Behavioral Economics

Berg, Nathan

2010

Behavioral Economics

Nathan Berg

nberg@utdallas.edu

University of Texas-Dallas

Introduction: Definitions and Naming Problems in Behavioral Economics

Definition of Behavioral Economics

Behavioral economics is the subfield of economics that borrows from psychology, empirically tests assumptions used elsewhere in economics, and provides theories that aim to be more realistic and closely tied to experimental and field data. In a frequently cited survey article, Rabin (1998) describes behavioral economics as "psychology and economics," which is a frequently used synonym for behavioral economics. Similarly, Camerer (1999) defines behavioral economics as a research program aimed at reunifying psychology and economics.

Reunification is a relevant description because of the rather tumultuous relationship between psychology and economics in the arc of economic history. A number of preeminent founders of important schools of economic thought, including Adam Smith, wrote extensively on psychological dimensions of human experience and economic behavior, while later economists sometimes sought explicitly to exclude psychology from economic analysis. For example, Slutsky, whose famous equation is taught to nearly all upper-level microeconomics students, sought to erect a boundary excluding psychology from economics: "[I]f we wish to place economic science upon a solid basis, we must make it completely independent of psychological assumptions" (Slutsky, 1915, p. 27).

Although historical accounts vary, one standard narrative holds that in the 20th century neoclassical economists made an intentional break with psychology in contrast to earlier classical and institutional economists who actively integrated psychology into their writings on economics (e.g., Bruni and Sugden, 2007). In 20th century economics' break with psychology, one especially important source is Milton Friedman's (1953) essay. In it, Friedman argues that unrealistic or even obviously untrue assumptions—especially, the core assumption used throughout much of contemporary economics (including much of behavioral economics) that all behavior can be modeled as resulting from decision makers solving constrained optimization problems—are perfectly legitimate, so long as they produce accurate predictions. Friedman put forth the analogy of a billiards player selecting shots "as if" he or she were solving a set of equations describing the paths of billiards balls based on Newtonian physics. We know that most expert billiards players have not studied academic physics and therefore do not in fact solve a set of equations each time they set up a shot. Nevertheless, Friedman argues that this model based on manifestly wrong assumptions should be judged strictly in terms of the predictions it makes and not the realism of its assumptions.

In contrast to Friedman's professed lack of interest in investigating the realism of assumptions, behavioral economists have made it a core theme in their work to empirically test assumptions in economic models and modify theory according to the results they observe. In spite of this difference with neoclassical economists like Friedman, behavioral economists frequently use as-if arguments to defend behavioral models, leading some methodological observers to see more similarity than contrast in the behavioral and neoclassical approaches (Berg and Gigerenzer, forthcoming).

Bounded Rationality

The term bounded rationality, coined by Nobel laureate, Herbert Simon, is strongly associated with behavioral economics, although there appears to be far less agreement on the term's meaning. The neoclassical model assumes that economic man, or *homo economicus*, is infinitely self-interested, infinitely capable of processing information and solving optimization problems, and infinitely self-disciplined or self-consistent when it comes to having the willpower to execute one's plans—whether those plans concern how much junk food to eat or how much to save for retirement. In contrast, much of behavioral economics focuses on limits, or bounds, on one or more of these three assumptions. Thus, bounded self-interest, bounded information processing capacity, and bounded willpower are three guiding themes in the behavioral economics literature.

Bounded self-interest enjoys widespread interest in behavioral economics, which has proposed numerous models of so-called social preferences to address a number of observations from human experiments that appear to falsify the assumption that people maximize their own monetary payoffs. A decision maker with social preferences cares about the material or monetary payoffs of others as well as his or her own, although the manner in which concern for others' payoffs is expressed can take a variety of forms. For example, a person with social preferences might be happier when others are worse off, which is sometimes described as spite; happier when others are better off, which is sometimes described as altruism; prefer equal over unequal allocations of money, which is sometimes described as inequality aversion; prefer allocations in which the sum of all people's payoffs is maximized, which is sometimes described as a preference for social welfare; prefer allocations in which the least well-off person has a larger payoff, which is sometimes described as a Rawlsian preference; or prefer allocations of resources in which his or her payoff is large relative to others, which is sometimes described as a competitive preference (Charness and Grosskopf, 2001). Common to all these variations and many other forms of social preferences is that people are not generally indifferent between two allocations of payoffs for all members in a group just because their own monetary payoff is the same. This violates the common neoclassical assumption that people are infinitely self-interested, because it would imply that people are indifferent so long as their own material payoffs are held constant.

Bounded information-processing capacity is another active area within behavioral economics, which would have looked very much out of place in the mainstream economics literature only three decades ago. Topics in this area include limited memory, limited attention, limited number of degrees of perspective-taking in strategic interaction, limited perceptual capacity, distorted beliefs, and decision and inference processes that violate various tenets of logic and probability theory (see Camerer, 2003, for examples).

Bounded willpower, often described as time-inconsistency or dynamic inconsistency, is another large and growing part of the behavioral economics research program. In the neoclassical optimization model, decision makers choose a sequence of actions through time by selecting the best feasible sequence, with virtually no mention of the costs associated with implementing that plan over the course of people's lives. If there is no new information, then the neoclassical inter-temporal choice problem is decided once and for all before the first action in the sequence is taken. Unlike the neoclassical model's assumption that acting on the optimal plan of action through time is costless, behavioral economists studying bounded willpower focus squarely on the tension between what a person wants him- or herself to do tomorrow versus what he or she actually does. This tension can be described as subjective inconsistency concerning what is best for oneself at a specific point in time which, contrary to the neoclassical assumption, changes as a function of the time at which the decision is considered. One can think of planning now to start working on a term paper tomorrow but then tomorrow deciding to do something else instead—and regretting it after the fact.

Empirical Realism

Proponents of bringing psychology more deeply into economics argue that it is necessary to depart from the assumptions of *homo economicus* to achieve improved empirical realism (i.e., more accurate descriptions of economic behavior, and better predictions following a change in policy or other economic conditions). In an article published in *Journal of Business* titled "Rationality in psychology and economics," Simon (1986, p. S209) writes:

> The substantive theories of rationality that are held by neoclassical economists lack an empirically based theory of choice. Procedural theories of rationality, which attempt to explain what information people use when making choices and how information is processed, could greatly improve the descriptive and forecasting ability of economic analysis.

Improving the empirical realism of economic analysis is a primary and ongoing motivation in behavioral economics, frequently stated in the writings and presentations of behavioral economists.

Example of reference-point dependent utility functions

Similarly to Simon, Rabin (1998) argues that theories and experimental results from psychology enrich mainstream economics. But Rabin's idea about how behavioral economists can bring more empirical realism into economics is much more narrowly circumscribed than Simon's, with Rabin essentially arguing that behavioral economics should proceed within the utility maximization model of neoclassical economics. Despite their occasional claims to radical or revolutionary methodological innovation, many behavioral economists side with neoclassical economists in viewing constrained optimization as a non-negotiable methodological tenet that defines and distinguishes economics from other disciplines. Rabin says that the new empirical content that behavioral economists bring to bear will help economics as a whole to more realistically describe people's utility functions.

For example, as mentioned earlier in the discussion of the neoclassical model's assumption of unbounded self-interest, this assumption is often interpreted to mean that consumers care only about their own levels of consumption and that workers care only about their own income, irrespective of what others are consuming or earning. Behavioral economics models, in contrast, allow utility to depend on the *difference* between one's own level of consumption or income and a reference-point level. The reference-point level might reflect what one is accustomed to or reflect a social comparison made with respect to the average level within a social group.

Thus, a worker with a behavioral reference-point-dependent utility function might prefer an annual salary of $90,000 at a company where the average worker earns $50,000 over a salary of $95,000 at a company where the average worker earns $200,000. In the standard economic model, only the worker's own payoffs should determine the ranking of job opportunities, holding all else equal, and not the comparison of one's own income with that of other workers. The reference-point-dependent utility function tries to reflect the observation that many normal, healthy, and socially intelligent people do in fact care about their own payoffs relative to others. For some workers, it may be worthwhile to trade off a few thousand dollars of their own salary for a work environment where the relative pay structure is more to their liking (e.g., a feeling of relative high status in the $90,000 job being subjectively worth more than the extra $5,000 of income at the $95,000 job).

It should be mentioned, however, that reference-point-dependent theories in behavioral economics are not entirely new. One finds interpersonal comparisons in Veblen's concept of conspicuous consumption from his classic *The Theory of the Working Class* (1899) and even earlier among some classical economists. Gigerenzer (2008) and Jorland (1987) have further questioned behavioral economists' historical reading of Kahneman and Tversky's (1979) prospect theory and its reliance on a reference point, which Gigerenzer and Jorland argue was already present in Daniel Bernoulli's original description of the expected utility function. According to Gigerenzer and Jorland, Bernoulli specified the argument of the expected utility function to be a *change* in, rather than absolute level of, wealth. This would imply that the reference point introduced by Kahneman and Tversky's prospect theory was in fact a re-introduction, interesting especially in light of Bernoulli and then Kahneman and Tversky's respective roles in the so-called repair program that began with expected value maximization (Güth, 2008; Gigerenzer, 2008, p. 90). Expected value maximization was, before Daniel Bernoulli, the unquestioned standard of rationality in gambling and games of chance. The St. Petersburg Paradox revealed the shortcomings of expected value maximization as a standard of rationality. Bernoulli introduced nonlinear transformations of payoffs--while retaining the weighting and summing operation underlying the mathematical expectation--to repair the rationality program which had begun as expected value maximization by putting forward expected utility maximization. Then came Allais' Paradox, which pointed to fundamental problems with expected utility theory. Once again, a repair was put forward by Kahneman and Tversky (1979) whose prospect theory introduced nonlinear transformations of probabilities to rationalize anomalous choice behavior while retaining the process of weighting and summing—this time, using transformed payoffs and transformed probabilities—but still analogous to the mathematical expectation. More recently, Brandstätter, Gigerenzer and Hertwig (2006) introduced alternative psychological models of risky choice based on heuristics that depend only on partial information about random payoffs.

Debates With Neoclassical and Within Behavioral Economics Concerning Realism of Assumptions

There is active debate between behavioral and non-behavioral economists—and among behavioral economists—about the extent to which empirical realism is being achieved by the behavioral economics research program. These two distinct layers of debate need to be untangled to appreciate the different issues at play and how behavioral economics is likely to influence public policy now that the Obama administration has recruited among its top advisers a number of behavioral economists, including Richard Thaler, Cass Sunstein and Daniel Kahneman.

When trying to convince neoclassical economists who are skeptical about the need for behavioral economics, behavioral economists point to the improved ability of their psychology-inspired models to fit data collected from a variety of sources, including experimental, macroeconomic, and financial market data. Skeptics from outside behavioral economics have questioned whether the deviations from neoclassical assumptions have any important consequences for the economy as a whole, suggesting that they might perhaps "average out" in the aggregate. Skepticism about the relevance of experimental data remains strong, with many doubts expressed about whether the college students who participate in economic experiments can be relied upon to teach us anything new about economics, and whether anything learned in one laboratory experiment can be generalized to broader populations in the economy—the so-called problem of external validity. Experimentalists have responded that the reason they carefully incentivize decisions by making subject payments dependent on their decisions is to make it costly for them to misrepresent their true preferences. Experimentalists have addressed the issue of external validity by going into the field with so-called field experiments, and by conducting experiments among different subpopulations, such as financial market traders, Japanese fishermen, and other groups of adult workers (e.g., Carpenter and Seki, 2006).

Within behavioral economics, a different debate takes place. Among behavioral economists, despite a shared commitment to borrowing from psychology and other disciplines, there remains tension over how far to move away from constrained optimization as the singular organizing framework of neoclassical theory and in much of behavioral economics, too. An alternative approach, advocated by a minority of more psychology- and less economics-inspired behavioral economists, seeks to break more substantially with neoclassical economics, dispensing with optimization theory as a necessary step in deriving equations that describe behavior. Constrained optimization, whether in behavioral or neoclassical economics, assumes that decision makers see a well-defined choice set; exhaustively scan this set, plugging each possible action into a scalar-valued objective function, which might include parameters intended to capture psychological phenomena; weigh the costs and benefits associated with each action, which includes psychic costs and benefits; and finally choose the element in the choice set with the highest value according to the objective function. There is very little direct evidence of people making decisions—especially high stakes decisions, such as choosing a career, buying a house, or choosing whom to marry—according to the constrained optimization process just described. In many real-world decisions such as those just mentioned, the choice set is impossibly large to clearly define and exhaustively search through. In other settings such as choosing a life partner or whom to marry, constrained optimization would be seen by some to violate important social norms.

Instead, critics such as Gigerenzer and Selten (2001) attempt to base theory directly on empirical description of actual decision processes. Like other economists, these critics use equations to describe behavior. However, their behavioral equations skip the step of deriving behavioral equations as solutions to constrained optimization problems. To these researchers, theorizing and observing how decision makers deal with the overwhelmingly high-dimensional choice sets they face, quickly searching for a good-enough action and discarding the rest, is a fundamental scientific question of primary importance. Herbert Simon referred to such threshold-seeking behavior as *satisficing* as distinct from optimizing.

Indeed, some leading voices in the applied area of marketing have recently discovered that they can make improved predictions about the way customers search for information and make purchase decisions by abandoning the optimization model. One reason why optimization does not explain observed consumer behavior is that it takes too long. For example, when shopping for a cell phone with a choice set containing 100 possible phones, the process of scoring each phone according to an objective function that depends on a vector of 15 phone features requires exhaustive search through all features of all phones, which is 100×15 = 1,500 pieces of information. Similarly, a complete ranking of all 100 phones would require consideration of all possible pairs, of which there are thousands. Very few people actually shop for cell phones like this (Yee, Dahan, Hauser and Orlin, 2007). Instead, data about customers' information search suggests that a handful of threshold criteria are imposed (e.g., the desired phone should weigh less than 10 ounces, have a flip-top, and cost less than $200), which effectively reduces the choice set of 100 to a manageable set of just a few remaining options, about which the customer will look up several more features and then make a choice. Notice that the satisficing or threshold-based conditions which allow for no compensating trade-offs are effective precisely because they quickly shrink the choice set to a manageable size and require only a small subset of the information about all 100 phones' features.

Thus, the perspective of the Gigerenzer and Selten heuristics school of thought is that being smart requires strategies of simplification—using simple heuristics that are appropriately matched to the environments in which they are used. This alternative normative concept that asks how well matched decision procedures are to the environments in which they are used is referred to by Gigerenzer and Vernon Smith as *ecological rationality* and stands in contrast to axiomatic rationality that dominates in both neoclassical and behavioral economics. Normative evaluation of the way in which experts play chess illustrates the difference between these distinct normative concepts.

German mathematician, Ernst Zermelo, proved that there exists an optimal strategy in chess. But the combinatorics of many possible paths of play in a game of typical duration lead to a strategy space for each player that contains more elements than there are atoms in the universe. Although we know it exists, no computer can possibly compute the optimal chess strategy. Therefore, when describing the actual behavior of real-world chess champions (or expert computer programs that play chess), the debate in behavioral economics amounts to the following. Is the interesting scientific question whether chess champions fail to optimize (even when they win)? Or is the relevant scientific challenge to describe the thought processes that enable people to win?

A more general statement of this problem of scientific relevance is whether to emphasize that human behavior differs from the prescriptions given by solutions to constrained optimization problems—or whether to focus on describing real human decision processes together with factors in the environment that enable people to succeed or fail. Many heuristics that are widely regarded as sub-optimal in behavioral economics could very well enjoy an alternative interpretation as brilliantly successful tools for making fast decisions in high-dimensional environments. As it stands now, much of the behavioral economics literature focuses on documenting deviations and biases with respect to the optimization model, which is equivalent to simply reporting that even grand master chess champions fail to play the optimal strategy in chess. An alternative empirical approach based on heuristics would study grand masters and identify the rules of thumb that they use to play the game as effectively as they do. The growing prominence of scholars and publications addressing these issues reveals the ongoing importance of debates over methodology and economic history (e.g., Gigerenzer, Todd and the ABC Group, 1999; Gilboa, Postlewaite and Schmeidler, 2004; Starmer, 2004, 2005; Heifetz, Shannon and

Spiegel, 2007; Bruni and Sugden, 2007; Caplin and Schotter, 2008; Hertwig and Hoffrage, forthcoming).

Critics of the universal constrained optimization model that dominates in neoclassical and behavioral economics contend that a huge—and unrealistic—step of simplification is typically required when setting up the optimization problem: summarizing everything a person cares about with a utility function, or everything that the managers of a firm base their decisions on with a simple profit function. The methodological debate then becomes whether decision scientists learn more by assuming decision makers optimize in a pre-simplified world, or whether optimization can be productively abandoned in favor of direct empirical description of the simplifying strategies people actually use in real-world environments that are typically many times more complex than the game of chess.

Methodological Pluralism

Another theme in behavioral economics derives from its willingness to borrow from psychology and other disciplines such as sociology, biology and neuroscience. To appreciate why methodological pluralism is characteristic of behavioral economics, one should recall that in neoclassical economics there is a singular behavioral model applied to all problems as well as a number of prominent efforts in economic history to expunge influence from other social sciences such as psychology and sociology. Although the structure of choice sets and the objective functions change depending on the application, contemporary economists typically apply the maximization principle to virtually every decision problem they consider. Consumer choice is modeled as utility maximization; firm behavior is modeled as profit maximization; and the evaluation of public policy is analyzed via a social welfare function whose maximized value depends systematically on parameters representing policy tools. In contrast, commitment to improved empirical description and its normative application to policy problems motivates behavioral economists, in many cases, to draw on a wider set of methodological tools, although the breadth of this pluralism is a matter of debate as indicated in the previous section.

Naming Problems

The term "behavioral economics" is generally attributed to George Katona (1951). Behavioral economists sometimes joke that the name of their subfield is redundant, since economics is a social science in which the objects of study depend directly on human behavior. "Isn't all economics supposed to be about behavior?," the quip goes. Despite the appearance of a 'distinction with no distinction' inherent in its name, proponents of behavioral economics argue that there is good reason for the explicit emphasis on accurate description of human behavior as indicated by the word "behavioral" in behavioral economics.

In psychology, there is a sharp distinction between the terms "behaviorist" and "behavioral." Behaviorism refers to research and researchers that draw on the work of B. F. Skinner in hypothesizing that most behavior can be explained in terms of adaptation to past rewards and punishments. Behaviorism rejects investigation of mental states or other psychic determinants of behavior. Thus, behaviorism is more similar to neoclassical economics because both schools of thought rely on a singular story about what underlies observed behavior while expressing overt antipathy toward the inclusion of mental states, cognitive processing, or emotion in their models. One frequently finds mistaken references to behavioral economists as "behaviorists" in the popular press, whereas "behavioralists" would be more accurate.

Another, albeit more minor, point that occasionally gives rise to confusion is the multiple uses of "psychology and economics" as a descriptor of an academic subfield, which can indicate subtly different communities of researchers depending on whether it is regarded as a subfield of economics or of psychology. The subfield of psychology referred to as "psychology and economics" overlaps in terms of subject matter and actively contributing scholars. Behavioral economists are sometimes regarded by psychologists, however, to overstate the extent to which the work of behavioral economists is actually informed by the broad and heterogeneous research programs within psychology.

Behavioral Economics and Experimental Economics

Strong connections between behavioral and experimental economics can be seen in behavioral economists' reliance on experimental data to test assumptions and motivate new theoretical models. There nevertheless remains a distinction to be made (Camerer and Loewenstein, 2004). Some experimental economists do not identify with behavioral economics at all, but rather place their work firmly within the rational choice category, studying, for example, the performance of different market institutions and factors that enhance the predictions of neoclassical theory. Experimental economics is defined by the method of experimentation whereas behavioral economics is methodologically eclectic. The two subfields have subtly different standards about proper technique for conducting lab experiments and very different interests about the kinds of data that are most interesting to collect. Therefore, it is incorrect to automatically place experimental work under the heading of behavioral economics. In the other direction, there are many behavioral economists working on theoretical problems or using non-experimental data. Thus, although behavioral and experimental economists frequently work complementarily on related sets of issues, there are strong networks of researchers working in the disjoint subsets of these subfields as well.

Frequently Discussed Violations of Internally Consistent Logic

This section describes several well-known violations of the rational choice model based on reasoning that allegedly suffers from internal inconsistency. The following example about deciding where to buy a textbook illustrates the kind of inconsistencies that are frequently studied in behavioral economics. Readers are encouraged to decide for themselves how reasonable or unreasonable these inconsistencies in fact are.

Suppose you are shopping for a required textbook. A bookstore across the street from where you work sells the book for $80. Another bookstore, which is 15 minutes away by car or public transportation, sells the book for only $45. Which do you choose: A) Buy the book at the nearby store for $80, or B) Buy the book at the farther-away store for $45?

Just as standard theory does not prescribe whether it is better to spend your money buying apples versus oranges, so, too, standard economic theory takes no stand on which choice of stores is correct or rational. But now consider a second choice problem.

Suppose you are buying a plane ticket to Europe. The travel agent across the street from where you work sells the ticket for $1,120. Another travel agency, which is 15 minutes away by car or public transportation, sells the same ticket for $1,085. Which do you choose: C) Buy the ticket from the nearby agency for $1,120, or D) Buy the ticket at the farther-away store for $1,085?

Considered in isolation, either A or B is consistent with rationality in the first choice problem, and either C or D can be rationalized in the second choice problem—as long as these problems are considered alone. Internal consistency requires, however, that a rational person choosing A in the first problem must choose C in the second problem, and that a rational person choosing B in the first problem must choose D in the second problem.

Based on extensive data from pairs of choices like the ones just described, many people prefer B in the first problem (i.e., the $35 saved on the cheaper textbook justifies spending extra time and money on the 30-minute round-trip commute) while preferring C in the second problem (i.e., the $35 saved on the cheaper airline ticket does not justify spending extra time and money on the 30-minute round-trip commute). According to axiomatic rationality, this pair of choices is inconsistent and therefore irrational. Yet many competent and successful people, without any obvious symptoms of economic pathology, choose this very combination of allegedly irrational (i.e., inconsistent) decisions.

One explanation is that some people weigh the $35 savings in percentage terms relative to total price. An $80 textbook is more than 75% more expensive than a $45 textbook, whereas a $1,120 planet ticket is less than 5% more expensive than a $1,085 ticket. Nevertheless, the logic of the cost-benefit model of human behavior at the core of rational choice, or neoclassical, economics regards dollars saved—and not percentages saved—as the relevant data.

Choosing A over B, in the eyes of a neoclassical economist, reveals an algebraic inequality:

utility of saving $35 > disutility of a 30-minute round-trip commute.

Choosing D over C reveals another algebraic inequality:

utility of saving $35 < disutility of a 30-minute round-trip commute.

Thus, choosing B over A and C over D leads to inconsistent inequalities, which violate the axiomatic definition of a rational preference ordering. One may justifiably ask, so what?

Skepticism over the importance of such violations of axiomatic rationality is discussed in a subsequent section under the heading *Rationality*.

Endowment Effect

Suppose you walk into a music store looking for a guitar. The very cheap ones do not produce a sound you like. And most of the guitars with beautiful sounds are priced thousands outside your budget. You finally find one that has a nice sound and a moderate price of $800. Given the guitar's qualities and its price, you are almost indifferent between owning the guitar and parting with $800, on the one hand, versus not owning it and hanging on to your money on the other. You go ahead and buy the guitar. After bringing it home, enjoying playing it, and generally feeling satisfied with your purchase, you receive a phone call the very next day from the music store asking if you would sell the guitar back. The store offers $1,000, giving you an extra $200 for your trouble. Would you sell it back?

According to the standard cost-benefit theory, if you were indifferent between the guitar and $800, then you should be more than happy to sell it back for anything over $800—as long as the amount extra includes enough to compensate for the hassle, time and transport costs of returning it to the store (and also assuming you haven't run into someone else who wants to buy the guitar and is willing to pay a higher price). Hoping to bargain for a higher offer from the music store, you might demand something far above $800 at first. But after bargaining, when facing a credible take-it-or-leave-it last offer, anything that gives you $800 plus compensation for returning to the store should leave you better off than holding onto the guitar.

Based on data showing the prevalence of the endowment effect, however, behavioral economists would predict that you probably will choose to hang onto the guitar even if the guitar store's offer climbed well over $1,000. The endowment effect occurs whenever owning something shifts the price at which one is willing to sell it upward to a significantly higher level than the price at which the same person is willing to buy it. In the neoclassical theory taught in undergraduate textbooks with demand curves and indifference curves, an important maintained assumption that is not frequently discussed in much depth is that, for small changes in a consumer's consumption bundle, the amount of money needed to just compensate for a reduction in consumption is exactly equal to the consumer's willingness to pay to acquire that same change in consumption.

This equivalence between willingness to accept money for a reduction in consumption and willingness to pay money to buy the amount by which consumption was reduced implies reversibility of demand and indifference curves, which the endowment effect opens up to question. Another description of this assumption is that there is a rock-solid stable relationship between quantities of consumption and their subjective valuations that does not change depending on what a person is currently consuming. When current consumption affects how I value all other possible combinations of consumption, then all bets are off—demand curves do not exist. Or to put it more precisely, a single person has many different demand curves, one for each point in the consumption space at which he or she consumes.

Kahneman, Knetsch and Thaler (1991) presented data showing that randomly assigned ownership systematically increases subjective valuations of goods and services. In one famous experiment, coffee mugs were randomly distributed to half the experimental subjects. Those who did not receive mugs were asked to submit bids to buy a mug, and those who owned a mug submitted offers to sell their mug. There was no negotiating or haggling. The experimenter collected bids and offers, and used these to find a market price at which the quantity supplied was equal to quantity demanded. The primary finding was that, despite the random assignment of ownership, the price at which owners were willing to sell (i.e., willingness to accept) was on average twice as large as buyers' willingness to pay. This led to far fewer transactions than one would predict using neoclassical theory.

In a related experiment, Carmon and Ariely (2000) showed that Duke University students who win the right to buy sports tickets in a university lottery valued these tickets—which they became the owner of by chance—roughly 14 times as much as students who had entered the lottery but did not win. Some researchers have linked the endowment effect to loss aversion, which refers to the phenomenon by which the psychic pain of parting with an object one currently owns is greater than the psychic gain from acquiring it. Thus, rather than ownership shifting the pleasure derived from a good or service upward, some experimental evidence suggests that an increase in pain at dispossessing oneself of a good or service generates the gap by which willingness to accept is significantly higher than willingness to pay. Other researchers have reported findings that cast doubt on the existence of the endowment effect, pointing to alternative explanations consistent with standard neoclassical theory (e.g., Hanemann, 1991; Shogren et al, 1994; Plott and Zeiler, 2007).

Preference Reversals

Lichtenstein and Slovic (1971) and Thaler and Tversky (1990) produced evidence that shook many observers' confidence in a fundamental economic concept—the preference ordering. These and other authors' observed preference reversals, which call into question the very existence of stable preferences that neoclassical analysis depends on, occurred in a variety of contexts: gamblers' valuations of risky gambles at casinos, citizens' valuations of public policies aimed at saving lives, firms' evaluations of job applicants, consumers' feelings toward everyday consumer products, and savers' attitudes toward different savings plans.

In a typical experiment, a group of subjects is asked to choose one of two gambles: A) win $4 with probability 8/9, or B) win $40 with probability 1/9. When asked to choose between the two, the less risky gamble A, which provides a high probability of winning a small amount, is typically chosen over B, which pays slightly more on average but pays off 0 most of the time. Next, another group of experimental subjects is asked to assign a dollar value to both gambles, A and B, stating the amount of money they would be willing to pay for A and B, respectively. Most subjects typically choose A over B when asked to *choose*. But most subjects place a larger dollar valuation on gamble B when asked to *evaluate* in terms of money. Choosing A over B, while valuing B more highly than A in dollar terms, is a preference reversal. In neoclassical theory, a person with a stable preference ordering should produce identical rankings of A and B whether it is elicited as a pair-wise choice or in terms of dollar valuations. A preference reversal occurs when two modes of elicitation theorized to produce the same implicit rankings actually produce rankings that reverse one another.

One explanation is that when asked to choose, people focus on the risk of getting zero. Gamble A provides a lower risk of getting zero and therefore dominates B by this criterion. When asked to give a dollar valuation, however, people tend to focus on the amounts or magnitudes of payoffs more than the probabilities of their occurrence. Focused on the magnitude of the largest payoff, gamble B's 40 dominates gamble A's 4. These different thought processes—prioritizing risks of getting zero, or prioritizing the magnitude of the largest payoff, respectively—might be very reasonable approaches to decision making in particular contexts. Nevertheless, they violate the norms defined by the standard definition of a rational preference ordering (see Brandstätter, Gigerenzer and Hertwig's, 2006, alternative explanation based on their *priority heuristic*).

Conjunction Fallacy

In logic, the conjunction of two events, A and B, refers to the combined event where both A and B occur. In set theory, the conjunction of two sets, A and B, is the intersection $A \cap B$. The conjunction fallacy refers to an experimental result in which human subjects are said to violate one of the basic tenets of the logic underlying probability theory: that the probability of a conjunction of events is less than or equal to the probability of any of the events being conjoined. In symbols, the definition of probability requires that:

$$P(A \cap B) \leq P(A) \text{ and } P(A \cap B) \leq P(B).$$

The reason is that the intersection of A and B can be no larger than A, and no larger than B.

Psychologists and linguists have pointed out, however, that people use words like "and" in a sophisticated context-dependent manner, and that the everyday meanings of words such as "probability" and "chance" and "likelihood" can reasonably deviate from definitions that economists have in mind. Consider how interpretations based on narrow logic-based norms versus smart everyday interpretations of the word "and" differ in de-coding the following invitation sent to me (an economist) by my friend who is a brain surgeon who works at a for-profit hospital: "Dear Friends and Colleagues, You are invited to a party at my place this Friday night."

Should I interpret this message to mean that I am invited to the party? According to the meaning of "and" dictated by formal logic, I am not invited to this party. Yes, I am a friend—but not a colleague, since I do not work at the same location or in the same profession as the woman hosting the party. Interpreting "and" as a computer would interpret it, I am not in the set formed by the conjunction, Friends $\cap$ Colleagues, and therefore not invited.

Does this mean that the woman, who addressed the invitation to "Friends and Colleagues" while clearly intending to invite anyone who was a friend *or* a colleague, made a logical error? Scholars disagree. For those who interpret the woman's invitation addressed to "Friends and Colleagues" as evidence of socially intelligent linguistic sophistication, the view is that in everyday English "and" can mean logical *and* (i.e., the set conjunction operator, $\cap$)—or "and" can mean logical *or* (i.e., the set union operator,). It depends on context. Using "and" to mean logical *or* is not necessarily a mistake.

In contrast, Tversky and Kahneman (1983), who famously reported the conjunction fallacy, described it as a logical fallacy and encouraged readers to interpret it as a mistake. Tversky and Kahneman presented the following scenario to experimental subjects:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?
>
> 1. Linda is a bank teller.
> 2. Linda is a bank teller and is active in the feminist movement.

Given the discussion above, the reader will notice that option 2 describes the conjunction of two events and therefore, under usual definitions from formal logic and probability theory, the probability of option 1 must be greater or equal than the probability of option 2. Nevertheless, 85% of subjects chose option 2, which Tversky and Kahneman labeled as the conjunction fallacy. Similar experiments produce other conjunction fallacies.

For example, when one baseball team is known to be much better than the other, a majority of subjects will choose the conjunction of events, "Team A beats team B by the score 3-0," as more *likely* than "Team A beats team B."

Hertwig, Benz and Krauss (2008) show that, after recognizing the sophisticated and socially intelligent ways in which people use words like "and," "or" and "probable," the conjunction fallacy disappears. That is, rather than indicating shortcomings hardwired into humans' cognitive architecture, these authors show the conjunction fallacy to be an artifact of the gap between meanings attached to words in everyday language versus formal logic.

Measuring Risk and Time Preferences

Innovative techniques for measuring preferences along a number of dimensions have emerged as an interesting subset of behavioral and experimental economics. Given the widespread reach of expected utility theory in economics in and outside behavioral economics, Eckel and Grossman (2002, 2008) and Holt and Laury (2002) have designed experimental instruments for quantifying the extent to which people are risk averse or risk loving. The Eckel-Grossman instrument has proven useful in capturing interpersonal variation of risky choice in a wide variety of populations, including those with very limited experience interpreting numerical risk measures, thanks to its remarkable simplicity. Together with a parameterized expected utility function, their instrument produces quantifiable ranges for an individual's risk-aversion parameter based on a single choice from among six binary gambles, where each gamble has only two possible outcomes, each of which occur with 50% probability. Reconciling possible inconsistencies among risk measures generated by different instruments and exploiting information about subjects whose preferences appear inconsistent is an active area of ongoing research (Dave, Eckel, Johnson and Rojas, 2007; Berg, Johnson and Eckel, 2009).

Behavioral economists have also established commonly used techniques for measuring time preferences. In the standard formulation, a person's time preference, or impatience, can be identified as the extent to which he trades off larger cash flows arriving in the more distant future in favor of smaller cash flows that arrive earlier. Impatience is frequently quantified in terms of the subject discount rate, although this depends on auxiliary assumptions about the utility function. A typical set of experimental tasks that produce information about people's choices over cash flows arriving sooner versus later is the following series of five binary decisions:

Choice 1: → 65 arriving today
→ 68.25 [5% more] arriving in one year

Choice 2: → 65 arriving today
→ 78 [20% more] arriving in one year

Choice 3: → 65 arriving today
→ 98.5 [50% more] arriving in one year

Choice 4: → 65 arriving today
→ 130 [100% more] arriving in one year

Choice 5: → 65 arriving today
→ 195 [200% more] arriving in one year

An experimental subject's five choices are coded as zeros and ones, 0 indicating an "impatient" choice of 65 arriving today, and 1 indicating a "patient" choice of the later-arriving and larger cash flow. These binary choices are then assembled into sequences that can be ordered from least to most patient, with typical choice response sequences 00000 (maximally impatient), 00001, 00011, 00111, 01111, or 11111 (maximally patient).

It sometimes occurs that subjects provide non-monotonic responses, for example, choosing the patient option when it is 20% larger, but choosing the impatient option when it is

50% larger. Different researchers take various approaches to handling these non-monotonic responses. Some ask subjects who submit non-monotonic responses to reconsider their choices in an attempt to enforce monotonicity, while others interpret non-monotonicity as valuable evidence about alternative decision processes at work in generating choices over gambles, which implies that observed choice behavior is based on a process other than the standard of weighing rewards for waiting against the disutility of waiting.

Hyperbolic discounting is a form of time inconsistency frequently labeled in psychological terms (wrongly, in some critics' views) as myopia, short-sidedness, or even temptation. Alternative discounting formulas with additional parameters are introduced in this approach. To see the behavioral patterns of interest, suppose a subject first makes the five time-trade-off choices above, and then makes another five choices over identical time-tradeoff choices except that there is a front-end delay that shifts both cash flows forward into the future by an equal amount. In this modified sequence with front-end delay, say, of two years, the first decision is between 65 arriving two years from today versus 5% more arriving in three years. The second decision with front-end delay is between 65 arriving two years from now versus 20% more arriving in three years, and so on.

Given data collected from two choice sequences with identical durations *between* the arrival of cash flows but different front-end delays, subjects are referred to as time-inconsistent if their choices are relatively impatient with no front-end delay (tempted by the immediate gratification of $65 arriving today) and relatively more patient with greater front-end delay (following the intuition that, "Well, if I'm going to wait two years in any case, it doesn't hurt very much more to wait an additional year). The implication is that people cannot be modeled correctly as time trade-offs by applying a single discount factor in a time-separable utility function, which would imply that rankings of cash flows are invariant to changes in front-end delay. Loewenstein and Thaler (1989) describe the clear-cut predictions of the neoclassical theory of inter-temporal choice and the interesting anomalies that are observable in a wide variety of data. These anomalies include discount rates of more than 200% per year among those who buy kitchen and household appliances and, at the opposite extreme, negative discount rates implied by US taxpayers whose tax withholding is consistently more than the taxes they owe but choose to receive a zero-interest tax refund arriving much later than they would receive it if they reduced the amount their employers withhold from their paychecks.

Biased Beliefs

In contrast to the violations of internally consistent logic discussed in the previous section, this section introduces another broad theme in behavioral economics concerning subjective beliefs that are objectively incorrect. The gap between a subjective belief about the probability that an event will occur and the objective probability of its occurrence (assuming an objective probability exists) is referred to as bias. The very notion of biased beliefs depends on how well-calibrated subjective perceptions are to external benchmarks (i.e., objective frequencies of occurrence in the world), whereas the rationality assumptions discussed earlier are based solely on internal consistency and make no reference to external normative benchmarks when describing what it means to make a good decision. Studies of biased beliefs confront surprisingly subtle challenges, first, in measuring people's subjective beliefs and, second, in establishing the existence of proper benchmarks (in the form of objective probabilities) against which subjective beliefs can be compared.

"All the kids are above average" not as crazy as it sounds

One sometimes hears people who should know better mistakenly claim that, if most people's beliefs about an attribute of theirs is different from the average value of that attribute, then people's beliefs must be systematically wrong (e.g., nearly everyone reporting that they are better-than-average drivers in terms of safety). In a bell-curved or other symmetric probability distribution, gaps between what most people believe—the modal response—and the average might justifiably be interpreted as evidence of bias. However, in many real-world probability distributions, such as traffic accidents (where a few bad drivers are responsible for most of the accidents) or annual income (where a small number of very high-earning individuals pull average income well above median income), the sample average is surprisingly non-representative of most people.

Consider a society comprised of 999 people who have nothing and one person--call him Bill Gates--who owns $1 billion in wealth. The average person in this society is a millionaire, with average wealth = $1,000,000,000 / 1000 = $1 million. Nearly everyone in this society is poorer than average. Thus, when the modal belief about how wealthy a person is turns out to be significantly lower than average wealth, it implies no bias in beliefs. Realizing this, researchers attempt to carefully elicit beliefs about medians and other percentiles that pin down the value of some variable X below which a known percentage of the population falls (e.g., Camerer and Hogarth, 1999).

Bias implies existence of normative benchmarks

Bias in econometrics is defined as the difference between the expected value of an estimator and the true value of the number(s) being estimated. In econometrics as well as in everyday usage, asserting that there is a “bias” implies having made an unambiguous commitment to what the true or correct value is. The term bias is ubiquitous in behavioral economics, and much of its empirical and theoretical work concerns deviations from normative benchmarks that implicitly assert how people ought to behave. Given the observation that people deviate from a benchmark (typically an axiomatic definition of rationality or formal logic), there are at least two distinct reactions to consider.

Most behavioral economists have interpreted observed deviations from the assumptions of neoclassical economics as bias, implicitly asserting that neoclassical assumptions are undisputed statements defining what good, or smart, economic behavior ought to be. According to this view, people who deviate from the neoclassical benchmarks are making mistakes, which is equivalent to saying they are biased. This, in turn, motivates some authors to recommend prescriptive policy changes aimed at de-biasing the choices we make, inducing us to more closely conform to the axioms of economic rationality. Jolls and Sunstein’s (2006) article, “Debiasing Through Law,” is one example of behavioralists relying on neoclassical theory as their source for normative benchmarks. They are hardly alone. These authors also make the argument that laws can and should be used to help people conform more closely to neoclassical norms.

Alternative interpretations of observed deviations from neoclassical norms have been put forward by those who question whether the neoclassical model provides sound guidance for how we ought to behave and by those who fear the paternalistic implications of policies aimed at de-biasing choice. One alternative interpretation takes its point of departure from the observation that people who systematically violate neoclassical assumptions are also surviving quite successfully in their respective economic environments—they are going to college, holding down jobs, having children and grandchildren, etc. Thus, without glossing Pollyannic on the foibles of the human condition, humans' resounding adaptive successes must be acknowledged in any objective biological assessment. It is far from obvious that violating tenets of neoclassical economics reduces fitness or leads to any significant economic harms. Earlier, this article gave examples demonstrating that conforming with neoclassical assumptions is neither necessary nor sufficient for adaptive success in the environments that matter to people. Therefore, it follows that the benchmarks—and not the people violating the benchmarks—are errant. If this is the case, then social scientists should abandon neoclassical benchmarks as normative guideposts in favor of more meaningful measures of economic performance, happiness, health, longevity, and new measures of adaptive success that have yet to be proposed.

Optimism and Overconfidence

Behavioral economics and behavioral finance have produced a large literature on overconfidence, which has been defined in at least three distinct ways. Let X be a random variable representing a decision maker's payoff at some point in the future. For example, one can think of X as the value of an entrepreneur's business in several years time, or the value of an investor's stock portfolio. First-order overconfidence (i.e., overconfidence in beliefs about the mean of a random variable) is defined as a subjective expected value of X that is larger than the objective expected value E[X].

Second-order overconfidence concerns the variance or second moment of a random variable, often interpreted as an inflated belief in the informativeness of a noisy signal. Let S represent a random signal that is observable today and positively correlated with the random payoff X. Positive correlation between S and X can be expressed through the definition S = X +

e, where e is a random variable uncorrelated with X. Thus, the noisy signal S contains the true value (which will be revealed in the future) plus noise represented by e. Second-order overconfidence can be expressed mathematically as an inflated belief about the signal-to-noise ratio, var(X)/var(e). Second-order overconfidence therefore involves a distorted belief about the second moment of a noisy signal. This distortion implies that the subjective information content of the signal is greater than its objective information content. If X and S are jointly normal, it is possible to write the conditional expectation in a form that shows explicitly how this subjectively inflated signal-to-noise ratio would lead a person to excessively update his or her beliefs about X conditional on S (relative to a person with objectively correct beliefs). The conditional expectation is:

$$E[X|S] = E[X] + (S - E[S])/(1 + 1/(\mathrm{var}(X)/\mathrm{var}(e))).$$

The coefficient on S in the expression above is $1/(1 + 1/(\mathrm{var}(X)/\mathrm{var}(e)))$, which measures the sensitivity of the conditional expectation of X with respect to changes in S. It is straightforward to show that this coefficient is an increasing function of the signal-to-noise ratio, implying that second-order overconfidence (at least under the assumptions of joint normality) is equivalent to over-weighting the noisy signal S. One real-world example of this is when non-expert investors consult investment advisors, believing that the information they receive from experts is more correlated with the future than it really is. Another variation on these definitions of overconfidence is when inflated beliefs about random payoffs and the information content of noisy signals concerns random variables possessed by other people in a strategic interaction (Berg and Lien, 2005).

Social Preferences

The standard assumption of unbounded self-interest is often described as a hypothesis holding that people only care about their own monetary payoffs and are completely indifferent among allocations of payoffs to different people in a group as long as their own payoff is the same. Challenging this assumption, behavioral economists seeking to study the extent to which people care about the overall allocation of payoffs among participants in a strategic interaction have described their alternative hypothesis as "social preferences." Researchers studying social preferences have sought to remain as close to the standard utility maximization framework as possible, modeling and testing the implications of social preferences by introducing utility functions that depend on other people's monetary payoffs as well as one's own payoff. Two of the most famous experiments in behavioral economics, the Dictator Game and the Ultimatum Game, are discussed below. These are formulated as extremely simple two-player games in which the hypothesis that people maximize their own monetary payoff makes a clear prediction. After hundreds of experimental tests in many places and in the presence of different contextual factors, there is widespread consensus that real people's behavior typically violates the hypothesis of own-payoff maximization. Attempts to introduce new utility functions with parameters measuring the extent to which people put weight on factors other than their own monetary payoff (the so-called "social preferences terms" in the utility function) are discussed subsequently, as well as the fierce debates about rationality that these models have spawned.

Dictator Game

In the Dictator Game, one player is handed a resource endowment, say $10, and then decides how to split or allocate it between him or herself and the other player. The other player has no choice to make. There is typically no communication. The game is played anonymously and one time only to avoid motivating players to try appearing "nice" in the expectation of future reciprocation. The player making the decision, referred to as the Dictator, can keep all $10 for him or herself and give the other player $0. The Dictator can also choose a $9-$1 split, an $8-$2 split, a $5-$5 split, and so on. In versions of the game with unrestricted action spaces, the Dictator can keep any amount for him or herself K, $0 \leq K \leq 10$, leaving the other player with a monetary payoff of $10 - K$. The theory that players of games maximize their own monetary payoffs without regard for other people's payoffs makes a clear prediction in the Dictator Game: Dictators will choose to keep everything, maximizing their own monetary payoff at $K = 10$ and allocating 0 to the other player.

However, when real people play this game, the most common choice by Dictators is a 50-50 split, even when playing versions of the game with much larger monetary payoffs. This is a clear violation of the hypothesis that people maximize an objective function that depends only on one's own monetary payoffs, and it is typically interpreted as evidence in favor of social preferences. In other words, the gap between the predictions of standard economic theory and the data observed in experiments implies (although this point is open to alternative interpretations) that people care about the monetary payoffs of others. Note that caring about the payoffs of others does not imply altruism or benevolence, so that spiteful preferences which register increased psychic gain based on the deprivation of others is also a form of social preferences.

Ultimatum Game

In the Ultimatum Game, a Proposer receives an endowment, say $10, and then makes a proposed allocation. Re-using the symbol K to represent the amount the Proposer proposes to keep, the proposed allocation looks similar to the allocation in the Dictator Game: K for the Proposer, $0 \leq K \leq 10$, and $10 - K$ for the other player, sometimes referred to as the Responder. Unlike the Dictator Game, however, the Responder has a binary decision to make in the Ultimatum Game: whether to accept the Proposer's proposal or not. If the Responder accepts, then payoffs follow the proposal exactly. If the Responder declines the proposal, then both players receive 0. Once again, this game is typically played anonymously and only one time to limit the expectation of future reciprocation as a confounding motive when interpreting the results.

As long as the proposal includes any positive payoff for the Responder, a Responder who maximizes his or her own monetary payoff will choose to accept, because even a small amount is better than zero according to the money-maximization hypothesis. The subgame perfect equilibrium is for the Proposer to offer the smallest positive amount possible to the Responder and for the Responder to accept. For example, if payoffs are restricted to integer values, the strict subgame perfect equilibrium is uniquely defined by a proposal in which the Proposer keeps $9 and the Responder receives $1, and the Responder accepts this proposal even though it is far from an even split.

Contrary to the theoretical prediction of Proposers offering a $9-$1 split and Responders accepting it, the most common proposal in the Ultimatum Game is an even (or nearly even) 50-50 split. The Responder's behavior is especially interesting to students of social preferences, because Responders typically reject unfair offers even though it leaves them with 0 as opposed to a small positive amount. It is this willingness of Responder's to choose 0 by rejecting "unfair" offers of $9-$1 and higher that provides one of the most decisive pieces of evidence for social preferences. A common interpretation is that the Responder receives more utility from punishing the Proposer for having made an unfair proposal than he or she would get by accepting $1 and leaving the Proposer's unfair offer unpunished. This shows that the Responder does not make decisions solely on the basis of his or her own payoff, but is considering the payoffs of the other player. In a collaboration between economists and anthropologists, Heinrich, et al. (2001) report how widespread and heterogeneous other-regarding behavior is in different societies, additionally demonstrating that people's experiences in production (i.e., whether their payoffs are based on individual or group outcomes) predict differences in behavior in the Ultimatum Game.

Extending utility theory to incorporate social preferences

A well-known approach to modeling social preferences (Fehr and Schmidt, 1999) extends the neoclassical utility function (which typically depends only on one's own monetary payoffs) to include three components: utility from one's own payoff (as one finds in a neoclassical utility function), utility from the positive deviation between one's own payoff and other players' payoffs (i.e., the pleasure of doing better than others), and a third term placing negative weight on negative deviations from other players' payoffs (i.e., displeasure of doing worse than others). Some authors have introduced models that add similar "social preferences" terms to the utility function, for example, placing negative weight on highly unequal allocations, weighted by a parameter referred to as inequality aversion. Psychologists, mathematicians, quantitative sociologists and biologists working collaboratively have offered alternative explanations for social interactions in terms of heuristics (Hertwig and Hoffrage, forthcoming).

Critics of the social preferences program draw on distinct points of view. Binmore and Shaked (2007) argue that the tools of classical and neoclassical economics can easily take social factors into account and need not be set off from neoclassical economics under distinct "social preferences" or "behavioral economics" labels. Although Binmore and Shaked are correct that, in principle, neoclassical utility theory does not preclude other people's payoffs from entering the utility function, the assumption of unbounded self-interest is indeed a key tenet of neoclassical normative theory. The no-externalities assumption (i.e., people care only about their own payoffs, and their actions affect each other only indirectly through market prices) is crucial to the validity of the Fundamental Welfare Theorem, which states that competitive markets are socially efficient. It is difficult to overstate the role that this theoretical result has enjoyed in guiding public policy toward private versus government provision of services such as healthcare. Critics rightly point to this theory's reliance on the unrealistic assumptions of no externalities and no information asymmetries.

Binmore and Shaked, and psychologists such as Gerd Gigerenzer, have also criticized social preferences models because they contain so many free parameters that it becomes difficult to gauge how impressed one should be when it is reported that these models fit experimental data with a high R-squared. Allowing for so many free parameters (potentially two new utility parameters for each subject in the sample if subjects are not assumed to have identical utility functions) gives the appearance of explaining sample variance in a trivial way. Questions have also been raised about why proponents of social preference models have not done more to check that these free parameters are estimated consistently across different data sets—in practice, these parameters are allowed to take on new values every time they are estimated with new data.

Arguing from a different perspective, Berg and Gigerenzer (forthcoming) question whether adding more terms to the utility function and then labeling parameters in the utility function as "social preferences" parameters provides a genuinely new and more realistic explanation of behavior. At times, the case for social preferences using social preferences utility functions amounts to a logically circular labeling scheme. Proponents of social preference utility function maximization essentially argue that, although it is not realistic to assume individuals maximize a utility function depending on own payoffs alone, behavioral economics can add psychological realism by assuming that individuals maximize a more complicated utility function.

The decision process implied by the maximization of a social preferences utility function begins just like any neoclassical model—with exhaustive search through the decision maker's choice space. In the case of the Fehr-Schmidt model, it assigns benefits and costs to each element in that space based on a weighted sum of the intrinsic benefits of one's own payoffs, the psychic benefits of being ahead of others, and the psychic costs of falling behind others. Finally, the decision maker chooses the feasible action with the largest utility score based on a hypothesized process of weighted summation. If the weights on the "social preferences" that register psychic satisfaction or dissatisfaction from deviations between one's own and other players' payoffs are estimated to be different than zero, then Fehr and Schmidt ask us to conclude that they have produced evidence confirming the social preference model.

To see the circularity of the labeling scheme and why it is distinct from a genuine theoretical explanation, consider how the social preferences model is used to fit data from the ultimatum game. If Proposers in the ultimatum game share equally or Responders reject positive offers, the observed choices can only be fit within the utility maximization framework by applying non-zero weights to the "social preferences" terms in the utility function. Thus, the model is used to label or categorize individuals. One should then be skeptical about subsequently explaining the behavior as a consequence of "social preferences."

Although this approach has gained popularity among behavioral economists, it almost surely fails at bringing improved psychological insight regarding the manner in which social variables systematically influence choice in real-world settings. To gauge the intuitive appeal of this approach, one should think of a setting in which social variables are likely to loom large, and ask oneself whether it sounds reasonable that people deal with these settings by computing the benefits of being ahead of others, the costs of falling behind the others, and the intrinsic benefits of own payoffs—and, after weighting and adding these three values for each element in the choice set, choosing the best. Berg and Gigerenzer argue that this is not a genuine attempt at describing the psychological process underlying high stakes decisions in social settings but rather an as-if model staunchly in the tradition of Milton Friedman.

Rationality

In the popular press, behavioral economics is often portrayed as a branch of economics that points to systematic irrationality in human populations and in markets in particular. Titles like *Irrational Exuberance* (Schiller, 2000), *Predictably Irrational* (Ariely, 2008), and much of Nobel laureate Daniel Kahneman's work documenting deviations from axiomatic definitions of rationality, make it easy for non-experts to associate behavioral economics with irrationality. Indeed, many behavioral economists in their writing, and especially when describing their results verbally, use "rational" as a synonym for behavior that conforms to standard economic theory and "irrational" as a catch-all label for behavior that deviates from standard neoclassical assumptions.

One prominent voice in behavioral economics, David Laibson, advocates to aspiring behavioral economists that they avoid describing behavior as "irrational" and avoid the ambiguous term "bounded rationality." Laibson's admonition is, however, very frequently violated, leading to subtle paradoxes regarding the normative status of the neoclassical model within behavioral economics (Berg, 2003).

In neoclassical economics, a rational preference ordering is defined as any ranking scheme that conforms to the axioms of completeness and transitivity. In choice under uncertainty, many economists—including the seminal contributor Leonard Savage—argue for a strong normative interpretation of what has now become the dominant tool in economics for modeling choice under uncertainty (Starmer, 2005). The normative interpretation of expected utility theory asserts that choices over probabilistic payoff distributions that can be rationalized as having maximized any expected utility function are rational, and those that admit no such rationalization are irrational.

In choice problems that involve trade-offs over time, choices that can be rationalized as maximizing a time-consistent objective function are frequently referred to as "rational," and those that cannot as "irrational." In macroeconomics and beyond, "rational expectations" refers to beliefs that are objectively correct, and incorrect beliefs therefore are described as irrational. In game theory, researchers often describe behavior that conforms to a Nash Equilibrium as rational and all deviations from Nash strategies as irrational. An important intuitive problem arises in these uses of the term "rationality"—namely, that all parties in a strategic interaction in which more than one party is playing an allegedly "irrational" strategy may be strictly better off than would be the case if each party accepted economic "rationality" as a prescription for action.

Two distinct problems arise. First, there is behavior that conforms to axiomatic rationality but is manifestly bad or undesirable in many people's views. Second, there is behavior that is very reasonable to many people because it achieves a high level of performance, which nevertheless violates axiomatic rationality. Thus, axiomatic rationality is both too strong and too weak. Too strong because it rules out reasonable behavior as irrational, and too weak because it allows for behavior whose consequences for well-being are intuitively bad in many people's views.

Rational preferences impose the requirement of self-consistent choice but typically say nothing about how well choices work in the real world, a distinction that psychologists Hastie and Rasinski (1988) and Hammond (1996) describe as coherence (internally consistent) versus correspondence (well-calibrated to the world) norms. Thus, dropping out of college, walking past a pile of cash on the ground, becoming addicted to drugs, or even committing suicide can be rationalized (and regularly have been in the economics literature) as maximizing a rational preference ordering, because they can be made to satisfy internal coherence. Economic rationality here imposes nothing more than consistency.

A rational person can walk past $100 lying on the sidewalk, revealing (within the rational choice framework) that his or her disutility of stopping to lean over and pick up the money is greater than the benefit of the money. Rationality requires only that the person is consistent about this ranking, never stopping to pick up money on the sidewalk when the amount is $100 or less. In the other direction, a person who drops out of college and then—without anything else in his or her life circumstances changing significantly—decides to re-enroll is regarded as inconsistent and therefore irrational, even though many parents would no doubt regard this as, on the whole, good economic behavior.

In expected utility theory, a decision maker can be completely averse to risk or love risk taking, but not both. The requirement of rationality is that all risky choices are consistent. Thus, someone who always takes risks and perhaps is regarded by many as foolish and imprudent would pass the consistency requirement, conforming squarely with axiomatic rationality based on consistent foolishness. In the other direction, a person who buys health insurance (revealing him or herself to be risk averse) and also decides to start a new business (revealing him or herself to be risk loving) cannot easily be rationalized within expected utility theory because of the appearance of inconsistent choices in risky settings.

In time trade-off decision problems, consistency allows for both infinite impatience and infinite patience, but not in the same person. A person who, every payday, throws a party and spends all his or her money and then starves until next payday passes the consistency test and is therefore rational. Similarly, a person who never spends a dollar of income and deprives him- or herself of all consumption over most of the life course is consistent and therefore rational. Consistently strange behavior (e.g., blowing one's paycheck and starving thereafter until the next paycheck) can be rationalized, but inconsistent behavior—even when it seems to reflect a positive step in a person's maturing or taking responsibility for their well being—is labeled irrational because it violates consistency. If the person who previously blew his paycheck every payday decides to start saving money for his or her retirement, this would be inconsistent, although very reasonable to most people. These cases illustrate tension between what most people regard as sensible economic behavior and the surprising simultaneous tightness and looseness of rational choice as it is defined in economics as a criterion for normative evaluation.

The rationality of beliefs, when defined as beliefs about probabilities that are objectively accurate, is distinct from definitions of axiomatic rationality mentioned above. Rather than internal consistency, objective accuracy requires a correct calibration, or correspondence, to an externally measurable standard of performance. When investigating rational preferences, the question is never how close did a person come to consuming the "right things" as defined by some external standard such as nutrition, lifespan or offspring. Rather, rational preferences hinge only on internal consistency. In many instances, however, discussions of rationality of beliefs refer back again to internal consistency rather than objective accuracy.

The vast literature on non-Bayesian beliefs contains numerous examples of authors who refer to internally inconsistent beliefs (i.e., those that violate Bayes Law) as irrational. One easy-to-see measure of whether beliefs are internally consistent (in the way that behavioral economists have focused on) is whether subjective probabilities from an exhaustive list of mutually exclusive events sum to 1. For example, the events "today it rains" and "today it does not rain" are exhaustive and mutually exclusive, assuming there is a clear definition and measurement of "rain." Therefore, the logic of probability theory requires that the two probabilities associated with these events sum to 1.

Consider, however, a person who believes that the probability of rain is 0.30 and the probability of not rain is 0.71. Because the probabilities sum to more than 1, these beliefs would be described as irrational when using the coherence norm of internal consistency. If the objective probability of rain is, say, 0.29, then these beliefs are very close to the objective truth—that is, these beliefs correspond very well to the real world although they violate internal consistency. In the other direction, another person might believe that the probability of rain is 0.90 and the probability of not rain is 0.10, which is internally consistent and therefore rational although highly inaccurate. Hammond (1996) points out that people can be wrong about everything they believe and yet perfectly internally consistent. Berg, Biele and Gigerenzer (2008) investigate this tension between consistency versus accuracy (i.e., coherence versus correspondence) in the beliefs of male economists about prostate cancer.

The paradoxes surrounding rationality are perhaps most evident in game theory. In the Prisoner's Dilemma, for example, individual rationality as defined by the Nash Equilibrium of the game requires that both players opt for the socially destructive option—defecting instead of cooperating. When the two players behave irrationally by cooperating, however, both of them are strictly better off than if they both made the allegedly rational choice.

Why, then, is cooperating in Prisoners Dilemma described as irrational? Rationality, when used as a synonym for Nash strategies, is merely a stability concept and says nothing about the global desirability of the outcome—by individual or aggregate measures. As mentioned before, in Prisoner's Dilemma each player is individually better off when both players cooperate than when both players defect. But mutual cooperation is labeled as irrational because each player, in individual terms and holding the other players' actions constant, could do better still by defecting while the other cooperates. There are many other examples in game theory where the Nash Equilibrium yields very low payoffs for all players compared to alternative combinations of actions which, because they are not jointly stable with respect to single-player deviations, are referred to—unfortunately—as irrational.

Behavioral Economics: Prospects and Problems

The origins of behavioral economics are many, without clear boundaries or singularly defining moments (Hands, 2007; Heukelom, 2007). And yet, even a cursory look at articles published in economics today versus, say, 1980, reveals a far-reaching behavioral shift. One can cite a number of concrete events as markers of the emergence of behavioral economics onto a broader stage with wide, mainstream appeal. One might imagine that such a list would surely include Herbert Simon's Nobel Prize in 1978. But prior to the 1990s, behavioral work appeared very infrequently in flagship general interest journals of the economics profession. A concise and, of course, incomplete timeline of milestones in the recent rise of behavioral economics would include: Richard Thaler's "Anomalies" series, which ran in the *Journal of Economic Perspectives* starting in 1987; hiring patterns at elite business schools and economics departments in the 1990s; frequent popular press accounts of behavioral economics in *The Economist*, *New York Times* and *Wall Street Journal* in the last 10 years; and the 2002 Nobel Prize being awarded to experimental economist, Vernon Smith, and psychologist, Daniel Kahneman. The 1994 Nobel Prize was shared by another economist who is an active experimenter and leading voice in game theory and behavioral economics, Reinhardt Selten.

A striking element in the arguments of those who have successfully brought behavioral economics to mainstream economics audiences is the close similarity to Friedman's as-if methodology. In prospect theory, behavioral economics adds new parameters rather than psychological realism to repair and add greater statistical fit to an otherwise neoclassical weighting-and-summing approach to modeling choice under uncertainty. In the social preferences approach, behavioral economics adds parameters weighting decision makers' concern for receiving more, or less, than others do to an otherwise neoclassical utility function. In inter-temporal choice, behavioral models of time inconsistency add discounting parameters with non-exponential weighting schemes while hanging onto the assumption of maximization of a time-separable utility function. Frequently billing itself as a new empirical enterprise aimed at uncovering the true preferences of real people, the dominant method in the most widely cited innovations to emerge from behavioral economics can be perhaps better described as filtering observed choices through otherwise neoclassical constrained optimization problems, with augmented utility functions that depend on new functional arguments and parameters.

Behavioral economists' attempts to filter data through more complexly parameterized constrained optimization problems suggests more similarity than difference with respect to neoclassical economics. It will be interesting to see whether moves in the direction of neuroeconomics based on brain imaging data portend more radical methodological shifts (Camerer, Loewenstein and Prelec, 2005) or rather a strengthening of the core methodological tenets of neoclassical economics (Glimcher, 2003). There is a route not taken, or not yet taken, following Herbert Simon's call to abandon universalizing, context- and content-free characterizations of rational choice in favor of models that explicitly consider the interaction of decision processes and the different environments in which they are used—that is, ecological rationality (Gigerenzer, Todd, and the ABC Group, 1999).

Criticisms notwithstanding, a number of new and practical suggestions for designing institutions and intervening to help people change their behavior have emerged from the behavioral economics literature of recent decades. These include plans that encourage greater levels of retirement savings (Benartzi and Thaler, 2004), higher rates of organ donation (Johnson and Goldstein, 2003), controlling the amount of food we eat (Wansink, 2006), and new tools for encouraging greater levels of charitable giving (Shang and Croson, 2009). One recent example of behavioral economics being put into practice is President Obama's tax cut, which is disbursed as a small reduction in monthly tax withholding, thereby increasing workers' monthly take-home pay by a small amount each month instead of arriving in taxpayers' mailboxes as a single check for the year. In the rational choice theory, taxpayers' decisions about how much of the tax cut to spend should not depend significantly on whether one receives, say, 12 paychecks with an extra $50, or a single check for $600 for the entire year. But behavioral economists like Richard Thaler have advised the Obama administration that, in order to induce immediate spending (which is what most economists call for in response to a recession), it is important that taxpayers view tax cuts as an increase in income rather than wealth—in other words, that taxpayers put the tax cut proceeds into a "mental account" from which they are relatively more likely to spend (Surowiecki, 2009).

All indications suggest that more empirical findings and theories from behavioral economics will make their way into public policy and private organizations aiming to influence the behavior of workers and consumers. Whether these tools will be regarded as benevolent interventions that make our environments better matched to our cognitive architecture or an Orwellian shift toward psychology-inspired paternalism is currently under debate (Thaler and Sunstein, 2003; Berg and Gigerenzer, 2007). There would seem to be genuine cause for optimism regarding behavioral economists' widely shared goal of improving the predictive accuracy and descriptive realism of economic models that tie economics more closely to observational data, while undertaking bolder normative analysis using broader sets of criteria that measure how smart, or rational, behavior is.

References

Ariely, D. (2008). *Predictably Irrational.* HarperCollins: New York.

Berg, N. (2003). Normative behavioral economics, *Journal of Socio-Economics, 32*, 411-427.

Berg, N., Biele, G., and Gigerenzer, G. (2008). Consistency versus accuracy of beliefs: Economists surveyed about PSA. Working Paper, University of Texas-Dallas.

Berg, N., Eckel, C., and Johnson, C. (2009). Inconsistency pays?: Time-inconsistent subjects and EU violators earn more. Working Paper, University of Texas-Dallas.

Berg, N., and Gigerenzer, G. (forthcoming). "As-if" behavioral economics: Neoclassical economics in disguise?, *History of Economic Ideas.*

Berg, N., and Gigerenzer, G. (2007). Psychology implies paternalism?: Bounded rationality may reduce the rationale to regulate risk-taking. *Social Choice and Welfare, 28*(2), 337-359.

Berg, N., and Lien, D. (2005). Does society benefit from investor overconfidence in the ability of financial market experts?. *Journal of Economic Behavior and Organization, 58*, 95-116.

Benartzi, S., and Thaler, R. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy, 112*, 164-187.

Binmore, K., and Shaked, A. (2007). Experimental Economics: Science or What?, London, ELSE Working Paper 263.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113*(2), 409-432.

Bruni, L., and Sugden, R. (2007). The road not taken: How psychology was removed from economics, and how it might be brought back. *Economic Journal, 117*(516), 146-173.

Camerer, C. (1999). Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences of the United States of America, 96*(19), 10575-10577.

Camerer, C. (2003). *Behavioral Game Theory: Experiments In Strategic Interaction*. New York, N.Y. Princeton, N.J.: Russell Sage Foundation; Princeton University Press.

Camerer, C., and Hogarth, R. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3), 7-42.

Camerer, C., and Loewenstein, G. (2004). Behavioral economics: Past, present and future. In *Advances in Behavioral Economics*, ed. Camerer, C., Loewenstein, G. and Rabin, M. Princeton: Princeton University Press.

Caplin, A., and Schotter, A. (2008). Foundations of Positive and Normative Economics: A Handbook. Oxford: Oxford University Press.

Carmon, Z., and Ariely, D. (2000). Focusing on the forgone: How value can appear so different to buyers and sellers. Journal of Consumer Research, 27 (3), 360-370.

Carpenter, J., and Seki, E. (2006). Competitive Work Environments and Social Preferences: Field experimental evidence from a Japanese fishing community, *Contributions to Economic Analysis & Policy Berkeley Electronic Press,* 5(2), Article 2.

Charness, G., and Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior and Organization,* 45(3), 301-328.

Dave, C., Eckel, C., Johnson, C., and Rojas, C. (2007), Eliciting risk preferences: When is simple better?. Working Paper, University of Texas-Dallas.

Eckel, C., and Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4), 281-295.

Eckel, C., and Grossman, P. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization*, 8(1), 1-17.

Fehr, E., and Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114,* 817–868.

Friedman, M. (1953). *Essays in Positive Economics*. Chicago, Ill.: University of Chicago Press.

Gigerenzer, G. (2008). *Rationality for mortals.* New York: Oxford University Press.

Gigerenzer, G., and Selten, R. (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge: MIT Press.

Gigerenzer, G., Todd, P. M., and ABC Research Group. (1999). *Simple Heuristics that Make us Smart*. New York: Oxford University Press.

Gilboa, I., Postlewaite, A., and Schmeidler, D. (2004). Rationality of belief, or: Why Bayesianism is Neither Necessary nor Sufficient for Rationality, Cowles Foundation Discussion Papers 1484, Cowles Foundation, Yale University.

Güth, W. (2008). (Non-) behavioral economics - A programmatic assessment. *Journal of Psychology*, 216, 244-253.

Hanemann, W. M. (1991). Willingness to pay and willingness to accept: How much can they differ?. *American Economic Review*, 81(3), 635–647.

Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York: Oxford University Press.

Hands, D. W. (2007), "Economics, Psychology, and the History of Consumer Choice Theory" Available at SSRN: http://ssrn.com/abstract=988125.

Hastie, R., and Rasinski, K. A. (1988). The concept of accuracy in social judgment**.** In D. Bar-Tal and A. W. Kruglanski: *The Social Psychology of Knowledge.* New York, NY; Paris, France: Cambridge University Press; Editions de la Maison des Sciences de l'Homme, pp. 193-208.

Heifetz, A., Shannon, C., and Spiegel, Y. (2007). What to maximize if you must. *Journal of Economic Theory, 133*(1)*,* 31-57.

Hertwig, R., and Hoffrage, U., and the ABC Research Group (forthcoming). *Social Heuristics That Make Us Smart*. Oxford University Press.

Heukelom, F. (2007). Kahneman and Tversky and the origin of behavioral economics. Tinbergen Institute Discussion Papers 07-003/1, Tinbergen Institute.

Heinrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91, 73-78.

Hertwig, R., Benz, B., and Krauss, S. (2008). The conjunction fallacy and the many meanings of "and". *Cognition*, 108, 740-753.

Holt, C., and Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644-1655.

Johnson, E. J., and Goldstein, D. G. (2003). Do defaults save lives? *Science, 302*, 1338-1339.

Jolls, C., and Sunstein, C. R. (2006). Debiasing through law. *Journal of Legal Studies*, 35, 199-241.

Jorland, G. (1987). The Saint Petersburg Paradox 1713 – 1937. In Krüger, L., Daston, L. and Heidelberger, M. (Eds.). (1987). *The probabilistic revolution: Vol. I. Ideas in history* (pp. 157-190) Cambridge, MA: MIT Press.

Kahneman, D., Knetsch, J., and Thaler, R. (1991). The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives,* 5(1), 193-206.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

Loewenstein, G., and Thaler, R. (1989). Intertemporal choice. *Journal of Economic Perspectives* 3(4), 181-193.

Plott, C., and Zeiler, K. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review,* 97(4), 1449-1466.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature, 36*(1), 11-46.

Shang, J., and Croson, R. (2009). Field experiments in charitable contribution: The impact of social influence on the voluntary provision of public goods. *Economic Journal,* 119, 1-17.

Shiller, R. (2000). *Irrational Exuberance.* Princeton, NJ: Princeton University Press.

Shogren, J. F., Shin, S. Y., Hayes, D. J., and Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *American Economic Review*, 84(1), 255–270.

Simon, H. (1986). Rationality in psychology and economics. *Journal of Business,* 59(4), S209-24.

Slutsky, E. E. ([1915] 1952), "Sulla Teoria del Bilancio del Consonatore," *Giornale degli Economisti*, 51, 1-26. Translated in *Readings in Price Theory*, G. J. Stigler and K. E. Boulding (eds.), Homewood, IL: Richard D. Irwin, 1952, 27-56.

Starmer, C. (2004). Friedman's risky methodology. Working Paper, University of Nottingham.

Starmer, C. (2005). Normative notions in descriptive dialogues. *Journal of Economic Methodology, 12*(2), 277-289.

Surowiecki, J. (2009). A smarter stimulus. *The New Yorker,* January 26, 2009, (http://www.newyorker.com/talk/financial/2009/01/26/090126ta_talk_surowiecki).

Thaler, R. H., and Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review, 93*(2), 175-179.

Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Veblen, T. ([1899], 1994). *The Theory of The Leisure Class*. New York: Dover Publications.

Wansink, B. (2006) *Mindless Eating: Why We Eat More than We Think*. New York: Bantam Books.

Yee, M., Dahan, E., Hauser, J. R., and Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, *26*(4), 532-549.