Behavioral Economics*

Sanjit Dhami† **Ali al-Nowaihi**‡

11 November 2010

Abstract

Behavioral economics is one of the fastest growing fields in economics. It is motivated by findings from psychology (and sociology) and allows for experimental, field and neuroeconomic evidence. Behavioral research over the past 4 decades has shown that two central assumptions in mainstream economics, namely, full rationality and purely selfish preferences are far less universal than was originally believed. In this essay, we bring out the role of alternative behavioral decision theories, emotions, framing effects, mental accounting, behavioral models of time discounting, new solution concepts in behavioral game theory, neuroeconomics and the debate on liberalism and paternalism, among others.

Keywords: anomalies of expected utility; prospect theory; rank dependent utility; heuristics; mental accounting; framing; emotions; intrinsic reciprocity; inequity aversion; anomalies of exponential discounting; hyperbolic discounting; neuroeconomics; quantal response equilibrium; level-k models; liberalism versus paternalism

* **A version of this paper is forthcoming in the Encyclopedia of Human Behavior to be published by Elsevier. The article has been written to conform to the required format. Hence, the presence of a glossary; no references in the text; a list of a maximum of 15 references for further reading (but the format does not allow them to be cross-referred to in the text); no footnotes, no quotes, no italics etc. Hence, while all our assertions in this paper can be backed by extensive references, the format did not allow us to do so (apologies to all offended authors). The material in this paper is probably of use for someone who is not very familiar with the area of behavioral economics, yet requires a quick, readable, overview of the field.**

† Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522086. Fax: +44-116-2522908. E-mail: Sanjit.Dhami@le.ac.uk.

‡ Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522898. Fax: +44-116-2522908. E-mail: aa10@le.ac.uk.

Glossary

Utility function: A function that assigns to bundles of goods, some level of human satisfaction that can be numerically measured.

Reference point: A level of outcome such that any outcome which is greater is coded as a gain and any that is lower is coded as a loss.

Loss aversion: The tendency of losses to bite more than gains of an equivalent magnitude.

Probability weighting function: A function that captures the subjective probability assigned by an individual to various events.

Heuristics: Rules of thumb used by individuals that substitute for fully optimizing behavior based on classical statistics.

Intrinsic reciprocity: The desire to reciprocate based on the notion that one is hard-wired to do so.

Social preferences: The presence of bundles of goods of others, in addition to one's own, in one's utility function.

Discount factor: A function of time that captures the degree to which current consumption is more important than future consumption.

Neuroeconomics: The use of the methods of neuroscience in economics.

Behavioral game theory: The application of limited rationality to strategic human interaction.

Liberalism versus Paternalism: The study of the policy response to less than perfect rationality.

---

1. Introduction

Traditional economics is based on the assumptions of fully-rational and self interested behavior. A very large body of evidence casts serious doubts on both these assumptions. A serious consideration of the psychological evidence for human motivation in economic phenomena and the construction of new theoretical models based on this evidence has given rise to Behavioral Economics. This new field is growing at a rapid pace and gaining increased acceptance in mainstream economics. We attempt a brief overview of behavioral economics and the insights that it offers.

2. Decision Theories

For simplicity, let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of real numbers, which we interpret as levels of wealth (n can be any positive integer). We assume that $x_1 \leq x_2 \leq ... \leq x_n$. Economists use the term lotteryfor the following entity

$$L = (x_1, p_1; x_2, p_2; ...; x_n, p_n) \,, \tag{2.1}$$

where outcome, x_i, occurs with probability, $p_i \geq 0$, and $\sum_{i=1}^{n} p_i = 1$. Thus, a lottery is simply a description of all possible outcomes that could arise, alongwith their respective probabilities. When the probabilities are verifiable independently of the decision maker, we have a situation of *risk*. On the other hand, if the probabilities can be inferred only from the actions of the decision maker, then we have a situation of *uncertainty*. For instance, while objectively verifiable probabilities of the risk from the MMR vaccine were shown to be low in almost all scientific studies, many individuals assigned high subjective probabilities, leading them to refuse the vaccination. Denote the set of all lotteries of the form in (2.1) by $\mathcal{L}$. Which lottery in $\mathcal{L}$ should be chosen?

2.1. Expected utility theory

The most widely used theory in economics under risk is the expected utility theory (EU). Under axioms on individual preferences, known as the *axioms of rationality*, it can be shown that, for each decision maker, there exists a strictly increasing function, $u : \mathbb{R} \to \mathbb{R}$, such that the utility to the decision maker from outcome, x_i, is $u(x_i)$. Furthermore, for any lottery $L = (x_1, p_1; x_2, p_2; ...; x_n, p_n)$, the expected utility to the decision maker is

$$EU(L) = \sum\nolimits_{i=1}^{n} p_i u(x_i) \tag{2.2}$$

The decision maker is assumed to choose the lottery that gives the highest expected utility. EU allows for a rich pattern of attitudes to risk. However, EU incorporates risk preferences through the shape of the utility function (rather than through the probabilities).

Definition 1 *(Risk aversion): A decision maker is said to be risk averse if he/she prefers the expected value of the lottery to playing the lottery. The converse behavior is risk loving and indifference is risk neutrality.*

It can be shown that risk-neutrality/risk-aversion/risk-seeking are equivalent, respectively, to the utility function, u, being linear/concave/convex, respectively. Risk aversion, arising from a concave utility function, underpins important explanations of economic phenomena in most areas of economics. These include the study of insurance, the design of contract schemes in prinicipal-agent problems, the optimal rates of taxes, determination of wage rates etc. A lottery L_p first order stochastically dominateslottery L_q if the distribution function of L_p lies everywhere below that of L_q. Decision makers under EU do not choose stochastically dominated options.

The axioms of rationality that underpin EU appear intuitive and plausible. However, they are only assumptions about human behavior. In particular, they should not be confused with the notions of rationality as used in logic, psychology or law. And, in fact, a large number of rigorous tests have rejected these axioms. This induced a major effort to propose auxiliary assumptions to EU in order to save it, but without much success.

2.2. Behavioral decision theories

Behavioral economics provides several alternative, axiomatically founded, theories that explain the evidence better than EU, yet are tractable. The EU formula in (2.2) has two salient features. (1) It is linear in probabilities. (2) Utility is defined over final wealth levels. Behavioral alternatives have relaxed both these assumptions.

Rank dependent utility (RDU) relaxes the linearity in probability assumption. However, the most satisfactory decision theory in economics, and the one with the strongest psychological foundation, is the Nobel Prize winning work on prospect theory (PT) due to the seminal paper in 1979 by Daniel Kahneman and Amos Tversky. However, PT allows for violations of first order stochastic dominance. Following the insights of RDU, PT was modified to cumulative prospect theory (CP). CP does not allow for the choice of stochastically dominated options. Since RDU is a special case of CP, it suffices to explain CP.

In CP (like PT but unlike EU and RDU) the carriers of utility are not final levels of wealth but deviations of these from a reference point; a well established idea from psychology. For instance, dip your right hand in cold water and your left hand in warm water. After a few seconds put them both in lukewarm water. Your right hand then feels warm and your left hand feels cold, although they are both gauging the same temperature. The status quo turns out to be a useful reference point in many applications. But some other possible choices could be an expected level, a fair level or a legal entitlement. If the outcome turns out to be superior (respectively inferior) to the reference point, then the decision maker is in the domain of gains (respectively losses).

CP also incorporates experimental and field evidence, which suggests that utility is evaluated differently in the domain of gains and losses. First, losses bite more than equivalent gains (loss aversion). Second, the utility function is concave for gains and convex for losses. Thus, in each domain, there is diminished sensitivity to gains/losses. Tversky and Kahneman propose the following (axiomatically founded) utility function that captures all these facts.

$$v(x) = \begin{cases} x^{\gamma} & \text{if } \; x \geq 0 \\ -\lambda(-x)^{\theta} & \text{if } \; x < 0 \end{cases} \tag{2.3}$$

where x is the deviation of some outcome relative to the reference point. γ, θ, λ are constants; $0 < \gamma < 1$, $0 < \theta < 1$. $\lambda > 1$ is known as the *coefficient of loss aversion*. Experimental results initially suggested $\gamma \simeq \theta \simeq 0.88$ and $\lambda \simeq 2.25$. However, subsequent research showed that the parameter of loss aversion is variable and depends on the context. The properties of the value function are illustrated in figure 2.1. Notice that the reference point is at zero and the function is relatively steeper for negative values. It is also concave for positive values but convex for negative values.

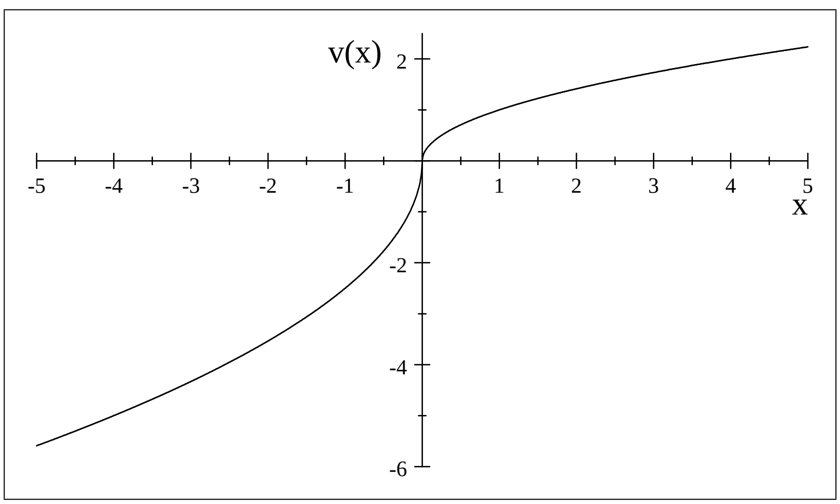

Figure 2.1: The utility function under CCP

The second main feature of CP (which it shares with PT and RDU but not EU) is that probabilities are transformed in a non-linear manner, using the device of a probability weighting function(PWF). Many PWF's have been proposed in the literature. The most satisfactory is the one by Drazen Prelec.

Definition 2 *By the Prelec function we mean the probability weighting function* $w(p)$: $[0,1] \to [0,1]$ *given by*

$$w(0) = 0,\ w(1) = 1, \tag{2.4}$$

$$w(p) = e^{-\beta(-\ln p)^{\alpha}},\ 0 < p \le 1,\ \alpha > 0,\ \beta > 0. \tag{2.5}$$

A plot of the Prelec function is shown in Figure 2.2. For the particular parameter values chosen the Prelec function is consistent with the empirically verified inverse S-shape.

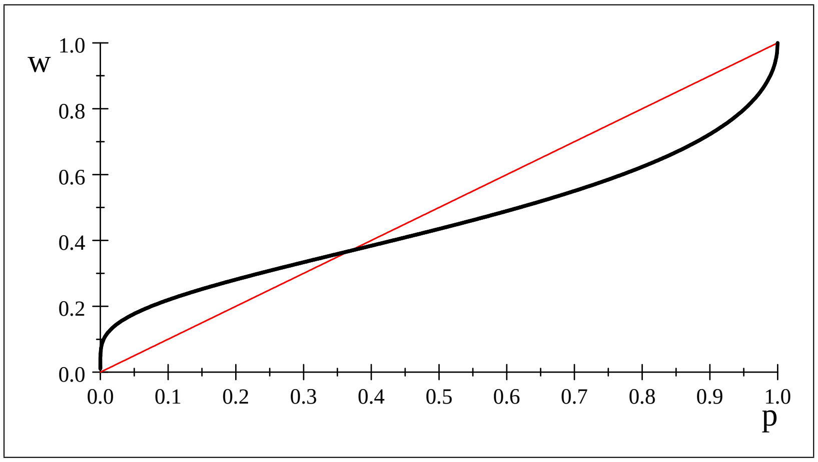

Figure 2.2: The Prelec function for $\beta = 1$ and $\alpha = 0.5$.

Let w^{+}, w^{-}, respectively, be the PWF's for the domain of gains and losses. There is evidence that $w^{+}(p) = w^{-}(p) = w(p)$. Define $\mathcal{L}_P$ as the set of lotteries formed by modifying (2.1) such that the reference outcome is subtracted from each of the individual outcomes. Also suppose that there are a total of $m+n$ outcomes such that there are m outcomes in the domain of loss and n outcomes in the domain of gains. More formally, we have the lottery

$$L = (\delta x_{-m}, p_{-m}; \delta x_{-m+1}, p_{-m+1}; ...; \delta x_{-1}, p_{-1}; \delta x_0, p_0; \delta x_1, p_1; \delta x_2, p_2; ...; \delta x_n, p_n), \tag{2.6}$$

where $\delta x = x - x_0$ and the restriction on probabilities is given by

$$\Sigma_{i=-m}^{n} p_i = 1,\ p_i \ge 0,\ i = -m, -m+1, ..., n, \tag{2.7}$$

and the restriction on outcomes is given by

$$\delta x_{-m} \le \delta x_{-m+1} \le ... < \delta x_{-1} \le \delta x_0 = 0 \le \delta x_1 \le \delta x_2 \le ... \le \delta x_n. \tag{2.8}$$

Definition 3 *For CP, the decision weights,* π_i, *are defined as follows:*
$\pi_n = w^{+}(p_n)$

$\pi_i = w^+\left(\Sigma_{j=i}^{n}\, p_j\right) - w^+\left(\Sigma_{j=i+1}^{n}\, p_j\right), i = 1, 2, ..., n-1,$
$\pi_{-m} = w^-\left(p_{-m}\right),$
$\pi_j = w^-\left(\Sigma_{i=-m}^{j}\, p_i\right) - w^-\left(\Sigma_{i=-m}^{j-1}\, p_i\right), j = -1, -2, ..., -m+1.$

A decision maker using CP maximizes a well defined objective function, called the value function, which we now define.

Definition 4 *(The value function under CP): The value of the lottery $L \in \mathcal{L}_P$ to the decision maker is given by*

$$V\left(L\right) = \Sigma_{i=-m}^{n} \pi_i v\left(\delta x_i\right). \tag{2.9}$$

Recall that $v\left(0\right) = 0$. Hence it is immaterial what value is chosen for π_0. This is why π_0 is undefined in Definition 3.

RDU is the special case of CP when all outcomes are in the domain of gains. EU is the special case of RDU when $w\left(p\right) = p$ ($\alpha = \beta = 1$ in Definition 2).

CP allows for a rich set of attitudes to risk, that depend not only on the curvature of the utility function but also on the shape of the PWF. A four-fold classification of risk is predicted by CP. In particular, decision makers are risk averse for moderate probability moderate gains (as in EU) but (unlike EU) they are risk seeking for low probability high gains. By contrast, decision makers are risk averse for low probability high losses (as in EU) but (unlike EU) are risk seeking for moderate probability moderate losses. This four-fold pattern of attitudes to risk is confirmed in laboratory experiments and explains a wide range of economic phenomena that EU is unable or finds it difficult to explain. See, for instance, the Asian disease example in the section on mental accounting below for separate attitudes to risk in the domain of gains and losses.

One criticism made in the recent work by al-Nowaihi and Dhami is that all standard PWF's used in RDU and CP are extremely steep at the end points (in the sense that $\lim\limits_{p\to 0} \frac{w(p)}{p} = \infty$, and $\lim\limits_{p\to 1} \frac{1-w(p)}{1-p} = \infty$). This makes very low and very high probabilities extremely salient to decision makers and leads RDU and CP to make predictions inconsistent with a range of important observed phenomena. For example, CP (and RDU) predicts far too much insurance (as compared to the evidence) against low probability natural hazards, such as earthquakes, floods and hurricanes, if insurance firms make zero profits (and at least some insurance even if insurance firms make positive profits). Another example is that CP (and RDU and EU) predicts that the imposition of the severest possible punishments (e.g., capital punishment), even with vanishingly small probability of detection, would deter all crime. This is contrary to the evidence. Other similar violations come from a range of other phenomena such as red traffic light running, driving and talking on mobile phones and not taking up adequate health checks.

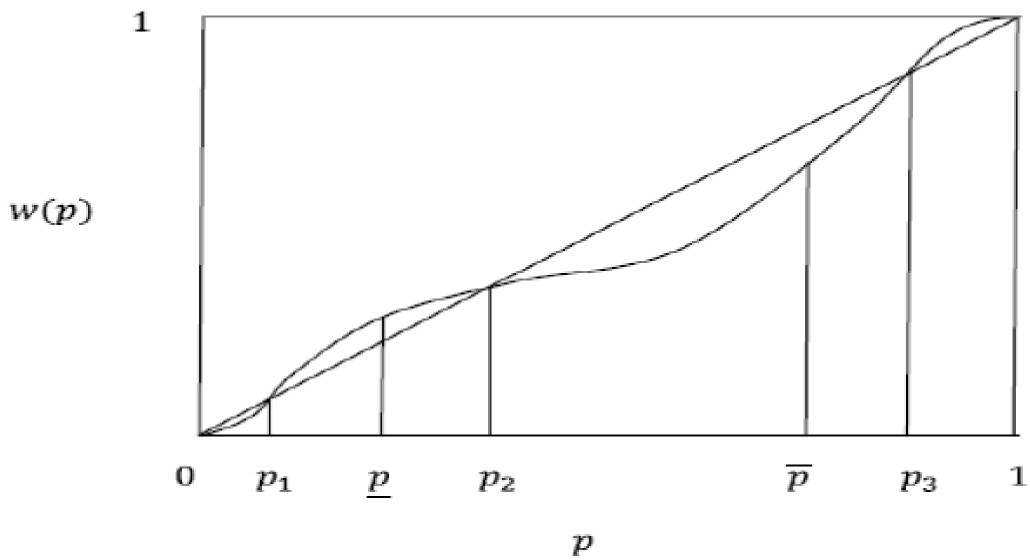

Figure 2.3: The composite Prelec weighting function (CPF).

al-Nowaihi and Dhami account for these problems by proposing composite cumulative prospect theory (CCP). This differs from CP in that it uses a composite Prelec probability weighting function(CPF), illustrated in Figure 2.3. It is made up of segments from three Prelec functions over the respective ranges $\left[0, \underline{p}\right]$, $\left(\underline{p}, \overline{p}\right]$, and $(\overline{p}, 1]$. Low probabilities are hugely underweighted in the sense that $\lim\limits_{p \to 0} \frac{w(p)}{p} = 0$. This allows CCP to predict human behavior, for instance, towards insurance and law enforcement, that is in better conformity with the evidence. CCP is much more general than the account that we have given here; it can do everything that EU, RDU, CP can do, but the converse is false.

Collectively, PT, CP and CCP successfully explain a wide range of otherwise unexplained phenomena. These include why it is hard to find a taxi on a rainy day in New York, asymmetric price elasticities, tax evasion problems, demand for insurance, the Allais paradox, the St. Petersburg paradox, paradoxes in the economics of crime and punishment, the endowment effect, the disposition effect, the equity premium puzzle, asset pricing, incentive schemes, etc.

3. Heuristics and Biases

The heuristics and biasesresearch program commenced with seminal papers by Tversky and Kahneman in 1971, 1974. Its basic premise is that individuals are not fully rational, but they are not irrational either. Tversky and Kahneman took the middle ground that individuals solve problems by relying on some fast(in terms of time) and frugal(in terms of information acquisition and processing) heuristics. Because such heuristics do not optimize in the classical sense, the performance of the heuristics is not necessarily optimal.

For instance, individuals often judge distance of an object by the heuristic of how clear it is, sharper objects are thought to be closer. This heuristic is fast and frugal. While this heuristic often works well, it might also result in biased and misleading perceptions. For instance, when conditions are hazy, distances might be overestimated. Conversely, when conditions are very clear, distances might be underestimated.

The heuristics and biases approach demonstrates a systematic departure of actual behavior from that prescribed by the laws of classical statistics. Rather, individuals employ a range of judgementheuristics. The point or interval estimates arising from these judgement heuristics are often in conflict with classical inferences. Even statistically sophisticated researchers or experienced decision makersoften rely on these judgement heuristics.

An important heuristic is the representativeness heuristic. When individuals use this heuristic they tend to assume that the properties of a small sample that they are familiar with is characteristic of the large population from which the sample was drawn. For example, an eminent colleague told us that up to quite late in his childhood he believed that most people in the world were Jewish members of the British Communist Party, because that was the circle of family and friends he grew up in. This is sometimes referred to as the law of small numbers. At other times, individuals exhibit the reverse bias, conservatism, i.e., they underestimate the likelihood of a sample from a given distribution. For example, people typically under estimate the likelihood of getting a consecutive run of 3 heads in 10 random tosses of a fair coin. These heuristics explain many puzzles such as the gambler's fallacy (avoid betting on a number that has come up last time), hot hands and winning streaks (because a basketball player has scored the last few shots, he is even more likely to score this one).

The behavior of humans is inconsistent with Bayes' rule, especially when they are presented with probabilistic information (as opposed to frequency format information). In particular, individuals engage in base rate neglect. For example, consider the following question. "There is a rare condition whose probability is one in a million. There is a test whose success rate is 99%. You tested positive. What is the probability that you have this condition?" The typical answer is 99%. The true answer is 0.01%.

Anchoringis one of the most robust and important heuristics that people use. In situations where people do not know the answer to a question, they are unduly influenced by an initial suggestion or anchor. Interestingly, the anchor could be completely random and unrelated to the answer. For example, a roulette wheel is spun. The resulting number is, say, 50. The subjects are then asked "how many countries are members of the UN?". The answers are very frequently found to be close to 50. Furthermore, once individuals fix an anchor in their minds, they adjust too slowly, and insufficiently, towards the target.

There is also a range of other judgement heuristics. These include hindsight-bias, false consensus, attribute substitution, conjunction fallacy, availability heuristic, affect heuristicand taking a necessary condition to be sufficient.

4. Mental accounting and framing

The idea of mental accounts is a significant and important part of behavioral economics with wide ranging implications. Richard Thaler, whose name is most closely identified with this concept in economics, defines mental accounting as follows. "I wish to use the term mental accountingto describe the entire process of coding, categorizing, and evaluating events." Elsewhere, he provides another definition that is useful in financial situations: "mental accounting is the set of cognitive operations used by individuals and households to organize, evaluate, and keep track of financial activities."

Let us begin with two features of mental accounts.

1. Money is not fungible across the accounts. So, for instance, one might put aside money in a separate account for children's education. This would be unnecessary if money were fully fungible across mental accounts.

2. The objective of the individual is not necessarily to maximize total financial wealth. An important objective that is highlighted in the mental accounts literature is to limit the size of loss in individual accounts.

Much of traditional theory makes the frame-invarianceassumption. By contrast, individual decisions are sensitive to the framing of problems. In a famous experiment, Kahneman and Tversky use a representative sample of physicians and pose them the following question framed in two alternative, but classically equivalent, ways that, respectively, stress the positive and the negative aspects. The question posed was the following.

Example 1 : *Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:*
Positive Framing: If program A is adopted, 200 people will be saved. If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. Which of the two programs would you favor?
Under positive framing, the solution is framed in terms of "lives saved". The result was that 72% of physicians chose A, the safe strategy and 28% chose program B, the risky strategy.

Negative Framing : If program C is adopted, 400 people will die. If program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. Which of the two programs would you favor?
For negative framing, only 28% of the physicians voted for the safe strategy C while 72% of them opted for D, the risky strategy.

The choice in each of the two frames (positive and negative) is identical. Thus, if A is preferred to B then C must be preferred to D (or vice versa). While the aim of Kahneman and Tversky was to show that individuals are risk averse for gains and risk loving for losses (see the discussion of prospect theory, above), the re-framing of the same question leads to completely different choices. Framing matters, and it can determine which mental account an item of expenditure or income is put into.

The evidence suggests that the marginal propensity to consume out of distinct mental accounts could be different. Thus, the effect of salary composition, for instance, between wage and bonus could have an effect on the employee's saving decision. Different pension plans that have otherwise identical effects in standard economics on account of perfect fungibility of money have distinct effects in the presence of non-fungibility and different marginal propensities to consume from different mental accounts. Hence, in the mental accounting framework, the effects of pension plans depend on which mental account is used to pay for the pension contributions.

Why does the ticking of the taxi meter reduce one's pleasure from the taxi ride? Why do we normally prefer to prepay for a vacation? Why do we often prefer to pay for consumer durables in installments? These simple facts fall outside the ambit of the workhorse life cycle model in macroeconomics. Prelec and Lowenstein argue that consumption that has already been paid for can be enjoyed as if it were free in the current period, which potentially increases its current enjoyment, hence, explaining prepayment of vacations. Similarly the thought of future payments reduces current utility from the consumption of a good/service (hence, the unpleasantness arising from a ticking taxi meter). The framework suggested by Prelec and Loewenstein can also explain the growing use of debit cards, use of tokens at casinos, dependence of mortgage payments on property values, popularity of fixed fee payment schemes, Club Med vacations and gift giving among couples, etc.

Individuals tend to break down complex problems into simpler ones that require less cognitive ability. Read, Loewenstein and Rabin term this as choice bracketing. An army General who focusses only on the battle at hand (narrow bracketing) might loose sight of winning the war (broad bracketing). Recall the problem of finding a taxi on a rainy day in New York. One possibility is that taxi drivers set a daily target for earnings (narrow bracketing) rather than an income target over a longer period (broad bracketing). The target is more quickly achieved on a busy rainy day leading to too many quits, too soon. The optimal for a taxi driver is, however, to work longer during rainy days because customers are more plentiful.

Choice bracketing also takes the form of deciding on the horizon over which one would evaluate one's portfolio. So, for instance, if one holds stocks and evaluates one's portfolio too often then one would observe too many upward and downward movements in the return to equities relative to bonds, creating feelings of anxiety. Equities might then require a higher premium, relative to the predictions of standard theory, if they are to be held in a portfolio. This explains an important puzzle, the equity premium puzzle, namely, an excess of return on equities over bonds that cannot be justified on the grounds of risk alone. Choice bracketing can also explain strategies used at alcoholics anonymousand the desire for risk aggregation. Alcoholics under treatment are usually asked to take one day at a time (narrow bracketing) because the possibility of lifetime abstinence (broad bracketing) is too daunting. Similarly, when a range of risks are aggregated (broad bracketing) they might well seem less onerous because a low outcome under one form of risk could be mitigated by a high outcome under another form of risk.

In standard economic theory, actual choice is the result of well-defined initial preferences and well-defined subjective probability distributions over uncertain events. By contrast, Ariely, Loewenstein and Prelec show that the initial framing of problems or situations can have an inordinate effect on ones initial preferences. In turn, this determines the individual's subsequent choices. One is then lead to question the very notion of humans having well defined initial preferences and subjective probability distributions over events. This helps to explain a range of phenomena, such as the initial hype before mega sporting events, reluctance of firms to cut wages and inter-industry wage differentials, among others.

An assortment of other mental accounting phenomena exist. An important category of such problems arises when individuals take actions to (mentally) justify sunk costs that have been incurred in the past. Hence, there is a tendency to hold on to loss making stocks for too long and sell winners too early. The reason is that individuals create a mental account for a particular category of expenditure. Sunk costs push this mental account into the red, which causes anxiety. Individuals then take wasteful actions to avoid closing accounts that are in red.

Recent research has also highlighted the mental accounting of goals. Individuals often form goals and targets, e.g., completing term papers, referee reports, quitting an addiction, loosing a pre-specified amount of weight by a certain date and so on. While goals can be motivational, they can also be counterproductive. On account of loss aversion, missed goals impose large losses, and the individual might also engage in emotional repair following a missed goal, etc.

Mental accounting has also be applied to retirement savings. Of particular importance is the role of status-quo options. It has been found that if the default is that the individual is signed-into a pension plan (as opposed to the converse) then the take-up of pension plans increases. The number of options offered in an investment plan is also important. Too many options can make choice more difficult. Individuals also use the diversification heuristicto choose among alternative pension vehicles, which motivates individuals to spread their money evenly among the available options even when the prediction of standard economics is to engage in an uneven spread.

5. Emotions

In economics, it has become standard to focus on emotionless deliberation. However, much recent research shows that this view is not supported by the evidence. Emotions are central to humans and affect much human deliberation.

We often worry about an uncertain future. Will my pension contributions be sufficient for my retired life? Will early retirement turn out to be a good decision? Should I buy or sell my stocks in the current economic environment? These and other temporal economic decisions produce a feeling of apprehension and unease: We shall simply refer to these feelings as anxiety. Anticipation of a (stressful) future event can create anxiety that in some cases is worse than the event itself. Think, for instance, of an impending visit to a dentist for treatment. Such feelings of anxiety have two features. (1) They are caused by anticipation of future events, and (2) they are aversive in the sense that we dread such future events. As a result, one might delay a visit to, say, the dentist, with detrimental consequences for health.

Future events need not, however, always be unpleasant. We might savor a pleasurable experience such as waiting for a dessert at the end of a meal. Or, we might delay a pleasurable vacation. Pleasant events provide examples where an individual seems to exhibit a negative discount rate; a preference for the future rather than the present. Insofar as feelings of anxiety and savoring impact temporal choice in important ways, we need to amend the standard model.

In one model by Caplin and Leahy, uncertainty is introduced into a model of anticipal utility. In an interesting application they derived equilibrium prices of assets whose returns are ex-ante random. The randomness in the returns creates a current feeling of anxiety. Hence, the ownership of a risky stock reduces its price and increases its return relative to the no-anxiety case and relative to assets whose returns are not too random, such as bonds. This provides one possible explanation for the equity premium puzzle.

Most dynamic models assume that future tastes are known. However, tastes can change over time due to a variety of reasons, e.g., habit formation, moods, social environment, maturity etc. Projection bias arises when individuals make predictions of their future tastes that are not realized in equilibrium. Evidence shows that people are disproportionately influenced by their current tastes in making predictions of their future tastes. For instance, when ordering food at the beginning of a meal we must predict how hungry we will be at the end of the meal. We often overestimate our future hunger and order more than we can eat. One can also make forecasting errors in mainstream economics but, in the dominant rational expectations view, there are no such errors in the absence of random shocks and, on average, such errors are zero. Both features are violated under projection bias.

Individuals who have projection bias make time inconsistent choices, without being aware of it. Under the influence of projection bias, individuals might end up buying goods in a hot state that they might regret later. In a hot state they might end up marrying or proposing a marriage/divorce without sufficient deliberation. Public policy in some cases provides a cooling-off period before which applications for marriage and divorce cannot be filed. Projection bias also provides an alternative explanation for why individuals get addicted. First, people underestimate the (negative) effect of current consumption on future utility. Second, individuals might under-appreciate the formation of habit, i.e., addiction with the harmful product. By contrast, in mainstream economics, individuals know the entire time path of costs and benefits of addiction and if they choose to be addicts, they do so rationally.

6. Fairness and Reciprocity

Traditional economic theory relies on the assumption of self-interested behavior (or selfish preferences). This is contradicted by the evidence. A substantial fraction of individuals exhibit social preferences, i.e., care about the consumption and well being of others. These preferences are more general than simply altruism (unconditional kindness) and spite (unconditional unkindness). In practice, however, altruism is unable to explain a very wide range of behaviors. For instance, it does not explain why many individuals are conditional cooperators, i.e., cooperate only when others cooperate. Intrinsic reciprocity, on the other hand, takes the form of responding to kindness (unkindness) with kindness (unkindness) even in the absence of obvious longer term gains.

Purely selfish preferences are unable to explain a range of phenomena from many diverse areas such as collective action, contract theory, the structure of incentives, political economy and the results of several experimental games. Many of these phenomena can be accounted for by a model which recognizes that individuals have social preferences.

1. Wages are usually not lowered in a recession by firms for fear that this action could be construed by workers as unfair and can affect their morale and effort levels.

2. Voters voluntarily vote, when pure self interest might dictate them not to take the

trouble of voting (almost all single votes are not pivotal, except in small committees, and voting is costly).

3. In an experiment, most respondents thought that a hardware store's actions to increase the price of shovels following a snowstorm was unfair. However, they did not perceive an increase in price following an increase in the cost of inputs as unfair.

4. Why does bargaining stall, with adverse consequences in a range of situations, such as civil litigation, labor unrest, domestic strife, and clashes among religious and ethnic groups? Individuals confuse between what serves their self interest and what is fair. This self-serving bias can lead to a bargaining impasse.

Social preferences have been found in a range of experimental games that allow for small group or bilateral interaction. These experiments include the gift exchange game, the ultimatum game, the dictator game, the trust game, the public good game, among others. The most commonly replicated of these experiments is the ultimatum game.

The ultimatum game is a game played between two players who could share a cake (or prize) of some unit size that is infinitely divisible. The first player, the proposer, proposes to give a share of the cake $s \in [0,1]$ to the second player, the responder. The responder can either accept the offer, in which case the respective payoffs of the proposer and the responder are: $y_1 = 1 - s$, $y_2 = s$. Or the responder can reject the offer, in which case the respective payoffs are $y_1 = 0$, $y_2 = 0$. Under selfish preferences the subgame perfect outcome(which is the appropriate notion of the game theoretic equilibrium here) is $y_1^* = 1$, $y_2^* = 0$. The basic idea being that if the responder does not engage in this game then he gets nothing, so he would accept the smallest indivisible unit of money (if this is a penny then $y_1^* = 99p$, $y_2^* = 1p$). The experimental results reject the prediction $y_1^* = 1$, $y_2^* = 0$. The gist of the results is as follows. (1) The mean offer is $s = 0.3$ to 0.4. (2) The median offer is $s = 0.4$ to 0.5. (3) There are rarely any unfair offers, $s = 0$ to 0.1 or over-fair offers $s > 0.5$. (4) Low offers are often rejected. These results continue to hold for high stake games.

In public good games, the purpose is to explain human cooperation in undertaking a task for the common good. All individuals simultaneously make a voluntary contribution to a common pool. The experimenter then returns a fraction of the total sum of money to each participant; this captures the essence of a public good. The prediction of the standard model with selfish preferences is that everyone should free ride, i.e., depend on others to make the contribution and make zero contributions themselves. The experimental results are that individuals start with a reasonably high level of contribution. However, in successive rounds the level of contributions drops off to a level that is consistent with the model of selfish preferences. Does the standard model stand vindicated then? Not necessarily. It does not explain why subjects start off with initially high levels of contribution. Furthermore, if these subjects start afresh then they again start with high levels of contributions. Negative reciprocity is a leading explanation. Most individuals start off by contributing reasonably high amounts but some do not. In subsequent rounds, negative reciprocity kicks in and everyone lowers their contribution amounts.

An important variant of the public good game is the public good game with punishment. What if contributors are allowed to punish the non-contributors at the end of the experiment. The prediction of standard theory is that individuals should treat bygones as bygones and not engage in any punishment. Anticipating this, non-cooperators will not be deterred. The results of experiments do not support this prediction. Allowing for the possibility of punishment, contribution levels reach almost the first-best level and non-contributors are heavily punished. The public good experiments indicate that subjects are conditional cooperators. They cooperate when they think that others will cooperate. Furthermore, their preferences allow for the material payoffs of others (social preferences), thus giving them an incentive to punish even in games with a finite horizon. The implications of this result are deeper than just the current context.

Humans take account of the role of intentions. Humans do not punish unkind behavior if it was unintentional. However, intentionally unkind behavior can get severely punished. These issues are formally modelled in psychological game theory.

Experimental evidence indicates that there is a very large fraction (roughly 40%-60%, depending on the experiment) of purely self-interested individuals. The behavior of these individuals accords well with the predictions of the selfish preferences model, even in bilateral interactions. An important and interesting issue for theoretical and empirical research is to examine the implications of heterogeneity of preferences in the population. A range of theoretical and experimental work indicates that even a minority of individuals with social preferences can significantly alter the standard predictions.

Conditional reciprocity, inequity aversion, and the role of intentions explain several phenomena in contract theory such as the following. Why don't incentives always work? Why do we often observe cooperation in finitely repeated relationships? Why do we observe bonus contracts? Why are contracts often incomplete? Too great a use of the carrot and stick policy is likely to be perceived as hostile intent by workers, who could conditionally reciprocate by reducing effort.

7. Time Discounting

Decisions that have a time dimension are central to many economic problems. The basic idea in most modern theories of time preference is the tension between immediate gratification and delayed, but greater, gratification. For example, if an individual or nation desires greater consumption in the future, then current consumption has to be reduced to increase saving and investment. This trade-off determines the consumption profile of individuals and nations. Models which are separable in time and outcome are sometimes called discounted utility models(DU).

Given two dates, t and $t+T$, let, $\mathbf{c}_t = (c_t, c_{t+1}, ..., c_{t+T})$, be a stream of dated consumption levels, or a consumption profile. A consumption profile could either be a historical record of the actual consumption levels, or a plan for future consumption formulated at a time, $t_0 \leq t$. The DU model states that the utility from this consumption profile is,

$$U(\mathbf{c}_t) = \sum_{\tau=0}^{T} D(t+\tau) u(c_{t+\tau}), \tag{7.1}$$

where D is the discount function and u is the single-period utility function or felicity. Note that D is a function of time only and u depends on time only through the dependence of c on time.

An important question is the following. Suppose that the consumption profile, $\mathbf{c}_t$, is optimal, in the sense that it maximizes (7.1) subject to the constraints facing the consumer. Is the consumption profile $\mathbf{c}_{t+s} = (c_{t+s}, c_{t+s+1}, ..., c_{t+T})$ also optimal in the sense that it maximizes $\sum_{\tau=s}^{T} D(t+\tau) u(c_{t+\tau})$? If this is the case then, $\mathbf{c}_t$ is said to be *time-consistent*. Generally, optimal consumption plans are not time-consistent. For example a consumer may, now, prefer to have two apples in 51 days time to one apple in 50 days time but, come day 50, the consumer may prefer one apple that day to two apples the following day.

A special case of the DU model is the exponential discounted utilitymodel (EDU) of Paul Samuelson, 1937. In this case the discount function takes the form

$$D(t) = e^{-\theta t},\ \theta > 0, \tag{7.2}$$

where θ is is a constant parameter known as the discount rate and $\delta = e^{-\theta}$ is known as the discount factor. Thus, we also have $D(t) = \delta^t$, $0 < \delta < 1$. This model has proved extremely tractable and useful for many purposes. In particular, optimal consumption plans are always time consistent if, and only if, the discount function is given by (7.2). However, EDU is rejected by the evidence.

Anomalies of EDU include the following. Under EDU, the discount rate between a time period of a given length remains the same irrespective of when the time period begins. However, the supposition of a constant discount rate is violated in many experiments; a finding that is known as the common difference effect. The most famous illustration of the common difference effect is Thaler's apples example, mentioned above. This led George Loewenstein and Drazen Prelec, in 1992, to propose the following hyperbolic discount functionthat can address the common difference effect:

$$D(t) = (1 + \alpha t)^{-\beta/\alpha}, \ \alpha > 0, \ \beta > 0. \tag{7.3}$$

Other anomalies of EDU include the following. Gains are more salient than losses (the sign effect), greater magnitudes are more salient than smaller ones (the magnitude effect), discounting over an interval depends on how that interval is subdivided (subadditive discounting), choices may appear to be cyclical (apparent intransitivity of preferences), choices may depend on the spread or monotonicity of the consumption sequence despite different sequences having the same discounted utility, and so on.

Loewenstein and Prelec explained the sign effect by proposing a per period utility function or felicity (u in (7.1)) with greater elasticity for losses as compared to gains. They explained gain-loss asymmetry by proposing that the elasticity of the felicity for losses exceeds that for gains. They explained the magnitude effect by assuming that the elasticity of the felicity is increasing. A function form for the felicity that incorporates all these features was given by Ali al-Nowaihi and Sanjit Dhami. The combination of such a value function with the discount function (7.3) may be called the hyperbolically discounted utilitymodel (HDU). Why do human and non-human subjects exhibit hyperbolic discounting? The work of Partha Dasgupta and Eric Maskin examines the evolutionary basis for this phenomenon. A wide range of puzzles that are largely unresolved with EDU can be explained by HDU. These include the following. Why does consumption track income so closely? Why do individuals under-save for retirement? Why is there a sharp drop in consumption at retirement? Why do individuals hold illiquid assets and credit card debt simultaneously? Why have national savings rates been declining? Why is marginal propensity to consume asset specific?

Unlike the EDU model, where the discount rate is constant across any two time periods of equal length, in the HDU model, the discount depends on the distance from the present, say, time t. In particular, under HDU, the discount rate over the interval $[t, s]$, $s > t$ is higher than the discount rate over the equal sized interval $[s, s + (s - t)]$. A higher discount rate implies a lower discount factor and, so, a lower weight on the future. In this sense, the decision maker is said to be present-biasedrelative to the EDU model.

A second class of models that allows for a departure of the present from the future (relative to the EDU model) is the model of multiple selves.In this approach, the same individual, over time, can be viewed as a collection of different selves, one for each time period. Furthermore, the current-self cannot perfectly control future selves by command. However, the current-self can potentially take actions that can alter the constraints faced by future-selves.

An important issue in the present-bias models is the degree of awareness that a current self has about the degree of present bias of one's own future selves. Experimental evidence is, at the moment, not conclusive about the degree of self awareness of the individual. It seems, however, that completely sophisticated behavior, i.e., full awareness of one's future self-control problems and completely naive behavior, i.e., no awareness of future self control problems are both unreasonable relative to the intermediate case of partial awareness.

There are several applications of present biased preferences. Applied to the life cycle model it turns out that the present-bias of individuals leads to greater current consumption and lower savings relative to the time consistent behavior found in the EDU model. The current selves can, however, take certain actions to bind the future selves.

One possibility is for the current self to invest in illiquid assets, such as housing, that constrains the ability of the future selves to engage in excessive consumption. This can explain, for instance, why individuals simultaneously hold illiquid assets, such as housing (certainly illiquid as compared to financial assets), and credit card debt, a phenomenon that EDU cannot explain. The current-self buys an illiquid asset to discipline the self-indulgent behavior of the future-selves. The future-self is now faced with mortgage payments (or a lack of cash from the purchase of the house) and, so, this alters his/her budget constraint. However, the easy availability of credit cards could induce the future-self to borrow and, so, successfully, relax the budget constraint. In so doing, the future self might believe that his/her future selves will be more responsible, which might well turn out to be a naive assumption.

8. Neuroeconomics

Neuroeconomics studies the neural activity in the human brain when one makes economic choices. One of the leading neuroeconomists, Colin Camerer, defines the subject matter in the following manner: "Neuroeconomics is a specialization of behavioral economics that plans to use neural data to create a mathematical and neurally disciplined approach to the microfoundation of economics."

There is a wide variety of techniques used in neuroeconomics to measure brain activity. Some of these techniques are so invasive, e.g., single neuron recordings, that they can only be used on animals. However, this is of value, as humans share a primitive mammalian brain with other mammals and so the lessons learnt from animal neuroeconomic experiments can provide valuable information. Other techniques, e.g., functional magnetic resonance imaging, FMRI, are less invasive and are able to measure electrical activity in the brain while pinpointing, with varying accuracy, to relatively small and specific areas. Establishing causality is an issue with many of these techniques. Such is the speed of progress in neuroscience that it is now possible to be more sure of direct causality using very new techniques such as transcranial magnetic stimulation, TMS, whose method can be described as follows. Suppose that we suspect that some region, R, of the brain is responsible for some economic relevant behavior, B. Then TMS can temporarily curtail/stimulate area R. If this curtails/enhances behavior B then there is indeed a possibility that region R is responsible for behavior, B.

It is also possible to fruitfully apply older techniques that rely on psychophysical measurements, such as pupil dilation, heart rate, skin conductance etc. For instance, pupil dilation is enhanced when one sends false or misleading information. This has obvious implications for economics, e.g., in predicting the truthfulness of messages in an asymmetric information game. Activation in the dorsal striatum (which is the brain's reward circuit) informs why the outcomes in the public good game with and without punishment are so different. Existing neoclassical models are unable to offer a convincing explanation.

Clearly, neuroeconomic variables have a direct role to play in economic models. But they can also be useful indirectly. Economic models are generally estimated by hypothesizing that some endogenous variable, y, is determined by a set of exogenous variables $x_1, ..., x_n, ..., x_N$ and some noise term, ε. Typically, data is available on only a subset of the variables, $x_1, ..., x_n$. The omitted variables are then, necessarily, subsumed within the noise term, ε. It is possible, however, that some neuroeconomic variables, $z_1, ..., z_M$, influence y and are correlated with the omitted variables and, hence, can be used as instruments for them.

Usually there are competing economic explanations for the same phenomenon. One can then ask which explanation is neurally plausible, in the sense of being consistent with known findings from neuroeconomics. Thus, potentially, neuroeconomics could deliver an additional criterion for selecting among competing explanations.

Our view is that the correct attitude is to regard behavioral economics as an extension of mainstream (neoclassical) economics into areas (and there are many) where the latter has empirically failed. The neuroeconomic findings are suggestive of a dual-system functioning of the brain in which system 2 (the neocortex) determines cognition and system 1 (the limbic system) determines emotion. Thus, one may hypothesize that decisions made by system 2 would be in accord with rational choice theories while the interaction between the two systems would be in accord with alternative behavioral theories. Of course, the two systems need not be in conflict. Thus both systems would, for example, urge us to breathe, eat and drink, seek shelter and avoid hazardous situations. And, indeed, neuroeconomic findings have supported a range of rational choice models. Support for the rational choice models, however, seems to arise largely in environments that are stationary and where evolution has had enough time to make an impact. Support for behavioral alternatives is more widespread; and include, for example, models of fairness and reciprocity, prospect theory, models of limited iterative thinking and hyperbolic time discounting.

9. Behavioral Game Theory

Game theory deals with situations in which a set of individuals (called players) take actions or play strategies that affect not only their own payoffs (or utilities) but also those of others. The fundamental concept in game theory is that of a noncooperative equilibrium, proposed by John Nash in 1950, hence also known as a Nash equilibrium. In such an equilibrium, each player plays a strategy that is best for himself in terms of maximizing payoffs or utility (i.e., a best reply), given the equilibrium strategies of the other players. Nash proved that every finite game (finite number of players and finite number of actions) has a Nash equilibrium.

It is permitted for players to have a probability distribution over the play of their actions. To emphasize this point, sometimes strategies are referred to as mixed strategies. If each action is played with strictly positive probability, then the strategy is strictly mixed. By contrast, in a pure strategy, each action is played with probability 0 or 1. A strategy x dominates a strategy y, for player i, if x gives higher payoffs for i than y, irrespective of the strategies of the other players. One very simple method of searching for a Nash equilibrium is that of iterated elimination of dominated strategies. A dynamic gameis a game played over time in which players have the opportunity of observing and learning from the previous actions of other players. A simple way to find a Nash equilibrium of a dynamic game is by backward induction: Find a Nash equilibrium for the subgameplayed in the final period. Then find a Nash equilibrium in the penultimate game, assuming that players, come the final period subgame, will play the Nash equilibrium already found for that subgame. Carry on till the initial subgame. Backward induction has the added attraction that it does not allow threats that are not credible (i.e., those that would never be used). Such an equilibrium is called a subgame perfect equilibrium.

Behavioral game theory has developed as a response to the violations of the predictions of (mainstream) game theory. Contrary to the prediction of game theory, players do not engage in iterated deletion of dominated strategies if this involves more than $2-3$ steps of iteration. Within the context of 2 steps of iterated eliminations of dominated strategies, the outcome is sensitive to framing effects.

A test of the iterated deletion of dominated strategies is in the context of the Abreu-Matsushima mechanism design solution. This is a mechanism that is implementable by using very few steps of iterated elimination of dominated strategies. The results of Martin Sefton and Abdullah Yavas show empirically that the predictions of this mechanism fail. In conjunction with similar evidence from elsewhere, this casts doubt on the implementability of other mechanisms (most of which are even more complex).

The backward induction argument usually fails in experiments.

Multiple Nash equilibria arise in almost all areas of economics. An important question is whether players can coordinate (agree) on one of the equilibria (preferably the best for all, if such exists). Widespread failure to coordinate on any equilibrium (particularly the one that maximizes joint payoffs) is observed in experimental results. Pre-play communication does not necessarily enhance coordination. History plays a powerful role in influencing the degree of coordination. Only under some conditions, e.g., gradual growth of group sizes, successful coordination in small groups, knowledge of the history of such coordination, etc., does coordination emerge successfully. In experiments, if generations give public advice to their successors then coordination can be enhanced.

While players often seem to play mixed strategies they do not do so in the proportions predicted by a mixed strategy Nash equilibrium (MNE). Furthermore, while the aggregate proportions (across players) are sometimes found to be consistent with MNE, individual proportions of play are typically inconsistent with a MNE. Tennis and football professionals seem to play mixed strategies that are consistent with the evidence. However, the serial dependence of serves, in tennis for example, is inconsistent with the predictions of a MNE.

There is also a fascinating new literature that tests the predictions of standard game theory by directly attempting to infer the cognitive process when subjects play experimental games. Game theory does not make any assumptions on the cognitive process when subjects play games. Some traditionalists argue that a study of the cognitive process cannot confirm or reject anything in game theory. Consider, for instance, the work of Eric Johnson et al. They consider Rubinstein's alternative offers bargaining game played by 2 players over 3 rounds. The game is easily solved by backward induction. One would, therefore, expect that subjects will look to see the last round data such as payoffs, actions etc., then look at the second round data and finally, the first round data. The authors use MOUSELAB software to observe if the pattern of searches made by experimental subjects is backward, beginning with the third period data. It turns out that subjects search forward rather than backward, i.e., looking first at the first round data, then the second round data and finally, the third round data.

The Nash equilibrium concept and its refinements have helped to shed light on many interesting questions in economics. Furthermore, the machinery of game theory has forced economists to be completely explicit about who the players are, what moves are open to them, the sequence of their moves, what objectives they have, what information they acquire and when, etc. This level of rigor gives modern economics a tremendous advantage over pre-game-theoretic economics. However, the empirical evidence on the Nash equilibrium and its refinements is not very encouraging, especially when individuals interact in bilateral or small group settings. More positive evidence is found at the level of market outcomes and in large groups.

There are now several alternative models of behavioral game theory that perform better than a MNE. These include the concept of a quantal response equilibrium (QRE).

QRE has been able to explain a range of findings from experimental/field data that the classical equilibrium concepts have found troubling to explain. In a QRE, players play strictly mixed strategies. In contrast to notions in classical game theory, such as trembling hand perfection, the trembles do not, in the limit, tend to zero in a QRE. Trembles are fundamental and remain significantly largein a QRE. This avoids a range of problems that arise in classical theory regarding assigning beliefs about actions that are not played in equilibrium. The reason is that, in a QRE, all paths of play are ex-ante possible. The main requirement of a QRE is that players form beliefs that are correct in equilibrium. However, because players randomize over all their actions in a QRE, they do not necessarily play best replies.

Forming beliefs that are correct in equilibrium, as in QRE, is cognitively challenging. This requirement is relaxed in a second class of behavioral game theory models, the level-k models. In these disequilibrium models, players' beliefs need not be correct. However, players play best responses, conditional on these, possibly incorrect, beliefs.

Game theorists steeped in classical theory might well react indignantly and ask how players could possibly hold on to beliefs that turn out to be incorrect? There are two responses. (1) Level-k models are motivated by actual experimental evidence which indicates disequilibrium beliefs. (2) Even when a process of learning might correct the beliefs of players over time, there are still good reasons to pay attention to level-k models. First, many important decisions that economic agents make in real life are of the nature of one-off decisions or decisions that are not repeated often enough. Second, in many experiments where long term learning has a positive impact, the economic environment is held fixed. In real life, however, the economic environment is often too variable to permit the clean learning results of experiments.

Level-k models introduce heterogeneity among players by distinguishing between types of players, $L0$, $L1$, $L2$, ... The generic type is denoted by Lk. Type $L0$, the only non-strategic type, typically chooses to simply randomize among the various available actions. In general, any type Lk, $k \geq 1$ assumes that all other players are of type $Lk - 1$ only. In it's cognitive hierarchy version, type Lk, $k \geq 1$ assumes that there is a distribution of all the lower types of players. Fixing beliefs in this manner, each type of player then chooses a best response over its available actions. Type $L0$, sometimes known as the anchoring type, need not be physically present in the population. It can exist purely as a model of other players in the minds of type $L1$ players.

Consider the p-beauty contest. In the simplest such game, subjects are asked to guess numbers between 0 and 100. The subject whose guess is closest to 2/3 times the average wins a prize. The Nash equilibrium prediction is for each player to choose 0 (the reader should check this).

However, the level-k model predicts that people will play the p-beauty contest as follows. A type $L0$ player simply randomizes and so chooses the number 50. A type $L1$ player thinks that all others are type $L0$ so chooses the number $\frac{2}{3}(50) \approx 33$. A type $L2$ player thinks that all others are type $L1$ and so chooses the number $50\left(\frac{2}{3}\right)^2 \approx 22$, and so on. Thus, if players were indeed of types $L0$, $L1$, $L2$, ... then we would expect to see spikes in the data at the numbers $50\left(\frac{2}{3}\right)^0$, $50\left(\frac{2}{3}\right)^1$, $50\left(\frac{2}{3}\right)^2$, ... This is a remarkably precise and unambiguous prediction. Moreover, this prediction has been verified in several empirical works even when played with very high ability subjects.

Level-k models can explain a range of phenomena that are difficult to explain with other models. These include results from auctions and market entry games, among many others. The evidence from MOUSELAB experiments indicates that types $L1$ and $L2$, in the level-k models, make up 67-89 percent of the population. This data is hard to reconcile with QRE models suggesting that the level-k models have better psychological foundations.

There is a range of other behavioral game theory models. These include the noisy introspection model, models that consider the role of memory explicitly, models of coarse thinking and mental categorization, analogy based equilibrium models, cursed equilibrium and so on. Some are more speculative than others.

We have not touched on behavioral models of learning for two reasons. First, the technical requirements for understanding even the basics are beyond the scope of the current contribution. Second, while there are several promising behavioral models of learning that explain the result of several experimental games more progress is required before this can be presented as a comprehensive body of work.

10. Liberalism and Paternalism

A fundamental principle of classical liberalism is that people are the best judges of their own welfare. This lies, explicitly or implicitly, at the heart of traditional public economics. Thus the role of government is limited to upholding the law, enforcing contracts and correcting for market failures, when practicable. Of course, it has always been recognized that some well defined groups, children for example, are not fully able to judge their own interests. The opposite view that, in the main, people are not the best judges of their own welfare, is known as paternalism. Under paternalism, an institution is needed to judge people's welfare for them, for example a government, political party or religious establishment. The sad history of paternalistic systems is succinctly captured by the following quote from The Oxford Companion to Philosophy:

"paternalism [is the] power and authority one person or institution exercises over another to confer benefits or prevent harm for the latter regardless of the latter's informed consent ... paternalism ... increases the potential for the abuse of state power, arbitrary discrimination, tyranny, and civil strife ... liberalism ... provides the only way for the proponents of conflicting ways of life to live together ... liberalism [is] the only viable basis for peaceful coexistence in culturally and religiously plural societies."

However, there is strong empirical evidence from behavioral economics that people are, by and large, not rational optimizers, unlike the assumption under liberalism. A small selected sample of the violation of the rationality assumption is as follows. Many individuals do not utilize the full limits of their 401k pension plans despite the fact that the company will match the extra contributions. There is also inadequate diversification of portfolios. People often use a range of judgement heuristics to make decisions. The evidence is largely (but not always) inconsistent with the use of Bayes' rule, rational expectations, and even strict optimization. People become obese without any apparent liking for the condition of obesity or addicts who then often need to be helped by others to get rid of the habit. There is also an overutilization of payday loans. These are loans that are repaid when the next paycheck arrives; the short term interest rates on these loans are much higher than loans available elsewhere.

In some cases, the textbook rational behavior might not be observed on account of lack of information, or limited cognitive abilities. In other cases it could be due to a current-bias in preferences, which lead to self-control problems and dynamic inconsistency. Examples are various forms of addictive behavior, the desire to quit, and the subsequent relapses that are often difficult to reconcile by using the traditional rationality framework. People exhibit preference reversals. Preferences are often not well-defined, are often constructed by individuals, and dependent on the framing of information.

Shlomo Benartzi and Richard Thaler show that only 20% of the people in their sample prefer their own portfolio to the median portfolio. In other important contexts such as pensions, individuals sometimes use the $1/n$ heuristic to allocate their portfolio where n is the number of options available, hence, making outcomes very sensitive to what assets are available.

This evidence has disturbing implications for the debate between liberalism and paternalism. Aware of this, behavioral economists sometimes refer to paternalism in the traditional sense as heavy-handed paternalism. By contrast, the form of paternalism suggested by behavioral economists is sometimes call soft paternalismor light paternalism. It is this form of paternalism that we now review.

George Loewenstein and Emily Haisley describe light paternalism as follows: "... the common goal of these approaches is to steer human behavior in more beneficial directions while minimizing coercion, maintaining individual autonomy and maximizing choice to the greatest extent possible. Light paternalism aims to enhance decision making without restricting it."

Colin Camerer et al. view the relaxation of rationality in welfare economics as a natural progression of the relaxation of other restrictive assumptions in economics, such as perfect competition, perfect information, certainty etc.

At the heart of most proposals under the ambit of light paternalism is the need to distort individual decisions/choices as little as possible and use as little coercion as possible. The idea is that the mistakes (or misperceptions) of the boundedly rational types will be nudgedin the direction of a correction while the minimal intrusionassociated with light paternalism should have little effect on the actions of the fully rational types. To operationalize light paternalism, one needs to specify what a mistake/misperception is. Furthermore, in the light of these mistakes/misperceptions, how should we judge the outcomes of light-paternalistic policies? The latter question requires us to have some behavioral welfare criteria. The reason that a traditional welfare criteria would not be suitable is that actual choices might not be indicative of an individual's best interests. Thus, unlike the classical analysis we cannot use preference based measures of welfare such as equivalent variation and/or compensating variation to seek the direction in which individual welfare is improving. As Loewenstein and Haisley put it: "Clearly, it doesn't make sense to assess whether someone is committing an error using a measure that is premised on the assumption that people don't commit errors."

Botond Koszegi and Matthew Rabin argue that unless one imposes ancillary assumptions, it is impossible to infer preferences from choice even with unlimited data and under the assumption that people are 100% rational. Consider, for instance, the possibility that people have social preferences. Suppose that we observe a set of choices in which an individual shares with others when there is an option to do so. From this, should we conclude that the individual is happy to share with others? Or that she would feel guilty if she did not? Or that she would prefer not to have an option to share? These distinctions cannot be made based on the choice data. They show that "... the potential wedge between preferredoutcomes and implemented outcomes can take virtually any form."

Although additional evidence is already accumulating at a rapid rate, behavioral economics is a relatively new field and many important effects and channels are not fully understood yet. For that reason one needs to be cautious and careful in testing the implications of various paternalistic policies. Some have advocated careful small-scale pilot studies as well as field studies before the actual implementation of paternalistic policies. Colin Camerer et al. give yet more caveats for the cautious approach. They point out that in many cases the evidence suggests that not all people are subject to the relevant behavioral biases. Clearly regulation that is too stringent might have a negative effect on these rational types.

Many notions of behavioral paternalism have been advocated in the literature. These include libertarian paternalism due to Richard Thaler and Cass Sunstein, asymmetric paternalism due to Colin Camerer et al. and a pragmatic approach due to George Loewenstein and Emily Haisley.

11. Brief Conclusions

Behavioral economics is one of the most rapidly growing fields in economics and is gradually becoming mainstream. It relies on a strong respect for the evidence, falsifiability, and a recognition that experiments in the lab, and neuroeconomic evidence, can usefully complement field data. It draws its principles from several decades of research in psychology (and sociology). Behavioral economics has seriously questioned the rationality assumption in economics as well as the assumption of selfish preferences. While traditional economics did recognize the possiblity of less than perfect rationality, it assumed that these errors would cancel out in the aggregate. Behavioral economics has, instead, shown that these errors might cumulate and snowball in one direction and lead to outcomes that are very different from those predicted by traditional economics. Behavioral economics has also stressed the importance of emotions in making decisions, role of framing in influencing outcomes, provided richer theories of decision making under risk and over time and provided new solution concepts in strategic interactions. In this sense it has provided new theory that is both refutable, and based on empirically robust psychological principles to explain the emerging evidence.

12. Further Reading

1. Bowles, S. (2006) Microeconomics: Behavior, Institutions, and Evolution. Princeton: Princeton University Press.

2. Boyd, R, Bowles, S. Fehr, E. and Gintis, H. (2005) Moral Sentiments and Material Interests, MIT Press, Cambridge Massachusetts.

3. Camerer, C. (2003) Behavioral Game Theory: Experiments on Strategic Interaction, Princeton: Princeton University Press.

4. Camerer, C., Loewenstein, G. & Rabin, M. (Eds.). (2003) Advances in Behavioral Economics, New York and Princeton: Russell Sage Foundation and Princeton University Press.

5. Fehr, E. and Gächter, S. (2000) Fairness and Retaliation – The Economics of Reciprocity, Journal of Economic Perspectives 14: 159-181.

6. Frederick, S., Loewenstein, G. & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. Journal of Economic Literature. 40: 351-401.

7. Glimcher, P. Camerer, C. Fehr E. and Poldrack, R. (2008) Handbook of Neuroeconomics, Elsevier, North Holland.

8. Kahneman, D., and Tversky, A. (Eds.). (2000). Choices, values and frames. Cambridge University Press: New York.

9. Kahneman D., Tversky A., 1979. Prospect theory : An analysis of decision under risk. Econometrica. 47: 263-291.

10. Loewentein, G. (2007). Exotic Preferences: Behavioural Economics and Human Motivation. Oxford, England: Oxford University Press.

11. Loewenstein, G., Read, D. & Baumeister, R. (Eds.) (2003). Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice. New York: Russell Sage Foundation Press.

12. Thaler, R. H. (2005) Advances in Behavioral Finance II, Russell Sage Foundation, New York.

13. Thaler, R. H. (1993) The Winner's Curse: Paradoxes and Anomalies of Economic Life, Princeton University Press.

14. Thaler, Richard H., Sunstein, Cass R, Nudge: Improving Decisions on Health, Wealth, and Happiness. New Haven, CT: Yale University Press, 2008.

15. Vohs, K. D. Baumeister, R. F., & Loewenstein, G. (2007). Do emotions help or hurt decision making? A Hedgefoxian perspective. New York: Russell Sage Foundation Press.