LEGG MASON
CAPITAL MANAGEMENT

Mauboussin on STRATEGY

March 20, 2007

Michael J. Mauboussin

mmauboussin@lmcm.com

Explaining the Wisdom of Crowds

Applying the Logic of Diversity

Understanding diversity and leveraging its potential requires deeper understanding than we currently possess. We won't get far with compelling anecdotes and metaphors . . . We need a logic of diversity.

Scott Page
The Difference [1]

Source: istockphoto.

- **Much of the debate about the wisdom of crowds has relied on anecdotes.**
- **Scott Page's new book, *The Difference*, provides a framework for the logic of diversity.**
- **Categorizing a problem is a crucial first step in determining how best to solve it.**
- **We show how the wisdom of crowds works for three problem types, and provide detailed examples for two problems.**

"It's a Mystery"—Not

Debate about the wisdom of crowds—the idea a collective can solve problems better than most individuals within the group, including experts—has percolated in recent years. While enthusiasts [2] and detractors [3] have made their case, much of the marshaled evidence is anecdotal. Even when the idea's supporters specify the necessary conditions for the wisdom of crowds to succeed, there is rarely discussion of *how* it works. In an approving 2006 *New York Times* article, columnist Joe Nocera explained collective accuracy by plucking a Hollywood movie line: "It's a mystery." [4]

Fortunately, Scott Page's important new book, *The Difference*, introduces some much-needed rigor into why collectives do well and why they fail, why experts are often inferior to the crowd, and why diversity is important. Page not only carefully defines his terms, he also uses mathematical models to develop and apply theorems. These theorems illustrate the logic of diversity, removing a good deal of mystery from the wisdom of crowds.

In this discussion, we apply Page's models to three types of problems and provide real-world examples and data for a pair of them. Before moving into the analysis, three points bear emphasis: the importance of recognizing the problem type, the conditions under which the wisdom of crowds works, and why these ideas are so important for investors and decision makers.

Problem Type: Expert or Crowd?

First, understanding the type of problem you face is a crucial step in figuring out how to best solve it. For example, while companies often talk about the value of diversity within an organization, diversity is of no help—and indeed may be a hindrance—in solving many problems. If your plumbing is in need of repair, you'll be better off with a plumber than an English lit major, a Peace Corps volunteer, and an astrophysicist working together. Diversity typically becomes more important when the problem is complex and specifiable rules cannot solve it.

We address three distinct problems. The first is what we call a needle-in-the-haystack problem. Here, some people in the crowd know the answer while many, if not most, don't. The second is a state estimation problem, where one person knows the answer but the group does not. Finally, there is a prediction problem, where the answer has yet to be revealed.

Each problem type has a distinct set of issues, and the discussion of the wisdom of crowds often incorrectly conflates the problem types. So matching the problem type with the best means to solve the problem is a crucial, and almost always overlooked, step. [5]

Conditions for the Crowd to Be Wise

Even if you determine a collective is the best means to solve a problem, certain conditions must prevail for the crowd to be smart. These include diversity, aggregation, and incentives.

You can think of diversity as cognitive differences. Page unpacks diversity into four frameworks: [6]

- Perspectives: ways of representing situations and problems
- Interpretations: ways of categorizing or partitioning perspectives
- Heuristics: ways of generating solutions to problems
- Predictive Models: ways of inferring cause and effect

More often than not, when organizations discuss diversity, they refer to social identity diversity—gender, race, religion, age, etc. There's good reason to believe social identity diversity correlates with cognitive diversity, but they are certainly not the same. The ultimate goal is cognitive diversity. [7]

Aggregation means there's a way to bring the group's information together. Stock exchanges are a good example of aggregation mechanisms, as are polls and voting. Diversity without aggregation results in unused problem-solving potential.[8]

The final condition is incentives. The basic idea is you are rewarded for being right and penalized for being wrong. The payoffs can be monetary, as they are in markets, but they need not be. Payoffs can be measured with reputation, or even fitness (an ability to survive and propagate).

Incentives serve to grease the skids of collective accuracy in a couple of ways. [9] The first, which is very relevant for markets, is self selection. Active money managers tend to buy or sell when they believe they have an investment edge. This serves to improve predictive accuracy. [10]

The second way relates to rewards—your payoff is higher if your bet is away from the crowd. Just as a horse race handicapper makes more money by correctly anticipating a win by a long shot than by a favorite, so too investors earn higher returns by making non-consensus bets. This encourages greater diversity.

Why Should I Care?

This discussion of problem types, diversity, and collectives may seem far afield from what investors think and talk about from day to day. Are these concepts relevant for someone trying to generate excess returns? Our answer is an unmitigated yes, for lots of reasons.

To begin, the wisdom of crowds appears to be a viable and robust way to explain market behavior. At a high level, market efficiency prevails when the wisdom of crowds conditions are in place. Said differently, the market yields prices the economic textbooks predict without the constrictive assumptions associated with most economic models. [11] When the conditions are violated, markets can and will periodically veer from efficiency, accommodating inefficiencies, and even booms and crashes. This approach specifies the conditions under which markets will be efficient or inefficient.

This discussion also allows for a deeper understanding of how to solve problems—especially complex problems. It also highlights pitfalls that decision makers commonly fall into, especially in committee settings or in the process of deliberation. So these ideas may not only lead to better decisions but also to fewer decision-making failures.

Finally, these ideas have tremendous implications for organizations—from whom to hire, to how to create teams, to when to leverage collective wisdom. Managers can shed superficial ideas about diversity and think proactively about how to leverage diversity's value.

Problem One: Who Wants to Be a Millionaire?

Now let's turn to our problems. The first is a needle-in-a-haystack problem, where there is an answer, and some members of the crowd know what it is. It's like playing *Trivial Pursuit* with a huge group: some people are likely to know the answer to a particular question, and it will be different people for different questions. Diversity's value is easy to see in this context. What's remarkable is it doesn't take many people knowing the answer—or even having a better than random chance to guess the right answer—for the correct answer to emerge.

Jim Surowiecki presents a neat example of this problem based on the TV show *Who Wants to Be a Millionaire?* [12] In the show, a contestant is asked a series of consecutively difficult multiple-choice questions, with a payoff of $1 million for getting them all right. The producers added spice to the show by allowing baffled contestants to choose one of three options to help answer a question: eliminate two of the four possible answers (offering the contestant a fifty-fifty chance), call a predetermined "expert" for counsel, or poll the studio audience.

When called on, the experts provided the right answer a respectable two-thirds of the time. More surprising was that the audience—a group of folks with nothing better to do on a weekday afternoon—returned the correct answer over 90 percent of the time. The crowd smoked the expert.

How can we explain this result? Law professor Cass Sunstein offers one possibility based on the Condorcet Jury Theorem. [13] In its simplest form, the Condorcet Jury Theorem holds that if an average group member has better than a 50 percent chance of knowing the right answer and the answer is tabulated using majority rule, the probability of a correct answer rises toward 100 percent as the group size increases.

The Condorcet Jury Theorem has some useful applications in the social sciences, but we don't believe this is one of them. It turns out we don't need assumptions anywhere near as strong as those in the theorem to show why the crowd is smart in this instance.

To illustrate the point, we borrow Page's example from *The Difference*. [14] He hypothetically presents this question to a crowd:

Which person from the following list was not a member of the Monkees (a 1960s pop band)?

(A) Peter Tork
(B) Davy Jones
(C) Roger Noll
(D) Michael Nesmith

The non-Monkee is Roger Noll, a Stanford economist. Now imagine a crowd of 100 people with knowledge distributed as follows:

- 7 know all 3 of the Monkees
- 10 know 2 of the Monkees
- 15 know 1 of the Monkees
- 68 have no clue

In other words, less than 10 percent of the crowd knows the answer, and over two-thirds are culturally deprived of any Monkees knowledge. We assume individuals without the answer vote randomly. The Condorcet Jury Theorem, then, doesn't apply because only a small minority knows the answer.

In this case, the crowd will have no problem fingering Noll as the non-Monkee. Here's the breakdown:

- The 7 who know all the Monkees vote for Noll;
- 5 of the 10 who know 2 of the Monkees will vote for Noll;
- 5 of the 15 who know 1 of the Monkees will vote for Noll; and
- 17 of the 68 clueless will vote for Noll.

So Noll will garner 34 votes, versus 22 votes for each of the other choices. The crowd easily identifies the non-Monkee. In plain words, random errors cancel out and the correct answer rises to the surface. We could add even more clueless people without violating the result: while the percentage margin by which Noll wins would decline, he would be the selection nonetheless.

Now if it were this easy, the crowd would always get 100 percent instead of 90 percent. Two variables are key: the percentage of the crowd who know the answer and the degree of randomness in the answers. Of the two, randomness is more important than accuracy: a surprisingly small percentage of the population can know the answer and the group will be right with high randomness. Deviations from randomness will create less-than-perfect crowd answers, albeit still very good ones.

This type of collective problem solving is not limited to humans. Biologists have shown similar mechanisms guide decision making among certain animals, notably schooling fish and bee hives. The scientists observe that large groups require a small proportion of informed individuals to guide the group, and that only a very small proportion of the group needs to be informed for the group as a whole to achieve great accuracy. That this decision-making approach has stood evolution's test underscores its robustness. [15]

In the real world problems like this do exist, but collective voting is rarely used to solve them. However, technology has made this type of problem solving much easier. Search engines, which use rankings based on the wisdom of crowds, answer most routine questions like the identities of the Monkees very quickly and effectively. [16]

Sites like Innocentive.com create an exchange to address more challenging questions. Innocentive allows "seeker" companies to pose technical scientific questions to a population of "solver" scientists. The site matches a company's research needs with a large population of qualified scientists, increasing the likelihood the company will find a cost-effective solution.

Estimating a State: The Diversity Prediction Theorem and Jelly Beans

We now turn to the second type of problem, estimating a state. Here, only one person knows the answer and none of the problem solvers do. A classic example of this problem is asking a group to guess the number of jelly beans in a jar. We have been doing this experiment for over a decade at Columbia Business School, and the collective answer has proven remarkably accurate in most trials.

To explain why the jelly bean experiment works, we turn to one of the central ideas in Page's book, what he calls the *diversity prediction theorem*. [17] The theorem states:

Collective error = average individual error – prediction diversity

The mathematical foundation for the theorem is the use of squared errors as a measure of accuracy. Researchers in the social sciences and statistics frequently use squared errors, which have the benefit of avoiding negative and positive errors cancelling out. [18]

Average individual error combines the squared errors of all of the participants. In plain language, it captures the average accuracy of the individual guesses.

Prediction diversity combines the squared difference between the individuals and the average guess. Simply, it reflects the dispersion of guesses, or how different they are.

The collective error, of course, is simply the difference between the correct answer and the average guess. Page treats the diversity prediction theorem in depth in his book, and includes numerous examples.

The diversity prediction theorem has some crucial implications. The first is a diverse crowd will always predict more accurately than the average individual. So the crowd predicts better than the people in it. Not sometimes. Always.

Second, collective predictive ability is equal parts accuracy and diversity. [19] You can reduce collective error by either increasing accuracy by a unit or by increasing diversity by a unit. Both are essential.

Finally, while not a formal implication of the theorem, it is true that the collective is often better than even the best of the individuals. So a diverse collective *always* beats the average individual, and *frequently* beats everyone. And the individuals who do beat the collective generally change, suggesting they are more of a statistical vestige than super-smart people.

Because the theorem is based on math—and pretty basic math at that—it's always true. Still, the theorem's implications are not necessarily intuitive. That the crowd is better than we are is not a comforting thought.

Our 2007 jelly bean results illustrate the point. The average guess of the class was 1,151 while the actual number of beans was 1,116, a 3.1 percent error. Of the 73 estimates, only two were better than the average. Appendix A provides more detail on the diversity prediction theorem and its application to the jelly bean experiment we conducted this year. There's nothing unique about 2007; the results are the same year after year.

Prediction: "I'd Like to Thank…"

The final problem deals with prediction, where the answer is unknown and will be revealed in the future. Our example here is another experiment with the Columbia Business School students.

Here's what we do. A few weeks prior to the Academy Awards ceremony, we distribute a two-sided form. On the front page are the six most popular Academy Awards categories:

Best actor
Best actress
Best supporting actor
Best supporting actress
Best film
Best director

On the back are six less visible categories:

Best adapted screenplay
Best cinematography
Best film editing
Best music (original score)
Best documentary
Best art direction

We ask the students to contribute $1 to a pot (with the winner getting the proceeds—incentive) and to select independently who they believe will win each category. The goal is to win the pot, not to reveal sentimental favorites.

The 2007 results show the diversity prediction theorem at work. The consensus, defined as the modal selection in each category, got 11 of 12 correct, including all 6 on the back page. Two students tied for the most correct answers, each getting 9 of 12 correct, and the average student got just 5 of 12 right.

While we can't tie the collective error back directly to the 11 of 12 accuracy because of the multiple selections per category, the diversity prediction theorem illustrates how accuracy and diversity combine to produce a crowd-beating answer. In Appendix B, we provide details of the student guesses and show how the collective error does link precisely to the answer for each category.

This problem is almost like a combination of the first two. Like the Monkees example, some people probably have better predictive models than others (i.e., they know their pop culture), hence allowing the most likely answer to rise to the surface. The answer emerges as we combine a lot of diversity with a little predictive accuracy.

While the logic of diversity certainly provides some important insights and useful models, we by no means fully understand the wisdom of crowds. For example, while it is clear the collective will do better than the average individual, it's not clear why the collective is often so accurate. But these models and examples provide a concrete step in the right direction, and allow us to dismiss other explanations like the Condorcet Jury Theorem and the square-root law.

The Diversity Prediction Theorem and the Stock Market

So what does all of this mean for investors in the stock market? Here are some thoughts:

- *Wisdom of crowds and market efficiency*. A number of leading finance academics, including William Sharpe, Richard Roll, and Jack Treynor, have pointed to the wisdom of crowds as a plausible explanation for market efficiency. [20] We are enthusiastic advocates for this view as well. [21] The value of this approach is it reveals the conditions under which markets are likely to be efficient or inefficient.

- *The importance of cognitive diversity*. If correct, this approach underscores the major role of cognitive diversity. Diversity can be beneficial on an organizational (company, club, school) or an individual level. Empirical research shows that cognitively-diverse individuals outperform cognitively-focused individuals. [22] Conversely, homogenous investor behavior—a well-studied topic in social psychology and sociology—can lead to diversity breakdowns and a large collective error. [23]

- *There are no answers in markets*. Except in cases of mergers or acquisitions, there are no objectively correct answers in the stock market. So it's impossible to apply the diversity prediction theorem directly to the market. However, the theorem is of great value even if it offers some insight into the market's mechanism, including a well-placed emphasis on accuracy and diversity.

- *Risk*. While there is clearly a long-term relationship between risk and reward, the market's perception of risk vacillates. Introducing diversity may provide an important dimension of risk that currently lies outside the traditional mean-variance framework. Further, diversity reductions have a non-linear impact on the systems they operate within, adding to the analytical challenge. [24]

- *Behavioral finance*. General equilibrium models, the foundation for the mean-variance framework, assume agent rationality. Over the past few decades, behavioral finance has emerged as a research area that weds psychology and economics. Researchers have found that individuals often behave in suboptimal, albeit often predictable, ways. This assault on rationality is a challenge to classic theory. However, provided suboptimal behavior leads to diversity, markets can still be collectively smart even as investors are individually suboptimal. This condition is only violated if nearly all investors act in unison, which happens infrequently.

- *Divergence of opinion as a proxy for diversity*? One potentially fruitful line of research links the dispersion of analyst forecasts (a proxy for differences of opinion) with subsequent stock returns. The research shows stocks with greater dispersion have lower subsequent returns because only the most optimistic investors, those who by definition place a high value on the company, trade the stock, and more pessimistic investors do not trade (the model assumes supply is finite). Stated differently, diversity is clipped, leading to greater collective error in the form of overpricing. [25]

- *Humility*. The logic of diversity shows that in solving hard problems, the crowd is almost always going to be better than most people if diversity, aggregation, and incentives are operative. Investors aware of this reality will remain humble, while seeking occasions when the crowd's wisdom gives way to whim, and hence opportunity.

Appendix A: Jelly Beans and the Ox

The Jelly Bean Experiment

In January 2007, 73 students independently guessed the number of jelly beans in a jar. There was a $20 reward offered for the best guess, and a $5 penalty for the guess farthest from the correct answer.

Column A shows the individual guesses. The mean of these guesses, the consensus, was 1,151. The actual number of jelly beans was 1,116. So the consensus was off by 35 beans, or 3.1 percent.

We can tie these results to Page's diversity prediction theorem, which states:

Collective error = average individual error – prediction diversity

Statisticians square errors [e.g., $(-5)^{2} + (5)^{2} = 50$] to make sure positive and negative errors don't cancel out (e.g., -5 + 5 = 0).

Let's run through an example with Student 1. Her guess (Column A) was 250. Since the actual number was 1,116, her difference from actual (Column B) was -866. We then square -866 to get 749,956 (Column C). We calculate this squared difference from actual for each student, and take the average for the whole class. This is the *average individual error*. In this experiment, the average individual error was 490,949. The more accurate the individual guesses, the smaller the average individual error.

Next, we compare Student 1's guess (250) with the class's average guess (1,151). Her difference from average (Column D) was -901. We square -901 to get 811,801 (Column E). Again, we calculate the squared difference of the average for each student and then take the average for the class. This is *prediction diversity*. In this experiment, the prediction diversity was 489,692. The more dispersed the guesses, the larger the prediction diversity.

We can now bring together the individual error and prediction diversity to calculate the collective error:

Collective error = average individual error – prediction diversity

Collective error = 490,949 – 489,692

Collective error = 1,258

Note the square root of the collective error, $\sqrt{1{,}258}$, is approximately 35, or the difference between the consensus guess and the actual number of jelly beans in the jar.

Student	COLUMN A Guess	COLUMN B "Difference From Actual"	COLUMN C Squared "Difference From Actual"	COLUMN D "Difference From Average"	COLUMN E Squared "Difference From Average"
1	250	-866	749,956	-901	811,801
2	315	-801	641,601	-836	698,896
3	399	-717	514,089	-752	565,504
4	400	-716	512,656	-751	564,001
5	420	-696	484,416	-731	534,361
6	437	-679	461,041	-714	509,796
7	479	-637	405,769	-672	451,584
8	500	-616	379,456	-651	423,801
9	540	-576	331,776	-611	373,321
10	585	-531	281,961	-566	320,356
11	600	-516	266,256	-551	303,601
12	600	-516	266,256	-551	303,601
13	604	-512	262,144	-547	299,209
14	616	-500	250,000	-535	286,225
15	624	-492	242,064	-527	277,729
16	632	-484	234,256	-519	269,361
17	645	-471	221,841	-506	256,036
18	650	-466	217,156	-501	251,001
19	651	-465	216,225	-500	250,000
20	699	-417	173,889	-452	204,304
21	721	-395	156,025	-430	184,900
22	723	-393	154,449	-428	183,184
23	734	-382	145,924	-417	173,889
24	750	-366	133,956	-401	160,801
25	750	-366	133,956	-401	160,801
26	750	-366	133,956	-401	160,801
27	750	-366	133,956	-401	160,801
28	768	-348	121,104	-383	146,689
29	780	-336	112,896	-371	137,641
30	800	-316	99,856	-351	123,201
31	800	-316	99,856	-351	123,201
32	820	-296	87,616	-331	109,561
33	850	-266	70,756	-301	90,601
34	874	-242	58,564	-277	76,729
35	876	-240	57,600	-275	75,625
36	900	-216	46,656	-251	63,001
37	900	-216	46,656	-251	63,001
38	900	-216	46,656	-251	63,001
39	1,000	-116	13,456	-151	22,801
40	1,000	-116	13,456	-151	22,801
41	1,008	-108	11,664	-143	20,449
42	1,120	4	16	-31	961
43	1,120	4	16	-31	961
44	1,152	36	1,296	1	1
45	1,234	118	13,924	83	6,889
46	1,234	118	13,924	83	6,889
47	1,250	134	17,956	99	9,801
48	1,250	134	17,956	99	9,801
49	1,260	144	20,736	109	11,881
50	1,288	172	29,584	137	18,769
51	1,300	184	33,856	149	22,201
52	1,400	284	80,656	249	62,001
53	1,500	384	147,456	349	121,801
54	1,500	384	147,456	349	121,801
55	1,500	384	147,456	349	121,801
56	1,523	407	165,649	372	138,384
57	1,564	448	200,704	413	170,569
58	1,575	459	210,681	424	179,776
59	1,580	464	215,296	429	184,041
60	1,583	467	218,089	432	186,624
61	1,588	472	222,784	437	190,969
62	1,700	584	341,056	549	301,401
63	1,732	616	379,456	581	337,561
64	1,872	756	571,536	721	519,841
65	1,896	780	608,400	745	555,025
66	1,899	783	613,089	748	559,504
67	1,963	847	717,409	812	659,344
68	2,000	884	781,456	849	720,801
69	2,250	1,134	1,285,956	1,099	1,207,801
70	3,000	1,884	3,549,456	1,849	3,418,801
71	3,000	1,884	3,549,456	1,849	3,418,801
72	3,024	1,908	3,640,464	1,873	3,508,129
73	4,100	2,984	8,904,256	2,949	8,696,601
Average	1,151		490,949		489,692

"Actual" # of Jelly Beans	1,116
"Average Guess" of # Jelly Beans	1,151
Average Individual Error (Average of Column C)	490,949
Prediction Diversity (Average of Column E)	489,692
Collective Error ("Average of Column A"-"Actual")^2	1,258

CHECKS:

Collective Error = Average Individual Error - Prediction Diversity
1,258 = 490,949 - 489,692

√Collective Error = ABS ["Average Guess" - "Actual"]
√1,258 = ABS [1,151 - 1,116] ≈ 35

The Ox-Weighing Contest

Jim Surowiecki opens *The Wisdom of Crowds* with the story of Francis Galton and an ox-weighing contest. Over a century ago, Galton observed an ox-weighing contest at a county fair, where nearly 800 participants paid a sixpenny fee to guess the ox's weight in the hope of winning a prize for the best guess.

As Surowiecki relates, Galton found the average guess to be 1,197 pounds, nearly identical to the ox's actual weight of 1,198 pounds. What's more, Galton provided a subset of the guess data, allowing us to apply the diversity prediction theorem.

While the ox problem is very similar to the jelly bean problem in approach and result (and even the absolute amounts—1,198 pounds for the ox and 1,116 jelly beans—are similar), the diversity prediction theorem shows the nature of the crowds were vastly different. Recall:

Collective error = average individual error – prediction diversity

Consider the two equations together:

Jelly bean jar contest:
1,258 = 490,949 – 489,692

Ox-weighing contest:
0.62 = 2,956.05 – 2,955.43

What's intriguing is both the average individual error and prediction diversity were *much* lower in the ox-weighing contest. Perhaps part of the difference is sample size, as the jelly bean experiment had roughly one-tenth the sample size of the ox-weighing experiment. Another explanation is the people who participated in the ox-weighing contest had a better sense of the ox's weight, many of them being farmers or butchers.

Buckets	COLUMN A Guess	COLUMN B "Difference From Actual"	COLUMN C Squared "Difference From Actual"	COLUMN D "Difference From Average"	COLUMN E Squared "Difference From Average"
1	1,074	-124	15,376	-123.21	15,180.95
2	1,109	-89	7,921	-88.21	7,781.18
3	1,126	-72	5,184	-71.21	5,071.01
4	1,148	-50	2,500	-49.21	2,421.72
5	1,162	-36	1,296	-35.21	1,239.81
6	1,174	-24	576	-23.21	538.75
7	1,181	-17	289	-16.21	262.80
8	1,188	-10	100	-9.21	84.84
9	1,197	-1	1	-0.21	0.04
10	1,207	9	81	9.79	95.82
11	1,214	16	256	16.79	281.87
12	1,219	21	441	21.79	474.76
13	1,225	27	729	27.79	772.23
14	1,230	32	1,024	32.79	1,075.12
15	1,236	38	1,444	38.79	1,504.59
16	1,243	45	2,025	45.79	2,096.63
17	1,254	56	3,136	56.79	3,224.99
18	1,267	69	4,761	69.79	4,870.50
19	1,293	95	9,025	95.79	9,175.53
Average	1,197.21		2,956.05		2,955.43

"Actual" weight of ox	1,198
"Average Guess" of weight of ox	1,197.211
Average Individual Error (Average of Column C)	2,956.05
Prediction Diversity (Average of Column E)	2,955.43
Collective Error ("Average of Column A"-"Actual")^2	.623

CHECKS:

Collective Error = Average Individual Error - Prediction Diversity
.623 = 2,956.053 - 2,955.429

√Collective Error = ABS ["Average Guess" - "Actual"]
√.623 = ABS [1,197.211 - 1,198] ≈ .789

Source: Francis Galton, "Vox Populi," *Nature*, 75, March 7, 1907.

Appendix B: The Academy Awards Experiment

In early February 2007, the students were asked to participate in an experiment to predict the winners in 12 categories of the Academy Awards. They received a form with six well-known categories on the front, and six less-known categories on the back. The assumption is predictive models are generally more robust for the first six categories, which the data support. The students were asked to contribute $1 to a communal pot, with the student(s) with the most correct choices winning the pot. The small contribution and possibility of winning a larger sum add some incentive to answer as well as possible.

Forms in hand, we determined the modal selection for each category, which we call the group answer. For example, for the lead actor category 48 percent of the participants (26 of 54) selected Forest Whitaker, the eventual winner. The modal selection was as low as 26 percent (14 of 54) for best music and as high as 65 percent (35 of 54) for best documentary. A purely random distribution would allocate 20 percent of the votes to each nominee.

We can think of the student votes as expressing a probability of each nominee winning. Accordingly, a better way to look at the performance is to express the selections as subjective probabilities, and to compare actual outcomes to the subjective estimates over a large sample. No single year offers us sufficient data to do that.

The 2007 results strongly support the diversity prediction theorem. The collective correctly selected in 11 of the 12 categories. The best students (number 3 and 5) got 9 of the 12 categories right and split the pot of money. The average student was correct in 5 of the 12 categories. We have seen similar results consistently over the years.

Two of the categories are particularly interesting. For the supporting actor category, 30 percent of the class (the consensus) selected Djimon Hounsou, while 26 percent of the vote went to Alan Arkin (the eventual winner) and another 26 percent to Eddie Murphy. So while the class got this category wrong, it's easy to see that three nominees were very close. Best music had a similar dynamic: *Babel* (the consensus and winner) gathered 26 percent of the votes, but two other nominees, *Pan's Labyrinth* and *The Queen*, each received 22 percent of the vote. So while the consensus was right here, it was also close to a one-in-three chance.

This problem appears to be a combination of *Who Wants to Be a Millionaire?* and the jelly bean jar. Some students are likely to have better predictive models than others, and hence the likely winner will rise to the surface. But since it's a prediction—essentially estimating a future state—the problem has features similar to the jelly bean jar problem.

	COLUMN A (Guesses)												COLUMN B	COLUMN C
Student	Lead Actor	Lead Actress	Support Actor	Support Actress	Mot. Pic.	Directing	Screen-play	Cinema-tography	Film Editing	Music	Docu-mentry	Art Direction	Squared "Difference From Actual"	Squared "Difference From Average"
1	0	0	1	1	1	0	0	0	0	0	1	1	7.0	3.0
2	1	1	0	1	0	0	0	1	0	0	1	0	7.0	2.1
3	0	1	0	1	1	1	1	1	0	1	1	1	3.0	3.4
4	1	0	1	0	0	0	0	0	0	0	1	0	9.0	2.4
5	0	1	0	1	1	1	0	1	1	1	1	1	3.0	3.5
6	1	1	0	1	0	1	0	1	0	1	1	0	5.0	2.6
7	0	1	0	0	0	0	0	0	1	0	1	0	9.0	2.2
8	0	1	0	1	0	0	0	0	0	0	0	0	10.0	2.1
9	0	0	0	1	1	1	1	0	1	0	0	0	7.0	3.2
10	1	1	0	0	1	0	0	1	0	0	1	0	7.0	2.2
11	1	1	1	0	0	0	1	1	0	0	0	0	7.0	3.1
12	0	0	0	1	0	1	0	1	0	0	1	0	8.0	2.3
13	1	1	0	1	0	1	1	1	0	0	1	1	4.0	2.8
14	0	1	0	1	0	0	1	1	1	1	0	1	5.0	4.0
15	1	1	1	1	1	1	0	0	0	0	1	1	4.0	2.9
16	1	1	0	1	0	0	1	1	0	0	1	1	5.0	2.8
17	1	1	0	1	0	1	1	0	0	1	0	0	6.0	2.9
18	1	1	0	1	0	0	1	0	1	1	1	0	5.0	3.1
19	0	0	1	0	0	0	0	0	0	1	0	0	10.0	3.2
20	1	1	0	0	0	1	0	1	0	0	1	1	6.0	2.4
21	1	0	0	0	0	1	1	0	0	0	1	0	8.0	2.3
22	0	0	0	0	0	0	0	0	0	1	1	0	10.0	2.4
23	0	1	0	1	1	0	1	0	1	0	1	0	6.0	2.7
24	1	1	0	1	1	1	1	0	0	0	0	0	6.0	2.6
25	1	1	0	1	0	0	0	1	0	0	0	1	7.0	2.8
26	0	0	0	0	1	0	0	1	0	0	1	1	8.0	2.7
27	1	1	0	0	1	1	0	0	0	0	1	0	7.0	1.9
28	1	1	0	0	1	0	0	0	0	1	1	0	7.0	2.4
29	1	1	0	1	1	1	1	0	0	0	1	0	5.0	2.3
30	0	1	1	0	0	0	0	1	0	0	1	1	7.0	2.9
31	0	0	1	1	0	0	1	0	0	0	1	1	7.0	3.1
32	0	0	1	0	1	0	1	0	0	0	1	0	8.0	2.9
33	1	1	1	0	1	1	0	0	1	0	0	0	6.0	3.1
34	0	1	0	0	1	1	1	0	1	0	1	0	6.0	2.6
35	0	0	0	0	1	0	0	1	0	0	0	0	10.0	2.6
36	0	0	0	0	0	1	1	0	0	0	1	0	9.0	2.2
37	0	0	1	0	0	1	0	0	0	0	0	1	9.0	3.1
38	1	0	0	1	0	0	0	0	0	0	0	0	10.0	2.3
39	1	1	0	1	1	0	0	1	0	0	1	0	6.0	2.3
40	0	0	0	0	0	1	0	0	1	1	1	0	8.0	2.9
41	1	0	1	0	0	0	0	0	0	1	1	0	8.0	2.9
42	1	1	0	1	0	0	1	0	0	1	1	0	6.0	2.6
43	1	1	0	0	0	0	0	0	1	0	0	0	9.0	2.5
44	1	0	0	1	1	1	1	0	1	1	1	0	4.0	3.4
45	0	0	0	0	1	0	0	0	0	0	0	0	11.0	2.4
46	1	1	0	0	0	1	1	1	0	1	1	1	4.0	3.2
47	0	1	1	0	0	1	0	1	0	0	1	1	6.0	2.9
48	0	1	0	0	0	1	0	1	0	0	1	1	7.0	2.4
49	0	0	1	0	0	1	0	0	1	0	0	0	9.0	3.1
50	1	1	0	1	1	1	0	0	1	0	1	0	5.0	2.4
51	0	1	1	0	0	1	0	0	1	0	0	1	7.0	3.3
52	0	0	0	0	1	0	0	0	1	0	0	0	10.0	2.8
53	0	0	0	0	1	1	0	0	0	0	0	0	10.0	2.4
54	0	0	0	1	1	1	0	1	0	0	0	0	8.0	2.7
Avg.	0.48	0.59	0.26	0.46	0.43	0.50	0.35	0.37	0.28	0.26	0.65	0.31	7.1	7.0

Average Individual Error (Average of Column B)	7.056
Prediction Diversity (Average of Column C)	2.724
CHECK:	
Collective Error = Average Individual Error - Prediction Diversity	
Collective Error = 7.056 - 2.724	4.332

The Academy Awards Experiment—One Category

Category: "Lead Actor" Student	COLUMN A Guess	COLUMN B Squared "Difference From Actual"	COLUMN C Squared "Difference From Average"
1	0	1	0.232
2	1	0	0.269
3	0	1	0.232
4	1	0	0.269
5	0	1	0.232
6	1	0	0.269
7	0	1	0.232
8	0	1	0.232
9	0	1	0.232
10	1	0	0.269
11	1	0	0.269
12	0	1	0.232
13	1	0	0.269
14	0	1	0.232
15	1	0	0.269
16	1	0	0.269
17	1	0	0.269
18	1	0	0.269
19	0	1	0.232
20	1	0	0.269
21	1	0	0.269
22	0	1	0.232
23	0	1	0.232
24	1	0	0.269
25	1	0	0.269
26	0	1	0.232
27	1	0	0.269
28	1	0	0.269
29	1	0	0.269
30	0	1	0.232
31	0	1	0.232
32	0	1	0.232
33	1	0	0.269
34	0	1	0.232
35	0	1	0.232
36	0	1	0.232
37	0	1	0.232
38	1	0	0.269
39	1	0	0.269
40	0	1	0.232
41	1	0	0.269
42	1	0	0.269
43	1	0	0.269
44	1	0	0.269
45	0	1	0.232
46	1	0	0.269
47	0	1	0.232
48	0	1	0.232
49	0	1	0.232
50	1	0	0.269
51	0	1	0.232
52	0	1	0.232
53	0	1	0.232
54	0	1	0.232
Average	0.4814815	0.2688615	0.250

"Actual" Winner	1
"Average Guess" of Winner	.4818

Average Individual Error (Average of Column B)	0.518519
Prediction Diversity (Average of Column C)	0.249657
Collective Error ("Average of Column A"-"Actual")^2	0.268861

CHECKS:

Collective Error = Average Individual Error - Prediction Diversity
.2689 = .5185 - .2497

√Collective Error = ABS ["Average Guess" - "Actual"]
√.2689 = ABS [.4815 - 1] ≈ .5185

Endnotes

[1] Scott E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton, NJ: Princeton University Press, 2007), xxiv.
[2] For example, James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (New York: Doubleday and Company, 2004); Howard Rheingold, *Smart Mobs: The Next Social Revolution* (Cambridge, MA: Perseus Press, 2002); Cass R. Sunstein, *Infotopia: How Many Minds Produce Knowledge* (Oxford: Oxford University Press, 2006); Don Tapscott and Anthony Williams, *Wikinomics: How Mass Collaboration Changes Everything* (New York: Portfolio, 2006); Norman L. Johnson, "Diversity in Decentralized Systems: Enabling Self-Organizing Solutions," LA-UR-99-6281, 1999; Declan McCullagh, "Tech Lessons Learned from the Wisdom of Crowds," *News.com*, December 15, 2006.
[3] For example, Charles MacKay, *Extraordinary Delusions and the Madness of Crowds*, 1841 (New York: Three Rivers Press, 1995); Jaron Lanier, "Digital Maoism," *The Edge.org*, May 30, 2006.
[4] Joe Nocera, "The Future Divined by the Crowd," *The New York Times*, March 11, 2006.
[5] Michael J. Mauboussin, "Are You an Expert?" *Mauboussin on Strategy*, October 28, 2005.
[6] Page, 7-9.
[7] Surowiecki also discusses independence as a condition. We consider independence as a subset of diversity. If diversity prevails, so will independence, although the inverse is not true. Independence is a necessary but not sufficient condition for diversity.
[8] Surowiecki has a good discussion of aggregation in the context of the U.S. intelligence effort. See Surowiecki, 66-69.
[9] Page, 233-234.
[10] The strength of this argument is weakened by agency costs in the investment industry. For a good survey, see Alfred Rappaport, "The Economics of Short-Term Performance Obsession," *Financial Analysts Journal*, Vol. 61,3, May/June 2005, 65-79.
[11] This view is expressed by Nobel-winning economist William Sharpe. He says, "The basic argument [of *The Wisdom of Crowds*] is that if we have enough people even though they may be ill-informed and irrational coming to market, it is entirely possible the prices of assets, thereby true risks and rewards, are what you would get if they were all rational and well informed." See Ayse Ferliel, "Interview with William Sharpe," *Investment Adviser*, December 6, 2004.
[12] Surowiecki, 3-4.
[13] Sunstein, 25-32.
[14] Page, 183-185.
[15] Iain D. Couzin, Jens Krause, Nigel R. Franks, and Simon A. Levin, "Effective Leadership and Decision-Making in Animal Groups on the Move," *Nature*, February 3, 2005; Thomas A. Seeley, P. Kirk Visscher, and Kevin M. Passino, "Group Decision Making in Honey Bee Swarms," *American Scientist*, Vol. 94, May-June 2006.
[16] David Austin, "How Google Finds Your Needle in the Web's Haystack," *American Mathematical Society Feature Column*, December 2006. http://www.ams.org/featurecolumn/archive/pagerank.html
[17] Page, 205-209.
[18] In his classic paper on the jelly bean experiment and market efficiency, Jack Treynor suggests the model's accuracy "comes from the faulty opinions of a large number of investors who err independently. If their errors are wholly independent, the standard error in equilibrium price declines with roughly the square root of the number of investors." We believe the square-root law, which says the standard error of the mean decreases with the square root of N (number of observations), is an inappropriate explanation for the jelly bean (or market efficiency) problem. The square-root law applies to sampling theory, where there are independent observations that include the answer plus a random noise term. Over a large number of observations, the errors cancel out. An example is observing and measuring star luminosity. The underlying assumption behind the square-root law is the observations are independent and identically distributed around a mean. This is clearly not the case with either the jelly bean jar or markets. We believe the diversity prediction theorem is a more robust way to explain the wisdom of crowds in this case. See Jack L. Treynor, "Market Efficiency and the Bean Jar Experiment," *Financial Analysts Journal*, May-June 1987, 50-53.
[19] For another discussion of the usefulness of diversity in forecasting, see J. Scott Armstrong, "Combining Forecasts," in J. Scott Armstrong, ed. *Principles of Forecasting* (New York: Springer, 2001), 417-439.

[20] See Ferliel interview with Sharpe, Treynor, and Richard Roll, "What Every CFO Should Know About Scientific Progress in Financial Economics: What is Known and What Remains to be Resolved," *Financial Management*, Vol. 23, 2, Summer 1994, 69-75.
[21] Michael J. Mauboussin, "Capital Ideas Revisited," *Mauboussin on Strategy*, March 30, 2005.
[22] Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton University Press, 2005).
[23] Blake LeBaron, "Financial Market Efficiency in a Coevolutionary Environment," *Proceedings of the Workshop on Simulation of Social Agents: Architectures and Institutions, Argonne National Laboratory and University of Chicago*, October 2000, Argonne 2001, 33-51.
[24] Steven Strogatz, *Sync: The Emerging Science of Spontaneous Order* (New York: Theia, 2003), 53-59.
[25] Karl B. Diether, Christopher J. Malloy, and Anna Scherbina, "Differences of Opinion and the Cross Section of Stock Returns, *The Journal of Finance,* Vol. 57, 5, October 2002, 2113-2141; and Edward M. Miller, "Risk, Uncertainty, and Divergence of Opinion," *The Journal of Finance,* Vol. 32, September 1977, 1151-1168.

Resources

Books

Armstrong, J. Scott, ed., *Principles of Forecasting* (New York: Springer, 2001).

MacKay, Charles, *Extraordinary Delusions and the Madness of Crowds*, 1841 (New York: Three Rivers Press, 1995).

Page, Scott E., *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton, NJ: Princeton University Press, 2007).

Rheingold, Howard, *Smart Mobs: The Next Social Revolution* (Cambridge, MA: Perseus Press, 2002).

Strogatz, Steven, *Sync: The Emerging Science of Spontaneous Order* (New York: Theia, 2003).

Sunstein, Cass R., *Infotopia: How Many Minds Produce Knowledge* (Oxford: Oxford University Press, 2006).

Surowiecki, James, *The Wisdom of Crowds Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (New York: Doubleday and Company, 2004).

Tapscott, Don, and Anthony Williams, *Wikinomics: How Mass Collaboration Changes Everything* (New York: Portfolio, 2006).

Tetlock, Philip E., *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton University Press, 2005).

Articles and Papers

Austin, David, "How Google Finds Your Needle in the Web's Haystack," *American Mathematical Society Feature Column*, December 2006.

Couzin, Iain D., Jens Krause, Nigel R. Franks, and Simon A. Levin, "Effective Leadership and Decision-Making in Animal Groups on the Move," *Nature*, February 3, 2005.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina, "Differences of Opinion and the Cross Section of Stock Returns, *The Journal of Finance,* Vol. 57, 5, October 2002, 2113-2141.

Ferliel, Ayse, "Interview with William Sharpe," *Investment Adviser*, December 6, 2004.

Galton, Francis, "Vox Populi," *Nature*, 75, March 7, 1907.

Johnson, Norman L., "Diversity in Decentralized Systems: Enabling Self-Organizing Solutions," LA-UR-99-6281, 1999.

Lanier, Jaron, "Digital Maoism," *The Edge.org*, May 30, 2006.

LeBaron, Blake, "Financial Market Efficiency in a Coevolutionary Environment," *Proceedings of the Workshop on Simulation of Social Agents: Architectures and Institutions, Argonne National Laboratory and University of Chicago*, October 2000, Argonne 2001, 33-51.

Mauboussin, Michael J., "Capital Ideas Revisited," *Mauboussin on Strategy*, March 30, 2005.

_____., "Are You an Expert?" *Mauboussin on Strategy*, October 28, 2005.

McCullagh, Declan, "Tech Lessons Learned from the Wisdom of Crowds," *News.com*, December 15, 2006.

Miller, Edward M., "Risk, Uncertainty, and Divergence of Opinion," *The Journal of Finance,* Vol. 32, September 1977, 1151-1168.

Nocera, Joe, "The Future Divined by the Crowd," *The New York Times*, March 11, 2006.

Rappaport, Alfred, "The Economics of Short-Term Performance Obsession," *Financial Analysts Journal*, Vol. 61,3, May/June 2005, 65-79.

Roll, Richard, "What Every CFO Should Know About Scientific Progress in Financial Economics: What is Known and What Remains to be Resolved, *Financial Management*, Vol. 23, 2, Summer 1994, 69-75.

Seeley, Thomas A., P. Kirk Visscher, and Kevin M. Passino, "Group Decision Making in Honey Bee Swarms," *American Scientist*, Vol. 94, May-June 2006.

Treynor, Jack L., "Market Efficiency and the Bean Jar Experiment," *Financial Analysts Journal*, May-June 1987, 50-53.

07-0351