Behavioral Economics Comes of Age: A Review Essay on Advances in Behavioral Economics

WOLFGANG PESENDORFER

Advances in Behavioral Economics contains influential second-generation contributions to behavioral economics. Building on the seminal work by Kahneman, Strotz, Thaler, Tversky, and others, these contributions have established behavioral economics as an important field of study in economics. In this essay, I discuss aspects of the research strategy and methodology of behavioral economics, as exemplified by the contributions to Advances.

1. Introduction

The publication of Advances in Behavioral Economics (edited by Colin Camerer, George Loewenstein, and Matthew Rabin and published by the Russell Sage Foundation and Princeton University Press in 2004) is a testament to the success of behavioral economics. The book contains important second-generation contributions to behavioral economics that build on the seminal work by Daniel Kahneman, Amos Tversky, Richard H. Thaler, Robert H. Strotz, and others.

Behavioral economics is organized around experimental findings that suggest inadequacies of standard economic theories. The most celebrated of those are (1) failures of expected utility theory; (2) the endowment effect; (3) hyperbolic discounting and (4) social preferences. Most of the articles collected in Advances deal with one of these four topics.

Expected Utility: Expected utility theory assumes the independence axiom. The (stochastic version of the) independence axiom says that the frequency with which a pool of subjects chooses lottery \( p \) over \( q \) does not change when both lotteries are mixed with some common lottery \( r \). Experiments by Allais, Kahneman, Tversky, and others demonstrate systematic failures of the independence axiom. Chapter 4 in the book discusses this experimental evidence and the theories that address it.

Kahneman and Tversky (1979) develop prospect theory to address the failures of expected utility theory. They argue that when analyzing choice under uncertainty it is not enough to know the lotteries an agent is choosing over. Rather, we must know more about the subject's situation at the time he makes his choice. Prospect theory distinguishes between gains and losses from a situation-specific reference point. The agent evaluates gains and losses differently and exhibits first-order risk aversion locally around the reference point.
The **Endowment Effect**: In standard consumer theory, demand is a function of wealth and prices but does not depend on the composition of the endowment. Thaler, in his 1980 paper, coined the term “endowment effect” to describe the experimental finding that subjects value a good more if it is part of their endowment than if it is not. The endowment effect is addressed by assuming that agents treat additions to their endowments differently from subtractions. Hence, the endowment point is treated as a reference point and agents are assumed to have a kink in their valuations around the endowment point.

**Hyperbolic Discounting**: Standard dynamic decision theory assumes that intertemporal choices do not depend on the decision date. Whether the agent chooses consumption in the initial period or sequentially has no effect on the choice if the budget constraint is the same in both cases. Strotz (1955/56) develops a model of decision making that relaxes this assumption. As David Laibson (1997) points out, this model can be used to address experimental evidence of an “immediacy effect” in behavior: subjects have a tendency to choose earlier, smaller rewards over later, larger rewards when the earlier reward offers immediate consumption but reverse this preference when both rewards are delayed.

**Social Preferences**: Standard incentive theory assumes that the choices of an agent depend only on his own monetary payoff. This assumption has been challenged by a variety of experiments, perhaps most famously the experiments on ultimatum bargaining. Introduced by Werner Guth, Rolf Schmittberger, and Bernd Schwarze (1982), this experiment pairs subjects to play a simple bargaining game. The proposer makes an offer and the responder accepts or rejects. A rejection leaves both players with a zero payoff. Responders routinely reject small offers and therefore do not maximize their selfish monetary payoff. To address this evidence, behavioral economists have introduced models that allow players’ utility to depend on their own and their opponent’s monetary payoffs. Closest to standard models is the work of Ernst Fehr and Klaus M. Schmidt (chapter 9) in which the utility of a player depends on the monetary payoff of all players. There is additional evidence that players not only care about the material outcomes of their opponents but also about their opponent’s character. As a result, players may care about what other players reveal during the course of the interaction about their character. John Geanakoplos, David Pearce, and Ennio Stacchetti (1989) develop a framework that allows for such interdependence and Matthew Rabin (chapter 10) applies that framework.

As these examples illustrate, the focus of research in behavioral economics is on individual choice and the motives underlying that choice. Starting with an experimental finding that shows violations of standard economic assumptions, research in behavioral economics proceeds by introducing new variables that are used to “parameterize” deviations from standard models. In many cases, the new variable is used to describe a “bias” in decision making, i.e., some form of irrationality or systematic mistake.

This essay contains three observations on the direction and the focus of behavioral economics. The next section previews the three observations and sections 3–5 discuss them in more detail. Section 6 concludes.

2. **Three Observations**

(1) Much of behavioral economics builds on experimental evidence in which a new variable that is ignored in standard economic models is shown to “matter.” While the new variable may be observable in experimental settings, it is often unobservable when the researcher must deal with economic (field) data. A prominent example of this is the reference point in prospect theory. The reference point can be manipulated in experimental settings but is essentially unobservable outside the laboratory. This
makes it difficult to adapt prospect theory to economic contexts.

(2) Behavioral economics accommodates observed violations of economic theory by building the observed biases into the behavior of the individual. For example, if experimental evidence suggests that subjects underutilize their information, a behavioral model might deal with this observation by defining agent's actual posterior beliefs as an average of the correct posterior and the prior. But modeling devices that make sense for an unbiased decisionmaker may not make sense for a biased one. For example, why would individuals have priors and posteriors if they are destined to apply Bayes' law incorrectly? Similarly, why would individuals maximize an objective function if the objective function is the wrong one?

(3) The focus on biases and mistakes in decision making naturally leads to an exploration of the psychological sources of these biases. In many instances, behavioral economics turns into economic psychology. Observed behavior in experiments is taken as a window into the mind of the decision-maker and the theory explores how the person thinks and feels in a particular situation. Such theories are difficult to connect to economic data because their main insights are about psychological variables, that is, how the person thinks (i.e., deals with biases) and feels.

3. New Variables and Their Measurement

Behavioral economics argues that economists ignore important variables that affect behavior. The new variables are typically shown to affect decisions in experimental settings. For economists, the difficulty is that these new variables may be unobservable or even difficult to define in economic settings with economic data. From the perspective of an economist, the unobservable variable amounts to a free parameter in the utility function. Having too many such parameters already, the economist finds it difficult to utilize the experimental finding.

Successful innovations in economics find new variables that “matter” and, in addition, show how these variables can be identified and measured. In that case, the new theory also delivers new testable predictions that can be used to prove it wrong. Models of hyperbolic discounting (Strotz 1955; Laibson 1997) are examples of such innovations in behavioral economics.

3.1 Prospect Theory

An expected utility maximizer has a utility function over consumption lotteries and chooses a portfolio to maximize his utility. For prospect theory, the situation is more complicated. When an agent makes a decision, he does not necessarily take into account how this decision affects his consumption. He may view certain risky prospects in isolation or combine them with other risky prospects and evaluate this subset of lotteries separately. Utility depends on a reference point that partitions outcomes into gains and losses.

The variables of prospect theory are adapted to the setting of experiments where the researcher can manipulate the reference point. For example, the experimenter may frame a lottery \( L \) as (i) \( x \) with probability \( p \) and \( y \) with probability \( 1 - p \); or, alternatively, as (ii) a payment of \( z \) and a subsequent lottery of \( x - z > 0 \) with probability \( p \) and the lottery \( y - z < 0 \) with probability \( 1 - p \). The difference in the two lotteries is interpreted as a manipulation of the reference point.

When prospect theory is applied to economic settings, it is often impossible to identify the reference point. For example, in many economic applications the researcher won’t be able to determine whether (i) or (ii) is the correct framing of an asset corresponding to lottery \( L \). Behavioral economists deal with this ambiguity by treating the reference point as a free variable chosen to match the observed behavior of an application. Depending on the application, the reference point may be the endowment, the value of the endowment at a fixed date, the value at
which the agent previously bought an asset, or the expected earnings at the end of a working day. Colin F. Camerer (chapter 5) illustrates how prospect theory is applied to a variety of economic settings.

Shlomo Benartzi and Thaler (1997) suggest an explanation of the equity premium based on prospect theory. Investors’ utility depends on the annual return of their portfolio. In particular, investors are loss averse and compute gains and losses with respect to the value of their portfolio at the beginning of the planning horizon (approximately one year). In this model, the reference point is the value of the asset in the portfolio at the beginning of the year (the beginning of the “planning horizon”). If investors are sufficiently risk averse around the reference point, this formulation will imply a large equity premium. To match the observed equity premium, it is essential that gains and losses are computed over relatively short periods (annually rather than over five or ten years) and that the reference point adjusts over time. If the reference point stays constant (for example at the price the asset was purchased), then over time the effect of loss aversion fades.

Studies on stock trading (Terrance Odean 1998) and housing transactions (David Genesove and Christopher Meyer 2001) find that agents are less likely to sell assets that have incurred losses than assets that have incurred gains. Prospect theory explains this behavior by assuming that investors treat each asset as a separate decision and use the purchase price as a reference point. Investors are assumed to tolerate more risk when they try to recover a loss than when they try to increase their gains. As a result, investors may hold on to assets that have made losses even if those assets yield a lower expected return than some other asset in their portfolio.\(^1\) In this case, the reference point is the value of asset when it was purchased and does not adjust over time. Moreover, it is assumed that assets are treated in isolation so that the agent cannot use a different asset with superior return to recover losses.

Camerer, Linda Babcock, George Loewenstein, and Thaler (chapter 19) suggest that New York cab drivers use a daily income target as a reference point and stop working after the target is achieved. Income targeting is an extreme form of loss aversion with the target income level as a reference point. As Henry S. Farber (2005) points out, this implies that cab drivers quit early on days where it is easy to make money and quit late on days when it is hard to make money. Farber reexamines the evidence presented by Camerer et al. and finds that there is no evidence of a target-income behavior among New York cab drivers. Farber concludes that the primary determinant of stopping work is hours worked and that cumulative income is at most weakly related to stopping work. Farber (2004) estimates a more general utility function that allows intertemporal substitution but also includes loss aversion around a reference point. For that model, he finds evidence that a reference level may affect the labor supply decision on a given day. But, even for a given driver, the estimated reference level varies substantially from day to day. Farber concludes that “This [variation] seriously limits the predictive power of the reference point, and undermines the usefulness of the construct of the reference income level as a determinant of labor supply” (Farber 2004, p. 4).

In all these applications, we cannot observe variations in the reference point in the same way that experimenters can fix and manipulate the reference point. Therefore, the reference point becomes a parameter that is calibrated to match the observed data. But unlike risk aversion or the discount factor, the reference point need not be consistent across applications or even consistent across periods for the same application. Essentially, it captures a

\(^1\) Nicholas Barberis and Wei Xiong (2006) show that this argument is flawed. A formal model corresponding to the intuitive description by Camerer (chapter 5) may not be consistent with the evidence in Odean.
Applications of prospect theory are similar to models of habit formation. In habit models, a new parameter (typically a function of past consumption) is added to the utility function and calibrated to match observed data. This research seeks the “right” utility function for a particular application. In its most general form, the utility of a decision-maker depends on his choice $x$ and on a subjective state $s$ that, in turn, is a function of all other observable variables. The goal is to find the best specification of $u(x, s)$. Applications of prospect theory and habit formation are particular examples of that program.

The unobservability of the subjective state makes it more difficult to falsify the theory. It is hard to imagine that one could formulate a “puzzle” analogous to the equity premium puzzle for prospect theory. Ultimately, the theory allows too many degrees of freedom. (Of course, defenders of prospect theory might say that there can’t be a puzzle because prospect theory is right!)

3.2 Discounting

In their survey of evidence on time preference, Shane Frederick, Loewenstein, and Ted O’Donoghue (chapter 6) discuss evidence that subjects resolve the same intertemporal trade-off differently depending on when the decision is made. Consider a choice between $x$ at time $t$ and $y > x$ at time $t’ > t$. For appropriate choices of the parameters, subjects choose $(x, t)$ if the decision is made in period $t$ and $(y, t’)$ if the decision is made in period $t’ - 1$ or earlier.

Strotz, in his seminal 1955 article, proposes a solution concept (consistent planning) similar to modern subgame perfect equilibrium to solve dynamically inconsistent decision problems. As Strotz points out, the consequence of consistent planning is a preference for commitment. Laibson (1997) initiated research that seeks to demonstrate the usefulness of dynamically inconsistent preferences for (macro)economic analysis.

The behavior of Strotzian decisionmakers depends not only on their budget (wealth) but also on the availability of “commitment.” In economic applications, commitment can be achieved by holding illiquid assets or through regulation. Standard consumers have no demand for commitment and hence will (weakly) prefer liquid assets while Strotzian consumers may strictly prefer illiquid assets. The liquidity of an asset is often observable for economists.

Strotzian models add new variables to the analysis but also add new testable predictions. For example, we would expect agents to pay for commitment. Laibson, Andrea Repetto, and Jeremy Tobacman (2005) estimate the parameters of Strotzian consumers and find that they would greatly benefit from commitment: “Specifically, according to a comparison of value functions, at age 20, sophisticated quasi-hyperbolics would be willing to pay 2,000 USD to get rid of their credit cards immediately and never have access to them in the future. This begs the question of why only a tiny fraction of consumers cut up their credit cards” (Laibson, Repetto, and Tobacman 2005, p. 22). The Strotzian model can generate this type of “puzzle” because the basic ingredients are observable in many economic contexts.

In addition to evidence in favor of the Strotzian model, Frederick, Loewenstein, and O’Donoghue emphasize that other aspects of decision problems that economists routinely ignore “matter.” They present evidence that “gains are discounted more than losses; small amounts are discounted more than large amounts; greater discounting is shown to avoid delay of a good than to expedite its receipt” (p. 175–76). In typical economic applications the data do not allow a distinction between reduced gains and increased losses, or between decreasing delay and expediting receipt. Even at modest levels of aggregation, it is impossible to distinguish between several small and a single
large purchase. As a consequence, it is difficult to use economic data to calibrate utility functions that depend on those variables.

3.3 Calibrating from Experiments

Since novel behavioral variables such as the reference point can be manipulated in experiments, it seems tempting to use experimental data to estimate the parameters of utility functions and then apply these estimates to economic applications. The experimental evidence on discounting offers a good illustration of the difficulty of this approach.

The impatience displayed by subjects in experimental settings is striking. Thaler (1981) finds annual discount rates over 300 percent for time-delays of a month. A recent experiment by Benhabib, Bisin, and Schotter (2005) finds annual discount rates of 472 percent. These findings are striking because there are market prices for these intertemporal trade-offs. Real interest rates are typically in the low single digits—orders of magnitude off the experimental findings. Moreover, economic agents tend to hold assets that offer those low rates of return.2

A second striking feature of the experiments is the large cross-study differences in estimated discount rates. Frederick, Loewenstein, and O'Donoghue write “the spectacular cross-study differences in discount rates also reflect the diversity of considerations that are relevant in intertemporal choices and that legitimately affect different types of intertemporal choices differently” (p. 211). The argument by Frederick, Loewenstein, and O'Donoghue implies that experimental evidence cannot be used to calibrate utility functions that then are applied in (very different) economic contexts.

Rather, the calibration has to be done separately for every context. As a result, the experimental evidence offers little quantitative guidance for economists.

4. Humans as Stubborn Operators of Broken Machines

Behavioral economics grew out of a critique of standard economic assumptions. This tradition sometimes leads to a view of behavioral economics as a collection of “biases,” that is, violations of standard assumptions in economics. Behavioral models often take as a starting point a standard economic model and reinterpret the model as a description how the person thinks and feels. Next, an (often compelling) case is made that many of the assumptions are unrealistic because humans cannot perform the difficult mental tasks embodied in the formalism. The mistake or bias is typically modeled by assuming that some aspect of the optimization procedure in the decision model is done incorrectly. Typically, behavioral economists take great care in motivating the particular mistake and provide detailed evidence that humans indeed have trouble performing the task required by the model. Finally, the psychological and sometimes also the economic consequences of the model with the bias are explored. However, rarely do these theories ask whether—once the mistake is taken for granted—the original model makes sense. In other words, why would humans go to the trouble of maximizing objective functions, formulating complicated beliefs only to get things systematically wrong?

For example, consider the O'Donoghue–Rabin model of naive, dynamically inconsistent decisionmakers. These decisionmakers maximize a standard utility function but have incorrect beliefs about their own future behavior. The assumption that humans have wrong, even systematically biased, beliefs seems plausible and strikes a cord with most readers. At the same time, it seems highly implausible that individuals would go to the

---

2 Of course, there is substantial diversity in savings behavior. As Laibson et al. (2005) point out, many U.S. households hold credit card debt at high interest rates which suggests a great deal of impatience. On the other hand, savings behavior in Germany has puzzled researchers (see Axel Boersch-Supan et al. 2001) because it suggests a great deal of patience.
trouble of solving a dynamic optimization problem when solving this problem has no clear benefit. In a standard model, maximizing a utility function is simply a concise representation of how the agent behaves. But once the model is interpreted as a mental process, we must imagine that the decisionmaker actually performs the optimization. Since the decisionmaker is systematically wrong about future behavior there is no obvious benefit from maximizing the objective function as opposed to taking some other (perhaps arbitrary) action. The metaphor of an operator of a broken machine comes to mind. Think of a car owner who operates a car with a defective steering and yet behaves as if the steering were in perfect working condition. He spends great effort learning how to drive on a perfectly working car only to crash his defective vehicle at every opportunity.

Bayes’ rule is an implication of the standard model of decision making together with the existence of subjective probabilities and the separation of payoffs and beliefs. There are numerous experiments to describe how individuals process information incorrectly, that is, in ways inconsistent with Bayes’ rule. One way in which behavioral models deal with violations of Bayesian updating is to assume that individuals apply Bayes’ rule incorrectly or apply it to the wrong underlying stochastic process. The problem with this type of model is that by introducing the bias the model has lost the reason why individuals should hold probabilistic beliefs to begin with. Why go to the trouble of forming numerical representations of likelihoods when those representations are used incorrectly? While it is very plausible that agents make mistakes in information processing, it seems even more plausible that the remainder of the standard model is also wrong given that information is processed incorrectly.

Dealing with deviations from the standard model often requires more radical departures from standard theory. The chapter on case-based decision theory (chapter 25) is an example of such a departure. In that paper, Itzhak Gilboa and David Schmeidler develop an alternative to the standard expected utility framework. Their model replaces probabilities (subjective or objective) with a less demanding similarity function that is used to weigh the relevance of past experiences (cases) for every action. The paper offers an intriguing alternative to the familiar probabilistic models of decision making.

5. Behavioral Economics or Economic Psychology?

Traditional choice theoretic models and behavioral theories differ in their focus when analyzing “behavioral” phenomena. To illustrate the traditional perspective, consider the work of Larry G. Epstein and Stanley E. Zin (1989). In that paper, the authors provide a synthesis of non-expected utility theory and the Kreps–Porteus model of dynamic choice. Their purpose is to find a user-friendly formulation that can be taken to macroeconomic data. The theoretical work of Epstein and Zin formulates assumptions on individual behavior that allow deviations from expected utility and yet maintain the essential structure used in macroeconomics and finance.

The work by Epstein and Zin is motivated by psychological and experimental evidence that demonstrates violations of expected utility theory. Yet, it would not be considered a paper in behavioral economics. While Epstein and Zin deal with the Allais paradox, anxiety and other psychological phenomena, their focus is on the economic implications of these effects.

Theoretical work in behavioral economics often focuses on the psychology of the situation using the economic behavior as a window into the mind of the decisionmaker. For example, in chapter 3, Richard Thaler discusses his theory of mental accounting. Mental accounting is a description of the thought processes of an economic decisionmaker. It catalogues how decisionmakers
perceive and experience outcomes in various situations and how these decisions are evaluated.

Thaler (chapter 3, p. 83) notes that sunk costs play a role in decision making and he gives the example of season tickets to a theater. In an experiment, season ticket holders were randomly given a discount on their season ticket purchase. It turns out that the discount affects the subsequent probability of attendance—a fact that is in conflict with standard economic models. However, over time the difference in attendance diminishes and in the second half of the season the difference in behavior between individuals who paid the full price and those who got a discount disappears.

In Thaler’s model, the initial purchase of the season ticket opens an account in the mind of the decisionmaker with a negative balance. Every time the individual fails to attend the theater, he experiences a mental loss proportional to the price of the season ticket. The agent wishes to offset this loss with a corresponding gain from attending the theater and hence is more inclined to go to the theater if the price of the ticket is high. The fact that over time the effect of the sunk cost wears off is interpreted as a depreciation of the mental account that records the initial purchase of the ticket. In the second half of the season, the decisionmaker apparently closes his mental account for the season ticket. We know this because the sunk cost no longer plays a role.

Note the reversal in the roles of economics and psychology. The economic evidence is used to flesh out the details of a particular mental process that is operational for this particular case. If the effect of sunk costs did not wear off after half a season we would conclude that the mental account does not depreciate over the course of a year. Conversely, if there is no effect of the sunk cost we would conclude that the account is “closed” very quickly.

In their paper “Doing It Now or Later” (chapter 7), O’Donoghue and Rabin analyze the decisions of dynamically inconsistent agents. Their purpose is to capture “the human tendency to grab rewards and to avoid immediate costs in a way that our ‘long-run selves’ do not appreciate” (p. 223). The time-inconsistent agents are described by a sequence of “selves” and the decision problem under consideration serves to illustrate how the conflict between the selves is played out. O’Donoghue and Rabin ask how decisionmakers are harmed by their time-inconsistency. They compare naive and sophisticated agents with normal time-consistent agents and ask when the time-inconsistent agents consume too early or complete a task too late—where the norm is defined as the time-consistent decisions. Comparing naive and sophisticated decisionmakers they ask whether naivete or sophistication helps or hurts the agent’s well-being.

In this work, as in Thaler’s analysis of mental accounting, economic behavior is the starting point for psychological explorations. The main contribution is to provide an analysis of mental processes and their benefit or harm for an agent’s well-being. While the model reproduces the motivating economic evidence and suggests related behaviors in similar contexts, the development of an applicable economic model is not the focus of the analysis. Instead, the psychology of the particular situation takes center stage. This is in contrast to Epstein and Zin (1989) where the psychological evidence may motivate the attempt to generalize the preferences. However, the particular formulas are not meant as describing a mental process, nor is the psychology of decision making the focus of the analysis.

Standard economic models relate behavior in different situations. For example, the Epstein–Zin axioms describe how the decisionmaker behaves in simple situations and the formula derived in the representation theorem (applied to an economic decision problem) describes how decisions are made in more complicated economic problems. The theory does not ask what psychological
motives are behind the assumed behavior.

By contrast, economic psychology explores the psychological mechanisms behind economic behavior. An implication of this shift in focus is that the interesting observations of behavioral theories are often psychological, that is, about the mental processes behind behavior. The objective of a broadly applicable economic framework becomes less important, is left for future work or is viewed as a psychologically unrealistic goal.

6. Conclusions

Behavioral economics has reached the status of an established discipline. The central issues and theories discussed in Advances are familiar to most current PhD students in economics.

At the same time, behavioral economics remains a discipline that is organized around the failures of standard economics. The typical contribution starts with a demonstration of a failure of some common economic assumption (usually in some experiment) and proceeds to provide a psychological explanation for that failure. This symbiotic relationship with standard economics works well as long as small changes to standard assumptions are made. In that case, the behavioral evidence can be the impetus for small changes of standard models that leave the basic structure of the theory intact. Examples of this are theories that allow Allais-type behavior and loss aversion, or the models that allow a preference for commitment (such as Strotz model of consistent planning).

With the success of behavioral economics, more radical departures are being considered. In that case, the traditional blueprint—evidence plus small modification—is less compelling. For example, the economic model of individual behavior as the result of constrained optimization is not well suited to describe wrong or biased behavior. There is no “small” modification of the standard model that can deal convincingly with the hypothesis that people are wrong about their objective function or process probabilities incorrectly.

Behavioral economics emphasizes the context-dependence of decision making. A corollary of this observation is that it is difficult to extrapolate from experimental settings to field data or, more generally, economic settings. Moreover, not all variables that are shown to matter in some experiment are useful or relevant in economic applications. The question whether a particular variable is useful or even observable for economics rarely comes up in behavioral models, yet the success or failure of modeling innovations often depends on its answer.

References


