

## The Gauntlet

I accepted the job at Cornell about halfway through my time at Stanford, and

This fact was brought home to me at one early conference where I gave a talk on my recent work. During the question and answer period that followed, a well-known economist asked me a question: “If I take what you are saying seriously, what am I supposed to do? My skill is knowing how to solve optimization problems.” His point was that if I were right, and optimization models were poor descriptions of actual behavior, his toolkit would be obsolete.

His reaction was unusually candid. The more common response, for those who engaged at all, was to explain what I was doing wrong, and what obvious factors I had ignored. I soon had another list: reasons why economists could safely ignore behaviors such as those on the List. Among friends I would call this series of questions the Gauntlet, since any time I gave a talk about my work it felt like running a medieval gauntlet. Here are a few of the most important ones, along with the preliminary responses I had worked up at the time. To some extent people are still arguing about these points; you will see them reappear throughout the book.

## As if

One of the most prominent of the putdowns had only two words: “as if.” Briefly stated, the argument is that even if people are not capable of actually solving the complex problems that economists assume they can handle, they behave “as if” they can.

To understand the “as if” critique, it is helpful to look back a bit into the history of economics. The discipline underwent something of a revolution after World War II. Economists led by Kenneth Arrow, John Hicks, and Paul Samuelson accelerated an ongoing trend of making economic theory more mathematically formal. The two central concepts of economics remained the same—namely, that agents optimize and markets reach a stable equilibrium—but economists became more sophisticated in their ability to characterize the optimal solutions to problems as well as to determine the conditions under which a market will reach an equilibrium.

One example is the so-called theory of the firm, which comes down to saying that firms maximize profits (or share price). As modern theorists started to spell out precisely what this meant, some economists objected on the grounds that real managers were not able to solve such problems.

One simple example was called “marginal analysis.” Recall from [chapter 4](#) that a firm striving to maximize profits will set price and output at the point where marginal cost equals marginal revenue. The same analysis applies to hiring workers. Keep hiring workers until the cost of the last worker equals the increase in revenue that the worker produces. These results may seem innocuous enough, but in the late 1940s a debate raged in the *American Economic Review* about whether *real* managers actually behaved this way.

The debate was kicked off by Richard Lester, a plucky associate professor of economics at Princeton. He had the temerity to write to the owners of manufacturing companies and ask them to explain their processes for deciding how many workers to hire and how much output to produce. None of the executives reported doing anything that appeared to resemble “equating at the margin.” First, they did not seem to think about the effect of changes in the prices of their products or the possibility of changing what they paid to workers. Counter to the theory, they did not appear to think that changes in wages would affect either their hiring or output decisions much. Instead, they reported trying to sell as much of their product as they could, and increasing or decreasing the workforce to meet that level of demand.

Lester ends his paper boldly: “This paper raises grave doubts as to the validity of conventional marginal theory and the assumptions on which it rests.”

The defense team for the marginal theory was headed up by Fritz Machlup, who was then at the University of Buffalo but later joined Lester at Princeton, perhaps to continue the debate in person. Machlup brushed Lester’s survey data aside on the grounds that economists are not really interested in what people say they are doing. The theory does not require that firms explicitly calculate marginal costs and marginal revenues, he argued, but their actions nevertheless will approximate those predicted by the theory. He offered the analogy of a driver deciding when to pass a truck on a two-lane highway. The driver will not make any calculations, yet will manage to overtake the truck. An executive, he argued, would make decisions much the same way. “He would simply rely on his sense or his ‘feel’ of the situation . . . [and] would ‘just know’ in a vague and rough way, whether or not it would pay him to hire more men.” Machlup was highly critical of Lester’s data, but presented none of his own.

It is in the context of this debate that Milton Friedman, a young economist headed for fame, weighed in. In an influential essay called “The Methodology of Positive Economics,” Friedman argued that it was silly to evaluate a theory based on the realism of its assumptions. What mattered was the accuracy of the theory’s predictions. (He is using the word “positive” in his title here the way I use “descriptive” in this book, that is, as a contrast to normative.)

To illustrate his point, he traded Machlup’s driver for an expert billiard player. He notes that:

excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimum direction of travel, could estimate by eye the angles etc., describing the location of the balls, could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas. Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be expert billiard players.

Friedman was a brilliant debater and his argument certainly seemed compelling. For many economists at the time this settled the issue. The *AER* stopped publishing any more rounds of the debate it had been running, and economists returned to their models free from worry about whether their

assumptions were “realistic.” A good theory, it seemed, could not be defeated using just survey data, even if the defenders of the theory presented no data of their own. This remained the state of play some thirty years later, when I began to have my deviant thoughts. Even today, grunts of “as if” crop up in economics workshops to dismiss results that do not support standard theoretical predictions.

Fortunately, Kahneman and Tversky had provided an answer to the “as if” question. Both their work on heuristics and biases as well as that on prospect theory clearly showed that people did not act “as if” they were choosing in accordance with the rational economic model. When the subjects in one of Kahneman and Tversky’s experiments choose an alternative that is dominated by another one—that is, chosen in lieu of an alternative that is better in every way—there is no way they can be said to be acting *as if* they were making a correct judgment. There was also no way Professor Rosett’s wine-buying habits could be declared rational.

In homage to Friedman, whom I genuinely admired, I titled my first behavioral economics paper “Toward a Positive Theory of Consumer Choice.” The last section contained a detailed answer to the inevitable “as if” question. I too began with billiards. My main point was that economics is supposed to be a theory of everyone, not only experts. An expert billiard player might play as if he knows all the relevant geometry and physics, but the typical bar player usually aims at the ball closest to a pocket and shoots, often missing. If we are going to have useful theories about how typical people shop, save for retirement, search for a job, or cook dinner, those theories had better not assume that people behave as if they were experts. We don’t play chess like a grandmaster, invest like Warren Buffett, or cook like an Iron Chef. Not even “as if.” It’s more likely that we cook like Warren Buffett (who loves to eat at Dairy Queen). But a snappy retort to the “as if” critique was far from sufficient; to win the argument I would need hard empirical evidence that would convince economists.

To this day, the phrase “survey evidence” is rarely heard in economics circles without the necessary adjective “mere,” which rhymes with “sneer.” This disdain is simply unscientific. Polling data, which just comes from asking people whether they are planning to vote and for whom, when carefully used by skilled statisticians such as Nate Silver, yield remarkably accurate predictions of elections. The most amusing aspect of this anti-survey attitude is that many important macroeconomic variables are produced by

surveys!

For instance, in America the press often obsesses over the monthly announcement of the latest “jobs” data, with serious-looking economists asked to weigh in about how to interpret the figures. Where do these jobs numbers come from? They come from surveys conducted by the Census Bureau. The unemployment rate, one of the key variables in macroeconomic modeling, is also determined from a survey that asks people whether they are looking for work. Yet using published unemployment rate data is not considered a faux pas in macro-economics. Apparently economists don’t mind survey data as long as someone other than the researcher collected it.

But in 1980, survey questions were not going to overcome the “as if” grunt. There would need to be some proper data brought to bear that demonstrated that people misbehaved in their real-life choices.

## Incentives

Economists put great stock in incentives. If the stakes are raised, the argument goes, people will have greater incentive to think harder, ask for help, or do what is necessary to get the problem right. Kahneman and Tversky’s experiments were typically done with nothing at stake, so for economists that meant they could be safely ignored. And if actual incentives were introduced in a laboratory setting, the stakes were typically low, just a few dollars. Surely, it was often said, if the stakes were raised, people would get stuff right. This assertion, unsupported by any evidence, was firmly believed, even in spite of the fact that nothing in the theory or practice of economics suggested that economics only applies to large-stakes problems. Economic theory should work just as well for purchases of popcorn as for automobiles.

Two Caltech economists provided some early evidence against this line of attack: David Grether and Charlie Plott, one of my experimental economics tutors. Grether and Plott had come across research conducted by two of my psychology mentors, Sarah Lichtenstein and Paul Slovic. Lichtenstein and Slovic had discovered “preference reversals,” a phenomenon that proved disconcerting to economists. In brief, subjects were induced to say that they preferred choice A to choice B . . . and also that they preferred B to A.

This finding upset a theoretical foundation essential to any formal

economic theory, namely that people have what are called “well-defined preferences,” which simply means that we consistently know what we like. Economists don’t care whether you like a firm mattress better than a soft one or vice versa, but they cannot tolerate you saying that you like a firm mattress better than a soft one *and* a soft one better than a firm one. That will not do. Economic theory textbooks would stop on the first page if the assumption of well-ordered preferences had to be abandoned, because without stable preferences there is nothing to be optimized.

Lichtenstein and Slovic elicited preference reversals when they presented subjects with a pair of gambles: one a relatively sure thing, such as a 97% chance to win \$10, and the other more risky, such as a 37% chance to win \$30. They called the near sure thing the “p” bet, for high probability, and the more risky gamble the “\$” bet, since it offered a chance to win more money. First they asked people which gamble they preferred. Most took the p bet since they liked an almost sure win. For these subjects this means p is preferred to \$. Then they asked these p bet-loving subjects: “Suppose you owned the p bet. What is the lowest price at which you would be willing to sell it?” They also asked them the same question for the \$ bet. Strangely, a majority of these subjects demanded more to give up the \$ bet than the p bet, indicating they liked the \$ bet more. But this means they prefer the p bet to the \$ bet, and the \$ bet to the p bet. Blasphemy!

Grether and Plott wanted to know what was driving these weird results, and their leading hypothesis was incentives.\* If the bets were real, they conjectured, this nonsense would stop. So they ran the experiments for real money, and much to their surprise, the frequency and severity of the preference reversals actually increased. Raising the stakes made things worse.

This did not put an end to the incentive objection. But at least there was one paper to cite disputing the claim that money would solve all of the problems economists had with behavioral research. And, as we will see, this has been very much a recurring theme in the debate about the validity of experimental evidence.

## Learning

The style of experiment Kahneman and Tversky ran was often faulted as a “one-shot” game. In the “real world,” economists argued, people have

opportunities to learn. The idea is reasonable enough. We don't start out life as good drivers, but most of us do learn to drive without frequent mishaps. The fact that a clever psychologist can devise a question that will lure people in the lab into making a mistake does not necessarily imply that the same mistake would be made in the "real world." (Laboratories are thought to be unreal worlds.) Out there, people have had lots of time to practice their decision-making tasks, so they won't make the mistakes we see in the lab.

The problem with the learning story is that it assumes that we all live in a world like the Bill Murray movie *Groundhog Day*. Bill Murray's character keeps waking up and reliving the same day, over and over. Once he figures out what is going on, he is able to learn because he can vary things one at a time and see what happens. Real life is not as controlled as that, and thankfully so. But as a result, learning can be difficult.

Psychologists tell us that in order to learn from experience, two ingredients are necessary: frequent practice and immediate feedback. When these conditions are present, such as when we learn to ride a bike or drive a car, we learn, possibly with some mishaps along the way. But many of life's problems do not offer these opportunities, which raises an interesting point. The learning and incentives arguments are, to some extent, contradictory. This first occurred to me in a public debate of sorts that I had with the British game theorist Ken Binmore.

At a conference organized for graduate students, Binmore and I were each giving one lecture a day. I was presenting new findings of behavioral economics and although Binmore was presenting unrelated work, he took the opportunity at the beginning of each of his lectures to reply to the one I had given the day before. After my first lecture, Binmore offered a version of the "low stakes" critique. He said that if he were running a supermarket, he would want to consult my research because, for inexpensive purchases, the things I studied might possibly matter. But if he were running an automobile dealership, my research would be of little relevance. At high stakes people would get stuff right.

The next day I presented what I now call the "Binmore continuum" in his honor. I wrote a list of products on the blackboard that varied from left to right based on frequency of purchase. On the left I started with cafeteria lunch (daily), then milk and bread (twice a week), and so forth up to sweaters, cars, and homes, career choices, and spouses (not more than two or three per lifetime for most of us). Notice the trend. We do small stuff often

enough to learn to get it right, but when it comes to choosing a home, a mortgage, or a job, we don't get much practice or opportunities to learn. And when it comes to saving for retirement, barring reincarnation we do that exactly once. So Binmore had it backward. Because learning takes practice, we are more likely to get things right at small stakes than at large stakes. This means critics have to decide which argument they want to apply. If learning is crucial, then as the stakes go up, decision-making quality is likely to go down.

## Markets: the invisible handwave

The most important counter-argument in the Gauntlet involves markets. I remember well the first time Amos was introduced to this argument. It came during dinner at a conference organized by the leading intellectual figure at the Rochester business school where I had been teaching, Michael Jensen. At that time Jensen was a firm believer in both rational choice models and the efficiency of financial markets. (He has changed his views in various ways since then.) I think he saw the conference as a chance to find out what all the fuss around Kahneman and Tversky was about, as well as an opportunity to straighten out two confused psychologists.

In the course of conversation, Amos asked Jensen to assess the decision-making capabilities of his wife. Mike was soon regaling us with stories of the ridiculous economic decisions she made, like buying an expensive car and then refusing to drive it because she was afraid it would be dented. Amos then asked Jensen about his students, and Mike rattled off silly mistakes they made, complaining about how slow they were to understand the most basic economics concepts. As more wine was consumed, Mike's stories got better.

Then Amos went in for the kill. "Mike," he said, "you seem to think that virtually everyone you know is incapable of correctly making even the simplest of economic decisions, but then you assume that all the agents in your models are geniuses. What gives?"

Jensen was unfazed. "Amos," he said, "you just don't understand." He then launched into a speech that I attribute to Milton Friedman. I have not been able to find such an argument in Friedman's writings, but at Rochester at that time, people attributed it to Uncle Miltie, as he was lovingly called. The speech goes something like this. "Suppose there were people doing silly



things like the subjects in your experiments, and those people had to interact in competitive markets, then . . .”

I call this argument the *invisible handwave* because, in my experience, no one has ever finished that sentence with both hands remaining still, and it is thought to be somehow related to Adam Smith’s invisible hand, the workings of which are both overstated and mysterious. The vague argument is that markets somehow discipline people who are misbehaving. Handwaving is a must because there is no logical way to arrive at a conclusion that markets transform people into rational agents. Suppose you pay attention to sunk costs, and finish a rich dessert after a big dinner just because you paid for the dessert. What will happen to you? If you make this mistake often you might be a bit chubbier, but otherwise you are fine. What if you suffer from loss aversion? Is that fatal? No. Suppose you decide to start a new business because you are overconfident and put your chances of success at 90%, when in fact a majority of new businesses fail. Well, either you will be lucky and succeed in spite of your dumb decision, or you will muddle along barely making a living. Or perhaps you will give up, shut the business down, and go do something else. As cruel as the market may be, it cannot make you rational. And except in rare circumstances, failing to act in accordance with the rational agent model is not fatal.

Sometimes the invisible handwave is combined with the incentives argument to suggest that when the stakes are high and the choices are difficult, people will go out and hire experts to help them. The problem with this argument is that it can be hard to find a true expert who does not have a conflict of interest. It is illogical to think that someone who is not sophisticated enough to choose a good portfolio for her retirement saving will somehow be sophisticated about searching for a financial advisor, mortgage broker, or real estate agent. Many people have made money selling magic potions and Ponzi schemes, but few have gotten rich selling the advice, “Don’t buy that stuff.”

A different version of the argument is that the forces of competition inexorably drive business firms to be maximizers, even if they are managed by Humans, including some who did not distinguish themselves as students. Of course there is some merit to this argument, but I think it is vastly overrated. In my lifetime, I cannot remember any time when experts thought General Motors was a well-run company. But GM stumbled along as a badly-run company for decades. For most of this period they were also the largest

car company in the world. Perhaps they would have disappeared from the global economy in 2009 after the financial crisis, but with the aid of a government bailout, they are now the second largest automobile company in the world, a bit behind Toyota and just ahead of Volkswagen. Competitive forces apparently are slow-acting.

To be fair to Jensen, there is a more coherent version of his argument. Instead of arguing that markets force people to be rational, one can argue that market prices will still be rational, even if many individuals are decidedly Human. This argument is certainly plausible, perhaps even compelling. It just happens to be wrong. But how and why it is wrong is a long story that we will take up in Section VI.

For the field of behavioral economics to succeed, we needed answers to these questions. And in some quarters, we still do. But now, instead of snappy one-liners, it is possible to point to studies of real people interacting at high stakes in markets—even financial markets, where the invisible handwave would be expected to be most likely to be valid.



It was with the Gauntlet in my mind that I arrived at Cornell, in rural Ithaca, New York, in the fall of 1978. Ithaca is a small town with long, snowy winters, and not much to do. It was a good place to work.

While in California I had managed to finish two papers. One expounded on the List, and the other was called “An Economic Theory of Self-Control.” Writing the papers was the easy part; getting them published was another story. The first paper, mentioned earlier, “Toward a Positive Theory of Consumer Choice,” was rejected by six or seven major journals; I have repressed the exact count. In hindsight, I am not surprised. The paper had plenty of ideas, but little hard evidence to support them. Each rejection came with a set of referee reports, with often scathing comments that I would try to incorporate in the next revision. Still, I did not seem to be making any progress.

At some point I had to get this paper published, if for no other reason than that I needed to move on. Luckily, two open-minded economists were starting a new journal called the *Journal of Economic Behavior and Organization*. I guessed that they were anxious to get submissions, so I sent the paper to them and they published it in the inaugural issue. I had my first

behavioral economics publication, albeit in a journal no one had ever heard of.

If I were going to stay in academia and get tenure at a research-focused university like Cornell, I would have to start publishing regularly in top journals. I had returned from California with two ideas at the top of my list of topics to explore. The first was to understand the psychology of spending, saving, and other household financial behavior, what has now become known as mental accounting. The second was self-control and, more generally, choosing between now and later. The next two sections of the book take up those topics.

---

\* They favored this hypothesis even though Lichtenstein and Slovic (1973) had replicated their studies for real money on the floor of a casino in Las Vegas. Their dismissal of this evidence might be explained by another of their hypotheses. They also explicitly entertained the possibility that the perverse results were obtained simply because the experimenters were psychologists, who were known to deceive people in experiments. Needless to say, this hypothesis did not sit well with any psychologists who stumbled onto their paper.