely as we move from left field to right. Runners are more likely to score from first on doubles or second on singles to center field than to the corners. It is interesting to note that scoring from first on doubles is both less likely and less influenced by hit location than scoring from second on singles.

Levitt, Dan (2000). Speed scores and reaching base on errors. Retrieved from <http://www.baseballthinkfactory.org/btf/scholars/levitt/articles/speedscores.htm>

Levitt (2000), this time using individual-level data from 1980 Retrosheet files, found Speed Scores to correlate only .14 with percentage of times reaching base on error per opportunity to do so. Further, this relationship appeared to be an artifact of the number of ground balls hit, given that faster runners are more likely to hit ground balls (Speed Score and percentage of batted balls that are grounders were correlated .3), such that Speed Scores only correlated .04 with times reached base on errors as a percentage of non-basehit ground balls. In other words, faster runners do not reach base on errors more often because they are fast, but rather because they hit more grounders, which lead to more errors than fly balls.

Levitt, Dan (2006). Empirical analysis of bunting.

http://baseballanalysts.com/archives/2006/07/empirical_analy_1.php

Following are run expectancy and one-run probability charts for lineup positions, each league, based on 1977 to 1992 Retrosheet data.

Run Expectancies

AL				NL			
1	0	1	2	1	0	1	2
---	.553	.291	.100	---	.542	.294	.102
x--	.951	.567	.210	x--	.911	.530	.213
-x-	1.263	.753	.323	-x-	1.130	.720	.342
xx-	1.614	.966	.428	xx-	1.526	.868	.418
--x	1.395	.976	.399	--x	1.319	1.003	.399
x-x	1.840	1.242	.527	x-x	1.786	1.107	.506
-xx	2.182	1.456	.623	-xx	1.978	1.336	.621
xxx	2.365	1.621	.773	xxx	2.081	1.480	.722

2	0	1	2	2	0	1	2
---	.543	.297	.113	---	.530	.286	.104
x--	.966	.576	.253	x--	.977	.611	.251
-x-	1.214	.752	.346	-x-	1.180	.723	.333
xx-	1.599	1.028	.453	xx-	1.583	.979	.450
--x	1.435	1.012	.432	--x	1.368	.971	.394
x-x	1.865	1.286	.531	x-x	1.778	1.211	.523
-xx	2.100	1.487	.609	-xx	2.068	1.375	.570
xxx	2.434	1.685	.822	xxx	2.398	1.473	.732

3	0	1	2	3	0	1	2
---	.536	.305	.117	---	.517	.297	.118
x--	.945	.581	.268	x--	.928	.582	.278
-x-	1.192	.740	.385	-x-	1.129	.735	.395
xx-	1.609	1.002	.522	xx-	1.607	1.007	.518
--x	1.422	1.017	.400	--x	1.337	.993	.401
x-x	1.820	1.249	.574	x-x	1.831	1.266	.562
-xx	2.052	1.534	.674	-xx	2.031	1.518	.715
xxx	2.468	1.699	.867	xxx	2.402	1.720	.817

4	0	1	2	4	0	1	2
---	.488	.293	.118	---	.442	.274	.115
x--	.885	.567	.252	x--	.849	.553	.261
-x-	1.160	.711	.343	-x-	1.098	.719	.350
xx-	1.501	.962	.488	xx-	1.488	.961	.532
--x	1.318	.972	.412	--x	1.308	.958	.390
x-x	1.816	1.230	.530	x-x	1.741	1.247	.559
-xx	1.950	1.445	.644	-xx	1.864	1.426	.596
xxx	2.345	1.616	.863	xxx	2.457	1.615	.867

5	0	1	2	5	0	1	2
---	.452	.254	.107	---	.403	.224	.103
x--	.835	.537	.245	x--	.757	.494	.220
-x-	1.110	.706	.339	-x-	.925	.648	.340
xx-	1.453	.930	.463	xx-	1.336	.913	.452
--x	1.223	.946	.373	--x	1.159	.942	.389
x-x	1.674	1.200	.529	x-x	1.579	1.163	.496
-xx	1.900	1.353	.550	-xx	1.881	1.356	.607
xxx	2.301	1.601	.795	xxx	2.284	1.588	.775

6	0	1	2	6	0	1	2
---	.446	.231	.094	---	.370	.191	.079

x--	.791	.464	.220
-x-	1.059	.646	.336
xx-	1.415	.905	.459
--x	1.328	.951	.367
x-x	1.712	1.129	.518
-xx	2.016	1.340	.581
xxx	2.200	1.532	.755

x--	.725	.430	.210
-x-	.941	.585	.309
xx-	1.311	.851	.404
--x	1.095	.829	.342
x-x	1.435	1.106	.452
-xx	1.764	1.336	.531
xxx	1.997	1.536	.726

7	0	1	2
---	.439	.225	.083
x--	.800	.438	.201
-x-	1.076	.617	.310
xx-	1.408	.836	.419
--x	1.230	.888	.354
x-x	1.625	1.107	.453
-xx	1.852	1.360	.570
xxx	2.337	1.480	.753

7	0	1	2
---	.363	.183	.061
x--	.652	.388	.176
-x-	.913	.540	.261
xx-	1.293	.756	.385
--x	1.242	.749	.327
x-x	1.507	1.036	.419
-xx	1.718	1.220	.469
xxx	2.062	1.450	.717

8	0	1	2
---	.474	.226	.077
x--	.798	.461	.179
-x-	1.039	.609	.283
xx-	1.431	.804	.410
--x	1.419	.919	.347
x-x	1.674	1.105	.444
-xx	1.962	1.322	.561
xxx	2.289	1.465	.686

8	0	1	2
---	.397	.172	.054
x--	.678	.375	.127
-x-	.923	.485	.230
xx-	1.179	.694	.321
--x	1.212	.782	.274
x-x	1.514	.945	.429
-xx	1.620	1.157	.495
xxx	1.994	1.315	.661

9	0	1	2
---	.519	.263	.081
x--	.852	.480	.182
-x-	1.128	.641	.293
xx-	1.475	.927	.382
--x	1.423	.947	.341
x-x	1.725	1.145	.457
-xx	2.108	1.396	.513
xxx	2.386	1.533	.709

9	0	1	2
---	.450	.194	.050
x--	.739	.362	.125
-x-	1.022	.542	.181
xx-	1.238	.705	.230
--x	1.281	.753	.236
x-x	1.466	.891	.269
-xx	1.730	1.048	.387
xxx	1.930	1.219	.470

One Run Probabilities

AL	1	0	1	2
---	.302	.170	.067	
x--	.458	.292	.121	
-x-	.662	.436	.218	
xx-	.658	.427	.233	
--x	.827	.655	.295	
x-x	.872	.672	.302	
-xx	.874	.691	.293	
xxx	.867	.696	.340	

NL	1	0	1	2
---	.301	.173	.066	
x--	.426	.263	.120	
-x-	.606	.411	.232	
xx-	.653	.428	.228	
--x	.794	.666	.284	
x-x	.864	.662	.305	
-xx	.852	.669	.290	
xxx	.829	.661	.338	

2	0	1	2
---	.298	.176	.073
x--	.483	.306	.148
-x-	.665	.433	.236
xx-	.678	.466	.233
--x	.870	.687	.294
x-x	.880	.671	.296
-xx	.887	.725	.278
xxx	.889	.700	.347

2	0	1	2
---	.300	.171	.065
x--	.497	.320	.146
-x-	.659	.429	.223
xx-	.653	.433	.232
--x	.846	.653	.276
x-x	.855	.651	.299
-xx	.847	.696	.268
xxx	.882	.663	.323

3	0	1	2
---	.299	.185	.077
x--	.457	.308	.150
-x-	.649	.439	.252
xx-	.673	.460	.271

3	0	1	2
---	.301	.187	.078
x--	.470	.314	.156
-x-	.644	.436	.254
xx-	.668	.459	.262

--x	.842	.707	.288	--x	.854	.696	.291
x-x	.899	.688	.312	x-x	.889	.687	.302
-xx	.874	.731	.302	-xx	.888	.701	.315
xxx	.910	.702	.365	xxx	.879	.720	.350

4	0	1	2	4	0	1	2
---	.280	.182	.081	---	.271	.176	.083
x--	.433	.294	.141	x--	.444	.304	.149
-x-	.635	.427	.231	-x-	.632	.437	.234
xx-	.645	.445	.252	xx-	.638	.443	.260
--x	.830	.667	.290	--x	.811	.680	.278
x-x	.868	.668	.299	x-x	.866	.680	.308
-xx	.867	.706	.288	-xx	.862	.694	.268
xxx	.886	.682	.351	xxx	.908	.692	.352

5	0	1	2	5	0	1	2
---	.261	.163	.076	---	.245	.151	.073
x--	.405	.286	.136	x--	.396	.269	.132
-x-	.629	.416	.231	-x-	.581	.412	.235
xx-	.622	.435	.241	xx-	.630	.441	.236
--x	.842	.660	.263	--x	.794	.682	.288
x-x	.859	.654	.294	x-x	.855	.685	.284
-xx	.842	.672	.260	-xx	.853	.674	.278
xxx	.883	.691	.342	xxx	.878	.701	.326

6	0	1	2	6	0	1	2
---	.252	.148	.066	---	.221	.127	.060
x--	.394	.253	.128	x--	.383	.249	.128
-x-	.599	.405	.233	-x-	.559	.380	.218
xx-	.604	.427	.239	xx-	.606	.418	.224
--x	.793	.668	.270	--x	.756	.640	.266
x-x	.869	.636	.291	x-x	.832	.651	.278
-xx	.876	.654	.278	-xx	.838	.668	.261
xxx	.844	.668	.315	xxx	.848	.684	.307

7	0	1	2	7	0	1	2
---	.245	.140	.059	---	.211	.115	.046
x--	.394	.240	.117	x--	.350	.226	.113
-x-	.605	.380	.216	-x-	.556	.357	.200
xx-	.602	.403	.235	xx-	.592	.389	.222
--x	.814	.628	.266	--x	.784	.582	.267
x-x	.845	.643	.277	x-x	.830	.653	.257
-xx	.864	.690	.272	-xx	.815	.656	.236
xxx	.854	.668	.331	xxx	.831	.681	.322

8	0	1	2	8	0	1	2
---	.259	.134	.053	---	.220	.104	.038
x--	.393	.247	.106	x--	.360	.204	.082
-x-	.593	.379	.207	-x-	.537	.324	.168
xx-	.608	.392	.216	xx-	.549	.346	.194
--x	.855	.652	.262	--x	.759	.583	.222
x-x	.847	.627	.264	x-x	.810	.600	.286
-xx	.843	.651	.266	-xx	.727	.625	.248
xxx	.864	.656	.302	xxx	.840	.610	.306

9	0	1	2	9	0	1	2
---	.277	.154	.052	---	.240	.109	.030
x--	.423	.252	.108	x--	.397	.217	.072
-x-	.624	.386	.209	-x-	.585	.342	.133
xx-	.624	.425	.213	xx-	.580	.336	.136
--x	.822	.653	.266	--x	.781	.530	.194
x-x	.860	.646	.272	x-x	.734	.514	.165
-xx	.872	.678	.241	-xx	.770	.543	.197
xxx	.872	.662	.304	xxx	.808	.559	.218

Lg	BOP	Runners	All	SH Only	All Bunts
No Outs					
A	1	x--	.951	.848	.899
A	1	-x-	1.263	1.062	1.203
A	1	xx-	1.614	1.635	1.676
A	2	x--	.966	.753	.848
A	2	-x-	1.214	1.131	1.206
A	2	xx-	1.599	1.694	1.744
A	3	x--	.945	.769	.818
A	5	x--	.835	.702	.752
A	6	x--	.791	.642	.643
A	6	xx-	1.415	1.416	1.388
A	7	x--	.800	.664	.709
A	7	xx-	1.408	1.517	1.430
A	8	x--	.798	.714	.715
A	8	-x-	1.039	1.057	1.082
A	8	xx-	1.431	1.575	1.496
A	9	x--	.852	.802	.790
A	9	-x-	1.128	1.146	1.137
A	9	xx-	1.475	1.464	1.455
N	1	x--	.911	.878	.909
N	2	x--	.977	.784	.837
N	2	-x-	1.180	1.094	1.185
N	2	xx-	1.583	1.606	1.612
N	5	x--	.757	.800	.714
N	6	x--	.725	.683	.682
N	7	x--	.652	.575	.587
N	8	x--	.678	.619	.611
N	9	x--	.739	.769	.724
N	9	-x-	1.022	1.159	1.137
N	9	xx-	1.238	1.404	1.325
1 Out					
N	9	x--	.362	.380	.354
N	9	xx-	.705	.732	.724

BOP is batting order position. Note that the findings here look better for the sac bunt than standard run expectancy analysis in some circumstances. Also note all bunts versus sac bunts; need to interpret this in bunt for hits section.

TABLE 6 - Probability Results of Actual Bunts Compared to All Events

Lg	BOP	Runners	All	SH Only	All Bunts
No Outs					
A	1	x--	.458	.476	.478
A	1	-x-	.662	.681	.692
A	1	xx-	.658	.766	.729
A	2	x--	.483	.455	.474
A	2	-x-	.665	.736	.726
A	2	xx-	.678	.757	.738
A	3	x--	.457	.448	.451
A	5	x--	.405	.405	.400
A	6	x--	.394	.386	.372
A	6	xx-	.604	.686	.667
A	7	x--	.394	.386	.395
A	7	xx-	.602	.715	.654
A	8	x--	.393	.421	.404
A	8	-x-	.593	.664	.646
A	8	xx-	.608	.714	.681
A	9	x--	.423	.454	.440
A	9	-x-	.624	.708	.693
A	9	xx-	.624	.703	.656
N	1	x--	.426	.457	.459
N	2	x--	.497	.461	.472
N	2	-x-	.659	.741	.735
N	2	xx-	.653	.686	.670

N	5	x--	.396	.454	.403
N	6	x--	.383	.417	.419
N	7	x--	.350	.401	.387
N	8	x--	.360	.384	.376
N	9	x--	.397	.432	.405
N	9	-x-	.585	.723	.663
N	9	xx-	.580	.678	.616
1 Out					
N	9	x--	.217	.247	.226
N	9	xx-	.336	.339	.342

Levitt, Dan (n.d.2) Fielding opportunities by position based on pitcher hand.
https://www.baseballthinkfactory.org/btf/scholars/levitt/articles/fielding_opps.htm

No date, but probably 1984, very early in the discussion. Dan realized that pitcher handedness affected number of opportunities for fielder to make plays. His argument was the same as mine; more righty(lefty) pitchers mean more lefty(righty) batter and so more batted balls to the left(right) side of the field. Using 1980-1983 Retrosheet data. The first table is the person fielding a batted ball resulting in the first out of an inning.

Table 1 -- First "Out" By Position

Pos	<---LHP----->		<---RHP----->	
	"Outs"	Pct	"Outs"	Pct
1	5420	5.8%	12144	5.7%
2	1711	1.8%	4025	1.9%
3	7337	7.9%	22091	10.3%
4	14348	15.4%	39037	18.2%
5	14027	15.0%	26518	12.4%
6	18232	19.5%	37957	17.7%
7	9006	9.6%	23598	11.0%
8	12841	13.8%	28630	13.3%
9	10466	11.2%	20636	9.6%
Total	93388	100.0%	214636	100.0%

As expected, more for right(left)side of infield with righthanded(left-handed) pitcher). Dan was surprised that the opposite occurred for outfielders, my guess is that because flies to the opposite field tend to be hit with lower exit velocity and worse launch angle than pulled flies.

The second is which position fielded hits (not clear, but I am guessing hits on first batted ball of an inning. Note that here outfielders get batted balls as expected by pitcher handedness.

Table 2 -- Hits Fielded By Position

Pos	<---LHP----->		<---RHP----->	
	"Hits"	Pct	"Hits"	Pct
1	501	1.4%	1088	1.3%
2	42	0.1%	81	0.1%
3	294	0.8%	696	0.9%
4	572	1.6%	1594	2.0%
5	1225	3.4%	2358	2.9%
6	1289	3.6%	2679	3.3%
7	13397	37.0%	25195	30.9%
8	10322	28.5%	24108	29.6%

9	8544	23.6%	23666	29.1%
Total	36186	100.0%	81465	100.0%

Lindbergh, Ben (2016). Sabermetrics is killing bad dugout decisions.

<https://fivethirtyeight.com/features/sabermetrics-is-killing-bad-dugout-decisions/>

Based on data from Retrosheet and MLB, pitch outs averaged between 0.6 and 0.8 per team per game late during the 1980s and early 1990s, dipped to around 0.3 mid-1990s to mid 2000s, then down some more to about 0.1 in 2015. The webpost title says it all.

Lyle, Arlo (2007). Baseball prediction using ensemble learning.

https://arti.franklin.uga.edu/sites/default/files/inline-files/lyle_arlo.pdf

The most trustworthy attempt to compare the accuracy of offensive projection models that I have been able to find is a M.A. thesis by Lyle (2007). The author, applying Retrosheet data between 1973 and 2006, used the previous 162-game performance of batters to predict the next 162 game outcomes for six metrics. For four of the six (runs scored, doubles, homers, and RBI), PECOTA slightly outperformed his own method and significantly defeated ZiPS and MARCEL. Lyle's did the best with triples, with ZiPS second, and with hits, which PECOTA did not project.

Mains, Rob (2020). Some bunts are OK.

<https://www.baseballprospectus.com/news/article/61876/veteran-presence-some-bunts-are-ok/>

This was probably but not definitely from Retrosheet data: Only twelve players attempted 100 or more what were definitely attempts at bunt hits (bunts with bases empty) between 2003 and most of 2020 (this entry was dated September 15). Even with these players, the attempt was relatively rare, with the leader in percentage of plate appearances at only 8.3 (Willy Taveras). Only twelve (with ten overlapping the two lists) had forty or more successful attempts; but among these twelve, the success rate aka batting average on bunts for hits was exactly .400, ranging from Taveras (.476) to Dave Roberts (.328). This means that there are some players who have been quite good at it.

Mains, Rob (2022). Why they're going to keep swinging for the fences.

<https://www.baseballprospectus.com/news/article/72363/veteran-presence-why-theyre-going-to-keep-swinging-for-the-fences/>

As part of an ongoing project relating home runs with team winning average, Rob Mains (2022) ascertained from Retrosheet game logs that since 1969 home teams have consistently had winning averages of over .700 in those games in which they outhomered the away team, with (reading off charts) that figure at around .750 in the 1980s but up to about .800 in the 2010s. Away teams with more roundtrippers than

home teams have had analogous success, averaging perhaps .680 in the 1980s and .730 in the 2010s.

Mains, Rob (2022). Just when we had it figured out...

<https://www.baseballprospectus.com/news/article/79214/veteran-presence-stolen-bases-attempts/>

Rob Mains (2022) presented from very useful figures on stolen base success rates and breakevens from what is almost certainly Retrosheet data 1950 to 2022:

Runner on first, no outs – success rate was below 55 percent in the early 1950s and has risen since, to about 60 percent with a lot of annual fluctuation in the 1960s, between 65 and 70 percent around 1970-2000, 70 percent 2000-2020, then up to 75 percent 2021-2022. Breakevens were consistently between 70 and 75 percent, and 2021 and 2022 were the first years in which success rate topped it.

Runner on first, one out – success rates and breakevens about the same, but the former were still a bit below the latter in 2021 and 2022.

Runner on first, two out – success rates were somewhat higher, 60 percent early on, around 65 percent in the 1960s, up to between 70 and 75 percent 1980s through 2000s, and between 75 and 80 percent mid-2010s to 2022. Breakeven was much lower, generally between 65 and 70 percent, such that success rate reached breakeven by 1960s and have consistently topped it by an ever increasing amount since.

Runner on second, no outs – with a far lower sample size, success rates fluctuated wildly season to season, say 55 to 65 percent during the 1950s, mostly 65 to 75 percent since. With breakeven at 75-80 percent, teams have consistently been hurting themselves.

Runner on second, one out – success rate was lower, 55 to 65 percent through around 1980, 65 to 70 percent through 2000, and 70 to 75 percent afterward. Breakeven has actually gone down; 70 to 75 percent 1950s, around 70 percent through around 2005 and then between 65 and 70 percent since. So teams starting topping breakevens during some seasons in the 1980s through 2000 and consistently since.

Runner on second, two out – success rate was mostly 70 to 80 percent to 1980, then 80 to 90 percent since, and with breakeven at 85 to 90 percent, success rate approximated it since 1980.

Overall, in only three seasons had teams gained outs through steals; 2007 (101), 2021 (122), and 2022 (182). Looked at this way, teams finally got smart about steals in the 2020s.

Marchi, Max (2009). Guarding the lines. <https://tbt.fangraphs.com/guarding-the-lines/>

Top of the ninth, ahead by one run, average right-handed hitter at the plate; do you guard the line? Max Marchi's (2009) analysis (with 2008 Retrosheet and Gameday data) assumed only groundballs hit and assumed away triples as possible outcomes, but the logic should work with lefty hitters, liners, and with triples included. Based on proportion of groundball outs, singles, and doubles located at each infield batted ball location (+45 degrees to -45 degrees), Max calculated odds of 75.4 percent for outs, 22.3 percent for singles, and 1.9 percent for doubles with corner infielders in normal

position and 70.1 percent, 28.2 percent, and 1.3 percent for these outcomes with corner infielders playing the line. Then, by multiplying each of those with respective Win Expectancies given that situation and each of these outcomes, Max estimated total Win Expectancies of 84.3 percent for normal positioning and 83.2 percent for guarding the lines.

Marchi, Max (2010). Two dimensions of catching – and dealing with interactions.

<https://tft.fangraphs.com/two-dimensions-of-catching/>

Using Retrosheet data, Max Marchi (2010) devised an index for pitch blocking by dividing the sum of wild pitches and passed balls by the number of plate appearances with runners on base for each catcher/pitcher dyad, combining all of the data for (I assume) a league-year, and then using multilevel analysis to distinguish the impact of individual pitchers and catchers. Finally, he assigned a run value based on .3 runs per unblocked pitch. As would be expected, Hoyt Wilhelm and Charlie Hough ranked as the most responsible pitchers and Greg Maddux the least; the sage himself (Yogi Berra) as the best pitch blocking catcher. Max did the same with base stealing, with the third available factor (baserunner) added to the mix.

Marchi, Max (2012). The art of handling the pitching staff.

<https://www.baseballprospectus.com/news/article/16096/the-stats-go-marching-in-the-art-of-handling-the-pitching-staff/>

Marchi, Max (2012a). The hidden helpers of the pitching staff.

<https://www.baseballprospectus.com/news/article/16199/the-stats-go-marching-in-the-hidden-helpers-of-the-pitching-staff/>

Max Marchi (2012) used his multi-level analytic technique and Retrosheet data to, after removing the influence of batter, pitcher, and ballpark, estimate the amount that catchers impact on the outcomes of plate appearances, in effect devising an overall catcher evaluation system. To keep things simple, Max applied the average run value of different types of batted balls. Between 2008 and 2011, the amazing Jose Molina led the way with an estimated 103 runs saved despite being involved in about half the PAs of the closest competitors. Jason Kendall came in last at minus 80 runs. Dividing the data into even versus odd-numbered days allowed a guesstimate of reliability, with a decent correlation of 0.51.

Following up with data going back to 1948 (2012a), Tony Pena was the winner at 248 with the falsely-maligned Mike Piazza tied for third at 205; Molina was easily out front on a rate basis with 38 saved per 5000 PAs. Max also estimated that rookie catchers cost their teams about four runs, and catchers new to a team three runs, per 5000 PA. Each year spent with a team increased these figures by an average of 0.70 runs per 5000 PA. He was unable to locate any noticeable aging effects. Then, adding managers to the mix yielded analogous evaluations. Bobby Cox easily the best at 82 runs, but given his long tenure this only works out to “a couple of runs” per 5000 PA. The Cox effect may be largely due to his pairing with Leo Mazzone. The two together saved 3 runs per 5000 PA whereas Cox with other pitching coaches only coaxed 0.2

runs extra per 5000 PA. Overall, former MLB pitchers who became managers saved 0.50 and former MLB catchers 0.37 runs per 5000 PA, whereas other positions and managers never playing in the majors either saved or lost 0.11 at most.

Marchi, Max (2013). Catcher framing before PITCHf/x.

<https://www.baseballprospectus.com/news/article/20596/the-stats-go-marching-in-catcher-framing-before-pitchfx/>

Two years before this work, Max Marchi (linked to in the present article) had developed what was then a state-of-the-art multilevel model to estimate the impact of pitchers, catchers, batters, and umpires on ump calls for borderline pitches. In this piece, Max used 1988-2012 Retrosheet data to estimate an analogous model for pitches in the data set that were not swung at. This model when used on 2008 to 2012 data, for which there is PITCHf/x data, correlated at .72 with his earlier model, implying that it is probably of value for getting approximate figures for earlier catchers. However, it had a far smaller standard deviation, about 7.5 versus 13 for the PITCHf/x model. The latter means that less extreme, more conservative figures are produced, which is probably good given the very provisional status of specific catcher ratings.

Marchi, Max (2013). Who's ahead of whom?

<https://www.baseballprospectus.com/news/article/19716/the-stats-go-marching-in-whos-ahead-of-whom/>

A different kind of matchup question – at the beginning of the season, are hitters or pitchers ahead of the other? To answer it using 1991-2012 Retrosheet data, Max Marchi (2013) first partialled out the impact of temperature on run scoring, which increases by about 0.2 runs per 10 degree difference, so as to equalize its impact across the season. After doing that, Max calculated that run scoring went down about 0.60 runs between the first and sixtieth games of the season, implying that offense were indeed ahead of defense. Just to make sure it was pitchers who were behind, Max examined Defensive Efficiency Record (the percentage of balls in play on which a team successfully makes a play; see the Fielding Evaluation chapter) and noted no large difference among months.

Maynard, M. Travis, Christian J. Resick, Quinn W. Cunningham, and Marco S. Di Renzo (2017). Ch-ch-ch changes: How action phase functional leadership, team human capital, and interim vs. permanent leader status impact post-transition team performance. *Journal of Business and Psychology*, Vol. 32, pages 575-593.

Maynard, Resick, Cunningham, and Di Renzo (2017) examined 129 in-season managerial changes between 1974 and 2008, and noted that team performance improved after the change; which of course just means that mid-season managerial changes usually occur when a team is going through a particularly bad stretch, and the new manager benefits from regression to the mean. The authors seemed to realize this to an extent, noting that the relevant teams were bad to begin with and continued to display losing records after the change. The authors also noted that player performance

improvement was (of course) responsible for the improvement, and particularly when the newly-installed managers made more pitching changes during the games. These impacts were a bit stronger when the new manager was designated as permanent rather than interim. Retrosheet data was apparently used in compiling team winning percentage before and after the managerial change.

McCotter, Trent (2008). Hitting streaks don't obey your rules. *Baseball Research Journal*, Vol. 37, pages 62-70.

Pavitt, Charlie (2009). Hitting streaks and psychology. *Baseball Research Journal*, Vol. 38 No. 1, pages 6-7.

McCotter, Trent (2009). Reply. *Baseball Research Journal*, Vol. 38 No. 1, pages 7-8.

Albert, Jim (2008). Long streaks. *Baseball by the Numbers*, Vol. 18 No. 4, pages 9-13.

Albert, Jim (2010). Great streaks. *Baseball Research Journal*, Vol. 39 No. 2, pages 58-62 and 64.

McCotter, Trent (2009). Reply. *Baseball Research Journal*, Vol. 38 No. 1, pages 7-8.

McCotter, Trent (2010). Hitting streaks don't obey your rules. *Chance*, Vol. 23 No. 4, pages 52-57.

Some work by Trent McCotter has continued the debate concerning the reality of hitting streaks. McCotter's method was as follows: Using Retrosheet data from 1957 through 2006, he recorded the number and length of all batting streaks starting with one game along with the total number of games with and without hits in them. He then compared the number of streaks of different lengths to what occurred in ten thousand random simulated permutations of the games with/without hits in them. There was a consistent and highly statistically significant pattern across all lengths starting at five for more real-life streaks than in the simulations. Trent concluded that hitting streaks are not random occurrences.

Although nobody challenged Trent's analysis as such, there has been some criticism of other aspects of his work. His first attempts at explaining these patterns (batters facing long stretches of subpar pitching or playing in a good hitting ballpark, and streaks occurring more often in the warmer months) were proposed, found no evidence for the first, and claimed the second and third to be unlikely, but never empirically evaluated (although all could be). He instead opted for untestable speculations concerning a change in batter strategy toward single hitting and just the existence of a hot hand. I called him on these, and he responded with helpful analyses inconsistent with the second and third of the testable explanations and basically punted on the untestable ones. Jim Albert (2008) lauded the method and replicated it, but this time restricting the sample to five seasons of Retrosheet data studied separately (2004 through 2008). Again, real streaks occurred more often than in the random permutations, but only three out of twenty comparisons (for 5 or more, 10 or more, 15 or more, and 20 or more, for each of the five seasons) were significant at .05 and a fourth at .10, leading Jim to question the practical significance of Trent's results. This initiated a debate in the *Baseball Research Journal* Volume 39 Number 2, in which Jim questioned the practical significance of Trent's findings giving the huge sample size Trent used, Trent defended the huge sample size as necessary to tease out streaks buried in noisy data, and Jim challenged and Trent upheld Trent's use of the normal

distribution as the basis for comparison. A later paper (McCotter, 2010) added nothing substantive to the debate.

Mejdal, Sig (2000). The recipe for a stolen base. *By the Numbers*, Vol. 10 No. 3, pages 20-22.

Loughlin, Thomas M. and Jason L. Bargen (2008). Assessing pitcher and catcher influences on base stealing in Major League Baseball. *Journal of Sports Sciences*, Vol. 26 No. 1, pages 15-20.

Given the steal attempt, what are the factors that determine its odds of success? Sig Mejdal (2000) made a nice attempt at answering this question. Mejdal began with the reasonable premise that the possibilities include the baserunner's speed, catcher's throwing ability, speed of pitcher's delivery, umpire play-judgment tendencies, and the stadium surface (turf is easier to run on than grass). One confound is between catcher and pitcher, as a particularly good or poor throwing catcher would make it appear that the pitchers he works with are better or worse than average, whereas a staff populated by pitchers particularly quick or slow at delivering the ball to the plate would make it seem that their catcher is better or worse than average. Thus it looks as if the probability of successful stolen bases against particular catchers and the probability against certain pitchers are seriously dependent on one another. However, using three years of Retrosheet data, Mejdal found that an attempt to correct the catcher's successful steal percentage by adjusting it by the average percentage of pitchers teamed up did not lead to significantly different numbers than merely computing the catcher's percentage across those years, so he used the simpler measure. Mejdal then corrected the pitcher's percentage by computing the percentage for all the catchers they have worked with, comparing the two percentages, and then using the difference between the two to represent the pitcher. To use his example, if pitcher Joe Schmo was paired up with catchers that averaged a 60 percent steal rate and his own steal rate was 40 percent, then Mejdal credited Joe with a 20 percent "stolen base value." Mejdal's method, in essence, given precedence to the catcher by presuming that his successful steal percentage, when taken over a long enough time frame, is a valid measure of ability, and that pitcher's percentage should be determined within their catchers' context.

Mejdal then entered measures for the relevant factors into a multiple regression equation predicting successful steal rate. Unfortunately, he failed to provide data on the overall predictive power of the five factors. Of that variance in successful steal percentage that was accounted for by the equation, 36 percent was attributed to the baserunner, 34 percent to the pitcher, 19 percent to the catcher, 11 percent to the surface, and absolutely none to the umpire. It is particularly interesting that the pitcher was found to be almost twice as influential as the catcher, as the correction described above in a sense gave the catcher a greater "opportunity" to influence the results.

Using Retrosheet data from 1978 through 1990, Loughlin and Bargen (2008) demonstrated that differences in catchers' ability to control the "running game," as measured by success steals divided by attempts, and of pitchers' ability to hold runners, as measured by attempted steals divided by opportunities, are statistically significant; which they claim nobody had done previously. The variation among pitchers was greater than that for catchers, which is consistent with Mejdal's division of responsibility

just mentioned.

Menéndez, Héctor D., Miguel Vázquez and David Camacho (2015). Mixed clustering methods to forecast baseball trends. In David Camacho, Lars Braubach, Salvatore Ventcinque and Costin Badica (Eds.), *Intelligent Distributed Computing VIII* (pages 175-184). Heidelberg, Germany: Springer.

Soto Valero, C. (2016). Predicting win-loss outcomes in MLB regular season games – A comparative study using data mining methods. *International Journal of Computer Science in Sport*, Vol. 15 No. 2, pages 91-112.

Menéndez, Vázquez and Camacho (2015) and Soto Valero (2016) used Retrosheet data in methodological studies attempting to predict the outcome of games; neither have substantive import.

Mills, Brian M. (2017). Policy changes in major league baseball: Improved agent behavior and ancillary productivity outcomes. *Economic Inquiry*, Vol. 55 No. 2, pages 1104-1118.

Using PITCHf/x data, Mills (2017) concluded that the average strike zone as called by umps had expanded on the bottom by three inches between 2008 and 2014, resulting in three times as many called strikes in the zone between 18 and 21 inches off the ground. Both pitcher and batters appear to have noticed the change, with the proportion of pitches in that zone increasing from about 22 percent to about 27½ percent, and swings on pitches in that zone from about 31 percent to about 34½ percent. This change favors the pitchers, as when a batter swings at pitches in that zone, the odds of making contact are 73%, putting a ball in play 48%, and getting a hit 26% lower than for pitches above it. Using Retrosheet data, Mills noted a relationship between this change and run production per game over this interim.

Morey, Leslie C. and Mark A. Cohen (2015). Bias in the log5 estimation of outcome of batter/pitcher matchups, and an alternative. *Journal of Sports Analytics*, Vol. 1 No. 1, pages 65-76.

Morey and Cohen (2015) argued that applying the log5 method to batter/pitcher matchups may result in biased findings because the method presumes a mean probability of .500, which will occur across teams but not for batting indices. Simulations for the 1996 through 2013 seasons based on data downloaded from Retrosheet and Lahman's database resulted in BA (around .300) and HR (almost 8 per 100 ABs) consistently too high, with the bias more pronounced as true performance becomes more extreme. The first author's alternative method is better, although in this case producing underestimates.

Palmer, Pete (2014). Stolen bases and caught stealing by catchers: Updating Total Player Rating. *Baseball Research Journal*, Vol. 43 No. 1, pages 23-25.

Thanks to historical information that became available thanks to Retrosheet, Pete

has been able to add stolen base/caught stealing data to TPR for catchers; incidentally, his list of the top 20 all-time in controlling the running game is consistent with catchers' reputations, with Ivan Rodriguez leading the pack.

Palmer, Pete (2017). Intentional walks revisited. *By the Numbers*, Vol. 27 No. 1, pages 16-25.

Pete Palmer's well known run-value figures, popularized in *The Hidden Game of Baseball*, were, due to absence of sufficiently-detailed, estimated with the presumption that the likelihood of all relevant events is independent of base-out situation. In 2017, Pete used 1946-2015 Retrosheet data to determine the actual run values of the following events:

Single .453	Home Run 1.413	Unintentional Walk about .31
Double .752	Out -.241	Intentional Walk .157
Triple 1.038		

The figure for unintentional walks is an estimate, as Pete actually provided a combined value of .298. The reason that intentional walks are so much lower than unintentional is that the former tend to occur in circumstances in which their impact of runs is less, particularly with runners on second, third, or both those bases, occurring in more than two percent of relevant cases (the highest is 2nd and 3rd with one out; more than 12 percent). IBBs are given in fewer than one percent in all other circumstances.

Pete also examined the IBB as a strategic tool. With the exception of when designated hitters are available, the IBB has been most often used for the #8 batter due to the weakness of the upcoming #9. Even so, it usually works in favor of the team at bat. For example, with runners on second and third and two out, it has historically decreased expected runs by .033 for that inning but increased it by .113 for the following inning, given that the #1 position is then likely to lead off in that next inning. Walking a stronger batter to face a weaker one is also usually a loser for the defensive team, as the next batter must be considerably weaker (e.g., at least a bit below average when the batter that is walked is among the upper one-sixth in performance) to be worth the tradeoff. And walking a batter to get the platoon advantage is also not worth it, as the advantage normally does not offset the value of the extra baserunner.

Palmer, Pete (2018). Relief pitching strategy: Past, present, and future? *Baseball Research Journal*, Vol 47 No. 1, pages 45-52.

Pete Palmer (2018) offered a far-reaching discussion of some of the implications that the growth in relief pitcher usage has implied for the game. Beginning with that growth., Pete calculated that the percentage of time in which a team's save leader entering the game with their team ahead but with win probability percentages of less than 50 percent due to the base-out-inning situation, has plummeted from 23 percent during the 1980s to 10 percent during the 1990s to 4.7 percent during the 2000s up to 2017. That is because only about 3 percent of them occur in the ninth inning, which has more and more become the only time the save leader (a better term than closer given previous usage patterns) appears.

In evidence relevant to the myth of the proven closer, since 1961 the difference in

save percentage for a team's save leader versus other pitchers has increased, but is not as large as some might think. In the 1960s, the difference was about 4 percent with a one-run lead in the ninth inning; by the 2000-2017 interim it had increased to about 9 percent. Yet, and this is critical, even now the success rate of non-closers with a one-run lead in the 9th was more than 76 percent for visiting teams and more than 83 percent for home teams. These jump to about 89 percent for visitors and 92 percent for home with a two run lead, and over 95 percent with a three run lead, with corresponding decreases in the disadvantage they have to closers in this regard.

The increase in number of pitchers per team is of course linked with the decrease in the number of position players on the 25 man roster. This has restricted the number of substitutions managers can make with the matter. In the 1960s, there were an average of 233 fielding substitutions, 211 pinch-hitters, and 40 pinch-runners per team per season; between 2011 and 2017, these figures had dropped to 197, 183, and 28 respectively. Platooning has also dropped. Defined as a circumstance in which, for a position, a team has at least 70 starts by a lefty hitter and 30 starts by a righty hitters against opposite handed starting pitchers. Using Retrosheet data, Pete noted that platooning was almost non-existent at the beginning of the 20th century, the proportion of platooned positions had increased to about 20 percent from 1958 to 1990, but was down to about 14 percent by 2017.

Palmer, Pete (2021). Career park effects for individual players. *By The Numbers*, Vol. 30 No. 1, pages 9-13.

Pete Palmer (2021), using Retrosheet data, computed individual player OPS values for home and away games, divided the former by the latter, and then multiplied by 100, in so doing producing a career park effect figure for each player. These what I will call OPS park ratios (combining two labels Pete proposed) need to be distinguished from Pete's general park factors as they measure individual player/ballpark fit. Not surprisingly, Rockies players dominate the top ten., with Charlie Blackmon (134) the highest ever at the time of Pete's work. Nonetheless, while dominating at home (1.054 and 1.072), Larry Walker and Todd Helton's respective road figures (.857 and .867) show that they excelled everywhere. Gil McDougald's 80 (OPS of .680 home and .847 road) was the lowest of anyone with 3500 at bats by an astounding eight points, making him the player with the worst ever ballpark fit and demonstrating that his honors (five-time All-Star and five-time recipient of MVP votes) were deserved.

Panas, Lee (2010). *Beyond Batting Average*. self-published, available at lulu.com

This book is a summary of sabermetric research, concentrating on player evaluation measures but short on material relevant to strategy. Panas used data from several sources; from Retrosheet, he computed a run expectancy chart for 2005-2008 (Chapter 5, Linear Weights), some RBI percentage rankings (Chapter 6, Situational Hitting), an example for a measure of baserunning performance (Chapter 7, Baserunning), and figures on BA and SA on batted ball type (Chapter 9, Fielding Independent Pitching).

Patt, Emily-Anne and James Stockton (2024). Noisy judgments: A probability surface-based analysis of umpire variability. MIT Sloan Sports Analytics Conference.

The authors used 5,307,386 pitches called by umpires between 2008 and 2022, with raw data from Statcast and umpire game assignments from Retrosheet, to formulate individual called strike zone (CSZ) for umpires constructed as probability distributions for calls based on specific pitch location. With that data, they determined that, over the entire time span, umpire accuracy for varied from 70 to 90 percent, averaging 84 percent, but most around 84 percent; and accuracy (as we already know) has steadily increased over time. Also not news but good to see reiterated, during that span the CSZ narrowed from 63 to 54 centimeters (24.8 to 21¼ inches, with inside pitches toward righty batters responsible for most of it; the top of the CSZ went up 2 centimeters (¾ inch), whereas the bottom went down 9 centimeters (3½ inches; this is not the place to describe in detail, but the latter in particular is one of the causes of the recent surge in strikeouts). Other findings worth noting: Batters and pitchers have individual CSZ's independent of their height; there is now more evidence supporting both the effect of pitch framing (Jose Molina still rules) and one-half of the strongest known bias in umpire calls, the tendency to call borderline pitches as balls with two strike counts (they did not examine the corresponding tendency for strike calls with three-ball counts).

Pemstein, Jonah, and Sean Dolinar (2015). A new way to look at sample size.
<https://blogs.fangraphs.com/a-new-way-to-look-at-sample-size/>

Using 2009 through 2014 Retrosheet data, Jonah Pemstein and Sean Dolinar (2015) include graphs and lists showing Cronbach's alphas for every tenth PA from 10 through EITHER 400, 500 or 600 for 14 metrics. The 600's are
1B% - 0.73, 2B% - 0.36, 3B% - 0.57, HR% - 0.80, BB% - 0.85, HBP% - 0.62, K% - 0.92, wOBA - 0.61
The following go up to 500: BA - 0.48, ISO - 0.76, OBA - 0.60, SLG - 0.63,
The following go up to 400: BABIP - 0.45, BABIPcon - 0.46

Petti, Bill (2014). The value of (in)consistent play in major league baseball.
<https://tbt.fangraphs.com/the-value-of-inconsistent-play-in-major-league-baseball/>

Past research suggested that teams that were more consistent in scoring and less consistent with giving up runs from game to game were more successful than their opposites, all else being equal. Bill Petti examined this issue in greater detail. First, he used Retrosheet game outcomes from 1971 through 2012 to calculate Gini coefficients for runs scored and allowed per team per season. (The Gini coefficient is used by economists to measure variation in a data set, with 0 indicating no variation and 1 the upper limit. It has been used extensively in the study of changes in MLB competitive balance over time.) Gini coefficients for runs scored correlated with runs scored per game at -0.59 and those for runs allowed correlated with runs allowed per game at -

0.64, implying that teams scoring and giving up more runs tended to be more consistent at both from game to game. In addition, he noted actual winning average to correlate at -0.43 between the coefficient for runs scored and $+0.49$ with that for runs allowed, implying once again that consistent run scoring and inconsistent run allowing is related to team winning.

Bill then computed Pythagpat (a variation of the Pythagorean formula) winning average estimates and found that including Gini coefficients in regression equations improved predictions of team winning average over using Pythagpat alone, although by less than 1 percent. However, the Gini coefficients were more successful than Pythagpat at predicting wins over and under the Pythagpat prediction, which allowed him to estimate that optimal consistency in preventing and scoring runs would amount to a two-win advantage over average consistency, controlling for runs scored and allowed.

Phillips, David C. (2011). You're hurting my game: Lineup protection and injuries in major league baseball. *Journal of Quantitative Analysis in Sports*, Vol. 7 Issue 3, Article 7.

Phillips (2011) performed the most thoughtful study of protection to date, with results analogous with other studies. He realized that a study of protection based on player movement within a batting order (e.g., moving a cold hitter to a different spot in the lineup) leads to ambiguous findings, because any change in the performance of that hitter could be due to the change in subsequent batter or to random changes in that player's performance irrelevant to who is batting behind. In response, Phillips looked at differences in performance for a given player remaining in the same lineup position based on changes in the next batter caused by injury. Based on Retrosheet data from 2002 through 2009 and limited to protectors with an OPS of at least .700 for a minimum of 200 plate appearances (in other words, hitters good enough to count as potential protectors), Phillips noted that injuries to protectors resulted in an overall OPS decrease of 28 points at that lineup position due to a weaker replacement. With the weaker replacement, the hitter being protected tended to receive a lot more intentional walks but fewer extra base hits (but no more hits, as additional singles compensating), indicative of the expectation that a non-protected hitter will be pitched around more often. These two tendencies pretty much cancelled one another out, resulting in little overall protection effect.

Phillips, David C. (2017). Stopping on nine: Evidence of heuristic managerial decision-making in major league baseball pitcher substitutions. *Southern Economic Journal*, Vol. 84 No. 2, pages 577-599.

Phillips (2017) examined 1992-2012 Retrosheet data to see if there has been a tendency to remove starting pitchers before their pitch count crosses a number that ends in zero. Although any such tendency was weak in the first decade of the study, there was a two percent increase in the number of times relief pitchers entered when the starter reached a pitch count ending in nine. However, the bias was weaker the closer the game score, implying that managers are less concerned with pitch counts

and more with immediate strategy in those games. Finally, the bulk of the bias was for pitchers in their first three seasons, showing that managers were more concerned with in protecting the arms of the relatively young. An additional tidbit was that 80 percent of starter pitch counts were between 69 and 125 in 1992 and 78 and 114 in 2012, evidence that managers were concerned with protecting both starters (decrease in the higher number) and the bullpen (increase in the lower number, meaning fewer innings for relievers) from overwork.

Pinheiro, Ryan, and Stefan Szymanski (2022). All runs are created equal: Labor market efficiency in major league baseball. *Journal of Sports Economics*, Vol. 23 No. 8, pages 1046-1075.

Here is a season-by-season run expectancy matrix for 1996-2015 (apologies that it is not lined up correctly):

Season	Walk	Single	Double	Triple	Home	Run	Out
1996	0.331	0.485	0.784	1.105	1.403	-0.302	
1997	0.307	0.465	0.761	1.083	1.393	-0.284	
1998	0.312	0.469	0.780	1.014	1.400	-0.285	
1999	0.311	0.477	0.789	1.059	1.408	-0.302	
2000	0.332	0.482	0.765	1.085	1.406	-0.307	
2001	0.298	0.460	0.778	1.084	1.380	-0.283	
2002	0.303	0.466	0.755	1.052	1.398	-0.279	
2003	0.307	0.466	0.775	1.080	1.391	-0.284	
2004	0.307	0.462	0.786	1.041	1.396	-0.287	
2005	0.295	0.458	0.768	1.056	1.412	-0.277	
2006	0.317	0.467	0.766	1.070	1.389	-0.290	
2007	0.310	0.468	0.798	1.044	1.406	-0.289	
2008	0.312	0.460	0.772	1.081	1.405	-0.281	
2009	0.304	0.459	0.762	1.004	1.392	-0.278	
2010	0.299	0.451	0.763	1.076	1.404	-0.266	
2011	0.289	0.442	0.736	1.064	1.392	-0.255	
2012	0.284	0.441	0.747	1.039	1.396	-0.257	
2013	0.285	0.439	0.740	1.035	1.371	-0.250	
2014	0.283	0.437	0.739	1.054	1.400	-0.245	
2015	0.303	0.442	0.743	1.031	1.386	-0.257	
mean	0.304	0.459	0.764	1.056	1.396	-0.277	
standard deviation	0.014	0.014	0.018	0.027	0.010	0.018	

Note the year-to-year stability. The authors then ran regressions showing that these consistently accounted for 93 to 94 percent of variance in each team's runs scored for those seasons. Finally, they used the run expectancy data to calculate run value for individual non-pitchers with at least 130 AB, and then related the individual values to salaries. Salary were roughly consistent with run values, with a slight improvement in

2005-2016 over 1996-2004 due to better valuing walks following (in their opinion) the publication of *Moneyball*.

Pinto, David (2007). Analyzing the umpires.

<https://www.baseballprospectus.com/news/article/6533/the-big-picture-analyzing-the-umpires/>

Using the Retrosheet record, David Pinto (2007) estimated the winning average of the team with the better record against the team with a worse record in two-team matchups using the following formula:

$$\frac{\text{winning average of better team} \times (1 - \text{winning average of worse team})}{(\text{winning average of better team} \times [1 - \text{winning average of worse team}]) + (\text{winning average of worse team} \times [1 - \text{winning average of better team}])}$$

and then compared these results with the records of individual umpires in such matchups from 2000 to 2006. As one would expect, there was a distribution of umpires such that the better team consistently won more often than the formula would predict for some and less often for others. However, there was nothing apparent in the data to suggest that any of this was intentional influence on game outcomes. In addition, relevant to the home field advantage, David noted that the overall estimate for the better team during these seasons was a winning average of .587, analogous to a 95-win season, but the actual home team record for the better team was .623, or 101 wins.

plen (2010). The leadoff walk. <https://community.fangraphs.com/the-leadoff-walk/>

Somebody calling themselves plen used 1952-2009 data to examine the odds of leadoff hitters scoring when they get on base. In order of raw number of occurrences, these were: singles 37.69 percent, walks 37.9 percent, hit by pitches 38.77 percent, errors 37.74 percent, strike three pitches getting past the catcher 37.24 percent, and catcher's interference 34.84 percent (with the latter occurring only 155 times, leaving the possibility that the lower figure is due to random variation).

Poling, Alan, Marc A. Weeden, Ryan Redner and T. Mary Foster. (2011). Switch hitting in baseball: Apparent rule following, not matching. *Journal of the Experimental Analysis of Behavior*, Vol. 96 No. 2, pages 283-289.

Poling, Weeden, Redner, and Foster (2011), looking at play-by-play data from Retrosheet via Baseball Reference, were apparently experimental psychologists of the behavioristic school. They wrote as if they were surprised to discover that switchhitters Mickey Mantle, Eddie Murray, and Pete Rose's "apparently chose handedness based on the rule 'bat opposite the pitcher,' not on differential consequences obtained in major league games." As this was inconsistent with previous data about the impact of reinforcement of past success/failure seen in basketball shot selection and American football play selection in specific and human behavior in general, they called for more

research into the variables that affect behavioral choice. Methinks that they really were not surprised by their findings.

Pope, Devin and Uri Simonsohn (2011). Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychological Science*, Vol. 22 No. 1, pages 71-79.

Hitting .300 is a goal for many hitters, and Pope and Simonsohn (2011) believed that the desire to do so can serve as motivation for hitters very close to that mark with a game or two left in the season to perform particularly well in those last couple of games. Examining Retrosheet data from 1975 through 2008 for all hitters with 200 or more at bats in a season (comprising a sample size of 8817), the authors showed that a higher proportion of players hitting .298 or .299 got a hit on their last plate appearance (.352) than players hitting .300 or .301 (.224). They were also, however, less likely to be replaced by a pinchhitter (.041 versus .197). The latter leads to an obvious bias; that hitters just over the .300 benchmark have less of an opportunity to drop under than hitters just under to move over it. Scott and Birnbaum (2010) demonstrate that a statistical correction for this bias removes this last at bat advantage, and in fact there is “nothing unusual about the performance of players on the cusp of .300” (page 3).

Powers. Scott, Trevor Hastie, and Robert Tibshirani (2018). Nuclear penalized multinomial regression application to predicting at bat outcomes in baseball. *Statistical Modeling*, Vol. 18 Nos. 4-5, pages 388-410.

Powers, Hastie, and Tibshirani (2018) presented a model for predicting the outcomes of specific plate appearances using 2015 Retrosheet data. All batters with at least 390 PA and all pitchers with at least 360 PA against included individually; the data for the rest combined with positions into an abstract “replacement level” player. The predictors were batter and pitcher tendencies, their handedness match or mismatch, the ballpark, and the home-field advantage. The relevant categories were strikeouts, walks, hit by pitches, homers, triples, doubles, singles, groundouts and flyouts. The model was designed to take advantage of the associations existing between these categories, which were computed using principal components factor analysis. For example, above average singles hitters also tended to ground out more than average, analogously for homers and strikeouts, and those flying out a lot tended to not ground out a lot.

In addition, the principal components analysis allowed the authors to present both trilogies of factors for describing batter skills. The first factor included negative loadings for strikeouts, walks, hit by pitches, and homers, and positive loadings for fly and ground outs, singles, doubles, and triples. Most of these loadings were very small, but nonetheless Three True Outcome type hitters were clearly being distinguished from contact hitters. The second factor included positive loadings for fly outs and homers and negative loadings for ground outs and singles. Implying a distinction based on vertical angle of batted ball, The third factor features a very strong negative loading for singles and a very strong positive loading for ground outs; and as these two were positively associated in the previous two factors, this seems to differentiate non-power

hitters with high and low batting averages.

Powers et al. did the same for pitchers, with the first two factors reflecting well-established distinctions. The first factor included a strong negative loading for strikeouts and positive loadings for singles, ground outs, and fly outs, clearly distinguishing strikeout from pitch-to-contact pitchers. The second factor featured a strong negative loading for fly outs and positive loadings for ground outs and singles. The third was not as clear cut, as ground outs and fly outs (and also homers) loaded positively and walks (but also singles) negatively, signaling some division between giving up walks versus batted balls.

Rockoff, David M. and Philip A. Yates (2009). Chasing DiMaggio: Streaks in simulated seasons using non-consistent at-bats. *Journal of Quantitative Analysis in Sports*, Vol. 5 Issue 2, Article 4.

Rockoff, David, and Philip Yates (2011). Joe DiMaggio done it again...and again and again and again? *Chance*, Vol. 24 No. 1, pages 14-18.

There have been numerous attempts to estimate the odds of a 56 game hitting streak, and in my opinion Rockoff and Yates (2008) is the best of all these attempts. Their idea was to simulate 1000 seasons of play using actual seasonal game-to-game performance for each of 58 years of Retrosheet data. Out of the 58,000 simulated seasons, a total of 30 (about .005%) included a hitting streak of 56 or more games. Interestingly, Ichiro's 2004 season included 5 of them. Using this data, the authors concluded that the odds of a streak of more than 56 games in any of the 58 seasons in the data set was about 2½ percent. In a follow-up (Rockoff & Yates, 2011), they performed 1000 simulated "baseball histories" under a number of different assumptions: the odds of a hit directly determined by player batting average, including the odds of a hit determined by a varying amount centered around the player batting average, and the odds of a hit partly determined by overall batting average but also by performance in 15 and 30 game stretches around each game under question. The latter two methods assume the existence of hot and cold streaks, which I think is an error. This is because, as will be described later in this chapter, the very existence of such streaks as anything other than the results of random processes is questionable. Part of the point of examining this topic in the first place should be to address whether hitting streaks are or not random, and so to presuppose that they are not leads to an invalid bias in favor of long streaks. As a consequence, the author(s) uncovered 85 56-game or greater streaks using the "batting average" approach, 88 using the "variation around batting average" approach, 561 using the "15 game" approach, and 432 using the "30 game approach." I only consider the first two to be defensible. To make this point more clearly, the simulated Joe DiMaggio equaled or bettered his real streak once using each of the two methods and twice using an "equal at bats" approach, but four and nine times respectively for the latter two methods. Anyway, Rockoff and Yates estimated that allowing streaks to carry over across two seasons would increase the overall number by about ten percent.

Roher, David (2007). Quantifying the impact of opponent quality. *By The Numbers*, Vol.

Does good pitching stop good hitting? Earlier work by Dan Levitt and Tom Hanrahan suggests not, but rather implies that good pitching is better than bad pitching at stopping good hitting, and good hitting is better than bad hitting at overcoming good pitching, but nothing more. However, they worked with aggregated data, which could mask relationships which only come to light when variation among player tendencies are considered. Happily, David Roher (2007) took this on. Using Retrosheet data from 2006, David calculated the relative value of each event for run production, measured pitcher quality by Fair Run Average and batter quality through Equivalent Average, and used those to measure the impact of opponent quality on both batter and pitcher performance. The result, which he called Opponent Quality Effect, had a good deal of variation across players – in other words, a big difference among players in how much their performance was affected by opponent quality – but absolutely no relationship with measures of pitching and batting performance. The conclusion is then the same as that from Dan and Tom's work.

Roley, Ross (2007). Runner's reluctance – part one.

http://baseballanalysts.com/archives/2007/11/runners_relucta.php

2007 Retrosheet data on runner advancement.

Less Than 2 Outs	Chances	Attempted Advance	Out Trying	Success Rate
1st to 3rd on single	1079	305	6	.98
2nd to home on single	1039	700	22	.97
1st to home on double	469	243	13	.95
1st to 2nd on flyball	1385	23	2	.91
2nd to 3rd on flyball	1008	360	11	.97
3rd to home on flyball	689	599	15	.97
Total	5669	2230	69	.97

2 Outs	Chances	Attempted Advance	Out Trying	Success Rate
1st to 3rd on single	699	262	3	.99
2nd to home on single	813	784	33	.96
1st to home on double	304	225	8	.96
Total	1816	1271	44	.97

Roley computed break-even points for each of these for 0, 1, and 2 outs, and of course in each case they were well below success rate. All were on graphs! see the original for details. In a second part

(http://baseballanalysts.com/archives/2007/11/runners_relucta_1.php)

Roley provided an estimate of net runs gained for different success rates over that break-even for sending runners home on hits to centerfielder for each 75 runners sent, approximately 1 percent of all the opportunities per season, while assuming an overall break-even point of 71 percent.

Rosciam, Chuck (2004). Professional thieves vs. the constabulary. *Baseball Research Journal*, No. 33, pages 81-83.

Based on a larger data set than analogous efforts (1963 and 1965-1968 A. L. games and 1969 to 2004 games for both leagues), Chuck provided stolen base success rates of 73.1 percent for second base, 71.6 percent for third base, and 37.4 percent for home. He also presented detailed indices for the most prolific base stealers and the catchers most successful at thwarting them for that period of time.

Ruane, Tom (1999). Stolen base strategies revisited. *By The Numbers*, Vol. 9 No. 1, pages 24-28.

Tom Ruane (1999), using raw game data for 1980 to 1989 compiled by Project Scoresheet and Retrosheet, found specifically for runner on first stolen base breakeven points of 70.9 percent success rate with no out, 70.4 percent for one out, and 67.1 percent for two outs. Tom also computed both run potential and probability of scoring both when a steal was and was not attempted from first on the next play, with the following differences:

	Run Potential			Odds of Scoring		
Outs	0	1	2	0	1	2
All runners	-.005	-.014	+.031	+.053	+.031	+.043
Fast runners	-.014	-.045	+.030	+.060	+.018	+.047

For example, looking at the first row, attempted steals from first lower run potential 1.4 percent with one out but raise it 3.1 percent with two outs. Trying to stealing second does increase the odds of scoring in all situations. The overall point, however, is how small these differences are. Interestingly enough, the speed of the base stealer has little impact. Using an informal method devised by Bill James (1987) for classifying base runner speed called Speed Scores, Tom Ruane computed the analogous figures only for the fastest runners (second row) and discovered them to be almost the same.

Ruane, Tom (2005). In search of clutch hitting. *Baseball Research Journal*, No. 34, pages 29-36.

In this study, which is also posted on the Retrosheet research page, Tom examined the difference between batting performance with runners in scoring position versus not, using Retrosheet data from 1960 through 2004 for all batters with at least 3000 career at bats during that interim. Based on each player's performance with runners on second and/or third versus not, Tom noted the difference between simulated

and actual outcomes and uncovered no systematic differences in the distribution of those differences across all of the players. As a methodological note, Tom thought to take all walks and sacrifice flies out of the data set, because the former is very dependent on base-out situation (much more likely with runners in scoring position but first base unoccupied) and the latter biases batting average with runners in scoring position (i.e., they do not count as at bats). Tom found that batters averaged 7 points higher in batting and 15 in slugging with no runners in scoring position, which is likely more accurate than earlier studies that failed to include these corrections.

Ruane, Tom (2005). Do some batters reach on errors more than others? *Baseball Research Journal*, No. 34, pages 113-120.

Replicating earlier work by Clifford Blau, Bill James, and Mark Pankin using Retrosheet data to analyze batters who made at least 2000 outs between 1960 and 2004, Tom noted that batters that get on base due to errors tend not surprisingly to be faster (causing the fielder to hurry and perhaps get careless), ground ball hitters (grounders result in more errors than flies) and righthanded hitters (more errors on grounders to the left side of the infield, probably due to the longer and more hurried throw). The effects are small, with the lefty/righty difference only at 3/10 or 4/10 of 1 percent and speed effect in the same range. This research is also available at the Retrosheet research page.

Ruane, Tom (n.d.). Strikeouts, grounders and fly balls. Retrieved from <http://www.baseballthinkfactory.org/btf/scholars/ruane/articles/goodout.htm>

Tom's analysis, based on 1982, 1983, and 1987 Retrosheet data, showed that the expected loss in runs during an inning from strikeouts was greater than that for flyouts and, in particular, groundouts, and that the difference among the three increases as the hitter becomes faster as measured by Bill James's "speed score" metric:

	Strikeouts	Fly outs	Ground outs
Slow runners	-.278	-.261	-.262
Average runners	-.276	-.257	-.244
Fast runners	-.268	-.254	-.230

These data imply that, relative to strikeouts and fly outs, the positive value of moving up baserunners when making outs through hitting the ball on the ground outweigh the negative value of hitting into double plays. These overall numbers mask huge situational differences. I illustrate with two extremes for average speed runners hitting with one out: With only a runner on first, the type of out barely mattered (strikeout, -.305; flyout -.303; groundout, -.327), although here double plays do slightly trump moving the runner up. With runners in second and third, it makes all the difference in the world (strikeout, -.825; fly out, -.438; ground out, -.302), reflecting in particular the possibility of the runner on third scoring with any batted ball and additionally the runner on second moving to third on a groundout.