This manuscript was written as an entry in the International Encyclopedia of the Social and Behavioral Sciences. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2000 by Sendhil Mullainathan and Richard H. Thaler. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
ABSTRACT

Behavioral Economics is the combination of psychology and economics that investigates what happens in markets in which some of the agents display human limitations and complications. We begin with a preliminary question about relevance. Does some combination of market forces, learning and evolution render these human qualities irrelevant? No. Because of limits of arbitrage less than perfect agents survive and influence market outcomes. We then discuss three important ways in which humans deviate from the standard economic model. Bounded rationality reflects the limited cognitive abilities that constrain human problem solving. Bounded willpower captures the fact that people sometimes make choices that are not in their long-run interest. Bounded self-interest incorporates the comforting fact that humans are often willing to sacrifice their own interests to help others. We then illustrate how these concepts can be applied in two settings: finance and savings. Financial markets have greater arbitrage opportunities than other markets, so behavioral factors might be thought to be less important here, but we show that even here the limits of arbitrage create anomalies that the psychology of decision making helps explain. Since saving for retirement requires both complex calculations and willpower, behavioral factors are essential elements of any complete descriptive theory.

Sendhil Mullainathan
Department of Economics
MIT
50 Memorial Drive, E52-380A
Cambridge, MA 02173
and NBER
mullain@mit.edu

Richard H. Thaler
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637
and NBER
richard.thaler@gsbpop.uchicago.edu
Introduction

It says something interesting about the field of economics that there is a sub-field called behavioral economics. Surely all of economics is meant to be about the behavior of economic agents, be they firms or consumers, suppliers or demanders, bankers or farmers. So, what is behavioral economics, and how does it differ from the rest of economics?

Economics traditionally conceptualizes a world populated by calculating, unemotional maximizers that have been dubbed Homo Economicus. In a sense, neo-classical economics has defined itself as explicitly “anti-behavioral”. Indeed, virtually all the behavior studied by cognitive and social psychologists is either ignored or ruled out in a standard economic framework. This unbehavioral economic agent has been defended on numerous grounds: some claimed that the model was “right”; most others simply argued that the standard model was easier to formalize and practically more relevant. Behavioral economics blossomed with the realization that neither point of view was correct. Empirical and experimental evidence mounted against the stark predictions of unbounded rationality. Further work made clear that one could formalize psychological ideas and translate them into testable predictions. The behavioral economics research program has consisted of two components: 1. Identifying the ways in which behavior differs from the standard model. 2. Showing how this behavior matters in economic contexts.

This paper gives a flavor of the program. We begin by discussing the most important ways in which the standard economic model needs to be enriched. We then illustrate how behavioral economics has been fruitfully applied to two important fields: finance and savings. But first, we discuss why the market forces and learning do not eliminate the importance of human actions.

Is Homo Economicus the Only One Who Survives?

Many economists have argued that a combination of market forces (competition and arbitrage) plus evolution should produce a world similar to that described in an economics textbook: do only the rational agents survive? Or, do the workings of markets at least render the actions of the quasi-rational irrelevant? These are questions that have been much studied in the past two decades, and the early impressions of many economists that markets would wipe out irrationality were, well, optimistic.

Consider a specific example: human capital formation. Suppose that a young economist, call him Sam, decides to become a behavioral economist, perhaps because Sam mistakenly thinks this will lead to riches, or because he thinks it is going to be the next fad, or because he finds it interesting and lacks the willpower to study “real” economics. Whatever the reason for the choice, let’s assume for the sake of argument that this decision was a mistake for Sam by any rational calculation. So, what will market forces do? Well, Sam may be poorer because of this choice than if he had sensibly chosen to study corporate finance, but he will not be destitute. Sam might even realize he could switch to corporate finance and make tons more money but is simply unable to resist the
temptation to continue wasting his time on behavioral economics. So, markets per se do not necessarily solve the problem: they provide an incentive to switch, but they cannot force Sam’s hand.

What about arbitrage? In this case, like most we study in economics outside the realm of financial markets, there is simply no arbitrage opportunity available. Suppose a wise arbitrageur is watching Sam’s choices, what bet can she place? None. The same can be said if Sam saves too little for retirement, picks the wrong wife, or buys the wrong car. None of these irrational acts generates an arbitrage opportunity for anyone else.

Indeed, economists now realize that even in financial markets there are important limits to the workings of arbitrage. First, in the face of irrational traders, the arbitrageur may privately benefit more from trading that helps push prices in the wrong direction than from trading that pushes prices in the right direction. Put another way, it may often pay “smart money” to follow “dumb money” rather than to lean against it (Haltiwanger and Waldman, 1985; Russell and Thaler 1985). For example, an extremely smart arbitrageur near the beginning of the tulip mania would have profited more from buying tulips and further destabilizing prices than by shorting them. Second, and slightly related, arbitrage is inherently risky activity and consequently the supply of arbitrage will be inherently limited (De Long, Shleifer, Summers and Waldman, 1990). Arbitrageurs who did decide to short tulips early would probably have been wiped out by the time their bets were proven to be “right”. Add to this the fact that in practice most arbitrageurs are managing other people’s money and, therefore judged periodically, and one sees the short horizons that an arbitrageur will be forced to take on. This point was made forcefully by Shleifer and Vishny (1997) who essentially foresaw the scenario that ended up closing Long Term Capital Management.

So, markets per se cannot be relied upon to make economic agents rational. What about evolution? An old argument that individuals who failed to maximize should have been weeded out by evolutionary forces, which presumably operated during ancient times. Overconfident hunters, for example, presumably caught less prey, ate less and died younger. Such reasoning, however, has turned out to be faulty. Evolutionary arguments can just as readily explain over-confidence as they can explain appropriate levels of confidence. For example, consider individuals playing a war of attrition (perhaps in deciding when to back down during combat). Here overconfidence will actually help. Seeing the overconfidence, a rational opponent will actually choose to back down sooner. As can be seen from this example, depending on the initial environment (especially when these environments have a game theoretic component to them), evolution may just as readily weed out rational behavior as it does weed out quasi-rational behavior. The troubling flexibility of evolutionary models means that they can just as readily argue for bounds on rationality.

The final argument is that individuals who systematically and consistently make the same mistake will eventually learn the error of their ways. This kind of argument has also not stood up well under theoretical scrutiny. First, the optimal experimentation literature has shown that there can be a complete lack of learning even in infinite horizons). The
intuition here is simple: as long as there are some opportunity costs to learning or to experimenting with a new strategy, even a completely “rational” learner will choose not to experiment. This player will get stuck in a non-optimal equilibrium, simply because the cost of trying something else is too high. Second, work on learning in games has formally demonstrated Keynes’ morbid observation on the “long run”. The time required to converge to an equilibrium strategy can be extremely long. Add to this a changing environment and one can easily be in a situation of perpetual non-convergence. In practice, for many of the important decisions we make, both arguments apply with full force. The number of times we get to learn from our retirement decisions is low (and possibly zero). The opportunity cost of experimenting with different ways of choosing a career can be very high.

The upshot of all these theoretical innovations has been clear. One cannot defend unbounded rationality on purely theoretical grounds. Neither arbitrage, competition, evolution, nor learning necessarily guarantees that unbounded rationality must be an effective model. In the end, as some might have expected, it must ultimately be an empirical issue. Does “behavior” matter? Before evaluating this question in two different fields of application, we explore the ways in which real behavior differs from the stylized neoclassical model.

**Three Bounds of Human Nature**

The standard economic model of human behavior includes (at least) three unrealistic traits: unbounded rationality, unbounded willpower, and unbounded selfishness. These three traits are good candidates for modification.

Herbert Simon (1955) was an early critic of modeling economic agents as having unlimited information processing capabilities. He suggested the term “bounded rationality” to describe a more realistic conception of human problem solving capabilities. As stressed by Conlisk (1996), the failure to incorporate bounded rationality into economic models is just bad economics—the equivalent to presuming the existence of a free lunch. Since we have only so much brainpower, and only so much time, we cannot be expected to solve difficult problems optimally. It is eminently “rational” for people to adopt rules of thumb as a way to economize on cognitive faculties. Yet the standard model ignores these bounds and hence the heuristics commonly used. As shown by Kahneman and Tversky (1974), this oversight can be important since sensible heuristics can lead to systematic errors.

Departures from rationality emerge both in judgments (beliefs) and in choice. The ways in which judgment diverges from rationality is long and extensive (see Kahneman, Slovic and Tversky, 1982). Some illustrative examples include overconfidence, optimism, anchoring, extrapolation, and making judgments of frequency or likelihood based on salience (the availability heuristic) or similarity (the representativeness heuristic).

Many of the departures from rational choice are captured by prospect theory (Kahneman and Tversky 1979), a purely descriptive theory of how people make choices under
uncertainty (see Starmer 2000 for a review of literature on non-EU theories of choice). Prospect theory is an excellent example of a behavioral economic theory in that its key theoretical components incorporate important features of psychology. Consider three features of the prospect theory value function. 1. It is defined over changes to wealth rather than levels of wealth (as in EU) to incorporate the concept of adaptation. 2. The loss function is steeper than the gain function to incorporate the notion of “loss aversion”; the notion that people are more sensitive to decreases in their well being than to increases. 3. Both the gain and loss function display diminishing sensitivity (the gain function is concave, the loss function convex) to reflect experimental findings. To fully describe choices prospect theory often needs to be combined with an understanding of “mental accounting” (Thaler, 1985). One needs to understand when individuals faced with separate gambles treat them as separate gains and losses and when they treat them as one, pooling them to produce one gain or loss.

A couple of examples can illustrate how these concepts are used in real economics contexts. Consider overconfidence. If investors are overconfident in their abilities, they will be willing to make trades even in the absence of true information. This insight helps explain a major anomaly of financial markets. In an efficient market when rationality is common knowledge, there is virtually no trading, but in actual markets there are hundreds of millions of shares traded daily and most professionally managed portfolios are turned over once a year or more. Individual investors also trade a lot: they incur transaction costs and yet the stocks they buy subsequently do worse than the stocks they sell.

An example involving loss aversion and mental accounting is Camerer et al’s (1997) study of New York City taxi cab drivers. These cab drivers pay a fixed fee to rent their cabs for twelve hours and then keep all their revenues. They must decide how long to drive each day. A maximizing strategy is to work longer hours on good days (days with high earnings per hour such as rainy days or days with a big convention in town) and to quit early on bad days. However, suppose cabbies set a target earnings level for each day, and treat shortfalls relative to that target as a loss. Then, they will end up quitting early on good days and working longer on bad days, precisely the opposite of the rational strategy. This is exactly what Camerer et al find in their empirical work.

Having solved for the optimum, Homo Economicus is next assumed to choose the optimum. Real humans, even when they know what is best, sometimes fail to choose it for self-control reasons. Most of us at some point have eaten, drank, or spent too much, and exercised, saved, or worked too little. Such is human nature, at least since Adam and Eve. (Well, make that Adam.) People (even economists) also procrastinate. We are writing this entry well after the date on which it was due, and we are confident that we are not the only guilty parties. Though people have these self-control problems, the are at least somewhat aware of them: they join diet plans and buy cigarettes by the pack (because having an entire carton around is too tempting). They also pay more withholding taxes than they need to (in 1997, nearly 90 million tax returns paid an average refund of around $1300) in order to assure themselves a refund, but then file their
taxes near midnight on April 15 (at the post office that is being kept open late to accommodate their fellow procrastinators.)

Finally, people are boundedly selfish. Although economic theory does not rule out altruism, as a practical matter economists stress self-interest as the primary motive. For example, the free rider problems widely discussed in economics are predicted to occur because individuals cannot be expected to contribute to the public good unless their private welfare is thus improved. In contrast, people often take selfless actions. In 1993, 73.4% of all households gave some money to charity, the average dollar amount being 2.1% of household income. Also, 47.7% of the population does volunteer work with 4.2 hours per week being the average hours volunteered. Similar selfless behavior is observed in controlled laboratory experiments. Subjects systematically often cooperate in public goods and prisoners dilemma games, and turn down unfair offers in “ultimatum” games.

Finance

If economists were polled twenty years ago and asked to name the domain in which bounded rationality was least likely to find useful applications the likely winner would have been finance. The limits of arbitrage arguments were not well understood at that time, and one leading economist had called the efficient markets hypothesis the best established fact in economics. Times change. Now, as we begin the 21st century finance is perhaps the branch of economics where behavioral economics has made the greatest contributions. How has this happened?

Two factors contributed to the surprising success of behavioral finance. First, financial economics in general, and the efficient market hypothesis in particular, generated sharp, testable predictions about observable phenomena. Second, there are great data readily available to test these sharp predictions. We briefly summarize here a few examples.

The rational efficient markets hypothesis makes two classes of predictions about stock price behavior. The first is that stock prices are “correct” in the sense that asset prices reflect the true or rational value of the security. In many cases this tenet of the efficient market hypothesis is untestable because intrinsic values are not observable. However, in some special cases the hypothesis can be tested by comparing two assets whose relative intrinsic values are known. One class of these is called “Siamese Twins”: two versions of the same stock that trade in different places.

A specific well-known example is the case of Royal Dutch Shell as documented in Froot and Dabora (1999). The facts are that Royal Dutch Petroleum and Shell Transport are independently incorporated in the Netherlands and England respectively. The current firm emerged from a 1907 alliance between Royal Dutch and Shell Transport in which the two companies agreed to merge their interests on a 60:40 basis. Royal Dutch trades primarily in the US and the Netherlands and Shell trades primarily in London. According to any rational model, the shares of these two components (after adjusting for foreign exchange) should trade in a ratio of 60:40. They do not; the actual price ratio has
deviated from the expected one by more than 35%. Simple explanations such as taxes and transactions costs cannot explain the disparity (see Froot and Dabora). This example illustrates that prices can diverge from intrinsic value because of limits of arbitrage. Some investors do try to exploit this mispricing, buying the cheaper stock and shorting the more expensive one, but this is not a sure thing, as many hedge funds learned in the Summer of 1998 (when, at the time hedge funds were trying to get more liquidity, the pricing disparity widened).

The Royal Dutch Shell anomaly is a violation of one of the most basic principles of economics: the law of one price. Another similar example is the case of closed-end mutual funds (Lee, Shleifer, and Thaler 1991). These funds are much like typical (open-end) mutual funds except that to cash out of the fund, investors must sell their shares on the open market. This means that closed-end funds have market prices that are determined by supply and demand, rather than set equal to the value of their assets by the fund managers as in an open-end fund. Since the holdings of closed-end funds are public, market efficiency would lead one to expect that the price of the fund should match the price of the underlying securities they hold (the net asset value or NAV). Instead, closed-end funds typically trade at substantial discounts relative to their NAV, and occasionally at substantial premia. Most interesting from a behavioral perspective is that closed-end fund discounts are correlated with one another and appear to reflect individual investor sentiment. (Closed-end funds are primarily owned by individual investors rather than institutions.) Lee, Shleifer and Thaler find that discounts shrink in months when shares of small companies (also owned primarily by individuals) do well, and in months when there is lots of IPO activity, indicating a “hot” market. Since these findings were predicted by their theory, they move the research beyond the demonstration of an embarrassing fact (price not equal to NAV) toward a constructive understanding of how markets work.

The second principle of the efficient market hypothesis is “unpredictability”. In an efficient market it is not possible to predict future stock price movements based on publicly available information. Many early violations of this had no explicit link to behavior. Thus it was reported that small firms, firms with low price earnings ratios earned higher returns than other stocks with the same risk. Also, stocks in general, but especially stocks of small companies have done well in January and on Fridays (but poorly on Mondays).

An early study by De Bondt and Thaler (1985) was explicitly motivated by the psychological finding that individuals tend to over-react to new information. For example, experimental evidence suggest that people tended to underweight base rate data (or prior information) in incorporating new data. De Bondt and Thaler hypothesized that if investors displayed this behavior, then stocks that had performed quite well over a period of years will eventually have prices that are too high. Individuals overreacting to the good news will drive the prices of these stocks too high. Similarly, poor performers will eventually have prices that are too low. This yields a prediction about future returns: past “winners” ought to underperform while past “losers” ought to outperform the market. Using data for stocks traded on the New York Stock Exchange, De Bondt and
Thaler found that the 35 stocks that had performed the worst over the past five years (the losers) outperformed the market over the next five years, while the 35 biggest winners over the past five years subsequently underperformed. Follow-up studies have shown that these early results cannot be attributed to risk (by some measures the portfolio of losers is actually less risky than the portfolio of winners), and can be extended to other measures of overreaction such as the ratio of market price to the book value of equity.

More recent studies have found other violations of unpredictability that have the opposite pattern from that found by DeBondt and Thaler, namely underreaction rather than overreaction. Over short periods of time, e.g., six months to one year, stocks display momentum—the stocks that go up the fastest for the first six months of the year tend to keep going up. Also, after many corporate announcements such as large earnings changes, dividend initiations and omissions, share repurchases, splits, and seasoned equity offerings, there is an initial price jump on the day of the announcement followed by a slow drift in the same direction for as long as a year or more (see Shleifer 2000).

These findings of underreaction are a further challenge to the efficient markets hypothesis, but also to behavioral finance. Do markets sometimes overreact, and sometimes underreact? If so, can any pattern be explained, at least ex post? Is there a unifying framework that can bring together these apparently opposing facts and ideas?

Work is only now beginning to attack this extremely important question. Several explanations have recently been offered (Shleifer 2000 summarizes them). All rely on psychological evidence (in one way or another) in unifying the facts. They all explain the anomalies by noting that underreaction appears at short horizons while overreaction appears at longer horizons, but each paper provides its own distinctive explanation. Which (if any) of these is the best one has yet to be decided. But the facts discovered so far, combined with these models, demonstrate the changing nature of finance. Rigorous empirical work building on psychological phenomena has given us new tools to unearth interesting empirical facts. Rigorous theoretical work, on the other hand, has been trying to address the challenge of incorporating these empirical facts into a psychologically plausible model.

The research discussed so far has been about asset prices and the controversy about the efficient market hypothesis. There is another stream of research that is just about investor behavior, not about prices. One example of this stream is motivated by mental accounting and loss aversion. The issue is whether investors are reluctant to realize capital losses (because they would have to “declare” the loss to themselves). Shefrin and Statman (1985) dubbed this hypothesis the “disposition effect”. The prediction that investors will be less willing to sell a loser than a winner is striking since the tax law encourages just the opposite behavior. Nevertheless, Odean (1998) finds evidence of just this behavior. In his sample of the customers of a discount brokerage firm, investors were more likely to sell a stock that had increased in value than one that had decreased in value. While around 15% of all gains were realized, only 10% of all losses are realized. This hesitancy to realize gains came at a cost. Odean shows that the stocks the loser stocks held under-performed the gainer stocks that were sold.
Savings

If finance was the field in which a behavioral approach was least likely, a priori, to succeed, saving had to be one of the most promising. The standard life-cycle model of savings abstracts from both bounded rationality and bounded willpower, yet saving for retirement is both a difficult cognitive problem and a difficult self-control problem. It is then, perhaps less surprising that a behavioral approach has been fruitful here. As in finance, progress has been helped by the combination of a refined standard theory with testable predictions and lots of data sources on household saving behavior.

One crisp prediction of the life-cycle model is that savings rates are independent of income. Suppose twins Tom and Ray are identical in every respect except that Tom earns most of his money early in his life (say he’s a basketball player), while Ray earns most of his late in life (say he’s a manager). The life-cycle model predicts that Tom the basketball player ought to save his early income to increase consumption later in life, while Ray the manager ought to borrow from his future income to increase consumption earlier in life. This prediction is completely unsupported by the data, which shows that consumption very closely tracks income over individuals’ life cycles. Furthermore, the departures from predicted behavior cannot be explained merely by people’s inability to borrow. Banks, Blundell and Tanner (1998) show, for example, that consumption drops sharply as individuals retire and their incomes drop. They have simply not saved enough for retirement. Indeed, many low to middle income families have essentially no savings whatsoever. The primary cause of this lack of saving appears to be self-control. One bit of evidence supporting this conclusion is that virtually all saving done by Americans is accomplished in vehicles that support what are often called “forced savings”, e.g., accumulating home equity by paying the mortgage and participation in pension plans. Coming full circle, one “forced” savings individuals may choose themselves may be high tax withhholdings, so that when the refund comes they can buy something they might not have had the willpower to save up for.

One of the most interesting research areas has been devoted to measuring the effectiveness of tax-subsidized savings programs such as IRAs and 401(k)s. The standard analysis of these programs is quite simple. Consider the original IRA program from the early 1980s. This program provided tax subsidies for savings up to a threshold, often $2000 per year, there was no marginal incentive to save for any household that was already saving more than $2000 per year. Thus, those saving more than the threshold should not increase their total saving, they will merely switch some money from a taxable account to the IRA for the tax gain. Moreover, everyone who is saving a positive amount should participate in this program for the infa-marginal gain. The actual analysis of these programs has shown that the reality is not so clear. By some accounts at least, these programs appear to have generated quite a bit of new savings. Some argue that almost every dollar of savings in IRAs appear to represent new savings In other words, people are not simply shifting their savings into IRAs and leaving their total behavior unchanged. Similar results are found for 401(k) plans. The behavioral explanation for these findings is that IRAs and 401(k) plans help solve self-control problems by setting
up special mental accounts that are devoted to retirement savings. Households tend to respect the designated use of these accounts, and their self-control is helped by the tax penalty that must be paid if funds are removed prematurely. (Some issues remain controversial. See the debate in the Fall 1996 issue of the Journal of Economic Perspectives.)

An interesting flip side to IRA/401(k) programs is that these programs also generated far less than the full participation one would have expected. Many eligible people do not participate, foregoing in effect a cash transfer from the government (and, in some cases, from their employer). O’Donoghue and Rabin (1999) present an explanation based on procrastination and hyperbolic discounting. Individuals typically show very sharp impatience for short horizon decisions, but much more patience at long horizons. This is often referred to as hyperbolic discounting by contrast with the standard assumption of exponential discounting, in which patience is independent of horizon. In exponential models, people are equally patient at long and short horizons. O’Donoghue and Rabin argue that hyperbolic individuals will show exactly the low IRA participation that we observe. Though hyperbolic people will want to eventually participate in IRAs (because they are patient in the long run), something always comes up in the short run (where they are very impatient) that provides greater immediate reward. Consequently, they may delay joining the IRA indefinitely.

If people procrastinate about joining the savings plan, then it should be possible to increase participation rates simply by lowering the psychic costs of joining. One simple way of accomplishing this is to switch the default option for new workers. In most companies, when employees first become eligible for the 401(k) plan they receive a form inviting them to join; to join they have to send the form back and make some choices. Several firms have made the seemingly inconsequential change of switching the default: employees are enrolled into the plan unless they explicitly opt out. This change has often produced dramatic increases in savings rates. For example, in one company studied by Madrian and Shea (2000) the employees who joined after switching the default to being in the plan were 50% more likely to participate than the workers in the year prior to the change. (They also find that the default asset allocation had a strong effect on workers choices. The firm had made the default asset allocation 100% in a money market account, and the proportion of workers selection this allocation soared.)

Along with these empirical facts, there has also been a bulk of theoretical work. To cite a few examples, Laibson (1997) and O’Donoghue and Rabin (1999) have both examined the effects of hyperbolic discounting on the savings decisions. They’ve highlighted that hyperbolic individuals will demand commitment devices (savings vehicles which are illiquid) and generally fail to obey the Euler Equation.

**Other Directions**

We have concentrated on two fields here in order to give a sense of what behavioral economics can do, but we do not want to leave the impression that savings and financial markets are the only domains in which a behavioral approach has been or could be
effective. In labor economics, experimental and empirical work has underlined the importance of fairness considerations in setting wages. For example, wages between industries differ dramatically, even for identical workers and most interestingly, even homogeneous workers (such as janitors) earn higher wages when they work in industries where other occupations earn more. In Law and Economics, we have seen the importance of “irrelevant” factors in a jury’s decision to sentence or in the magnitude of awards they give. In corporate finance, one can fruitfully interpret a manager’s to acquire or diversify as resulting from overconfidence. One could go on. There is much to be done.

References


