CH 1
The Wisdom of Crowds
pp. 3 - 22

THE WISDOM OF CROWDS

WHY THE MANY ARE SMARTER THAN THE FEW
AND HOW COLLECTIVE WISDOM SHAPES BUSINESS,
ECONOMIES, SOCIETIES, AND NATIONS

JAMES SUROWIECKI

DOUBLEDAY
New York London Toronto Sydney Auckland

1.

THE WISDOM OF CROWDS

I

If, years hence, people remember anything about the TV game show *Who Wants to Be a Millionaire?*, they will probably remember the contestants' panicked phone calls to friends and relatives. Or they may have a faint memory of that short-lived moment when Regis Philbin became a fashion icon for his willingness to wear a dark blue tie with a dark blue shirt. What people probably won't remember is that every week *Who Wants to Be a Millionaire?* pitted group intelligence against individual intelligence, and that every week, group intelligence won.

Who Wants to Be a Millionaire? was a simple show in terms of structure: a contestant was asked multiple-choice questions, which got successively more difficult, and if she answered fifteen questions in a row correctly, she walked away with $1 million. The show's gimmick was that if a contestant got stumped by a question, she could pursue three avenues of assistance. First, she could have two of the four multiple-choice answers removed (so she'd have at least a fifty-fifty shot at the right response). Second, she could place a call to a friend or relative, a person whom, before the show, she had singled out as one of the smartest people she knew, and ask him or her for the answer. And third, she could poll the studio audience, which would immediately cast its votes by computer.

Everything we think we know about intelligence suggests that the smart individual would offer the most help. And, in fact, the "experts" did okay, offering the right answer—under pressure—almost 65 percent of the time. But they paled in comparison to the audiences. Those random crowds of people with nothing better to do on a weekday afternoon than sit in a TV studio picked the right answer 91 percent of the time.

Now, the results of *Who Wants to Be a Millionaire?* would never stand up to scientific scrutiny. We don't know how smart the experts were, so we don't know how impressive outperforming them was. And since the experts and the audiences didn't always answer the same questions, it's possible, though not likely, that the audiences were asked easier questions. Even so, it's hard to resist the thought that the success of the *Millionaire* audience was a modern example of the same phenomenon that Francis Galton caught a glimpse of a century ago.

As it happens, the possibilities of group intelligence, at least when it came to judging questions of fact, were demonstrated by a host of experiments conducted by American sociologists and psychologists between 1920 and the mid-1950s, the heyday of research into group dynamics. Although in general, as we'll see, the bigger the crowd the better, the groups in most of these early experiments—which for some reason remained relatively unknown outside of academia—were relatively small. Yet they nonetheless performed very well. The Columbia sociologist Hazel Knight kicked things off with a series of studies in the early 1920s, the first of which had the virtue of simplicity. In that study Knight asked the students in her class to estimate the room's temperature, and then took a simple average of the estimates. The group guessed 72.4 degrees, while the actual temperature was 72 degrees. This was not, to be sure, the most auspicious beginning, since classroom temperatures are so stable that it's hard to imagine a class's estimate being too far off base. But in the years that followed, far more convincing evidence emerged, as students and soldiers across America

were subjected to a barrage of puzzles, intelligence tests, and word games. The sociologist Kate H. Gordon asked two hundred students to rank items by weight, and found that the group's "estimate" was 94 percent accurate, which was better than all but five of the individual guesses. In another experiment students were asked to look at ten piles of buckshot—each a slightly different size than the rest—that had been glued to a piece of white cardboard, and rank them by size. This time, the group's guess was 94.5 percent accurate. A classic demonstration of group intelligence is the jelly-beans-in-the-jar experiment, in which invariably the group's estimate is superior to the vast majority of the individual guesses. When finance professor Jack Treynor ran the experiment in his class with a jar that held 850 beans, the group estimate was 871. Only one of the fifty-six people in the class made a better guess.

There are two lessons to draw from these experiments. First, in most of them the members of the group were not talking to each other or working on a problem together. They were making individual guesses, which were aggregated and then averaged. This is exactly what Galton did, and it is likely to produce excellent results. (In a later chapter, we'll see how having members interact changes things, sometimes for the better, sometimes for the worse.) Second, the group's guess will not be better than that of every single person in the group each time. In many (perhaps most) cases, there will be a few people who do better than the group. This is, in some sense, a good thing, since especially in situations where there is an incentive for doing well (like, say, the stock market) it gives people reason to keep participating. But there is no evidence in these studies that certain people consistently outperform the group. In other words, if you run ten different jelly-bean-counting experiments, it's likely that each time one or two students will outperform the group. But they will not be the same students each time. Over the ten experiments, the group's performance will almost certainly be the best possible. The simplest way to get reliably good answers is just to ask the group each time.

A similarly blunt approach also seems to work when wrestling with other kinds of problems. The theoretical physicist Norman L. Johnson has demonstrated this using computer simulations of individual "agents" making their way through a maze. Johnson, who does his work at the Los Alamos National Laboratory, was interested in understanding how groups might be able to solve problems that individuals on their own found difficult. So he built a maze—one that could be navigated via many different paths, some shorter, and some longer—and sent a group of agents into the maze one by one. The first time through, they just wandered around, the way you would if you were looking for a particular café in a city where you'd never been before. Whenever they came to a turning point—what Johnson called a "node"—they would randomly choose to go right or left. Therefore some people found their way, by chance, to the exit quickly, others more slowly. Then Johnson sent the agents back into the maze, but this time he allowed them to use the information they'd learned on their first trip, as if they'd dropped bread crumbs behind them the first time around. Johnson wanted to know how well his agents would use their new information. Predictably enough, they used it well, and were much smarter the second time through. The average agent took 34.3 steps to find the exit the first time, and just 12.8 steps to find it the second.

The key to the experiment, though, was this: Johnson took the results of all the trips through the maze and used them to calculate what he called the group's "collective solution." He figured out what a majority of the group did at each node of the maze, and then plotted a path through the maze based on the majority's decisions. (If more people turned left than right at a given node, that was the direction he assumed the group took. Tie votes were broken randomly.) The group's path was just nine steps long, which was not only shorter than the path of the average individual (12.8 steps), but as short as the path that even the smartest individual had been able to come up with. It was also as good an answer as you could find. There was no way to get through the maze in fewer

than nine steps, so the group had discovered the optimal solution. The obvious question that follows, though, is: The judgment of crowds may be good in laboratory settings and classrooms, but what happens in the real world?

II

At 11:38 AM on January 28, 1986, the space shuttle *Challenger* lifted off from its launch pad at Cape Canaveral. Seventy-four seconds later, it was ten miles high and rising. Then it blew up. The launch was televised, so news of the accident spread quickly. Eight minutes after the explosion, the first story hit the Dow Jones News Wire.

The stock market did not pause to mourn. Within minutes, investors started dumping the stocks of the four major contractors who had participated in the *Challenger* launch: Rockwell International, which built the shuttle and its main engines; Lockheed, which managed ground support; Martin Marietta, which manufactured the ship's external fuel tank; and Morton Thiokol, which built the solid-fuel booster rocket. Twenty-one minutes after the explosion, Lockheed's stock was down 5 percent, Martin Marietta's was down 3 percent, and Rockwell was down 6 percent.

Morton Thiokol's stock was hit hardest of all. As the finance professors Michael T. Maloney and J. Harold Mulherin report in their fascinating study of the market's reaction to the *Challenger* disaster, so many investors were trying to sell Thiokol stock and so few people were interested in buying it that a trading halt was called almost immediately. When the stock started trading again, almost an hour after the explosion, it was down 6 percent. By the end of the day, its decline had almost doubled, so that at market close, Thiokol's stock was down nearly 12 percent. By contrast, the stocks of the three other firms started to creep back up, and by the end of the day their value had fallen only around 3 percent.

What this means is that the stock market had, almost immediately, labeled Morton Thiokol as the company that was responsible for the *Challenger* disaster. The stock market is, at least in theory, a machine for calculating the present value of all the "free cash flow" a company will earn in the future. (Free cash flow is the money that's left over after a company has paid all its bills and its taxes, has accounted for depreciation, and has invested in the business. It's the money you'd get to take home and put in the bank if you were the sole owner of the company.) The steep decline in Thiokol's stock price—especially compared with the slight declines in the stock prices of its competitors—was an unmistakable sign that investors believed that Thiokol was responsible, and that the consequences for its bottom line would be severe.

As Maloney and Mulherin point out, though, on the day of the disaster there were no public comments singling out Thiokol as the guilty party. While the *New York Times* article on the disaster that appeared the next morning did mention two rumors that had been making the rounds, neither of the rumors implicated Thiokol, and the *Times* declared, "There are no clues to the cause of the accident."

Regardless, the market was right. Six months after the explosion, the Presidential Commission on the *Challenger* revealed that the O-ring seals on the booster rockets made by Thiokol—seals that were supposed to prevent hot exhaust gases from escaping—became less resilient in cold weather, creating gaps that allowed the gases to leak out. (The physicist Richard Feynman famously demonstrated this at a congressional hearing by dropping an O-ring in a glass of ice water. When he pulled it out, the drop in temperature had made it brittle.) In the case of the *Challenger,* the hot gases had escaped and burned into the main fuel tank, causing the cataclysmic explosion. Thiokol was held liable for the accident. The other companies were exonerated.

In other words, within a half hour of the shuttle blowing up, the stock market *knew* what company was responsible. To be sure,

this was a single event, and it's possible that the market's singling out of Thiokol was just luck. Or perhaps the company's business seemed especially susceptible to a downturn in the space program. Possibly the trading halt had sent a signal to investors to be wary. These all are important cautions, but there is still something eerie about what the market did. That's especially true because in this case the stock market was working as a pure weighing machine, undistorted by the factors—media speculation, momentum trading, and Wall Street hype—that make it a peculiarly erratic mechanism for aggregating the collective wisdom of investors. That day, it was just buyers and sellers trying to figure out what happened and getting it right.

How did they get it right? That's the question that Maloney and Mulherin found so vexing. First, they looked at the records of insider trades to see if Thiokol executives, who might have known that their company was responsible, had dumped stock on January 28. They hadn't. Nor had executives at Thiokol's competitors, who might have heard about the O-rings and sold Thiokol's stock short. There was no evidence that anyone had dumped Thiokol stock while buying the stocks of the other three contractors (which would have been the logical trade for someone with inside information). Savvy insiders alone did not cause that first-day drop in Thiokol's price. It was all those investors—most of them relatively uninformed—who simply refused to buy the stock.

But why did they not want Thiokol's stock? Maloney and Mulherin were finally unable to come up with a convincing answer to that question. In the end, they assumed that insider information was responsible for the fall in Thiokol's price, but they could not explain how. Tellingly, they quoted the Cornell economist Maureen O'Hara, who has said, "While markets appear to work in practice, we are not sure how they work in theory."

Maybe. But it depends on what you mean by "theory." If you strip the story down to its basics, after all, what happened that January day was this: a large group of individuals (the actual and po-

tential shareholders of Thiokol's stock, and the stocks of its competitors) was asked a question—"How much less are these four companies worth now that the *Challenger* has exploded?"—that had an objectively correct answer. Those are conditions under which a crowd's average estimate—which is, dollar weighted, what a stock price is—is likely to be accurate. Perhaps someone did, in fact, have inside knowledge of what had happened to the O-rings. But even if no one did, it's plausible that once you aggregated all the bits of information about the explosion that all the traders in the market had in their heads that day, it added up to something close to the truth. As was true of those who helped John Craven find the *Scorpion,* even if none of the traders was sure that Thiokol was responsible, collectively they were certain it was.

The market was smart that day because it satisfied the four conditions that characterize wise crowds: diversity of opinion (each person should have some private information, even if it's just an eccentric interpretation of the known facts), independence (people's opinions are not determined by the opinions of those around them), decentralization (people are able to specialize and draw on local knowledge), and aggregation (some mechanism exists for turning private judgments into a collective decision). If a group satisfies those conditions, its judgment is likely to be accurate. Why? At heart, the answer rests on a mathematical truism. If you ask a large enough group of diverse, independent people to make a prediction or estimate a probability, and then average those estimates, the errors each of them makes in coming up with an answer will cancel themselves out. Each person's guess, you might say, has two components: information and error. Subtract the error, and you're left with the information.

Now, even with the errors canceled out, it's possible that a group's judgment will be bad. For the group to be smart, there has to be at least some information in the "information" part of the "information minus error" equation. (If you'd asked a large group of

children to buy and sell stocks in the wake of the *Challenger* disaster, it's unlikely they would have picked out Thiokol as the culprit.) What is striking, though—and what makes a phrase like "the wisdom of crowds" meaningful—is just how much information a group's collective verdict so often contains. In cases like Francis Galton's experiment or the *Challenger* explosion, the crowd is holding a nearly complete picture of the world in its collective brain.

Perhaps this isn't surprising. After all, we are the products of evolution, and presumably we have been equipped to make sense of the world around us. But who knew that, given the chance, we can collectively make so *much* sense of the world. After all, think about what happens if you ask a hundred people to run a 100-meter race, and then average their times. The average time will not be better than the time of the fastest runners. It will be worse. It will be a mediocre time. But ask a hundred people to answer a question or solve a problem, and the average answer will often be at least as good as the answer of the smartest member. With most things, the average is mediocrity. With decision making, it's often excellence. You could say it's as if we've been programmed to be collectively smart.

III

Truly successful decision making, of course, demands more than just a picture of the world as it is. It demands in addition a picture of the world as it will (or at least as it may) be. Any decision-making mechanism therefore has to be good under conditions of uncertainty. And what's more uncertain than the future? Group intelligence may be good at telling how many jelly beans are in a jar or remembering the year Nirvana released *Nevermind.* But how does it perform under conditions of true uncertainty, when the right answer is seemingly unknowable—because it hasn't happened yet?

Robert Walker's entire career depends on the answer to that

question. Walker is the sports book director at the Mirage Hotel and Casino in Las Vegas, which means that every week he fields thousands of bets in sports ranging from pro football to Ivy League basketball. For all those games, Walker has to offer a line (or point spread), which lets bettors know which team is favored to win and by how many points. The way the line works is simple. Say the Giants are favored this week by three and a half points over the Rams. If you bet on the Giants, they have to win by four points or more for you to win the bet. Conversely, if you bet on the Rams, they have to lose by three points or less (or win), for you to walk away with the casino's money. In other sports, bets are framed in terms of odds: if you bet on the favorite, you might have to put down $150 to get $100 back, while if you bet on the underdog, you'd have to lay down $75 to win $100.

As a bookmaker, Walker's job is not to try to pick what team will win. He leaves that to the gamblers, at least in theory. Instead, his job is to make sure that the gamblers bet roughly the same amount of money on one team as on the other. If he does that, then he knows that he will win half the bets he's taken in and lose the other half. Why would Walker be satisfied with just breaking even? Because bookies make more money on every bet they win than they lose on every bet they get wrong. If you place a point-spread bet with a bookie, you have to put up $11 to win $10. Imagine there are only two bettors, one who bets on the favorite and the other who bets on the underdog. Walker takes in $22 ($11 from each of them). He pays out $21 to the winner. The $1 he keeps is his profit. That slim advantage, which is known as the vigorish, or the vig, is what pays the bookie's bills. And the bookie keeps that advantage only when he avoids having too much money riding on one side of a bet.

To keep that from happening, Walker needs to massage the point spread so that bets keep coming in for both teams. "The line we want is the line that'll split the public, because that's when you

start earning that vig," he said. In the week before the 2001 Super Bowl, for instance, the Mirage's opening line had the Baltimore Ravens favored by two and a half points. But soon after the line was posted, the Mirage booked a couple of early $3,000 bets on Baltimore. That's not much money, but it was enough to convince Walker to raise the point spread to three. If everyone wanted to bet on Baltimore, chances were the line wasn't right. So the line moved. The opening line is set by the bookmaker, but it shifts largely in response to what bettors do—much as stock prices rise and fall with investor demand.

In theory, you could set the opening line wherever, and simply allow it to adjust from there automatically, so that the point spread would rise or fall anytime there was a significant imbalance between the amounts wagered on each side. The Mirage would have no problem doing this; its computerized database tracks the bets as they come in. But bookies place a premium on making the opening line as accurate as possible, because if they set it badly they're going to get stuck taking a lot of bad bets. Once a line opens, though, it's out of the bookie's hands, and a game's point spread ends up representing bettors' collective judgment of what the final outcome of that game will be. As Bob Martin, who was essentially the country's oddsmaker in the 1970s, said, "Once you put a number on the board, it becomes public property."

The public, it turns out, is pretty smart. It does not have a crystal ball: point spreads only weakly predict the final scores of most NFL games, for instance. But it is very hard for even well-informed gamblers to beat the final spread consistently. In about half the games, favorites cover the spread, while in the other half underdogs beat the spread. This is exactly what a bookie wants to have happen. And there are no obvious mistakes in the market's judgment—like, say, home teams winning more than the crowd predicts they will, or road underdogs being consistently undervalued. Flaws in the crowd's judgment are found occasionally, but

when they are they're typically like the one documented in a recent paper that found that in weeks fifteen, sixteen, and seventeen of the NFL season, home underdogs have historically been a good bet. So you have to search hard to outperform the betting crowd. Roughly three-quarters of the time, the Mirage's final line will be the most reliable forecast of the outcomes of NFL games that you can find.

The same is true in many other sports. Because sports betting is a kind of ready-made laboratory to study predictions and their outcomes, a host of academics have perused gambling markets to see how efficient—that is, how good at capturing all the available information—they are. The results of their studies are consistent: in general, in most major sports the market is relatively efficient. In some cases, the crowd's performance is especially good: in horse racing, for instance, the final odds reliably predict the race's order of finish (that is, the favorite wins most often, the horse with the second-lowest odds is the second-most-often winner, and so on) and also provide, in economist Raymond D. Sauer's words, "reasonably good estimates of the probability of winning." In other words, a three-to-one horse will win roughly a quarter of the time. There are exceptions: odds are less accurate in those sports and games where the betting market is smaller and less liquid (meaning that the odds can change dramatically thanks to only a few bets), like hockey or golf or small-college basketball games. These are often the sports where professional gamblers can make real money, which makes sense given that we know the bigger the group, the more accurate it becomes. And there are also some interesting quirks: in horse racing, for instance, people tend to bet on long shots slightly more often than they should and bet on favorites slightly less often than they should. (This seems to be a case of risk-seeking behavior: bettors, especially bettors who have been losing, would rather take a flyer on a long shot that offers the possibility of big returns than grind it out by betting on short-odds favorites.) But on the whole, if bettors aren't collectively foreseeing the future, they're doing the next best thing.

IV

Recently I decided I needed—*this minute!*—the exact text of Bill Murray's *Caddyshack* riff about toting the Dalai Lama's golf bag. The punch line of the riff is "So I got that going for me, which is nice" and the Dalai Lama, in Murray's telling, likes to say "Gunga galunga." So I went to Google, the Internet search engine, typed in "going for me" and "gunga," and hit the search button. A list of 695 Web pages came back. First on the list was an article from GolfOnline, which included the second half of the riff. That was okay, but third on the list was a Web site for something called the Penn State Soccer Club. The goalie, a guy named David Feist, had posted the entire monologue. The search took 0.18 seconds.

Then I needed to check out the Mulherin paper on the *Challenger* that I discuss above. I couldn't remember the author's name, so I typed in " 'stock market' challenger reaction": 2,370 pages came back. The first one was an article by *Slate*'s Daniel Gross about the Mulherin paper. The third was Mulherin's own Web site, with a link to his paper. That search—which, remember, did not include Mulherin's name—took 0.10 seconds. A few minutes later, my search for the lyrics to a Ramones song about Ronald Reagan visiting the Bitburg cemetery took 0.23 seconds, and the first item on the list had what I needed.

If you use the Internet regularly, these examples of Google's performance will not surprise you. This is what we have come to expect from Google: instantaneous responses with the exact page we need up high in the rankings. But if possible, it's worth letting yourself be a little amazed at what happened during those routine searches. Each time, Google surveyed billions of Web pages and

picked exactly the pages that I would find most useful. The cumulative time for all the searches: about a minute and a half.

Google started in 1998, at a time when Yahoo! seemed to have a stranglehold on the search business—and if Yahoo! stumbled, then AltaVista or Lycos looked certain to be the last man standing. But within a couple of years, Google had become the default search engine for anyone who used the Internet regularly, simply because it was able to do a better job of finding the right page quickly. And the way it does that—and does it while surveying three billion Web pages—is built on the wisdom of crowds.

Google keeps the details of its technology to itself, but the core of the Google system is the PageRank algorithm, which was first defined by the company's founders, Sergey Brin and Lawrence Page, in a now-legendary 1998 paper called "The Anatomy of a Large-Scale Hypertextual Web Search Engine." PageRank is an algorithm—a calculating method—that attempts to let all the Web pages on the Internet decide which pages are most relevant to a particular search. Here's how Google puts it:

> PageRank capitalizes on the uniquely democratic characteristic of the web by using its vast link structure as an organizational tool. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. Google assesses a page's importance by the votes it receives. But Google looks at more than sheer volume of votes, or links; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

In that 0.12 seconds, what Google is doing is asking the entire Web to decide which page contains the most useful information, and the page that gets the most votes goes first on the list. And that page, or the one immediately beneath it, more often than not is in fact the one with the most useful information.

Now, Google is a republic, not a perfect democracy. As the description says, the more people that have linked to a page, the more influence that page has on the final decision. The final vote is a "weighted average"—just as a stock price or an NFL point spread is—rather than a simple average like the ox-weighers' estimate. Nonetheless, the big sites that have more influence over the crowd's final verdict have that influence only because of all the votes that smaller sites have given them. If the smaller sites were giving the wrong sites too much influence, Google's search results would not be accurate. In the end, the crowd still rules. To be smart at the top, the system has to be smart all the way through.

V

If allowing people to bet on sporting events effectively creates a kind of machine that's good at predicting the outcome of those events, an obvious question follows: Wouldn't people betting on other kinds of events be equally good, as a group, at predicting them? Why confine ourselves to knowing what the chances are of Los Angeles beating Sacramento if there's a way we could know what the chances are of, say, George W. Bush beating John Kerry?

We do have a well-established way of knowing what George W. Bush's chances are: the poll. If you want to know how people are going to vote, you just ask them. Polling is, relatively speaking, accurate. It has a solid methodology behind it, and is statistically rigorous. But there's reason to wonder if a market such as the betting market—one that allowed the people participating in it to rely on many different kinds of information, including but not limited to polls—might at the very least offer a competitive alternative to Gallup. That's why the Iowa Electronic Markets (IEM) project was created.

Founded in 1988 and run by the College of Business at the University of Iowa, the IEM features a host of markets designed to predict the outcomes of elections—presidential, congressional, gu-

bernatorial, and foreign. Open to anyone who wants to participate, the IEM allows people to buy and sell futures "contracts" based on how they think a given candidate will do in an upcoming election. While the IEM offers many different types of contracts, two are most common. One is designed to predict the winner of an election. In the case of the California recall in 2003, for instance, you could have bought an "Arnold Schwarzenegger to win" contract, which would have paid you $1 when Schwarzenegger won. Had he lost, you would have gotten nothing. The price you pay for this kind of contract reflects the market's judgment of a candidate's chances of victory. If a candidate's contract costs 50 cents, it means, roughly speaking, that the market thinks he has a 50 percent chance of winning. If it costs 80 cents, he has an 80 percent chance of winning, and so on.

The other major kind of IEM contract is set up to predict what percentage of the final popular vote a candidate will get. In this case, the payoffs are determined by the vote percentage: if you'd bought a George W. Bush contract in 2000, you would have received 48 cents (he got 48 percent of the vote) when the election was over.

If the IEM's predictions are accurate, the prices of these different contracts will be close to their true values. In the market to predict election winners, the favorite should always win, and bigger favorites should win by bigger margins. Similarly, in the vote-share market, if George W. Bush were to end up getting 49 percent of the vote in 2004, then the price of a George W. Bush contract in the run-up to the election should be close to 49 cents.

So how has the IEM done? Well, a study of the IEM's performance in forty-nine different elections between 1988 and 2000 found that the election-eve prices in the IEM were, on average, off by just 1.37 percent in presidential elections, 3.43 percent in other U.S. elections, and 2.12 percent in foreign elections. (Those numbers are in absolute terms, meaning that the market would have been off by 1.37 percent if, say, it had predicted that Al Gore would get 48.63 percent of the vote when in reality he got 50 percent.)

The IEM has generally outperformed the major national polls, and has been more accurate than them even months in advance of the actual election. Over the course of the presidential elections between 1988 and 2000, for instance, 596 different polls were released. Three-fourths of the time, the IEM's market price on the day each of those polls was released was more accurate. Polls tend to be very volatile, with vote shares swinging wildly up and down. But the IEM forecasts, though ever-changing, are considerably less volatile, and tend to change dramatically only in response to new information. That makes them more reliable as forecasts.

What's especially interesting about this is that the IEM isn't very big—there have never been more than eight hundred or so traders in the market—and it doesn't, in any way, reflect the makeup of the electorate as a whole. The vast majority of traders are men, and a disproportionate—though shrinking—number of them are from Iowa. So the people in the market aren't predicting their own behavior. But their predictions of what the voters of the country will do are better than the predictions you get when you ask the voters themselves what they're going to do.

The IEM's success has helped inspire other similar markets, including the Hollywood Stock Exchange (HSX), which allows people to wager on box-office returns, opening-weekend performance, and the Oscars. The HSX enjoyed its most notable success in March of 2000. That was when a team of twelve reporters from *The Wall Street Journal* assiduously canvassed members of the Academy of Motion Pictures Arts and Sciences in order to find out how they had voted. The Academy was not happy about this. The organization's president publicly attacked the *Journal* for trying to "scoop us before Oscar night," and the Academy urged members not to talk to reporters. But with the *Journal* promising anonymity, more than a few people—356, or about 6 percent of all members—disclosed how they had filled out their ballots. The Friday before the ceremony, the *Journal* published its results, forecasting the winners in the six major Oscar categories—Best Picture, Best Direc-

tor, Best Actor and Best Actress, Best Supporting Actor and Best Supporting Actress. And when the envelopes were opened, the *Journal*'s predictions—much to the Academy's dismay—turned out to be pretty much on target, with the paper picking five of the six winners. The HSX, though, had done even better, getting all six of the six right. In 2002, the exchange, perhaps even more impressively, picked thirty-five of the eventual forty Oscar nominees.

The HSX's box-office forecasts are not as impressive or as accurate as the IEM's election forecasts. But Anita Elberse, a professor of marketing at Harvard Business School, has compared the HSX's forecasts to other Hollywood prediction tools, and found that the HSX's closing price the night before a movie opens is the single best available forecast of its weekend box office. As a result, the HSX's owner, Cantor Index Holdings, is now marketing its data to Hollywood studios.

One of the interesting things about markets like the IEM and the HSX is that they work fairly well without much—or any—money at stake. The IEM is a real-money market, but the most you can invest is $500, and the average trader has only $50 at stake. In the HSX, the wagering is done entirely with play money. All the evidence we have suggests that people focus better on a decision when there are financial rewards attached to it (which may help explain why the IEM's forecasts tend to be more accurate). But David Pennock—a researcher at Overture who has studied these markets closely—found that, especially for active traders in these markets, status and reputation provided incentive enough to encourage a serious investment of time and energy in what is, after all, a game.

As the potential virtues of these decision markets have become obvious, the range of subjects they cover has grown rapidly. At the NewsFutures and TradeSports exchanges, people could bet, in the fall of 2003, on whether or not Kobe Bryant would be convicted of sexual assault, on whether and when weapons of mass destruction would be found in Iraq, and on whether Ariel Sharon

would remain in power longer than Yasir Arafat. Ely Dahan, a professor at UCLA, has experimented with a classroom-decision market in which students bought and sold securities representing a variety of consumer goods and services, including SUVs, ski resorts, and personal digital assistants. (In a real-life market of this kind, the value of a security might depend on the first-year sales of a particular SUV.) The market's forecasts were eerily similar to the predictions that conventional market research had made (but the classroom research was much cheaper). In the fall of 2003, meanwhile, MIT's *Technology Review* set up a site called Innovation Futures, where people could wager on future technological developments. And Robin Hanson, an economics professor at George Mason University who was one of the first to write about the possibility of using decision markets in myriad contexts, has suggested that such markets could be used to guide scientific research and even as a tool to help governments adopt better policies.

Some of these markets will undoubtedly end up being of little use, either because they'll fail to attract enough participants to make intelligent forecasts or because they'll be trying to predict the unpredictable. But given the right conditions and the right problems, a decision market's fundamental characteristics—diversity, independence, and decentralization—are guaranteed to make for good group decisions. And because such markets represent a relatively simple and quick means of transforming many diverse opinions into a single collective judgment, they have the chance to improve dramatically the way organizations make decisions and think about the future.

In that sense, the most mystifying thing about decision markets is how little interest corporate America has shown in them. Corporate strategy is all about collecting information from many different sources, evaluating the probabilities of potential outcomes, and making decisions in the face of an uncertain future. These are tasks for which decision markets are tailor-made. Yet companies have remained, for the most part, indifferent to this

source of potentially excellent information, and have been surprisingly unwilling to improve their decision making by tapping into the collective wisdom of their employees. We'll look more closely at people's discomfort with the idea of the wisdom of crowds, but the problem is simple enough: just because collective intelligence is real doesn't mean that it will be put to good use.

A DECISION MARKET is an elegant and well-designed method for capturing the collective wisdom. But the truth is that the specific method that one uses probably doesn't matter very much. In this chapter, we've looked at a host of different ways of tapping into what a group knows: stock prices, votes, point spreads, pari-mutuel odds, computer algorithms, and futures contracts. Some of these methods seem to work better than others, but in the end there's nothing about a futures market that makes it inherently smarter than, say, Google or a pari-mutuel pool. These are all attempts to tap into the wisdom of the crowd, and that's the reason they work. The real key, it turns out, is not so much perfecting a particular method, but satisfying the conditions—diversity, independence, and decentralization—that a group needs to be smart. As we'll see in the chapters that follow, that's the hardest, but also perhaps the most interesting, part of the story.