GAME THEORY

When it comes to athletic prowess, don't believe your eyes.

BY MALCOLM GLADWELL

The first player picked in the 1996 National Basketball Association draft was a slender, six-foot guard from Georgetown University named Allen Iverson. Iverson was thrilling. He was lightning quick, and could stop and start on a dime. He would charge toward the basket, twist and turn and writhe through the arms and legs of much taller and heavier men, and somehow find a way to score. In his first season with the Philadelphia 76ers, Iverson was voted the N.B.A.'s Rookie of the Year. In every year since 2000, he has been named to the N.B.A.'s All-Star team. In the 2000-01 season, he finished first in the league in scoring and steals, led his team to the second-best record in the league, and was named, by the country's sportswriters and broadcasters, basketball's Most Valuable Player. He is currently in the midst of a four-year, seventy-seven-million-dollar contract. Almost everyone who knows basketball and who watches Iverson play thinks that he's one of the best players in the game.

But how do we *know* that we're watching a great player? That's an easier question to answer when it comes to, say, golf or tennis, where players compete against one another, under similar circumstances, week after week. Nobody would dispute that Roger Federer is the world's best tennis player. Baseball is a little more complicated, since it's a team sport. Still, because the game consists of a sequence of discrete, ritualized encounters between pitcher and hitter, it lends itself to statistical rankings and analysis. Most tasks that professionals perform, though, are surprisingly hard to evaluate. Suppose that we wanted to measure something in the real world, like the relative skill of New York City's heart surgeons. One obvious way would be to compare the mortality rates of the patients on whom they operate—except that substandard care isn't necessarily fatal, so a more accurate measure might be how quickly patients get better or how few complications they

have after surgery. But recovery time is a function as well of how a patient is treated in the intensive-care unit, which reflects the capabilities not just of the doctor but of the nurses in the I.C.U. So now we have to adjust for nurse quality in our assessment of surgeon quality. We'd also better adjust for how sick the patients were in the first place, and since well-regarded surgeons often treat the most difficult cases, the best surgeons might well have the poorest patient recovery rates. In order to measure something you thought was fairly straightforward, you really have to take into account a series of things that aren't so straightforward.

Basketball presents many of the same kinds of problems. The fact that Allen Iverson has been one of the league's most prolific scorers over the past decade, for instance, could mean that he is a brilliant player. It could mean that he's selfish and takes shots rather than passing the ball to his teammates. It could mean that he plays for a team that races up and down the court and plays so quickly that he has the opportunity to take many more shots than he would on a team that plays more deliberately. Or he might be the equivalent of an average surgeon with a first-rate I.C.U.: maybe his success reflects the fact that everyone else on his team excels at getting rebounds and forcing the other team to turn over the ball. Nor does the number of points that Iverson scores tell us anything about his tendency to do other things that contribute to winning and losing games; it doesn't tell us how often he makes a mistake and loses the ball to the other team, or commits a foul, or blocks a shot, or rebounds the ball. Figuring whether one basketball player is better than another is a challenge similar to figuring out whether one heart surgeon is better than another: you have to find a way to interpret someone's individual statistics in the context of the team that they're on and the task that they are performing.

In "The Wages of Wins" (Stanford; $29.95), the economists David J. Berri, Martin B. Schmidt, and Stacey L. Brook set out to solve the Iverson problem. Weighing the relative value of fouls, rebounds, shots taken, turnovers, and the like, they've created an algorithm that, they argue, comes closer than any previous statistical measure to capturing the true value of a basketball player. The algorithm yields what they call a Win Score, because it expresses a player's worth as the number of wins that his contributions bring to his team. According to their analysis, Iverson's finest season was in 2004-05, when he was worth ten wins, which made him the thirty-sixth-best player in the league. In the season in which he won the Most Valuable Player award, he was the ninety-first-best player in the league. In his worst season (2003-04), he was the two-hundred-and-twenty-seventh-best player in the league. On average, for his career, he has ranked a hundred and sixteenth. In some years, Iverson has not even been the best player on his own team. Looking at the findings that Berri, Schmidt, and Brook present is enough to make one wonder what exactly basketball experts—coaches, managers, sportswriters—know about basketball.

Basketball experts clearly appreciate basketball. They understand the gestalt of the game, in the way that someone who has spent a lifetime thinking about and watching, say, modern dance develops an understanding of that art form. They're able to teach and coach and motivate; to make judgments and predictions about a player's character and resolve and stage of development. But the argument of "The Wages of Wins" is that this kind of expertise has real limitations when it comes to making precise evaluations of individual performance, whether you're interested in the consistency of football quarterbacks or in testing claims that N.B.A. stars "turn it on" during playoffs. The baseball legend Ty Cobb, the authors point out, had a lifetime batting average of .366, almost thirty points higher than the former San Diego Padres outfielder Tony Gwynn, who had a lifetime batting average of .338:

So Cobb hit safely 37 percent of the time while Gwynn hit safely on 34 percent of his at bats. If all you did was watch these players, could you say who was a better hitter? Can one
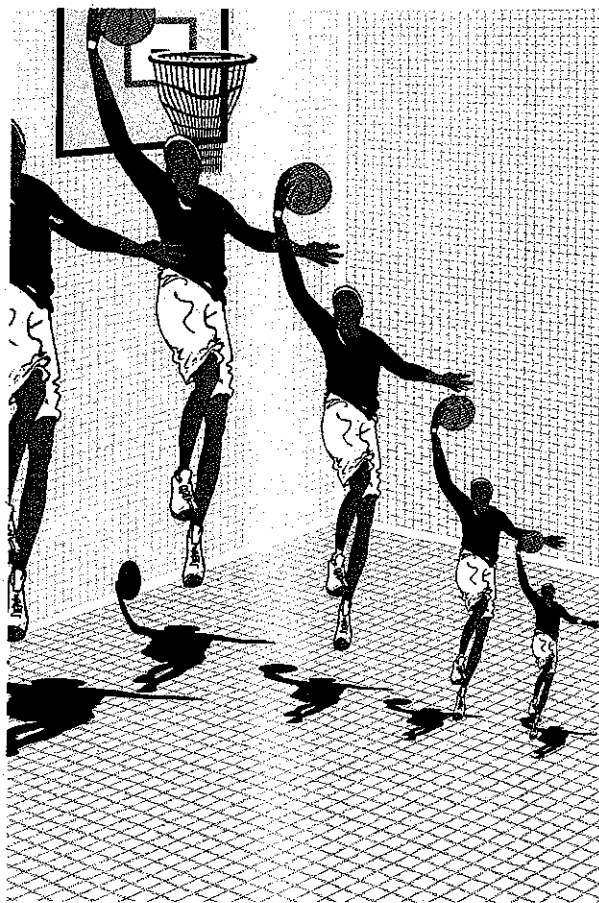
really tell the difference between 37 percent and 34 percent just staring at the players play? To see the problem with the non-numbers approach to player evaluation, consider that out of every 100 at bats, Cobb got three more hits than Gwynn. That's it, three hits.

Michael Lewis made a similar argument in his 2003 best-seller, "Moneyball," about how the so-called sabermetricians have changed the evaluation of talent in baseball. Baseball is sufficiently transparent, though, that the size of the discrepancies between intuitive and statistically aided judgment tends to be relatively modest. If you mistakenly thought that Gwynn was better than Cobb, you were still backing a terrific hitter. But "The Wages of Wins" suggests that when you move into more complex situations, like basketball, the limitations of "seeing" become enormous. Jermaine O'Neal, a center for the Indiana Pacers, finished third in the Most Valuable Player voting in 2004. His Win Score that year put him forty-fourth in the league. In 2004-05, the forward Antoine Walker made as much money as the point guard Jason Kidd, even though Walker produced 0.6 wins for Atlanta and Boston and Kidd produced nearly twenty wins for New Jersey. The Win Score algorithm suggests that Ray Allen has had nearly as good a career as Kobe Bryant, whom many consider the top player in the game, and that the journeyman forward Jerome Williams was actually among the strongest players of his generation.

Most egregious is the story of a young guard for the Chicago Bulls named Ben Gordon. Last season, Gordon finished second in the Rookie of the Year voting and was named the league's top "sixth man"—that is, the best non-starter—because he averaged an impressive 15.1 points per game in limited playing time. But Gordon rebounds less than he should, turns over the ball frequently, and makes such a low percentage of his shots that, of the N.B.A.'s top thirty-three scorers—that is, players who score at least one point for every two minutes

on the floor—Gordon's Win Score ranked him dead last.

The problem for basketball experts is that, in a situation with many variables, it's difficult to know how much weight to assign to each variable. Buying a house is agonizing because we look at the size, the location, the back yard, the proximity to local schools, the price, and so on, and we're unsure which of those things matters most. Assessing heart-attack risk is a notoriously difficult task for similar reasons. A doctor can analyze a dozen different factors. But how much weight should be given to a patient's cholesterol level relative to his blood pressure? In the face of such complexity, people construct their own arbitrary algorithms—they assume that every factor is of equal importance, or randomly elevate one or two factors for the sake of simplifying matters—and we make mistakes because those arbitrary algorithms are, well, arbitrary.

Berri, Schmidt, and Brook argue that the arbitrary algorithms of basketball experts elevate the number of points a player scores above all other consider-

Expert intuitions about performance can be wildly off.

ations. In one clever piece of research, they analyze the relationship between the statistics of rookies and the number of votes they receive in the All-Rookie Team balloting. If a rookie increases his scoring by ten per cent—regardless of how efficiently he scores those points—the number of votes he'll get will increase by twenty-three per cent. If he increases his rebounds by ten per cent, the number of votes he'll get will increase by six per cent. Every other factor, like turnovers, steals, assists, blocked shots, and personal fouls—factors that can have a significant influence on the outcome of a game—seemed to bear no statistical relationship to judgments of merit at all. It's not even the case that high scorers help their team by drawing more fans. As the authors point out, that's only true on the road. At home, attendance is primarily a function of games won. Basketball's decision-makers, it seems, are simply irrational.

It's hard not to wonder, after reading "The Wages of Wins," about the other instances in which we defer to the evaluations of experts. Boards of directors vote to pay C.E.O.s tens of millions of dollars, ostensibly because they believe—on the basis of what they have learned over the years by watching other C.E.O.s—that they are worth it. But so what? We see Allen Iverson, over and over again, charge toward the basket, twisting and turning and writhing through a thicket of arms and legs of much taller and heavier men—and all we learn is to appreciate twisting and turning and writhing. We become dance critics, blind to Iverson's dismal shooting percentage and his excessive turnovers, blind to the reality that the Philadelphia 76ers would be better off without him. "One can play basketball," the authors conclude. "One can watch basketball. One can both play and watch basketball for a thousand years. If you do not systematically track what the players do, and then uncover the statistical relationship between these actions and wins, you will never know why teams win and why they lose." ♦