

# Introduction

In the fall of 2012, Craig Biggio was up for election to baseball's Hall of Fame. A longtime Houston Astro and seven-time Major League All-Star, Biggio seemed like a good candidate. Still, he fell short of the votes required for induction that year, and then again the following year. Supporters of his case put forward evidence to convince the skeptics: numbers concerning doubles hit, batting averages recorded, runs scored; accounts of his loyalty and his apparently steroid-free record; descriptions of his scrappiness and versatility. The debates were not all that different from countless others in which quality is assessed. How do we know how to separate the good surgeons from the bad, or the great teachers from the merely good?

We're told there's been a modern revolution in how we should approach these questions. All humans, we've learned, suffer from unconscious flaws in how we see and think. As a result, we need to gather lots of data about a situation—ideally, numerical data—aggregate them, and analyze them statistically. Only by shackling ourselves to objective data and thereby limiting our own subjective biases and idiosyncrasies can we arrive at reliable knowledge.<sup>1</sup> For Biggio's candidacy, that meant considering his playing statistics, dispassionately assessing his numbers, and comparing them

## 2 INTRODUCTION

to those of his peers. It wasn't enough to recall him in action or to have fond associations with him: we needed numerical evidence to know if he was worthy of the honor.

The idea that there are seemingly irreconcilable approaches to judging quality in baseball was reinforced by Michael Lewis's 2003 book *Moneyball: The Art of Winning an Unfair Game* and the 2011 feature film based on it. The book and film stirred interest far beyond baseball fans, because Lewis was supposedly describing a general solution to the problem of valuation, especially under financial constraints. Indeed, many took the book as providing a larger lesson. The Harvard Business School used it as a case study in the cultivation of leadership and innovation. Others saw "moneyball" methods as offering new ways to replace tradition-bound fields such as politics (*Moneyball for Government*), law ("Moneyball Sentencing"), education ("Is 'Moneyball' the Next Big Thing in Education?"), and criminal justice ("Lessons for Policing from *Moneyball*").<sup>2</sup>

There's a scene about a half-hour into the film that dramatizes the stakes. Billy Beane, the protagonist and general manager of the Oakland Athletics, enters a room full of scouts trying to figure out which prospects to draft. Beane brings with him his new assistant—an Ivy League graduate with a degree in economics. Beane wants to figure out how the club is going to replace three top players from the previous season who had signed more lucrative contracts elsewhere.

The surface issue is money: if the club had enough of it there wouldn't be a problem. But the ultimate question is how to maximize what the team can do with limited resources. The scouts suggest they should get the best players they can afford. Beane's new approach—what Lewis called "moneyball"—is to buy outcomes, not players, because outcomes are cheaper. Beane is looking only for runs on offense and outs on defense. In order to rationally allocate limited resources, Beane argues, they must first turn prospects into statistical aggregates. Only the numbers matter.

It isn't difficult to tell who's on which side: The scouts are old; Beane and his new assistant are young. Scouts don't know the

word “aggregate”; Beane says they should all be “card counters” at the blackjack table. Scouts forget to carry the one when they calculate; Beane’s guy can manipulate numbers on the fly. Scouts know what other clubs think about prospects; Beane knows who gets on base. Scouts want to talk about who is “on the weed” or going to strip clubs; Beane says “on-base percentage” is all they’re allowed to discuss. Scouts talk about people; Beane talks about statistics.

This rhetorical division emphasizes a clear distinction between forms of expertise. The scouts think Beane is ignoring the wisdom and experience they represent. They talk about the age and condition of bodies as well as the way players behave on and off the field because ultimately the game is played by fallible humans. Beane redirects the conversation to statistical measures of performance. He wants to know the numbers. The distinction is between scorers and scouts, those who analyze the numbers and those who assess the bodies, but also between analytics and intuition, objectivity and instinct, rationality and superstition.<sup>3</sup> However expressed, scorers and scouts are understood to approach the evaluation of prospects in fundamentally different ways.

Curiously absent from the scene is the fact that every prospect was previously measured and quantified by the scouts themselves. Scouts may not talk about their work as a process of putting numbers on players, but fundamentally that’s what they do. Each prospect would have had at least one, and likely many, scouting reports written about him, reports that included scouts’ numerical judgments about his present and future abilities. If scouts were so focused on body types and emotional temperaments, it is odd that they bothered to write up reports that calculated a single number for each prospect: an overall future potential, or OFP. At the same time, Beane’s numbers come from human “scorers”—from the efforts of fallible statisticians, database creators, and official scorers. They also could not have existed, let alone have been interpreted, without an immense amount of human labor and expertise. The numbers scouts deploy are obviously different sorts of numbers than Beane’s, but it still seems strange that there could be such a

#### 4 INTRODUCTION

stark divide in how scouts and scorers approach the world if they both produce numbers that can be used to create a draft list and ultimately put a single price on signing each player. Are scouts' methods really all that distinct from scorers'? The answer, I suspected, would cut to the heart of what it means to produce knowledge useful for making value judgments and predictions in settings far beyond the baseball diamond.

---

I am a baseball fan, but I am also a historian. When I began this book I planned only to compare how scorers and scouts evaluate prospects, but I soon realized this was a special case of more general concern in academic fields that occupy my attention. Scouts and scorers document, categorize, and describe the past. They collect data and make judgments about that data in order to make decisions in the present and predictions about the future. Though scorers' and scouts' work is highly consequential, what they do is not all that different from what many of us do everyday: they try to make reliable decisions on the basis of what they know. Though I thought writing about scorers and scouts would be an occasion to release my inner baseball fandom, it turned out instead to be an opportunity for analyzing how reliable knowledge is made.

My topic is baseball, but this is a book about data in the modern world. As the scene in *Moneyball* suggests, not all data are created equal. The numerical data Beane brings to bear on the selection of players are presumed to be precise and objective, and thereby distinct from people knowledge, craft knowledge, or subjective knowledge. Scouting data, conversely, are portrayed as inescapably bound by tradition, culture, and history—that is, bound by the fallibility of humans.

It doesn't make much sense to distinguish numerical data from human data, however, if we think about the word itself. "Data" comes from a form of the Latin verb *dare*, to give. Data are "that which have been given." They didn't originally need to

be numerical, objective, or even true. They were simply the principles or assumptions that were conventionally agreed upon so that an argument could take place. Data were that which could be taken for granted. Over time, of course, the meaning has shifted, so that now we tend to think of data as the *result* of an investigation rather than its premise or foundation. In either sense of the term, data take effort to establish and have to be made useful. There is no natural category of “raw” data; data only exist in context.<sup>4</sup>

For Beane and others interested in performance statistics or data analytics within baseball, the primary complaint wasn’t that there were no numbers before they got involved, but that the *wrong* sort of numbers had been collected. Since at least the seminal publication of *The Hidden Game of Baseball* in 1984, it has been commonplace to distinguish useful and powerful “new” statistics from “old” or “traditional” statistics.<sup>5</sup> These analysts are quick to remind us that not all statistics are useful, but we often forget the corollary of that assertion: the very act of calling something “data” is a claim about its relevance for a particular argument. “Runs batted in” is a statistic, whether regarded as “old” or “new.” It is data, though, only if someone wants to win an argument with it. It’s possible, though perhaps mistaken, to imagine facts or numbers existing without people, but it is impossible to imagine data without people.

Books on the new data sciences characteristically spend little or no time discussing the human labor by which data are made. There is often acknowledgement that it matters who collects the data and how they collect them, but the belief, explicitly stated or not, is that with enough sophistication in processing and analyzing, any faults or improprieties in collection can be managed. We can transcend the problems and individual idiosyncrasies of data’s origins by collecting enough data.<sup>6</sup> Though it is possible to measure players’ abilities or performance without thinking about the origins of the data, if we want to know how scorers and scouts come to know what they know—not just find out who is right—then we have to think more carefully about how they create data.

## 6 INTRODUCTION

We can begin by simply asking how facts like Biggio's skill level or hit totals become stable, credible, and reliable. We rarely consider the trust we put in established statistics or how teams come to agreement on whether one player or another should be drafted first. As in any field, technical specifications and practices, politics, education, and social norms shape the creation of knowledge. Yet these factors have become invisible over time. Historian Paul Edwards, in his study of the history of climate models, notes that the difference between "settled knowledge" and a "controversial claim" is ultimately a difference in whether or not the support structure behind each fact is visible. To be a fact means to be supported by an infrastructure, but established, settled facts have made the infrastructure invisible enough that they can seem natural and eternal. Facts are controversial when we can see the infrastructure supporting them.<sup>7</sup> To understand how scoring and scouting knowledge works, I realized I needed to uncover the structures—the labor, technologies, and practices—behind them.

What I discovered was that historically the ways scorers and scouts produced knowledge and established facts were not all that different. Human expertise was required to collect, standardize, and verify performance statistics. Moral considerations determined what data to keep, while complex bureaucratic measures managed scorers' judgments. Scouts were fixated on accurate measures of performance and value. Over time they increasingly had to express their judgments with numbers. They, too, relied on complex bureaucracies and technologies to collect, standardize, and verify their data. Over the last half of the twentieth century scouts and scorers increasingly shared a goal of turning players into numbers.

Any claimed division between scouting's judgment-based subjectivities and scoring's data-based objectivities doesn't have a strong purchase, historically. That is not to say that such distinctions can't be made; in fact, assertions that one process is more objective than another or that one practice minimizes subjective bias can still play important roles in debates. *Moneyball* and similar

narratives have presented scouts and scorers as fundamentally divided in part because it makes a good story, a modern-day parable about the power of data and rationality to overcome superstition and guesswork. Parables, like myths, are important cultural markers, as anthropologists have told us for as long as there have been anthropologists. They are ways of organizing social norms and of communicating and maintaining them. But they are not necessarily accurate portrayals of how things work.

Stark divisions between subjective and objective modes, between intuition and measurement, and between different forms of expertise seem inappropriate when we look at how scouts and scorers have acquired knowledge over the years. There are many different ways claims can be made objective, and all of them—trained judgment, regulation and rule following, disinterestedness, mechanization, intersubjectivity, consensus formation—have been used by scouts at one point or another. Similarly, classical markers of subjectivity—judgments of taste and morality, deferral to authority or charisma, management of bodies—have also been applied to scoring practices.<sup>8</sup> Like scorers, scouts are overwhelmingly white and male, and yet they typically treat their bodies as irrelevant to the data they produce, even as their knowledge remains inextricably a product of their own observations. Neither scouts nor scorers care much about such philosophical distinctions, but in practice both groups are deeply concerned with solving the problems of reliably measuring and evaluating people.<sup>9</sup>

One reason baseball is such a good topic for thinking about the practices of evaluation and the nature of data is that performance statistics have been recorded on paper for nearly as long as games have been played, and interested observers have used these records from the beginning to measure and predict excellence. Early clubs—amateur social organizations in which baseball first flourished—nearly always had a scorekeeper and a scorebook in which statistics were recorded. When the sports reporter Henry Chadwick wrote some of the first manuals on baseball in the 1860s, he also noted the importance of scoring, placing the scorekeeper

## 8 INTRODUCTION

as one of the few people allowed on the field with the players. His desire was that the game of baseball be made American, scientific, and manly, and he believed that the best way to achieve this goal would be through careful recordkeeping.

Just as important as its historic connection with recordkeeping is the fact that baseball analytics has become nearly synonymous with data analytics generally. Nate Silver's rapid rise from independent baseball analyst to the *New York Times* and ESPN's payroll, as well as to *Time Magazine's* list of the "100 Most Influential People," seemingly proved that thinking about baseball data provides the skills to think about data in many domains. More striking, perhaps, is that baseball is portrayed as an "ideal home" for data analysis. Though otherwise critical of the use of data-driven mathematical models, Cathy O'Neil's *Weapons of Math Destruction* praises baseball's use of statistical algorithms and numerical analyses as healthy, fair, transparent, and rigorous.<sup>10</sup> If baseball is the paradigmatic example of the expertise and benefits that modern data science can provide, then we ought to consider the extent to which baseball really does represent the replacement of one way of ascertaining quality with another.

---

This book is a history of how scorers and scouts know what they know about baseball. The first four chapters cover the history of official scoring and the creation of baseball statistics while the remainder explores the history of scouting. Both parts trace the people, practices, and technologies used to translate the movement of bodies into reliable knowledge. The technologies involved certainly include the high-speed electronic computer, but I am also interested in the more mundane yet pervasive technologies—pencils, papers, scouting reports, stopwatches, and scoresheets—that have enabled data to be collected. Tools like scoresheets and scouting reports are not simply data-recording devices; they *create data* by enabling the relevant aspects of baseball to be made visible

and durable.<sup>11</sup> These basic tools are easily forgotten but essential to examining how scorers and scouts know what they know.

This book is neither a thorough history of data analysis in baseball nor a comprehensive account of official scoring or scouting. It is not meant to explore whether stats or scouts are more important to running a baseball team; that it takes both forms of knowledge is obvious to those who manage clubs.<sup>12</sup> Rather, I draw on the history of scoring and scouting, of statistical databases and scouting reports, to show that the attempt to create reliable data about the value of individual players looks quite similar on either side of the claimed scouting–scoring divide. In some ways scouts and scorers make an odd pairing; there certainly are fundamental differences in what they do. Scouts are single-mindedly focused on the future, on finding metrics and heuristics that will enable them to make predictions about who will succeed in the coming months and years. Scorers are more retrospective, collecting data on an ongoing basis while also finding statistics from the past that will help them analyze quality and strategy in the present. Both scorers and scouts, however, are focused on making characterizations and judgments about quality, on finding ways of measuring the abilities of players. Their practices are in many regards remarkably similar.

The book's first four chapters reveal the labor that goes into creating the data behind modern analytical claims. These numbers are powerful, but they are deeply tied to the processes of their creation, collection, and dissemination. These processes have been forgotten or actively ignored in accounts that simply treat the numerical data as reliable and stable. I use “scorers” and “scoring” as broad terms, covering those who are involved in the creation and maintenance of statistical data about baseball, regardless of the end uses to which individuals might apply that data.

The final three chapters focus on scouts. Scouts can perform many different roles for a club, though in general they fall into the categories of professional, advance, and amateur. Professional scouts are typically responsible for evaluating players in the minor

## 10 INTRODUCTION

leagues or on other teams for acquisition. Advance scouts determine tendencies of future opponents, ferreting out their strengths and weaknesses so that the best strategy can be deployed against them. The focus here will be on amateur scouts, those evaluating nonprofessional players who might currently be in high school, in college, or playing other sports entirely.<sup>13</sup> This is the “hard case,” the scouting practice seemingly the furthest removed from scoring; amateur competition is typically so inferior to that of the minor and major leagues as to render the performance statistics of amateurs useless for most clubs.

Instead of emphasizing their differences from scorers, however, I will show how amateur scouts also have tried to make reliable determinations of the value of players. Scouts like to talk about themselves as loners, as renegade hunters looking for diamonds in the rough. But ultimately they are cogs in a giant bureaucratic machine, producing written reports of what they’ve seen, reports that are turned into quantified evaluations of players. Scouts are hunters of data, recorders of data, and compilers of data. They have elaborate systems of how to see, measure, and evaluate players. They deploy tools and technologies to help quantify skills and ultimately reduce predictions of future performance to a single number.

These chapters rely not only upon memoirs, archival records, and interviews with scouts but also upon the thousands of scouting reports deposited in the library of the National Baseball Hall of Fame in Cooperstown, New York, and placed online in 2014 as part of the Diamond Mines exhibit.<sup>14</sup> This collection is by no means complete, with entire teams and scouts missing from its rolls. Scouting reports are ephemeral—they simply disappear when a general manager or scout goes through his papers and figures that decade-old reports on now-retired minor league players have no value. Nevertheless, the collection in Cooperstown is extensive enough to provide a real sense of how scouting reports have been used over the years. It is just one slice through the history of scouting, but it is a revealing one.

---

Biggio was elected to the Hall of Fame in 2015, his third time on the ballot. In the end, most commentators considered the decision uncontroversial. The statistics, after all, seemed to speak for themselves: he had 3,060 hits and the most doubles in history by a right-handed batter, just ahead of the turn-of-the-century star Nap Lajoie. After his election, though, I wondered how exactly we knew that Biggio—let alone Lajoie—had precisely that many hits and why we had such high confidence in these numbers. At the same time, I wondered how scouts had seen Biggio, how they had described him and his abilities, and whether they had predicted he would become a Hall of Famer.

No less a statistical authority than Bill James, the author of the influential *Baseball Abstracts*, once called Craig Biggio his favorite player. He explained this opinion in 2008:

[Biggio] was the player who wasn't a star, but who was just as valuable as the superstars because of his exceptional command of a collection of little skills—getting on base, and avoiding the double play, and stealing a base here and there, and playing defense. Here was the guy who scored 120 runs every year because he hit 45 or 50 doubles every year and walked 70 to 90 times a year and led the majors in being hit with the pitch and hardly ever grounded into a double play and somehow stole 25 to 50 bases every year although he really had very average speed.

James also praised the parts of Biggio's career that didn't show up in the box score, the way his move from catcher to second baseman "required something that you don't often see, an exceptional level of determination, dedication and adaptability." Given the choice between drafting a future Tom Glavine, Ken Griffey, or Frank Thomas, James declared that he would still take Biggio: "Maybe that's not what the numbers say is the right answer, but Biggio was the guy who would do whatever needed to be done.

## 12 INTRODUCTION

Makes it a lot easier to build a team.” James concluded with a note of sadness that Biggio’s career had been like a movie that went on too long. He didn’t “admire” the fact that Biggio “hung around” just to get 3,000 hits—it’s “like the director can’t find the ending so it goes on for another half hour.”<sup>15</sup>

Seamlessly—and characteristically—in this summation James wove together numbers and narratives, subjective judgments and objective facts. The original sabermetrician—a term he coined to unite both the “Society for American Baseball Research,” known as “SABR,” and “measurement”—he refused to make easy distinctions between quantitative and qualitative data. He treated them interchangeably, as reliable, established ways of evaluating a player, and as a basis for making a case for Biggio’s worth. If a guru of statistical analysis, one who was supposedly a crucial inspiration for the claim that data analytics should replace traditional ways of judging value in baseball, didn’t make stark divisions between scoring and scouting, surely it is worth thinking far more carefully about how both scorers and scouts come to know what they know about baseball.