

Moneyball: The Computational Turn in Professional Sports Management

by Eric S. Hintz, PhD (hintze@si.edu)

Historian, Lemelson Center for the Study of Invention and Innovation

National Museum of American History, Smithsonian Institution

Professional sports teams have become very big businesses. The most valuable clubs of 2019—the Dallas Cowboys (\$5.0 billion), the New York Yankees (\$4.6 billion), and Real Madrid (\$4.2 billion)—had valuations comparable to Fortune 500 firms such as JetBlue Airlines (\$5.0 billion) and Goodyear Tire & Rubber (\$4.2 billion).¹ They are late adopters, but these teams have recently joined other big businesses in deploying the techniques of “scientific management,” “operations research,” and “decision sciences” to gain a competitive advantage.²

Sports analytics encompasses a set of data management technologies and computational techniques for gathering and interpreting observable statistical data about athletes and game play. Team executives use statistical analysis to evaluate players for drafting, trades, and contract negotiations. Coaches use analytics to understand competitors’ tendencies, to develop in-game strategies, and to identify areas for player improvement.³

Many casual observers have become aware of “big data” and computational analytics through their applications in professional sports. Most notably, Michael Lewis’s bestselling book *Moneyball* (2003)—and the 2011 film adaptation starring Brad Pitt—described how Oakland Athletics general manager Billy Beane deployed analytical techniques to find high-performing but undervalued baseball players who helped his low-payroll team compete successfully against wealthier clubs.⁴ *Moneyball*’s publication signaled a growing enthusiasm for sports analytics that

**This material is based upon work supported by the National Science Foundation under Grant DRL-2005404. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

¹ Kurt Badenhausen, “The World’s 50 Most Valuable Sports Teams 2019,” *Forbes*, 22 July 2019, <https://www.forbes.com/sites/kurtbadenhausen/2019/07/22/the-worlds-50-most-valuable-sports-teams-2019/#39187bf9283d>; “Fortune 500,” *Fortune*, 2019, with market valuations as of 29 March 2019, <https://fortune.com/fortune500/2019/>.

² Frederick Winslow Taylor, *The Principles of Scientific Management* (New York: Harper & Bros., 1911); Daniel Nelson, *Frederick W. Taylor and the Rise of Scientific Management* (Madison, WI: University of Wisconsin Press, 1980); Saul I. Gass and Arjang A. Assad, *An Annotated Timeline of Operations Research: An Informal History* (New York: Kluwer Academic Publishers, 2005).

³ Benjamin C. Alamar, *Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers* (New York: Columbia University Press, 2013).

⁴ Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (New York: W. W. Norton, 2003). For a critique of certain artistic licenses in the book and film, see Benjamin Baumer and Andrew Zimbalist, *The Sabermetric*

was already in train, as several sports teams were already establishing analytics departments and hiring statisticians.

In the mid-2000s, a handful of practitioners attempted to document the explosive growth of sports analytics and place it in historical context.⁵ A few scholars have also traced the historical evolution and current statistical practices within individual sports.⁶ However, as Andrew Baerg noted in 2013, “a broader history of sports analytics remains yet to be written.”⁷ Indeed, historians of business, technology and computing have rarely examined sports as an important domain for data processing and quantitative analysis, despite the fact that analytics has gained wide notoriety largely because of its applications in sports.⁸

Drawing on published primary sources and contemporary news coverage, this paper traces the evolution and gradual professionalization of the sports analytics community, which emerged from an eclectic group of post-World War II operations researchers, freelance journalists, and internet hobbyists. I also review the modern practice of sports analytics in the United States, focusing on three professional sports—baseball, basketball, and American football. I argue that the computational turn in professional sports has had a profound impact on the way teams are managed and the way the games themselves are now played.

Revolution: Assessing the Growth of Analytics in Baseball (Philadelphia, PA: University of Pennsylvania Press, 2014), 1-22.

⁵ M. B. Wright, “Fifty Years of OR in Sport,” *The Journal of the Operational Research Society* 60, Supplement 1: Milestones in OR (May 2009): s161-s168; James J. Cochran, “The Emergence of Sports Analytics,” *Analytics*, 1 February 2010, <https://pubsonline.informs.org/doi/10.1287/LYTX.2010.01.06/full/>; Benjamin Alamar and Vijay Mehrotra, “Beyond ‘Moneyball’: Rapidly Evolving World of Sports Analytics, Part I,” *Analytics*, 3 October 2011, <https://pubsonline.informs.org/doi/10.1287/LYTX.2011.05.05/full/>; Alamar and Mehrotra, “Sports Analytics, Part II,” *Analytics*, 5 December 2011, <https://pubsonline.informs.org/doi/10.1287/LYTX.2011.05.05/full/>; Alamar and Mehrotra, “Analytics & Sports, Part III,” *Analytics*, 2 April 2012, <https://pubsonline.informs.org/doi/10.1287/LYTX.2012.02.03/full/>; Michael J. Fry and Jeffrey W. Ohlmann, “Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications,” *Interfaces* 42 no. 2. (March-April 2012): 105-108; B. Jay Coleman, “Identifying the ‘Players’ in Sports Analytics Research,” *Interfaces* 42 no. 2 (March-April 2012): 109-118.

⁶ On baseball, for example, see Alan Schwarz, *The Numbers Game: Baseball's Lifelong Fascination with Statistics* (New York: Thomas Dunne, 2004); Benjamin Baumer and Andrew Zimbalist, *The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball* (Philadelphia, PA: University of Pennsylvania Press, 2014); and Christopher Phillips, *Scouting and Scoring: How We Know What We Know About Baseball* (Princeton, NJ: Princeton University Press, 2019). On basketball, see Yago Colás, “The Culture of Moving Dots: Toward a History of Counting and of What Counts in Basketball,” *Journal of Sport History* 44, no. 2 (Summer 2017): 336-349. There are several popular, journalistic accounts of analytics in American football and soccer—including many cited below—but, to my knowledge, no scholarly, historical accounts.

⁷ Andrew Baerg, “Sport, Analytics, and the Number as a Communication Medium,” in *Routledge Handbook of Sport Communication*, ed. Paul M. Pedersen (New York: Routledge, 2013), 75.

⁸ An exception is Jameson Otto, Sara Metz, and Nathan Ensmenger, “Sports Fans and Their Information Gathering Habits,” in *Everyday Information: The Evolution of Information Seeking in America*, ed. William Aspray and Barbara M. Hayes (Cambridge, MA: MIT Press, 2011).

A Brief History of the Sports Analytics Community

The origins of sports analytics should be understood in the context of the long history of statistical innovations. Since biblical times, monarchs and emperors have conducted censuses to inform decisions about taxation, infrastructure improvements, and military interventions. This approach was formalized during the nineteenth century, when “statistics” emerged as the mathematical discipline practiced by “statists” (i.e., government officials). During this period, bureaucrats in Europe and the United States enthusiastically collected an “avalanche of printed numbers” about their citizens and developed dozens of metrics concerning births, deaths, age, race, gender, health, crime, and employment.⁹ Meanwhile, Frederick Winslow Taylor and other disciples of “scientific management” conducted time and motion studies of industrial workers, calculated their productivity rates, and suggested workplace interventions to improve their efficiency.¹⁰ Amidst this general zeal for statistics, nineteenth-century sports journalists timed horses at the racetrack and tracked hits and runs at the baseball diamond to enliven their reporting.¹¹

However, the idea of analyzing sports statistics *to gain a competitive advantage* did not occur until after World War II, and it was not championed by coaches or sports executives.¹² Rather, the first analytics pioneers emerged from three distinct sub-communities: 1) operations researchers; 2) freelance sports journalist; and 3) internet hobbyists and bloggers.

⁹ Alfred W. Crosby, *The Measure of Reality: Quantification in Western Society, 1250-1600* (New York: Cambridge University Press, 1997); quotation from Ian Hacking, *The Taming of Chance* (Cambridge, UK: Cambridge University Press, 1990), 2-3.

¹⁰ Nelson, *Frederick W. Taylor*; Robert Kanigel, *The One Best Way: Frederick Winslow Taylor and the Enigma of Efficiency* (New York: Viking, 1997). In addition, scientists deployed statistical thinking to explain electro-magnetism, heredity, and suicide rates, while engineers used cost-benefit projections to justify their construction projects; see Theodore M. Porter, *The Rise of Statistical Thinking, 1820-1900* (Princeton, NJ: Princeton University Press, 1986); and Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, NJ: Princeton University Press, 1996).

¹¹ Kenneth Cohen, *They Will Have Their Game: Sporting Culture and the Making of the Early American Republic* (Ithaca, NY: Cornell University Press, 2017); Alan Schwarz, *The Numbers Game: Baseball's Lifelong Fascination with Statistics* (New York: Thomas Dunne, 2004). Indeed, sports historian Allen Guttman has observed that the “numeration of achievements” became one of the defining characteristics that differentiated modern sports from their pre-modern, quasi-religious predecessors. See Allen Guttman, *From Ritual to Record: The Nature of Modern Sports* (New York: Columbia University Press, 1978), 50.

¹² Baseball executive Branch Rickey is an exception that proves the rule. On his innovative use of statistics, see Branch Rickey, “Goodby to Some Old Baseball Ideas,” *Life*, 2 August 1954, 78-89; Richard J. Puerzer, “Branch Rickey’s Innovative Approach to Baseball Management,” *NINE: A Journal of Baseball History and Culture* 12, no. 1 (Fall 2003): 72-87.

“Operations research” (OR) emerged during World War II as a new science of warfare, in which statisticians analyzed their military’s operational data and suggested tactical adjustments. For example, by accounting for the weather or time of day, operations researchers could help their air forces optimize the lethal efficiency of aerial bombing runs.¹³ As a diversion from their military work, postwar OR practitioners applied their analytical tools to various sports. In a 1954 paper, for example, US defense analyst Charles M. Mottley observed that football resembled “ground combat” and suggested that coaches, like generals, could use OR methods to “make significant improvements in team performance.” In an analysis of 400 running plays, Mottley argued that football teams could maximize yardage gained by balancing end runs and runs into the line of scrimmage.¹⁴ Likewise, in a 1959 *Operations Research* article, the Royal Canadian Air Force’s George R. Lindsey empirically demonstrated that right-handed batters had a higher batting average against left-handed pitchers (and vice versa) because they could see the baseball better. He suggested that managers should substitute different players into the lineup depending on the “handedness” of the opposing pitcher, a now common practice known as “platooning.”¹⁵

Second, an eclectic community of amateur hobbyists and freelance sports journalists developed new statistical metrics to animate their research and writing. During the 1960s, freelancer Bob Davids published statistically informed baseball columns for *The Sporting News*, *Washington Post*, and *Chicago Tribune*. In the spring of 1971, Davids sent a letter to thirty-two like-minded “statistorians” and proposed forming an association. Sixteen charter members founded the Society for Baseball Research (SABR) in August 1971.¹⁶

¹³ Maurice W. Kirby, *Operational Research in War and Peace: The British Experience from the 1930s to 1970* (London: Imperial College Press, 2003); Erik P. Rau, “Combat Scientists: The Emergence of Operations Research in the United States during World War II (PhD diss., University of Pennsylvania, 1999); M. Fortun and S. S. Schweber, “Scientists and the Legacy of World War II: The Case of Operations Research (OR),” *Social Studies of Science* 23, no. 4 (November 1993): 595-642; Stephen P. Waring, “Cold Calculus: The Cold War and Operations Research,” *Radical History Review* 63 (Fall 1995): 29-51.

¹⁴ Charles M. Mottley, “The Application of Operations Research Methods to Athletic Games,” *Journal of the Operations Research Society of America* 2 no. 3 (1954): 335-38.

¹⁵ George R. Lindsey, “Statistical Data Useful for the Operation of a Baseball Team,” *Operations Research* 7, no. 2 (March-April 1959): 197-207.

¹⁶ Dick Thompson and Tom Hufford, “A History of SABR,” Society for American Baseball Research, <https://sabr.org/about/history/1>, viewed 17 December 2019.



Figure 1: Charter members at SABR's first meeting on August 10, 1971, at the National Baseball Library, Baseball Hall of Fame, Cooperstown, New York. Source: <https://sabr.org/about/founders>

Bill James was not one of SABR's founders, but he eventually became its most influential and prominent member. In his freelance articles for *Baseball Digest* and *The Sporting News*, James used statistics to challenge received wisdoms and showed a knack for asking probing questions. For example, how did a pitcher's run support impact his won-loss record? How did hitters fare in day versus night games? Between 1977 and 1988, James published a preseason subscription newsletter, *Baseball Abstracts*, and introduced several new metrics, including Runs Created and Win Shares. Along the way, James became the poster boy for "sabermetrics"—the use of statistical methods to glean new insights into baseball—a term he coined in 1980.¹⁷

Baseball executives ignored James's thinking until the 1990s, but he inspired countless statisticians working across several sports. Bob Carroll, a Pennsylvania high school teacher and freelance author, gathered five fellow hobbyists at the Pro Football Hall of Fame in Canton, Ohio, and founded the Professional Football Researchers Association in June 1979. In *The Hidden Game of Football* (1988), Carroll and his co-authors developed new strategic metrics, such as "expected points," that forecast scoring probabilities for different field positions and

¹⁷ Dan Okrent, "He Does It By the Numbers," *Sports Illustrated*, 24 May 1981, 43-51; Scott Gray, *The Mind of Bill James: How a Complete Outsider Changed Baseball* (New York: Doubleday, 2006); Schwarz, *The Numbers Game*, 111-132.

game situations.¹⁸ That same year, Rob Bellotti published *Basketball's Hidden Game* (1988) and introduced “points created,” the basketball analog to James’s runs created. In another nod to James’s *Baseball Abstracts*, Dave Heeren published several *Basketball Abstracts* (1988-1994) and introduced TENDEX, a linear weights model that added positive events (e.g., points, rebounds) and subtracted negative events (e.g., turnovers) to rate a player’s total performance.¹⁹

During the 1990s and early 2000s, the sports analytics community began to flourish among hobbyists and bloggers on the nascent World Wide Web. Caltech graduate Dean Oliver was among the first statisticians to transition from traditional books and subscription newsletters to personal web pages and user groups as a cheaper and more accessible publishing platform. In 1990-1991, Oliver began posting his basketball analyses on a personal web site he called *The Journal of Basketball Studies*, and later on the Usenet group rec.sports.basketball. His 1991 working paper, “New Measurement Techniques and a Binomial Model of the Game of Basketball” introduced the so-called “Possession Scoring System,” which recognized each team’s possessions as basketball’s fundamental unit of analysis. Oliver defined a team’s offensive and defensive rating as the number of points scored or allowed per possession, a now-standard measure of efficiency.²⁰

It is important to note that these operations researchers, freelance journalists, and internet bloggers were quirky iconoclasts and outsiders. Generally, professional scouts, coaches, and team executives derided the sports analysts since most of them had never played the games. “This guy has never played baseball,” Detroit Tigers manager Sparky Anderson said dismissively of Bill James. “I don’t think he knows very much about it.”²¹

¹⁸ “Tribute to PFRA Founder Bob Carroll (1936-2009),” Professional Football Researchers Association, 2009, <http://www.profootballresearchers.org/bob-carroll-tribute.html>, viewed 17 December 2019; Bob Carroll, Pete Palmer, and John Thorn, *The Hidden Game of Football* (New York: Grand Central Pub, 1988).

¹⁹ Robert S. Bellotti, *Basketball's Hidden Game: Points Created, Boxscore Defense and Other Revelations* (Night Work Publishing, 1988); Dave Heeren, *The Basketball Abstract* (New York: Prentice Hall, 1988).

²⁰ For his self-published newsletter *Basketball Hoopla* 1988-89 and several other studies, see several pages on Oliver’s still-active *Journal of Basketball Studies*: “About JoBS,” (<http://www.rawbw.com/~deano/about.html>); “Basketball Hoopla,” (<http://www.rawbw.com/~deano/hoopla/index.html>); and “All Articles,” (<http://www.rawbw.com/~deano/artall.html>). In particular, see “New Measurement Techniques and a Binomial Model of the Game of Basketball” (1991) at <http://www.rawbw.com/~deano/articles/bbalpyth.html>. All sites viewed on 6 September 2020. University of North Carolina men’s basketball coach Frank Maguire anticipated Oliver in defining points per possession as a normalized metric; see Frank Maguire, *Defensive Basketball* (New York: Prentice Hall, 1959), 12-14.

²¹ Anderson, quoted in Schwarz, *The Numbers Game*, 131.

The tide began to turn in the early 2000s. When author Michael Lewis described James's influence on the Oakland Athletics and published *Moneyball* in 2003, it signaled a growing enthusiasm for analytics and inspired a hiring spree for statisticians among professional teams. In 2004, the Los Angeles Dodgers recruited Oakland's 31-year-old assistant GM Paul DePodesta to be its new general manager, even though he had never played or coached professional baseball. That same year, the Seattle Supersonics hired Dean Oliver as a full-time basketball analyst. By 2012, twenty-six of thirty Major League Baseball teams and twenty-four of thirty NBA basketball teams employed at least one person dedicated to analytics.²²

Sports analytics has blossomed into a full-blown academic subdiscipline. In 2006, Boston-based analysts Daryl Morey (Boston Celtics) and Jessica Gelman (Kraft Analytics Group, New England Patriots) co-founded the MIT Sloan Sports Analytics Conference; it annually attracts 2,000 participants.²³ The two main statistical societies—the Institute for Operations Research and Management Sciences (INFORMS) and the American Statistical Society—have each added sections devoted to sports analytics. Sports analysts have founded three peer-reviewed journals to share their non-proprietary methods, including the *Journal of Quantitative Analysis in Sports* (2005), the *International Journal of Sports Science and Engineering* (2007), and the *Journal of Sports Analytics* (2015).²⁴ In 2016, Syracuse University launched the US's first bachelor's degree program in sports analytics to prepare students for jobs in professional sports management.²⁵

By the 2010s, the methods of sports analytics had evolved from a quirky pastime practiced by operations researchers, freelance journalists, and internet bloggers, into a legitimate

²² Table 2: "Breakdown of MLB Front Offices by Number of Full-Time Employees Working Primarily in Analytics, 2012," in Baumer and Zimbalist, *The Sabermetric Revolution*, 26; also see pp. 87-91; Glockner, *Chasing Perfection*, 18-27; Chris Ballard, "Measure of Success," *Sports Illustrated*, 24 October 2005, 87-91.

²³ "About – Sloan MIT Sports Analytics Conference," <http://www.sloansportsconference.com/about/>, viewed 9 September 2020.

²⁴ Cochran, "The Emergence of Sports Analytics;" Alamar and Mehrotra, "Beyond 'Moneyball': Rapidly Evolving World of Sports Analytics, Part I."

²⁵ Alex Reimer, "Syracuse University Will Launch First Sports Analytics Degree in the U.S.," *Forbes*, 11 May 2016, <https://www.forbes.com/sites/alexreimer/2016/05/11/syracuse-university-will-launch-first-sports-analytics-degree-in-the-u-s/#1f1f146b73b4>, viewed 9 September 2020. Recent textbooks include Thomas A. Severini, *Analytic Methods in Sports: Using Mathematics and Statistics to Understand Data from Baseball, Football, Basketball, and Other Sports* (New York: CrC Press, 2015); Thomas W. Miller, *Sports Analytics and Data Science: Winning the Game with Methods and Models* (Old Tappan, NJ: Pearson Education, 2016); Lorena Martin, *Sports Performance Measurement and Analytics: The Science of Assessing Performance, Predicting Future Outcomes, Interpreting Statistical Models, and Evaluating the Market Value of Athletes* (Old Tappan, NJ: Pearson Education, 2016).

academic sub-discipline with expertise that was highly valued by several professional sports teams. As *Sports Illustrated* reported in 2005, this “burgeoning cult of independent statheads and academics” gave their teams a competitive advantage by identifying “trends, talent, and value that no one else sees.”²⁶

Collecting, Aggregating, Retrieving, Calculating, and Securing Sports Data

How do sports analysts conduct their work? The first prerequisite is to record and store accurate records of the games. These practices have changed significantly over time. In general, however, human scorekeepers (increasingly aided by cameras, radar, and other high-tech equipment) must record the movements and actions of the players. Depending on the sport, scorekeepers record a variety of game data—including the number of pitches, hits, runs, errors, points, assists, blocks, passes, receptions, or carries—that are deemed to be meaningful in the context of recreating the game or evaluating the contributions of individual players. The scorekeepers then inscribe those data onto paper or computerized scoresheets, along with any metadata or tags that might aid in subsequent categorization. The scoresheets for individual games are then stored—in a scorebook, paper ledger, file cabinet, or a relational database—so they can be subsequently retrieved, summarized into a “box score” for an individual game, or aggregated into season or career totals for teams and individual players. Indeed, as historian Christopher Phillips has observed, it requires significant human labor, carefully negotiated scorekeeping and inscription practices, and reliable data storage and retrieval systems to confidently claim that baseball player Nap Lajoie hit 30 doubles in 1907.²⁷

Who collects the data? In 1913, brothers Al and Walter Elias founded the Elias Sports Bureau to provide next-day baseball statistics to various newspapers that were unwilling to wait several days for the National League to share its official scoresheets. The brothers hired their own scorekeepers in each of the major league cities; those scorers kept their own (unofficial) scoresheets and reported them over the news wires to Elias’s central office in New York City. From this “centre of calculation,” the Bureau then transmitted their daily box scores and the

²⁶ Ballard, “Measure of Success,” 87.

²⁷ On information systems and data management, see Alamar, *Sports Analytics*, 24-34, 79-90; Christopher Phillips, *Scouting and Scoring: How We Know What We Know About Baseball* (Princeton, NJ: Princeton University Press, 2019), 13-32; Christopher Phillips, “The Bases of Data,” *Harvard Data Science Review* 1 no. 2 (1 November 2019), <https://doi.org/10.1162/99608f92.5c483119>.

leaders in various statistical categories to dozens of subscribing newspapers around the country.²⁸

Today, Elias is the official statistician for Major League Baseball (MLB), the National Football League (NFL), the National Basketball Association (NBA), the National Hockey League (NHL), Major League Soccer (MLS), the Arena Football League, and the Women's National Basketball Association (WNBA). Over time, Elias has compiled more and different kinds of game data to satisfy their subscribers, and by extension readers and fans. After World War II, for example, Elias gradually added assists (1946), rebounds (1950), and blocks and steals (both 1973) to complement a player's point totals in basketball's standard NBA box score.²⁹

Nevertheless, sports analysts regularly complained that Elias's official statistics were not detailed enough for their sophisticated statistical analyses. In 1988, football analyst Bob Carroll called Elias's NFL's stats "maddeningly incomplete." He lamented, for example, that Elias kept no records on the number and average length of kickoffs, a startling omission given that football scoring was so closely tied to starting field position.³⁰

To compensate, the amateur statisticians formed their own volunteer networks to collect more granular data. In 1984, Bill James organized Project Scoresheet and trained 200 volunteer scorekeepers to use an enhanced baseball scoresheet and notational system that tracked more detailed play-by-play data on things like pitch count, baserunning, and fielding. Project Scoresheet gradually evolved into the for-profit firm, STATS, Inc., which competes directly with Elias to provide stats to subscribers like *USA Today* and ESPN.³¹ Similarly, in 2002-2003, Roland Beech founded the website 82games.com and organized a network of 100 volunteer statheads to chart possession-level data while watching NBA basketball games at home on TV. This approach was somewhat precarious. In one of Dean Oliver's internet posts, he recalled charting an NBA game, and hoping that TNT, the television network, wouldn't switch its broadcast to a different game and render his chart incomplete.³²

²⁸ Schwarz, *The Numbers Game*, 25-41; on "centres of calculation, see Bruno Latour, *Science in Action: How to Follow Scientists and Engineers Through Society* (Cambridge, MA: Harvard University Press, 1987), 215-257.

²⁹ On the growth of the Elias Sports Bureau, see the obituary of its longtime CEO, Seymour Siwoff: Richard Goldstein, "Seymour Siwoff, 99, a Trailblazer in Sports Statistics," *New York Times*, 2 December 2019, A25; on the expansion of basketball statistics, see Colás, "The Culture of Moving Dots," 337-338.

³⁰ Bob Carroll, Pete Palmer, John Thorn, *The Hidden Game of Football* (New York: Grand Central Pub, 1988), 36-52, quotations on pp. 37, 46.

³¹ Phillips, *Scouting and Scoring*, 100-135; Schwarz, *The Numbers Game*, 172-194.

³² Glockner, *Chasing Perfection*, 21-23; Ballard, "Measure of Success," 90; "New Visitor's Guide to 82games.com," 82games.com, <http://www.82games.com/newuser.htm>, viewed 12 September 2020.

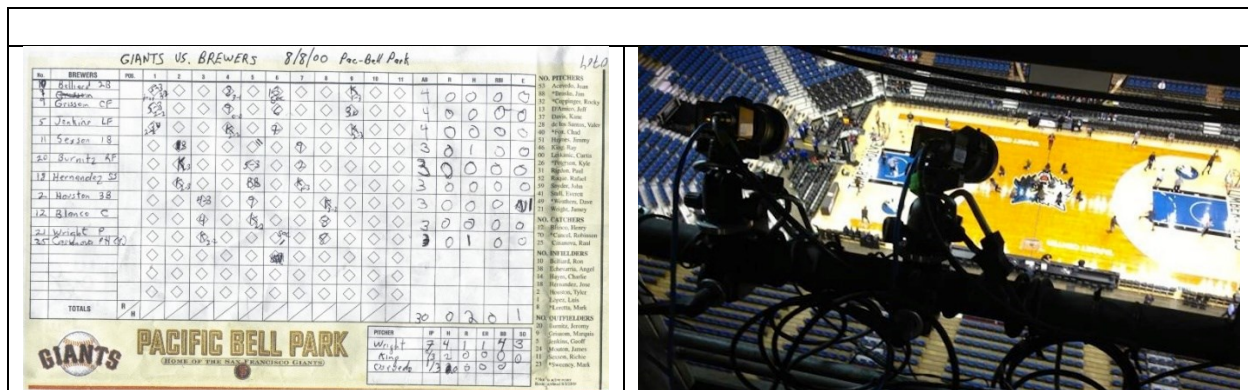


Figure 2 (left): Armed with a baseball scoresheet, a pencil, and carefully negotiated inscription practices, a human scorekeeper can capture a great deal of data about a baseball game. Source: [Wikimedia Commons](#)

Figure 3 (right): SportVU cameras installed in the catwalks of an NBA arena automate the collection of approximately 800,000 discrete observations of the ball and players during an NBA basketball game. Source: [NBA.com](#)

As demand for advanced metrics spread from the hobbyist communities to the professional teams, the for-profit statistical bureaus (i.e., Elias and STATS) began offering their subscribers more robust play-by-play data. They also began drawing on radar, video cameras, tracking chips, and software to automate the process of data collection. In 2013, the NBA signed a licensing deal to install SportVU tracking cameras in the catwalks of every arena. Twenty-five times per second, SportVU can plot the position of all ten players (and three referees) in two dimensions, and the basketball in three dimensions. At 800,000 discrete observations per game, SportVU could reveal how fast and far a player moved during a game; the number of times he touched the ball, the number of passes he made, the precise location and trajectory of every shot, and how closely he defended his opponent.³³

Major League Baseball (MLB) and the National Football League (NFL) also joined the panoptic revolution. In 2014, MLB announced installation of the Statcast tracking system in its stadiums. Like SportVU, Statcast blanketed a baseball diamond with a combination of high-speed cameras and doppler radar. At 40,000 frames per second, Statcast can track the precise

³³ Zach McCann, "Player tracking transforming NBA analytics," ESPN.com, 18 May 2012, https://www.espn.com/blog/playbook/tech/post/_id/492/492, viewed 15 September 2020. In 2016, the NBA ended its contract with STATS/SportVU in favor of competitors Sportradar and Second Spectrum, which provided the raw data and analysis, respectively. See Danny Ecker, "Stats loses NBA deals for data, player-tracking," *Crain's Chicago Business*, 22 September 2016, <https://www.chicagobusiness.com/article/20160922/BLOGS04/160929935/stats-loses-nba-deals-for-data-player-tracking>, viewed 12 September 2020.

position and speed of every player on the field, and the speed and flight path of the baseball accurate to one inch and one mile per hour. Strategically, Statcast data could establish that a pitcher tended to fatigue and lose velocity after 70 pitches, suggesting when a coach might substitute for a relief pitcher. When cross-referenced to play-by-play data, Statcast could also reveal how hitters fared against various pitch types.³⁴ Similarly, in 2014, the NFL partnered with Zebra and Wilson to embed RFID (radio-frequency identification) tracking microchips in each players' shoulder pads and in the football itself. By placing scanners around the stadium, team analysts and broadcasters could know exactly how far a pass traveled in the air, the pass receiver's speed and separation from his defender, the precise field position of the catch, and how far he ran before being tackled.³⁵ By tracking every movement of the ball and players, these automated, panoptic data collection systems created an avalanche of "big data" amenable to intense statistical analysis.

The data storage and calculating tools of sports analysts have reflected the technologies of their times. In the 1970s, sabermetrician Pete Palmer painstakingly transferred reams of printed statistics onto 30,000 punched cards, then used a CDC 6600 mainframe to calculate season and career totals for individual players and teams. In the 1980s, Project Scoresheet volunteers used Apple II personal computers to code in their enhanced play-by-play data. Using a 300 baud modem, they transferred their data to a DEC MicroVAX minicomputer, which calculated updated aggregate season and career totals using Palmer's FORTRAN batch code.³⁶ With the emergence of personal computers, Excel, Lotus 1-2-3, and other spreadsheets soon became the killer app for sports analytics computations. Excel can easily perform simple high-to-low data sorts, plus run regressions and dynamic, stochastic calculations such as Markov chains and

³⁴ Albert Chen, "The Metrics System: How MLB's Statcast is creating baseball's new arms race," *Sports Illustrated*, 26 August 2016, <https://www.si.com/mlb/2016/08/26/statcast-era-data-technology-statistics>, viewed 15 September 2020; Phillips, *Scouting and Scoring*, 126-135. Statcast was preceded by PITCHf/x, which only tracked the ball (not the players); see Mike Fast, "What the Heck is PITCHf/x?" in *The Hardball Times Baseball Annual 2010* (ACTA Publications, 2009) and available online at <http://baseball.physics.illinois.edu/FastPFXGuide.pdf>, viewed 15 September 2020. In 2020, Statcast received a technological boost by deploying the Hawk-Eye tracking system; see Ben Jedlovic, "Introducing Statcast 2020: Hawk-Eye and Google Cloud," 20 July 2020, <https://technology.mlblogs.com/introducing-statcast-2020-hawk-eye-and-google-cloud-a5f5c20321b8>.

³⁵ National Football League, "NFL NextGen Stats," <https://operations.nfl.com/the-game/technology/nfl-next-gen-stats/>, viewed 12 September 2020.

³⁶ Schwarz, *The Numbers Game*, 140.

15					
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
17	Intercept	0	#N/A	#N/A	#N/A
18	RET TD	3.172538521	1.2791899	2.48011536	0.014520952
19	PENDIF	-0.056152517	0.019530575	-2.8751082	0.004779821
20	PY/A	61.67756012	3.957309602	15.5857303	1.28665E-30
21	RY/A	26.44027931	6.303341114	4.19464516	5.26202E-05
22	TO	-2.771793664	0.473496142	-5.8538886	4.27337E-08
23	DPY/A	-67.49861999	5.396830578	-12.507085	1.73149E-23
24	DRY/A	-22.78529422	7.309963853	-3.1170187	0.002286651
25	DTO	3.48978828	0.50392494	6.92521446	2.3113E-10
26					

Figure 4: The Excel spreadsheet is the computational tool of choice among sports analysts. This regression analysis suggests that passing yards per attempt (PY/A) far outweighs rushing yards per attempt (RY/A) and other indicators as the most important independent variable associated with wins in football. Source: Wayne Winston, *Mathletics* (2012), p. 129.

Monte Carlo simulations.³⁷ During the 2010s, it became common for scouts, coaches, and executives to consult Apple iPads and other handheld devices on the sidelines. And of course, those proprietary data systems must be secured. In July 2016, Chris Correa, the former director of scouting for the St. Louis Cardinals, was sentenced to 46 months in federal prison for unlawfully hacking into the Houston Astros' scouting records.³⁸

A Sampling of Advanced Metrics: What Wins Games?

Across all sports, statisticians have asked a few key questions: What aspects of game play are most crucial? What are the key performance indicators that drive wins and losses? And how might individual players and teams optimize their performance around those key drivers? A sampling of a few statistical innovations should provide a sense of the community's influence on the management of various professional sports.

Sabermetrician Bill James believed that traditional baseball metrics such as Batting Average (hits/at bats) did not adequately account for a player's total offensive contribution to scoring runs. In 1979, James introduced a new metric called Runs Created. The formula captured the positive impact of disciplined batters who reached base by drawing walks, plus the "total

³⁷ On the use of Excel, see Wayne L. Winston, *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football* (Princeton, NJ: Princeton University Press, 2009); and Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't* (New York: Penguin, 2012), 84.

³⁸ Michael McCann, "Breaking down Chris Correa's prison sentence for hacking Astros," *Sports Illustrated*, 18 July 2016, <https://www.si.com/mlb/2016/07/19/cardinals-chris-correa-hacks-astros-prison-sentence>, viewed 18 September 2020.

bases” logged by power hitters who slugged doubles, triples, and home runs. The formula also added the positive contributions of stolen bases, and subtracted a negative effect when baserunners were caught stealing. When James and his fellow sabermetricians analyzed the Runs Created data, they found that the contributions of baserunning paled in comparison to batting. Attempting to steal a base, they concluded, was not worth the risk of getting caught, losing the run scoring chance, and surrendering one of a team’s 27 outs. As James’s ideas gained traction in the 2000s, baseball teams began to covet power hitters; speedy base stealers fell out of favor.



Figure 5: Bill James with his calculator in 1981. His formula for Runs Created is on the big screen in Royals Stadium in Kansas City. Source: Photo by [Lane Stewart](#), from Dan Okrent, “He Does It by the Numbers,” *Sports Illustrated*, 24 May 1981, 43-51.

$$\text{Runs Created} = \frac{(H + BB - CS) * (TB + 0.7SB)}{AB + BB + CS}$$

H = Hits, BB = base on balls (i.e. walks, plus hit by pitch), TB = total bases
CS = caught stealing, SB = stolen base, AB = at bats

Similarly, basketball’s traditional field goal percentage metric (shots made/shots attempted) did not adequately account for the 50% higher value of 3-point vs. 2-point shots. Dean Oliver’s effective field goal percentage (eFG%) corrected for this oversight. While teams generally made a lower percentage of the longer 3-point shots, the extra point they earned made

them especially valuable.³⁹ Using several seasons of geospatial data from SportVU cameras, analyst Kirk Goldsberry mapped the average points per shot attempt for every area of the basketball floor. On average, dunks and layups near the basket earned 1.2 points per attempt. And even though shooters missed them more often, 3-point shots still yielded more than 1 point per attempt. By comparison, midrange shots earned only around 0.85 points per attempt, so NBA teams have started coaching their players to avoid them.⁴⁰

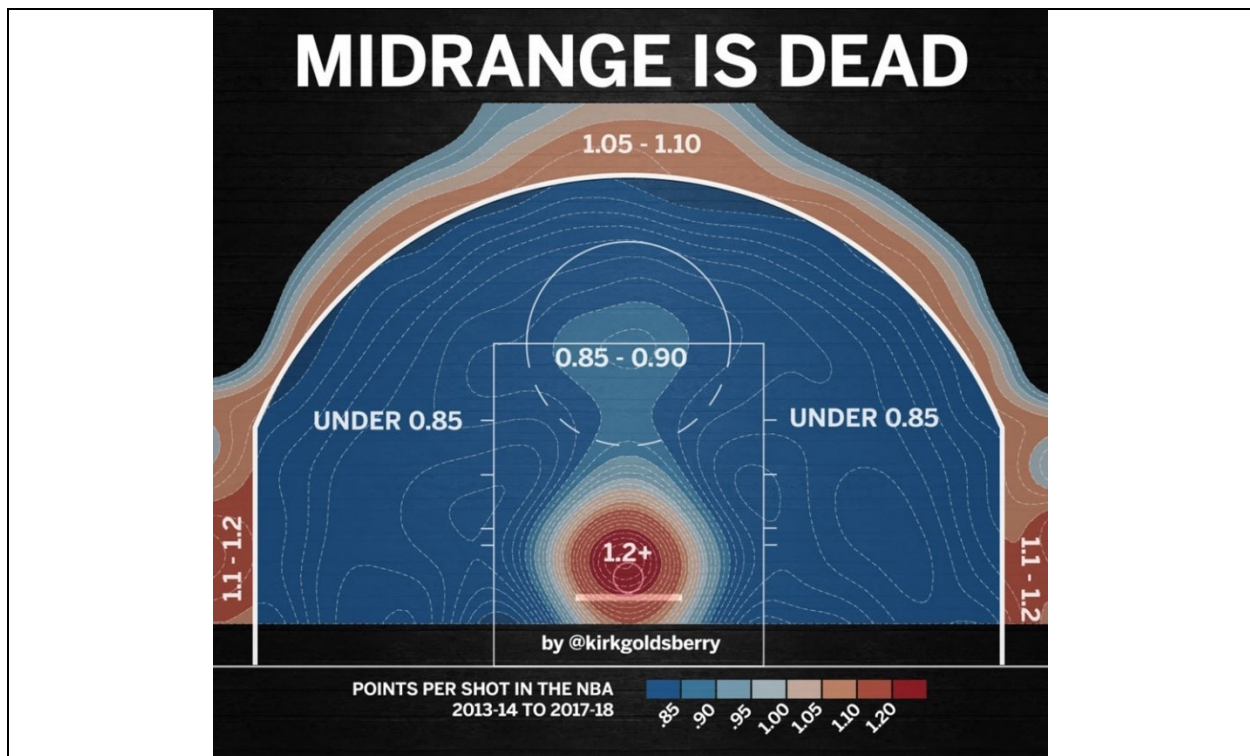


Figure 6: Dean Oliver’s effective field goal percentage (eFG%) metric helped basketball analysts better appreciate the strategic value of 3-point shots. Kirk Goldsberry’s subsequent geospatial mapping of all NBA shot attempts across several seasons highlighted the relative efficiency of dunks, layups, and 3-pointers, and the relative inefficiency of midrange shots. Sources: Dean Oliver, *Basketball on Paper* (2004); Kirk Goldsberry, *Sprawlball* (2019).

$$eFG\% = \frac{FGM + (0.5 * 3PFGM)}{FGA}$$

FGM = field goals made, 3PFGM = 3-pt field goals made, FGA = field goals attempted

³⁹ Oliver, *Basketball on Paper*, 63; Justin Kubatko, Dean Oliver, Kevin Pelton, and Dan T. Rosenbaum, “A Starting Point for Analyzing Basketball Statistics,” *Journal of Quantitative Analysis in Sports* 3, no. 3 (2007), Article 1.

⁴⁰ Kirk Goldsberry, *Sprawlball: A Visual Tour of the New Era of the NBA* (New York: Houghton Mifflin, 2019), 2-3.

In football, analyst Bud Goode used regression analysis to find that one “killer stat”—Yards Per Pass Attempt—was the most important independent variable associated with wins. Intuitively, this made sense; if offenses were consistently attempting and completing longer passes, they were more likely to reach the 10-yard line to gain within four downs, maintain possession, move down the field, and score.⁴¹ In 2006, UC Berkeley economist David Romer published an analysis of those crucial “fourth down” plays. Should teams 1) “go for it” and run an offensive play to earn a new set of downs; 2) kick a field goal; or 3) punt the ball away to pin the opponent deep in its own territory? Romer built a detailed probability function of expected point outcomes at different yard lines for both the offensive and defensive teams. He found that football teams had generally been too conservative on fourth down and suggested that coaches should “go for it” more often.⁴² In 2019, NFL teams attempted more fourth-down conversions (595) than ever before.⁴³

Conclusion: Analysis and Critiques

The computational techniques of scientific management, operations research, and decision sciences have become firmly entrenched in the management of professional sports. The sports analytics community emerged in the United States after World War II among operations researchers, freelance journalists, and internet bloggers. Although these analysts initially worked in obscurity, they eventually secured positions as executives and consultants with nearly every professional sports team. Like any innovation, the revolution in sports analytics has engendered praise for its value and critiques for its disruptions and unintended consequences. Ultimately, how should we evaluate the impact of analytics on the management of professional sports?

Not surprisingly, sports analysts have praised their craft as a progressive development, in which the principles of rationality and objectivity have triumphed over uninformed tradition and superstition. “‘It has always been done that way’ is not a good reason to do anything,” basketball

⁴¹ Bud Goode, “Killer Stat I,” BudGoodeSports.com, 14 September 2020 <http://mule.he.net/~budsport/pub/killer.php>, viewed 19 September 2020; for context, see Winston, *Mathletics*, 127-131.

⁴² David Romer, “Do Firms Maximize? Evidence from Professional Football,” *Journal of Political Economy* 114, no. 2 (2006): 340-365.

⁴³ Clark, “The NFL’s Analytics Revolution Has Arrived;” Fortier, “The NFL’s analytics movement has finally reached the sport’s mainstream;”

analyst Daryl Morey wrote in a 2007 tribute to Bill James.⁴⁴ Indeed, analytics has empirically tested—and overturned—various received wisdoms, such as the relative inefficiency of stealing bases in baseball. It has also suggested unconventional but winning strategies, such as high-volume three-point shooting in basketball. Ultimately, analytics are practically useful; they help win games. Analytics, Morey suggests, represent the “march of progress.”⁴⁵

But progress for whom? The analytical turn initiated a shift in epistemological authority in the front offices of professional sports teams, albeit with predictable resistance. Lewis’s *Moneyball* vividly described these tensions, such as Billy Beane’s heated arguments with manager Art Howe and scouting director Grady Fuson about which players the A’s should draft, play, and bench. As professional teams gradually learned to “trust in numbers” and let the “computer boys take over,” they elevated the influence of statisticians, most of whom had never played the games. At the same time, this analytical turn diminished the influence of players, coaches, and scouts, whose authority over in-game strategy and the evaluation of players usually relied on their embodied expertise and experience on the field.⁴⁶ “They’ve got all these super nerds as I call them, in the front office,” outfielder Jayson Werth complained in 2018. “I think it’s killing the game.”⁴⁷

Many professional athletes also worry about the specter of constant, Orwellian surveillance and the loss of their personal autonomy. In the nineteenth century, Frederick Winslow Taylor monitored the precise movements of industrial workers with a stopwatch and clipboard, then suggested factory-floor interventions to discipline their movements and optimize their efficiency. In the twenty-first century, high-speed cameras, doppler radar, and wearable microchips similarly monitor the movements of MLB, NBA, and NFL players, so that coaches and GMs may suggest interventions to discipline their in-game decisions and optimize their performance. Indeed, a professional employee-athlete has traded away some of his autonomy

⁴⁴ Morey quotation (p. 97) in Gregory F. Augustine Pierce ed., *How Bill James Changed Our View of Baseball: By Colleagues, Critics, Competitors and Just Plain Fans* (Skokie, IL: ACTA Sports, 2007).

⁴⁵ Daryl Morey and Sam Hinkie, “Moneyball and the Houston Rockets: The Rockets front office on the stats movement and how Michael Lewis’ book impacted the way they’ve built a franchise,” *Grantland*, 26 September 2011, <http://grantland.com/features/moneyball-houston-rockets/>, viewed 25 September 2020.

⁴⁶ Lewis, *Moneyball*; Porter, *Trust in Numbers*; Nathan Ensmenger, *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise* (Cambridge, MA: MIT Press, 2012).

⁴⁷ Werth, quoted in Aaron Timms, “Are super-nerds ruining US sports?” *The Guardian*, 16 August 2018, <https://www.theguardian.com/sport/2018/aug/16/sports-nerds-analytics-data>, viewed 30 September 2020.

each time he avoids a mid-range basketball shot in favor of the three-pointer recommended by his employers' analytics.⁴⁸

Professional athletes are also justifiably concerned that the analytics movement has elevated the prospects for certain players, while hastening the obsolescence of others. In baseball, power hitters have prospered, while speedy base-stealer have fallen out of favor. As football has increasingly valued the passing game, quarterbacks and receivers have eclipsed running backs in earning the highest salaries. In basketball, three-point shooting guards like Stephen Curry have excelled, while traditional low-post centers—like former All-Star Roy Hibbert—have found themselves unemployed and out of the league. This can be a vicious cycle. As certain skill sets become devalued by analytics, those players receive less playing time, record fewer of the right kinds of statistics, and are thus more likely to be cast aside. In short, analytics have real economic consequences for players when team executives valorize certain efficient skill sets and reject others.⁴⁹

Finally, many former players and fans have demonized analytics for ruining the aesthetic beauty of the games. They argue that strict adherence to analytics-inspired strategies have made players automatons and robbed the games of variety, spontaneity, and individual creativity. Many critics, for example, lament that baseball's new emphasis on power pitching, plate discipline, and power hitting has reduced most at bats to the "three true outcomes," i.e., a strikeout, a walk, or a home run. That means there are fewer batted balls in play, fewer baserunners, fewer stolen bases, fewer exciting defensive plays in the field, in short, fewer plays that make the game exciting for fans to enjoy. "I try to watch a baseball game, and I find it very difficult," former pitcher Rich "Goose" Gossage complained in 2018.⁵⁰ Yet, aesthetics judgements are always subjective. Some baseball fans undoubtedly enjoy watching strikeouts and home runs. Plus, analytics have arguably made football more fun to watch, as teams attempt more long passes, "go for it" on fourth down, and mix in the occasional fake punt or onside kick.

⁴⁸ On the "dataveillance" critique in sports, see Michael Atkinson, "Heidegger, Parkour, Post-Sport, and the Essence of Being," in *A Companion to Sport*, ed. David L. Andrews and Ben Carrington (West Sussex, UK: Blackwell, 2013), 363-364; and Andrew Baerg, "Big Data, Sport, and the Digital Divide: Theorizing How Athletes Might Respond to Big Data Monitoring," *Journal of Sport and Social Issues* 41, no. 1 (2017): 3-20.

⁴⁹ Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999); Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016); on Curry and Hibbert, see Goldsberry, *Sprawlball*, 178-179.

⁵⁰ Gossage, quoted in Scott Miller, "'I Find it Very Difficult' to Watch: Why MLB Greats Think Baseball is in Trouble," *Bleacher Report*, 20 August 2018, <https://bleacherreport.com/articles/2791455-i-find-it-very-difficult-to-watch-why-mlb-greats-think-baseballs-in-trouble>, viewed 30 September 2020.

For better or worse, analytics have become a permanent fixture in the management of professional sports. However, it is important to remember that statistical insights are merely one of several inputs—including professional instincts and the wisdom of experience—that factor into decision-making. In other words, analytical insights can influence—but never replace—human agency. Ultimately, individual players, coaches, and team executives decide which pitches to throw, which plays to call, and which players to draft.⁵¹

⁵¹ For meditations on technological determinism and human agency, see Leo Marx and Merritt Roe Smith eds., *Does Technology Drive History? The Dilemma of Technological Determinism* (Cambridge, MA: MIT Press, 1994).