"It is not a case of choosing those [faces] which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees."

Taxi Drivers and Beauty Contests

by Colin F. Camerer

The trading floor of the New York Stock Exchange. The British economist John Maynard Keynes likened playing the market to voting for the prettiest face in a beauty contest; hence the second part of this article's title.

I spent a year in New York City not long ago, and I took a lot of cabs. Most cabdrivers in New York are independent contractors. They rent the cab from a taxi company for $76—paid in advance—for 12 hours. They keep all the fares they collect, and they can call it quits and return the cab at any time before the 12 hours are up. Because Manhattan is so crowded, drivers usually just cruise the streets waiting for someone to hail them. Some days are especially good—when it rains or snows, during the holidays, or when a convention is in town, for example. Other days are bad—weekends, when fewer businesspeople are around; and the summer is slow because people leave Manhattan to escape the heat, humidity, and gunfire. So Linda Babcock (whose father, Charles Babcock [MS '58, PhD '62], was a professor of aeronautics and applied mechanics at Caltech until his untimely death in 1987) and George Loewenstein of Carnegie Mellon University, Richard Thaler of the University of Chicago, and I became curious about a simple question—how does the amount of hours a cabbie works vary with that day's average hourly earnings? There are two basic theories that might apply. One is called the law of supply. The other, which we crafted from bits and pieces of psychology, we call "daily income targeting."

The law of supply is the twin sister of the law of demand—people should want to sell more of something when the price is high than when the price is low, assuming everything else is constant. So you'll sell more of your labor hours when wages are high, and the so-called labor-supply curve slopes upward, as shown on the page after this one. The law of supply says that you should work a lot when it pays to do so, and when it doesn't pay, go home! Take time off.

The other theory, daily income targeting, was taught to us by the cabdrivers, many of whom are amateur philosophers, political scientists, and labor economists. (The late Harry Chapin appears

to have been familiar with it as well, as the opening couplet of his ballad "Taxi" bears witness: "It was raining hard on a Saturday / I needed one more fare to make my night.") Many of the drivers we talked to said they decided how long to work by setting themselves an income target every day—for example, they might want to earn $150 in cash in order to clear $75 beyond the rental fee—and when they reach that target, they quit. Target setting can be very motivating in unpleasant or tedious activities, like exercise. There's also substantial psychological evidence that people dislike losing a lot more than they like corresponding amounts of winning. This implies that drivers hate to quit before they reach the target, but once they reach it, they aren't very enthusiastic about trying to go beyond. So income targeting perversely predicts that cabbies are going to quit earlier on good days. If you want to make $150 and you're earning $25 an hour (which would be a pretty good day for these guys), you can go home after six hours. But on a bad day, when you're earning $15 an hour, you've got to drive ten hours. The labor-supply curve will be a hyperbola, which is also shown on the following page.

These two theories thus give very different predictions, which we tested in our study. We analyzed 3,000 observations of cabdrivers' behavior from the years 1988, 1990, and 1994. The data came from the New York City Taxi and Limousine Commission, ironically known as the TLC, which had collected it for other studies. These data were in the form of taximeter readings, and the TLC was kind enough to give them to us (for free!) on floppy disks—bureaucracy has its moments. When you get into a cab, the driver punches a button and a meter automatically records the number of miles driven and the amount of time spent sitting in traffic. From the meter records we could compute a driver's earnings, except for tips. Tips aren't recorded anywhere, so we left them out of our analysis, but
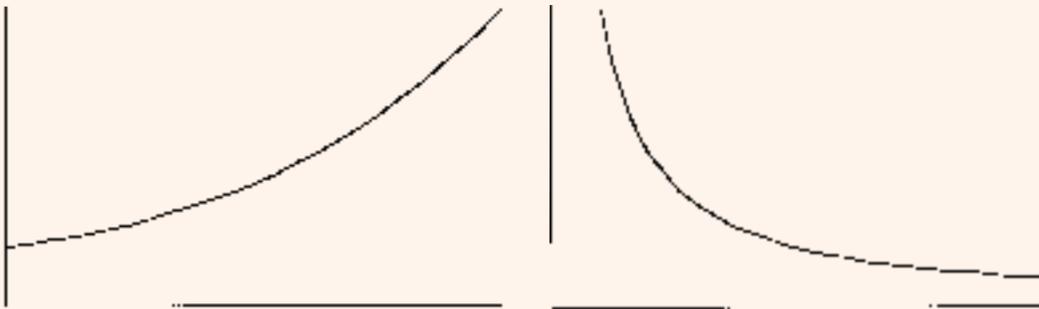
on average they're probably 10–15 percent of the driver's income and do not vary much from day to day. The meter data should represent the bulk of the driver's income accurately.
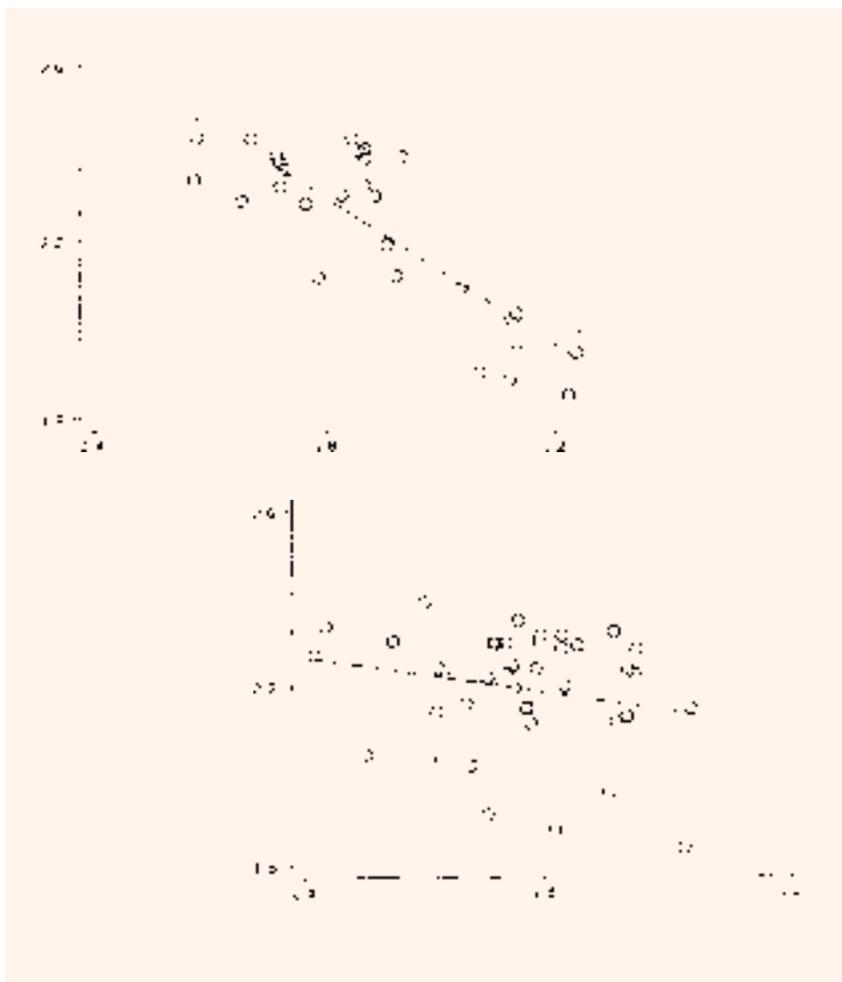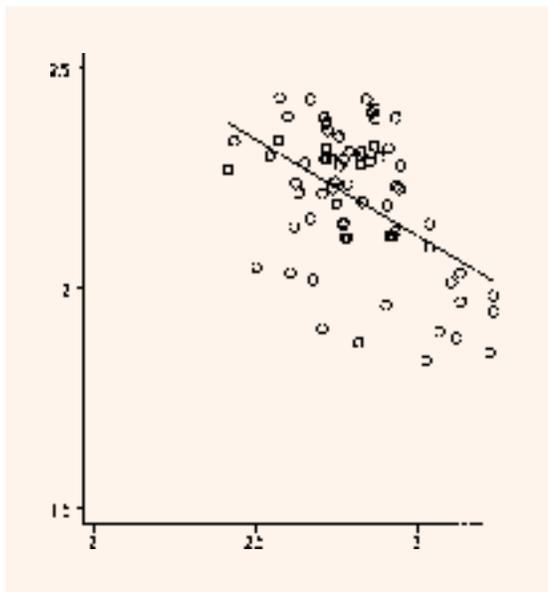
A scatter plot of some of our data is shown at the top of the opposite page. I should warn the faint of heart that it looks very messy, but as economic data go—particularly in areas like labor economics, where there are lots of external factors that influence the data—this is actually a pretty strong correlation. I hope you can see that the line of best fit through this cloud of data slopes downward. In fact, the slope is significantly negative to a confidence level of more than 99.9 percent. So the data clearly support the targeting theory rather than the law of supply.

There's an objection that can be raised here—in order to follow the law of supply, you've got to have a certain level of economic security. These guys may have to keep driving until they make $150 because they need the cash—they don't have enough savings to buy groceries and pay the rent if they quit early on slow days. The reason we don't think this explains our findings is because some drivers in our samples own their own cabs. In order to legally operate a cab in New York City, you have to own a taxi medallion—an ugly, plastic-metal thing that's pasted on the hood of the cab. These medallions are restricted in supply (there are only 11,387 of them, and that number has been fixed for 60 years), so they're quite valuable. They're worth about $150,000, yet 10 percent of the drivers in our samples own one personally. If we assume that the drivers who can afford to own a medallion have some cash in the bank, we might predict that they would behave differently than the renters. But both groups seem to behave about the same way.

Another important consideration is that cabdrivers vary in experience. Happily for us, New York City cabdriving licenses are numbered chronologically by date of first issue, so the person

Above: Some of the meter data were verified by examining "trip sheets" such as this one, in which drivers log (from left) pickup point, pickup time, destination, dropoff time, number of passengers, and fare.

with license number 14,682 got it just after the person with number 14,681. Therefore, we can sort drivers into high- and low-experience groups by their license numbers. Look at the difference in their labor-supply curves, as shown at left. Again, the data are noisy, but the low-experience drivers on the left have a slope very close to −1, which is what the income-targeting theory predicts. (A slope of −1 means that if your wage goes up by 10 percent, you cut your hours back by 10 percent to keep your income constant.) The high-experience drivers on the right still don't look much like they're obeying the law of supply, but it does appear as if experience is teaching them to make hay while the sun shines—to drive longer hours on good days.

This distinction between new and old drivers is important because about half of the cabbies in New York have been driving cabs for less than a year. In 1991, over 40 percent of all New York cabdrivers were born on the Indian subcontinent, 11 percent were from Africa, and another seven percent each were from the Caribbean, the Middle East, and the former Soviet Union. Only about 10 percent were born in the United States. The point is that driving a cab is an entry-level job for many immigrants, so there's a constant inflow and out-flow of new drivers. These inexperienced drivers may be using the income-targeting rule because they haven't yet figured out that they can do better by obeying the law of supply.

We learned two basic lessons from this study. The first is that cabbies would get an automatic raise of 8 percent if they drove the same number of hours every day, rather than knocking off early on the good days and working late on the bad days. If they obeyed the law of supply, they could earn 15 percent more income. The median annual wage of these drivers in 1995 was about $22,000 a year, so they could have made about $2,000 more per year by simply changing their driving habits.

The second, more important lesson is that

Which will You elect Miss Rheingold 1957 ?

Pick the girl who'll win a contract worth $50,000! Vote at any Rheingold store or tavern!

Rheingold EXTRA DRY

Photo courtesy of the John W. Hartman Center for Sales, Advertising and Marketing History, Duke University

The "Miss Rheingold" campaign, run by the J. Walter Thompson Co. for Liebmann Breweries, Inc. for over 25 years, is the best-known American example of a Keynesian beauty contest. At the height of its popularity, between 15 and 20 million votes were cast per year—a turnout second only to the Presidential elections.

perhaps we should be skeptical about simple economic principles like the law of supply. Most previous studies were inconclusive about whether the supply curve even went up or down, because most people's salaries change relatively rarely—once a year, perhaps. But cabdrivers earn a different hourly wage every day, and they can adjust the numbers of hours they drive, so there's enough variation in the data to see trends. That's why our study shows more clearly than ever before that for taxi drivers, the labor-supply curve slopes down, not up. During the Reagan years, supply-side economists argued that if income taxes were cut, the after-tax wage would rise. That's just simple arithmetic. And then, the argument went, people could earn more spending money by working an extra hour, so people *would* work extra hours, and everyone would be better off. Very logical. Our results suggest the opposite—if you were to lower the tax rates on cabdrivers to give them a higher after-tax wage, it looks as if they would drive *fewer* hours, not more.

Let me move on to beauty contests. Here I don't mean the Miss America pageant or the tryouts for Rose Parade queen, but you'll see in a moment where the term comes from. Imagine the following game: Everybody picks a number from 0 to 100. I compute the average of all your picks, and whoever's number is closest to two-thirds of that average wins. (We actually do this in experiments on students. The winner gets $20, so they think carefully before they choose.) Everyone wants to be at two-thirds of the average, but everyone else does, too, so the real goal of the contest is to guess what everyone else will guess.

This is like playing the stock market. The economist John Maynard Keynes remarked in the 1930s that the stock market is like a beauty contest. He had in mind contests that were popular in England at the time, where a newspaper would print 100 photographs, and people would write in and say which six faces they liked most. Everyone who picked the most popular face was automatically entered in a raffle, where they could win a prize. Keynes wrote, "It is not a case of choosing those [faces] which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees."

If you played this game repeatedly, your thoughts might run as follows. You'd assume that the starting average would probably be 50, so you'd guess 33. But then you'd say, hmmmm, if other people are as clever as I am, they will all pick 33, so I should pick 22. But if everyone else does that, too, I should pick two-thirds of 22. And if you carry this through infinitely many
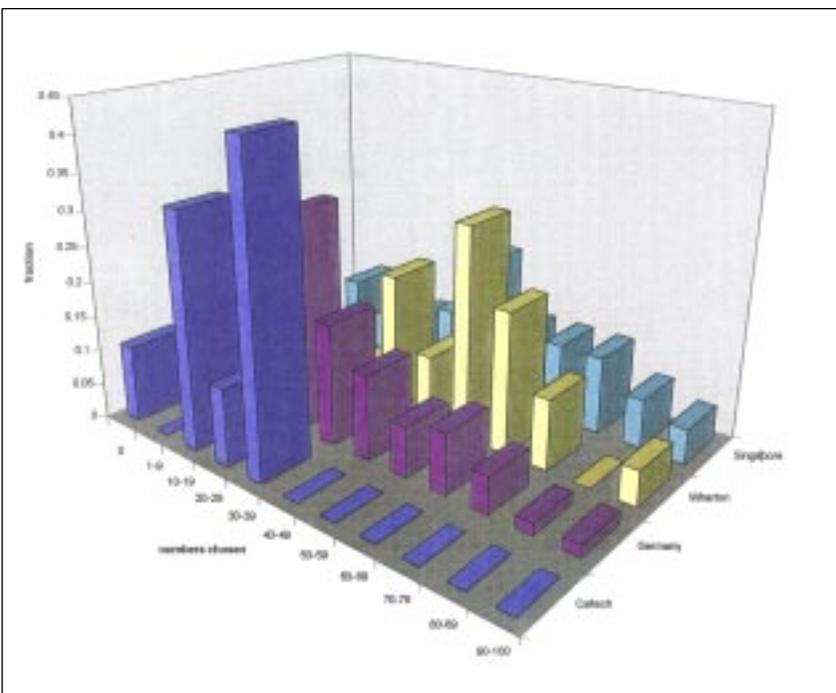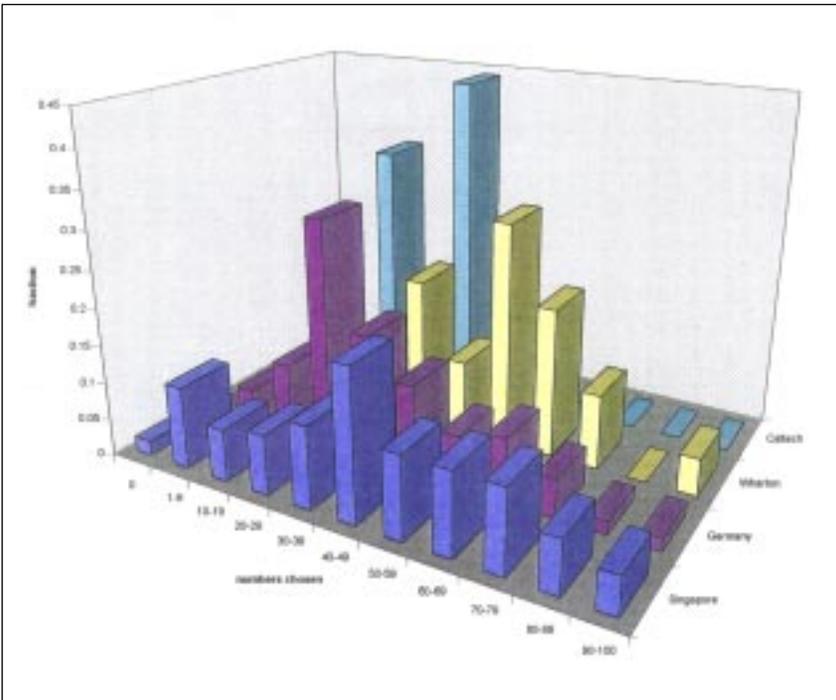
levels of reasoning to the logical end, you'll wind up picking zero. If I were speaking to a game-theory audience, people would nod profoundly, because zero is what game theory predicts for this situation. Game theory is the branch of social science that analyzes strategic interactions in mathematical terms. It was founded quite a long time ago, but it's had a slow fuse—only in the last 10 or 15 years has it come to the fore in reasoning about economics and political science. (In fact, people here at Caltech helped establish the use of game theory in political science, and still do quite a lot of it.)

So how do people actually behave? Do they pick zero? The data at left are from experiments on undergrads from Singapore, Germany, the Wharton School of Business at the University of Pennsylvania, and Caltech. The German data were collected by Rosemarie Nagel of the University of Pompeu Fabra. The Singaporean data were collected by Teck-Hua Ho and Keith Weigelt; they also collaborated with me on the Wharton data. (Ho and Weigelt, who are now on the faculty at Wharton, were both students of mine when I was there.) The average pick across all these experiments was around 40, so if you guessed about two-thirds of 40, or 27, you'd probably win. Notice that 40 is somewhat less than 50, so if we use these data to gauge how many steps of reasoning people are doing about other people's reasoning, some number from one to three seems reasonable. It's clearly not the game-theory prediction of infinity, but it also clearly demonstrates the performance of at least one step of reasoning.

We're now trying to refine this estimate of how many steps of reasoning seem natural, and how it varies with education and other factors. For example, no Caltech student chose above 40. Most Techers picked numbers between 30 and 40. Several picked in the neighborhood of 10 or 20, and 10 percent of them did, in fact, actually pick zero. The Caltech students and the German stu-

dents appear to have been reasoning one or two
steps more deeply than the Wharton students
and the students in Singapore.
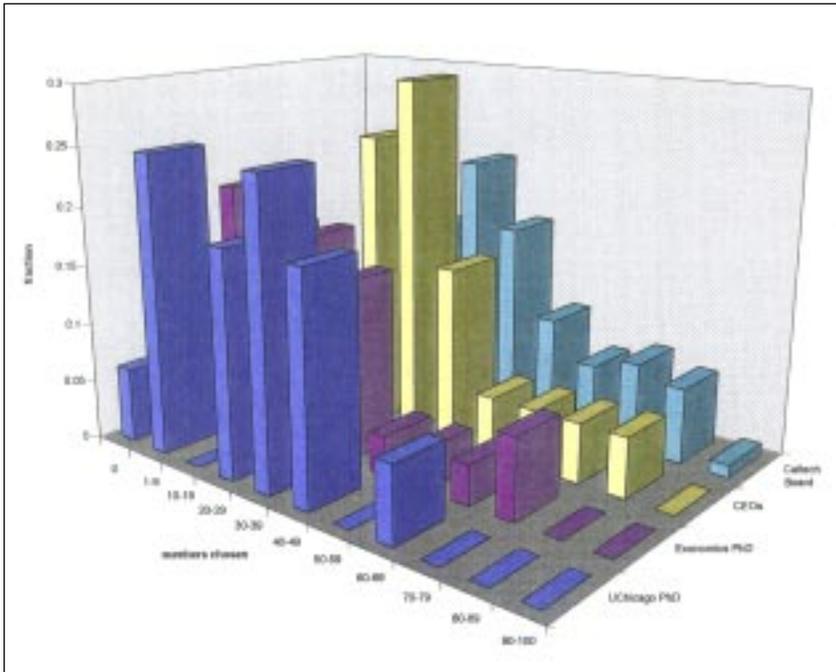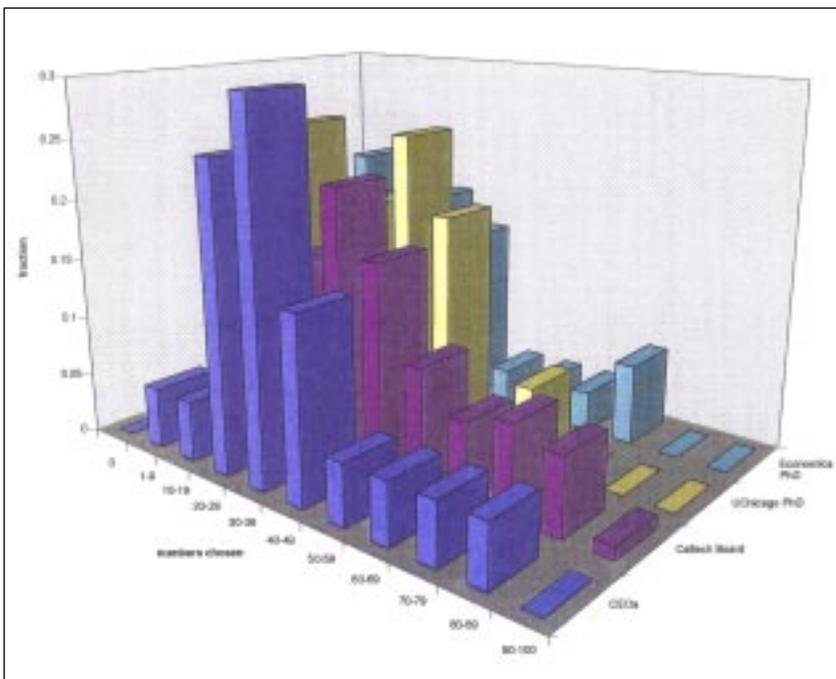
We've also conducted this experiment, more
informally, with other groups of subjects. (Repli-
cation with different groups of people is, of course,
essential if we want to generalize our findings to
all human beings.) The plot at left shows four
more groups. The front two groups in the top
figure are PhD students in economics (none from
Caltech), who may have had some exposure to
game theory. And, in fact, compared to the under-
grads in the previous plot (except for the Techers),
these PhD students do choose lower numbers.
The average pick here is around 25—one step
beyond the undergrads. The additional education
is doing something. The group labeled "Caltech
Board" is from an experiment I conducted when
I gave a talk at a meeting of Caltech's Board of
Trustees in the fall of 1995. There were about
80 or 90 people there, including spouses and some
people from the faculty and administration, and I
just couldn't resist the opportunity to see how
they would behave. The Caltech Board is a truly
amazing group that includes many extremely
successful businessmen, some billionaires, several
brilliant scientists, and two former judges. Notice
that they act pretty much like the college stu-
dents—the average pick is about 40. But a few
people do choose very low numbers, like zero.
And several people, who may have been confused
because I didn't explain the procedure as carefully
and thoroughly as I would have in a real experi-
ment, picked very high numbers. This was not
a well-run experiment, but the subject pool is
so unusual that I'll show it nonetheless.

The sample labeled "CEOs" is really remarkable.
We've seen that college students do not obey game
theory, which assumes that people are perfectly
rational. (This is hardly surprising to anybody
with teenagers in college.) So it's easy to criticize
our experiments by saying that what *really* matters

This strategy is sometimes called the "greater-fool theory," because even though you're a fool to pay as much as you did, you're betting that there's a greater fool just down the road. And if you're right, then of course you aren't being foolish.
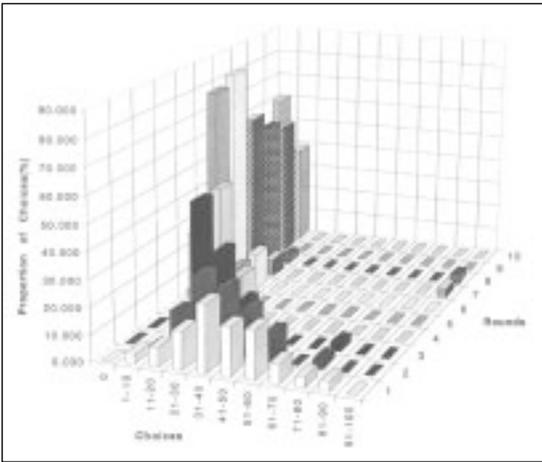
When people play the beauty-contest game for several rounds against the same group of opponents, the behavior quickly converges to what game theory predicts will happen.

is not what a bunch of college kids do, but whether the people who run large businesses behave according to economic theory. Well, the Caltech Board includes 20 chief executive officers, presidents, and corporate-board chairmen. These titans of industry are the "CEOs" sample. As you can see, none of them picked zero; and if any one of them had, that person would have lost. So they obviously knew who they were playing with. A few of them picked surprisingly high numbers, but the tallest spike is between 30 and 39, and there's another tall spike between 20 and 29. If you do the math, it turns out that they were reasoning about one step further than the other people at the meeting. The numbers they chose are statistically indistinguishable from the numbers the Caltech undergraduates and the econ PhD students chose.

The game-theory prediction was flat-out wrong. The same pattern emerged across three continents, both genders, and a tremendous variation in age, wealth, and educational background.

But what happens if we allow people to learn by announcing the winning number and repeating the game? Then we see a steady, slow convergence toward the game-theory prediction. The graph above shows what happened when the Singaporean students played a multi-round version of the game. After 10 rounds, about 50 or 60 percent of the students were choosing numbers between zero and 10. So game theory, which seemed so laughable at first, does predict what people will do with repetition. Again, psychology helps us understand what happens at first, and game theory tells us what will happen eventually as people learn. We need both to understand the entire picture.

This brings me to the stock market. That passage from Keynes describes a market in which investors care about what other investors will buy in the future. Here, you often pay more than a firm is worth, because you think that somebody else will pay even more later on. This strategy is
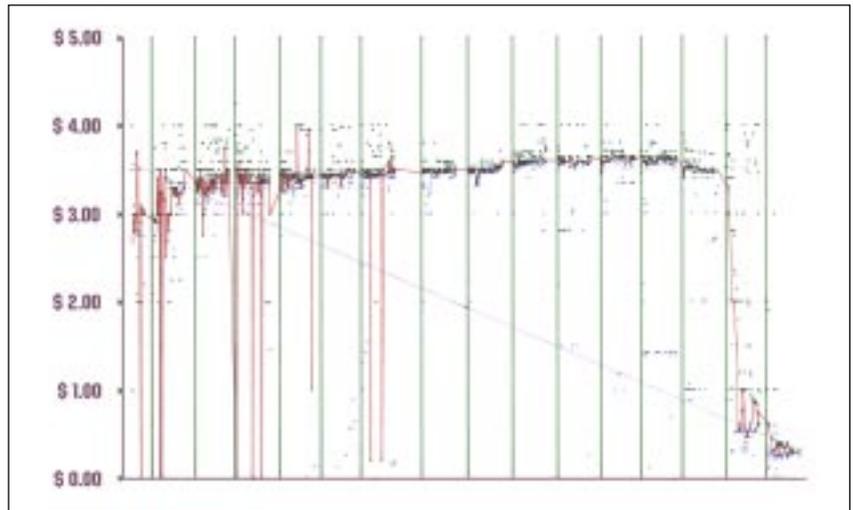
sometimes called the "greater-fool theory," because even though you're a fool to pay as much as you did, you're betting that there's a greater fool just down the road. And if you're right, then of course you aren't being foolish.

Economists call this a bubble. Prices rise simply because people expect them to rise, and it's a self-fulfilling prophecy right up to the moment when the bubble bursts. One famous example is that of tulip bulbs in Holland during the 1600s. People were paying several months' income for rare tulip bulbs. Thoroughbred horses in the 1970s, and L.A. real estate in the 1980s are other examples, as are booms in works by dead artists (who can't produce any more supply). The Japanese economy in the 1980s might be the most spectacular example in world history. However, a business-school professor who teaches about the stock market would probably be reluctant to admit that these episodes are bubbles, in the sense that I've defined the term. I'm asserting that people are consciously paying more than the intrinsic value of the asset, but the professor would probably say that we don't know its intrinsic value. How do you measure the intrinsic value of, say, Van Gogh's *Sunflowers*? Maybe it was a bargain at $50 million. Instead, most of the experts believe in the so-called efficient-market theory, which says that information about a stock's worth will quickly be reflected in its price.

It would be nice if we had an example to convince the experts who believe that markets are efficient. Until a couple of decades ago, people thought that economics, like astronomy, was not an experimental science—all you could do was study the data that the market provided. But in fact, many of the most interesting propositions in economics can be tested experimentally. About 10 years ago, Charles Plott, the Harkness Professor of Economics and Political Science, founded the Experimental Economics and Political Science Laboratory at Caltech. The whole thing is run

by computer and functions very much like a real market. (You can also study elections and other processes in it.) Each participant is isolated in a booth, and cannot communicate with other participants in any way except through the computer. People type in offers to sell x number of shares at such-and-such a price, or bids to buy, and all the offers and bids are displayed on everyone's computer screens. Players consummate trades with the push of a button. The computer records all offers, bids, and transactions sequentially; keeps track of who owns what; and calculates everyone's earnings. (Again, the students get paid real money, so we can be sure that they're taking this seriously and are giving it their best effort.) Everything is recorded as it happens, and software developed by Plott enables us to make a "movie" of how the market behaves, and analyze it in detail.

In these experiments, we created a market for an asset we invented whose value we chose. The students traded a share—a bond, if you will—for 15 five-minute periods. Each share paid a dividend of 24 cents at the end of each period, so if you held on to a share for all 15 periods, you'd earn $3.60. Everyone had a couple of shares to start with, and some money to buy more shares if they wanted to. The question we wanted to answer was, what would the price of the shares be? The efficient-market theory is very clear on this. It says that since everyone knew the share paid a total of $3.60 in dividends (we told them that, by the way—we gave them a table of dividends versus periods remaining), then the price of the share should be $3.60 in the first period. In the second period, the price should drop by 24 cents to $3.36, and so on.

Shown above is what real traders did in a typical experiment. The slanting purple line shows the shares' declining dividend value. Each dot is an attempt to sell or buy; all the completed transactions are connected by the red line. Dots above the red line are sellers asking too much, and dots

below the line are buyers offering too little. Notice that the price remains flat at around $3.50—even close to the end, where the efficient-market theory says the shares are worth less than a buck. (This is like those of you who bought a house in L.A. a few years ago, and refused to sell as the market collapsed.) The traders are trying to forecast whether the market will crash, or whether some nut will buy shares that are about to expire. And finally, of course, the market collapses.

We know that everyone knew a share's intrinsic value because we gave them a quiz before the first trading period began, so this is the clearest example of a bubble that you could possibly have. When we asked the subjects how it came about, they'd tell us a story that sounded very much like the greater-fool theory. They'd say, sure I knew the prices were way too high, but I saw other people buying and selling at high prices. I figured I could buy, collect a dividend or two, and then sell at the same price to some other idiot. And, of course, some of them were right. As long as they got out before the crash, they earned a lot of money at the expense of the poor folks who were left holding the bag.

We can see harbingers of the crash in what we've come to call nervousness in the market. Near the end, some people who think that the market has lost its mind will make extremely low bids. These people probably know that a lowball bid of a dollar won't be accepted when the going rate is three times that, so we think this is their way of expressing their surprise and warning everybody. It's the same as when somebody offers you $350,000 for the house you're desperately trying to sell for the half million you paid for it a few years ago. This is their way of politely saying you're nuts—your house isn't worth half a million.

After doing a number of such experiments, we've learned how to turn these bubbles on and off. To turn the bubble off, we bring the same group of subjects back and run the entire 15-

The traders are trying to forecast whether the market will crash, or whether

some nut will buy shares that are about to expire.

period market again. We usually see a smaller rise that crashes much earlier. And if we bring that same group back a third time, we hardly get any bubble at all. The market-price line now follows the intrinsic-value line very closely, so experienced traders do obey the efficient-markets theory. We can turn a bubble on by having had our subjects participate in a previous experiment in which we created inflation by adding money to the economy, just the way the government does. If prices rose in that earlier market—if they've lived through an inflationary experience—then we've planted a belief in their minds that prices will rise, like seeding clouds to make rain. Then, when we put them in the bubble experiment, prices *do* rise, because of this self-fulfilling prophecy based on their common experience. We don't always see bubbles—sometimes we see just what the efficient-market theory predicts, with prices sliding down along the intrinsic-value line. But bubbles are very common—the several of us doing this kind of research have observed about a hundred of them.

What does all this mean in the real world? Perhaps one-third of the market's trading volume is due to a handful of mutual funds and other large institutions. These portfolio managers may not behave rationally, either, although for other reasons. For example, they operate in a world where if they have one bad quarter—worse than everyone else—they may get fired. So they ask

their colleagues, what are you guys buying? They want to buy what the other guys buy, so they don't finish last. That, again, is very much like a Keynesian beauty contest, and I think the prevailing theories need to address it. Peter Bossaerts, an associate professor of finance here at Caltech, is actually working on this now. I should also point out that nothing I've said addresses the issue of stocks that haven't paid dividends yet, but may at some time in the future. This is a very common situation with growth stocks, such as those of startup companies in biotechnology, software, and other high-tech fields. The closest we've come to studying those was a couple of experiments where the dividend wasn't guaranteed—there was a large chance you'd get nothing, and a small chance you'd win big. We did see some things that looked like bubbles, but we haven't done much work in that area yet.

In conclusion, cabdrivers, beauty-contest games, and stock-market experiments have a common theme. Inexperienced cabdrivers, novice beauty-contest players, and traders participating in an experimental stock market for the first time don't seem to conform to standard economic theory, which assumes complete rationality by all participants. However, their actions are reasonably well explained by psychological theories that allow people to have normal, limited reasoning ability, and limited faith in others. The subjects of these experiments aren't dumb, but they're not perfectly brilliant, either, and they're not willing to bet a lot of money that other people are. But the behavior of experienced drivers, players who play the beauty contest over and over again, and traders who return to the stock market is often explained quite well by economic theories. Experimental observations help us figure out which theories are true, and which are false, and under what conditions. So we think that combining the best ideas in psychology and economics will make for the best social science of all. □

This research is very new, and there are many things we have yet to learn. We need help from cognitive psychology to understand what the people in our experiments are thinking. We need better pattern-recognition and data-analysis tools to help us look at the data and forecast when bubbles will start and crashes occur. Compared to other experimental sciences like physics, chemistry, and biology, the amount of work that's been done in experimental economics is relatively modest.