Rationality

Lawrence E. Blume[†‡] and David Easley[†]

June 2007

Cross references: EXPECTED UTILITY HYPOTHESIS, METHODOLOGICAL INDIVIDUALISM, SAVAGE'S SUBJECTIVE EXPECTED UTILITY MODEL, UNCERTAINTY, UTILITARIANISM AND ECONOMIC THEORY, UTILITY

[†] CORNELL UNIVERSITY.
[‡] THE SANTA FE INSTITUTE.

> I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, . . .
>
> Justice Potter Stewart, 378 U.S. 184, 197.

Rationality is for economists as pornography was to the U.S. Supreme Court, undefinable but nonetheless easily identified; and yet, like the Justices of the Court, no two economists share a common definition. This entry details some of the common meanings of individual (as opposed to social) rationality and discusses their uses. Our point of view is that of working economists rather than that of psychologists. Economics is committed to METHODOLOGICAL INDIVIDUALISM, the claim that social phenomena must be explained in terms of individual actions which in turn must be explained through individuals' motivations. This commitment requires a theory of human action. The *rationality principle*, that individuals act in their best interest as they perceive it, provides such a theory. In this entry we evaluate the rationality hypothesis and its alternatives from the perspective of how they explain social phenomena such as the behavior of a market. Our interest is in social life rather than in the psychology of an individual.

History and Description

The use of the rationality principle in economics certainly predates the utilitarianism with which it is so often conflated. Adam Smith ([1789] 1976, p. 19) describes, in his discussion of the division of labor, a tribe of hunters in which one person is particularly deft at making bows and arrows. 'He frequently exchanges them for cattle or for venison with his companions; and he finds at last that he can in this manner get more cattle and venison, than if he himself went to the field to catch them. From a regard to his own interest, therefore, the making of bows and arrows grows to be his chief business, . . .' Moving from intuition to analysis, however, requires a sharp understanding of what it means to regard one's own interest, and this has become a source of endless debate for rational-actor social scientists.

The utility-maximization version of rationality springs from the utilitarianism of Bentham and Mill. According to Bentham (1789, p. i.), 'Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne.' Although many thinkers toyed with utilitarian approaches to economic analysis, it was not until the 1870's, through the work of Jevons,

Menger, and Walras, that utility maximization began to assume the important role in economic analysis it has since held.For this trio, utility was a shortcut to a theory of value. Perhaps this is why they were not overly concerned with the issues of measurable utility and the possibility of interpersonal utility comparisons which so exercised their successors. Utility for Bentham, on the other hand, was a physical measure of pain and pleasure which could be computed according to his 'felicific calculus'. Although utility as a merely hedonic measure was rejected even by Mill, only in the 1930's, and after a half century's work beginning with Fisher (1892) and Pareto (1895) was it generally recognized that properties of demand derived from the shape of indifference curves, and so utility could admit a purely ordinal interpretation. This 'shift in emphasis away from the physiological and psychological hedonistic, introspective aspects of utility', as Samuelson (1947, p. 90-1) put it, led to the 'purging out of objectionable, and sometimes unnecessary, connotations . . . of the Bentham . . . variety.' The ultimate expression of this apsychological view is the theory of revealed preference, whose purpose is '. . . to develop the theory of consumer's behavior freed from any vestigial traces of the utility concept.' (Samuelson 1938a, p. 71.) The result is a mathematical structure that Edgeworth would have understood, interpreted in a manner completely foreign to his way of thinking.

Expected utility in the theory of choice under uncertainty is older than Benthamite utilitarianism. Both an expectation argument and a dominance (admissibility) argument for the existence of God were carefully laid out by Pascal ([1672] 1958, p. 233). These remarkable few paragraphs touch on many important issues in contemporary decision theory, including the principle of insufficient reason, the problem of infinite utility payoffs, and incomplete preferences: 'Yes; but you must wager. It is not optional. You are embarked. Which will you choose then?' Even the concept of marginal utility predates Bentham, in Gabriel Cramer's and Daniel Bernoulli's famous near-resolutions of the St. Petersburg paradox. But despite this early progress, the formalization of the modern theory of choice under uncertainty begins only with Wald (1939), who at one go describes the key structures of statistical decision theory: loss functions, *a priori* distributions, and Bayes, admissible, and minimax decision rules. Interest quickly coalesced, however, around the expected utility models described in the two great testaments of decision science, von Neumann and Morgenstern (1947) and Savage ([1956] 1972). Expected utility quickly became such a dominant paradigm for choice under uncertainty that research into alternatives was a backwater for twenty years. But criticisms of the expected utility models emerged almost before the ink was dry on the two manuscripts, in Allais' (1953) experiments and Cyert, Simon and Trow's (1956) empirical studies of firm behavior, and by the late 1970's behavioral economics and non-EU decision theory were active areas of research.

Psychological utilitarianism and decision theory are the two traditions which most inform the modern economist's thinking about 'rationality', and yet despite the long intellectual history of these ideas, no single vision of what it means to be a 'rational actor' has emerged. In the remainder of this entry we single out several sources of confusion and disagreement. We will discuss five models of

rationality.

General choice theory (GCT): A set A of alternatives is given, along with a family $\mathcal{B}$ of non-empty subsets of A. The set A is the set of possible alternatives and any element B in $\mathcal{B}$ is a set of feasible alternatives, a set from which the decisionmaker must choose. A *choice function* C assigns to each $B \in \mathcal{B}$ a nonempty subset of B, the objects chosen by the decisionmaker from the feasible set. In the theory of demand, for instance, A is the consumption set, $\mathcal{B}$ is the set of possible budget sets and the choice function is the demand function. A textbook treatment of the rational decisionmaker requires that she have *preference relation* $\succeq$ on A, and we understand $a \succeq b$ to mean that she finds a to be 'at least as good as' b. By 'preference relation' we mean a binary relation which is complete, all alternatives can be compared, and transitive. Transitivity means that if a is at least as good as b and b is at least as good as c, then a is at least as good as c. Then b is chosen from B, that is, $b \in C(B)$, if and only if $b \succeq a$ for all $a \in B$. Preference is the primitive expression of rationality. The role of utility is to provide a covenient *representation* of preference. A utility function u is a real-valued function on A, and u *represents* $\succeq$ if $u(a) \geq u(b)$ if and only if $a \succeq b$. While the decision theory toolkit of the working economist mostly specializes this model, much contemporary economic theory does not require this much of rationality. In particular, the completeness and transitivity assumptions can be done away with in general equilibrium theory, and utility theory can be developed without completeness. See, for instance, Aumann (1962), Chipman, Hurwicz, Richter, and Sonnenschein (1971), and Gale and Mas-Collel (1975).

Expected utility theory (EU): Expected utility is a specialization of GCT in which the set A and the preference relation have a specific structure. In EU theory, X is a finite set of prizes or outcomes, and the alternative set A is the set of all probability distributions on X. Preferences have the following representation: A *payoff function* v is a real-valued function on X, and any two probability distributions p and q in A are compared according to their expected values of v; that is $p \succeq q$ iff $E_p v \geq E_q v$. The content of this theory is that, geomerically speaking, indifference curves are parallel straight lines (hyperplanes). The first characterization of EU preferences was provided by von Neumann and Morgenstern (1947); today's standard axiomatic characterization of EU preference orders is due to Herstein and Milnor (1953).

Subjective expected utility theory (SEU): When we choose whether to play roulette or a slot machine, we are choosing among probability distributions. When we bet on the outcome of a horse or political race, we are betting on the realization of uncertain outcomes, but not objects to which probabilities are necessarily attached. Savage's ([1956] 1972) contribution was to provide a theory of what he called 'personal probability', a specialisation of GCT, here interpreted as a decisionmaker's degree of belief in the occurrence of some event. He characterized those preference relations which could be represented by the expectation of some payoff function with respect to a personal probability. In

Savage's subjective expected utility (SEU) theory, S is a set of states, such as the possible outcomes of the election. There is also a set X of outcomes. A bet on the election is a function which assigns an outcome to every state. Savage called such functions $f : S \to X$ *acts*, and the set of acts is the alternative set A. A preference relation $\succeq$ on the set A has an SEU representation if there is a payoff function v on outcomes X and a probability distribution p on states S such that $f \succeq g$ if and only if $E_p\{v(f(s))\} \geq E_p\{v(g(s))\}$.

Methodological individualism requires the analysis of social phenomena to be 'bottom-up', that is, to begin with individuals. It is a stronger statement to claim, however, that the description of the individual is entirely pre-social; that in economic models, for instance, that individuals come to the market with preferences and beliefs already formed. Most modern economists does not make this claim, and instead work with models in which the description of the individual is an equilibrium outcome. The two most prominent examples of this method are rational expectations equilibrium and non-cooperative game theory.

Rational Expectations (REE): The rational expectations hypothesis supposes a population of individuals solving decision problems which have a common state space, and furthermore that the state will be chosen according to the 'true distribution' μ, which is determined by the individuals' choices. The payoff $v(c,s)$ to a choice c depends on the state realization s, and preferences over choices are EU: $U_i(c) = E_\mu v\{c,s\}$. The hypothesis asserts that all beliefs will be correct; that is, that all SEU decisionmakers have preference representations in which the beliefs are in fact the probability distribution μ, and μ in turn is the distribution of states which is determined by their actions. Rational expections is a misuse of the adjective. Unfortuately it is probably too late to abandon the term. There is no connection between the rationality principle, which claims that individuals act in their perceived best interest, and the rational expectations hypothesis, which claims that those perceptions meet some ex ante standard of correctness. But so labelling a theory is certainly a nice rhetorical move for how it structures subsequent debate.

Non-cooperative Game Theory (NGT): A population of individuals chooses actions. Individual i's payoff to action c_i, $v(c_i, c_{-i})$ depends upon the choices c_{-i} of the others. He holds probabilistic beliefs about the actions of others, and evaluates a choice according to EU. The social construction of the individual is seen in the determination of beliefs. *Dominant strategies* are those which can be rationalized by *all* choices of beliefs. *Rationalizable strategies* are those which can be justified by some beliefs satisfying a belief restriction, that it be *common knowledge* that all members of the population are EU-rational with some beliefs. (See EPISTEMIC GAME THEORY: AN OVERVIEW.) Nash equilibrium requires, like REE, that everyones beliefs are correct. Various Nash equilibrium refinements also have belief interpretations. (See REFINEMENTS OF NASH EQUILIBRIUM.)

Rationality and Mind

The merits of the rational choice foundation of economics have been much discussed, both by its practitioners and by its critics. This discussion is often confused, in part because economists are not consistent in how they understand the contents of the rationality hypothesis. Economic theory holds two views of rationality. One is that rationality is consistency of choice, that the tools of choice theory are just an alternative encoding of certain choice functions; the other is that rationality is a theory of intentional behavior, in which beliefs and desires are meaningful constructs.

Revealed preference theory is the sharpest formulation of the consistency view. It takes demand as primitive and asks if it is *consistent* with the maximization of a preference order. It recovers desires from choice, and only to the extent that choices are different can two desires, preference orders, be distinguished. This view permeates the foundations of decision theory. For Savage ([1956] 1972, p. 17), 'It is possible that the person **prefers** $\mathbf{f}$ to $\mathbf{g}$. Loosely speaking, this means that, if he were required to decide between $\mathbf{f}$ and $\mathbf{g}$, no other acts being available, he would decide on $\mathbf{f}$.' In this account, preference is defined by choice. This means specifically that if the choice function C on a collection $\mathcal{B}$ of choice sets satisfies certain conditions, then there is a complete and transitive binary relation such that for every $B \in \mathcal{B}$, $C(B)$ contains exactly the elements of B which are maximal in B with respect to the relation. The binary relation is nothing more than an alternative description of C on $\mathcal{B}$. Suppose a new choice set $B' \notin \mathcal{B}$ is considered. What can we guess about the contents of $C(B)$? Knowing that the decisionmaker is consistent on $\mathcal{B}$ allows the observer to infer nothing at all about $C(B')$.

If revealed preference represents at all a psychology of choice, that psychology is a form of *radical behaviorism*. Radical behaviorism asserts that two mental states are distinguishable only to the extent that some observable behavior distinguishes them. Behaviors are all that one can theorize about. Samuelson writes, 'of a steady tendency toward the removal of moral, utilitarian, welfare connotations . . . ' and of 'the rejection of hedonistic, introspective, psychological elements.' (Samuelson 1938b). Although the behaviorist position seems extreme, the leading graduate microeconomics textbook writes approvingly of reveal preference, 'Perhaps most importantly, it makes clear that the theory of individual decision making need not be based on a process of introspection but can be given an entirely behavioral foundation.' (Mas-Colell, Whinston, and Green 1995, p. 5) Consistency is often justified as discipline. It requires a minimum of assumptions about the beliefs and desires of individuals, and minimizes the possibility of researchers' values and beliefs slipping unbidden into their analyses. It allows the data maximal scope to speak for itself.

Although received economics talks approvingly of rationality as mere consistency, this is not in fact what most economists do. Much of economics involves invisible hand explanations; aggregate

market behavior emerges from the decisions of many agents. Whether the invisible hand lifts the cup aloft or knocks it over, economic explanation entails explaining how it coordinates for good or ill the motives and interests of diverse individual actors. These kinds of questions call for explanations based on the motivations of economic actors, which purely behavioralist explanations cannot provide. So economists in practice take an intentional view.

The intentional view holds that rational choice theory is a commonsense or 'folk' (as opposed to 'scientific') psychology. Just as in our everyday transactions we use the language of beliefs and desires to interpret and forecast the behavior of others, so do economists interpret choice behavior. The investor *believes* that the asset price will be higher tomorrow. She *wants* greater wealth tomorrow. So she *acts* by purchasing the asset. In this view belief and desire are in fact mental states that are connected to action. The folk psychology is a theory of mind which is presumed by economists to be both adequate for a descriptive psychology of decision and accurate enough in its predictions of individual behaviors for the uses to which it is put. Although utility does not exist as a psychophysical quantity, rational choice models provide a representation of the mental states involved in judgment and decision. (The stronger claim that mental process is a more or less efficient utility maximization algorithm is a view held only by the strawman regularly beaten up by rationality's critics.)

The economist's folk psychology goes further than everyday folk psychology by specifying analytic representations of beliefs, desires, and how they interact. No matter what representation is ultimately chosen by the textbook economist, his folk psychology rests on two points. (1) Rationality is instrumental. Its concern is the efficient pursuing of ends by available means; not the sensibility of the ends. (2) Desire is not anchored by any other aspect of the decision problem; either the feasible set nor the context of choice. Formally, desires are captured by a preference ordering on possible objects of choice whose existence is independent of the feasible set or the context of choice. This is the content of GCT.

The tension between the demands for a parsimonious behavioral theory and the need for an intentional theory of choice is often resolved by holding that, of course beliefs and desires exist, but we economists have access to them only as they are revealed in observed choice behavior. In a recent critique of neuro-economics, two well-known theorists write, 'In standard economics, the testable implications of a theory are its content; once they are identified, the non-choice evidence that motivated a novel theory becomes irrelevant.' (Gul and Pesendorfer 2005, p. 6.) This view has a long history, perhaps with origins in the defense of marginal analysis against its early critics. Machlup (1946, p. 537) writes, 'Psychologists will readily confirm that statements by interviewed individuals about the motives and reasons for their actions are unreliable or at least incomplete,' and also raises the oft-heard incentive problem of eliciting survey data, that survey respondents may choose answers to meet their own goals, which may not include truth.

One source of confusion in evaluating claims for and againt the economist's psychology is the that the theory has both positive and normative components. According to Marshak (1950, p. 111), 'The theory of rational behavior is a set of propositions that can be regarded either as idealized approximations to the actual behavior of men or as recommendations to be followed.' Savage's early work with Milton Friedman (1948, 1952) was explicitly descriptive, but Savage ([1956] 1972) is just as explicitly normative. It is not surprising that a description of decision in terms of beliefs and desires should have a normative component which evaluates how well goals are achieved. Confusion arises, however, when the descriptive and prescriptive positions are inappropriately conflated to justify the rationality assumptions as a statement of fact. Many undergraduate microeconomics texts justify transitivity assumptions by a money pump argument as a prelude to demand theory. The Dutch book is used to defend probabilistic descriptions of belief. But both of these arguments are, at their source, explicitly normative. (See Davidson, McKinsey, and Suppes (1955, p. 146) and Ramsey (1931).)

A descriptive theory of choice which is grounded not in empirical reality but in logical deductions from normative principles, like Dutch books and money pumps, is not science, but metaphysics. Furthermore, normative justifications are implicitly introspective. A money pump argument really says, 'you wouldn't fall into this trap, would you?' Signficant empirical work in psychology (Nisbett and Wilson 1977), however, indicates that introspective evidence is simply unreliable. When individuals turn to review and justify their decisions, they may have no access to the mental states which guided their choice. On the other hand, it seems to us quite reasonable to build models of financial asset pricing which assume that traders are probabilistically sophisticated, on the supposition that traders who are not will either not long survive in the market or not, as a group, be large enough to have a significant effect on prices. Financial markets, unlike Dutch books, actually exist, and the claim that individuals with probabilistically incoherent beliefs do not fare well is a claim of fact, to be tested against market data.

The conflation of positive and normal concerns in decision theory is more fundamental than simple carelessness in an argument. In his criticism of the fact/value dichotomy, Putnam (2002) asks us to consider the word 'cruel'. He observes that the word often has both descriptive and normative content, and in most uses they cannot be separated. The same could be said of the adjective 'rational'. Marshak (1950, p. 111) illustrates this perfectly when he writes that the purpose of EU is, '. . . to describe the behavior of men who, it is believed, cannot be "all fools all the time,". . . .' When the word 'rational' is used to describe a system in which all agents hold accurate probabilistic beliefs, the implication is that someone holding inaccurate beliefs gets it wrong. REE is often informally defended by the assertion that if an economic actor's beliefs were incorrect, he would observe this and form new ones. The assertion is either a positive assertion, that actors do indeed have such beliefs, or a normative assertion, that they should hold such beliefs. The normative assertion is a metaphysical defense of the validity of the rational expectations hypothesis. The positive assertion is a claim of fact

whose validitiy could in principle be put to test, but testing the claim would in fact require so rich a set of ancillary maintained hypotheses that practically it is infeasible.

Given all the problems of the two views of rationality, one might wonder why economics needs a rational actor. Dennett (1971, p. 92) provides perhaps the best defense of belief/desire explanations. He contrasts what he calls the *design stance*, predicting behavior from an understanding of how an agent is designed, or built, with the *intentional stance*, attributing to the agent beliefs and desires, and predicting from them. The intentional stance is useful, he writes, 'Whenever we have reason to suppose the assumption of optimal design is warranted, and doubt the practicality of prediction from the design . . . stance.' Warranting the optimal design assumption means for Dennett not that the system actually be designed to achieve a fixed set of goals, but that this assumption is a useful first approximation. 'Not surprisingly', he observes, 'as we discover more and more imperfections . . . , our efforts at intentional prediction become more and more cumbersome and undecidable, for we can no longer count on the beliefs, desires, and actions going together that ought to go together. Eventually we end up, following this process, by predicting from the design stance; we end up, that is, dropping the assumption of rationality.' (p. 95.) This movement, from rationality to realism, is the motivation for taking behavior more seriously.

Rationality and Behaviors

Game theory and general equilibrium theory are 'system frameworks'. They imagine a collection of individual agents interacting in some systematic way, strategically in game theory, as described by the normal or extensive form of the game, and through markets in general equilibrium theory. In each case, the model produces an 'equilibrium' of the system. The first stage in the development of a system framework involves determining its consistency and internal coherency. That is, conditions which guarantee the existence of equilibrium. This analysis will be as abstract and general as possible, to encompass as large a repertory of behaviors as possible. The second stage is the application of the framework to derive useful statements about the world. This requires explicit behavioral assumptions about agent behavior and describing the resulting equilibrium. These statements — predictions about market or game behavior — can be examined empirically.

There are two difficulties with the received models of decision theory such as expected utility and dynamic programming in this kind of research program: First, as these models are formulated, behaviors are not accessible. For example, using expected utility to derive home bias in financial asset markets, that investors tend not to take positions in foreign assets, requires complicated assumptions about traders beliefs. Second, these models are insufficiently rich to capture all the behaviors

one might want to examine. For instance,the additively separable intertemporal expected utility model conflates time preference and risk aversion because the model is too thinly parametrized.

Behavioral economics is a research program which will, its proponents argue, replace rational actor models with a more psychologically informed view of human decision making. Much of behavioral economics, however, is less ambitious (and thus, perhaps, more useful). This work can be described as reformulating or extending rational actor models so as to make those observable behaviors whose implications we wish to examine more accessible. While much of this work is at the core of behavioral economics, many who do this work eschew the label; not only behavioral economists are interested in behavior. Here we discuss four categories of research which cover much work on behaviors, both by behavioral economics and by its critics.

Recontextualizing decision: GCT is a very parsimonious simplification of a decision problem. In modelling there is a tradeoff between behavioral accuracy and parsimony in the description of decision problems. In general equilibrium models, for example, behavioral accuracy may improve descriptive and explanatory power, but parsimony is required because individual decisions are only one piece of the analysis, and complicated models of individual behavior may generate only intractable market models.

One implication of GCT is that preferences are not choice-set dependent. Even in the early days of decision theory, important models such as minimax regret (Savage 1951) violated the requirement of a single preference order on a universal space of potential choices. Furthermore, many choice-set effects appear to be perfectly rational. Consider the behavior of a well-mannered but very hungry person at a dinner party. A plate is passed to him with three pieces of the main course, ordered in size such that $a < b < c$. Being both well-mannered and hungry, he chooses the second largest piece, b. Suppose now that the plate had been passed around the table in the other direction, so that when it comes to him there remains only a and b. Now according to his rule he chooses a. Is he called irrational by the GCT theorists at the table?

Kahneman and Tversky's (1979) prospect theory illustrates another way in which decision problems can be recontextualized. Here additional context, a *status quo*, is added to the description of the decision problems. Gambles are viewed as probability distributions over gains and losses relative to the status quo. Given a status quo, a preference order over all possible final outcomes exists, but that preference order varies with the status quo. There is, however, a stable preference order over the universe of all possible gains and losses; more context is added by redefining the objects of choice. A similar transformation is accomplished in Gul and Pesendorfer's (2004) model of choice with self-control problems. In the conventional infinite-horizon optimal consumption problem, the objects of choice are consumption paths. Gul and Pesendorfer, on the other hand, take the objects of choice to be pair consisting of a current period consumption and a decision problem to be

solved tomorrow. Gul and Pesendorfer's model is an example of a *menu choice model*. Although used somewhat earlier, the first formal development of such models was by Kreps (1979) to describe preferences for flexibility.

Constructing rationality: The economist's conventional view of market interaction posits a collection of individuals with well-formed preferences meeting in a market place. The preferences, along with endowments and technologies, are exogenous to the system. On the other hand, some attendees at a large outdoor concert are there because they like the music, while others are there because of the crowd. Teenagers' evaluation of clothing style has perhaps as much to do with who wears such clothes as with their cut and pattern. These are all examples of socially constructed preferences.

Socially constructed preferences are a part of conventional economic theory. Both NE and REE are models of socially constructed preferences. In each case desires are fixed, but beliefs adjust. However neither of these equilibrium concepts are particularly well-supported by belief adjustment (learning) processes. The literature on learning Nash equilibrium is huge, and the state of the art is that while one can construct learning dynamics that will find a Nash equilibrium, many intuitive learning processes will often fail. Blume and Easley (1982) show that rational equilibrium can easily fail to be reached by any reasonable learning process.

Restricting the socially determined component of preferences only to beliefs is an artificial constraint, and to limit social influence on preference formation to learning is to miss most of the interplay between the individual and the group. Manski (2000) observes that the implications of social interactions through learning and through tastes are distinct, and the difference is significant for policy analysis. Any theory of the interaction of desires requires a new set of primitives which describe the preference formation mechanism. One popular approach has been to model the evolution and workings of pro-social norms of cooperation and trust. Bowles (1998) is an engaging survey of this work. Much less has been done on the evolution and workings of anti-social norms, such as discrimination and stigmatization. Others have turned to biological metaphors. Here one might look at the population dynamics of rules or preferences on a game form or market where game or market outcomes (not utilities) determine the composition in the next round of the population's decision rules or preference orders (Güth (1998), Blume and Easley (1992, 2006)). Pro-social behavior such as reciprocity and altruism has also been investigated from the biological standpoint (Bergstrom (2002), Sethi and Somanathan (2003)). One lesson of this literature is that the nature of the interaction between agents is at least as important as the model of choice in determining system outcomes. About the embeddedness of economic action in social life, Granovetter (1985, p. 506) writes, 'The notion that rational choice is derailed by social influences has long discouraged detailed sociological analysis of economic life and led revisionist economists to reform economic theory by focusing on its naive psychology. My claim here is that however naive that psychology may be, this is not where the

main difficulty lies—it is rather in the neglect of social structure.'

The content of preferences and beliefs: It has been conventional in economic analysis to construe self-interest very narrowly. No 'other-regarding' values are expressed in preferences, and conventionally to do otherwise is frowned upon. For instance, it is hard to explain why an individual votes in an election by her effect on the outcome, without referring to the psychic rewards of the act of voting. Yet the claim that people vote because of norms of citizenship and the like is often regarded as 'nearly tautological'. (Ordeshook 1986, p. 50.) On the other hand, critics of economic man often incorrectly assert that rational actors are excessively self-interested; incorrectly, because the existence of preferences and the content of preferences are distinct issues. The rationality hypothesis does not preclude other-regarding desires. Interest in those externalities that arise from ethical concerns, social norms, and other social constructions, has increased enormously in the last decade. Much of the literature on social interactions is a study of the consequences of other-regarding preferences. Not surprisingly, other-regarding preferences usefully model both altruism and racism. This is not a fix for those critics who see the selfishness of traditional neoclassical models as a moral failing rather than a behavioral one.

A distinct problem which, unfortunately, has not been much addressed by behavioral economics, is the use of individual preferences in the economists version of moral philosophy. The same preferences which are revealed through shopping behavior at the grocery store are supposed to be informative for the ethical questions posed by welfare economics. One could, in fact, distinguish 'ethical preferences' from 'subjective preferences' as Harsanyi (1955) has done, and it would be interesting to know if social psychology has anything to say about the relationship between the two types of decision problems, individual and social, which economists address.

Different psychologies: Some economists look to replace the folk-psychology of beliefs and wants with something altogether different. Neuroeconomics is one such attempt, although the neuroeconomics literature seems to eschew drawing economic conclusions from imaging data. Unfortunately, the link between brain and mind is illusive. *Eliminative materialism* is a position taken by some cognitive scientists, which claims that beliefs and desires do not exist as mental states, and will have no place in an accurate account of the mind. Theoretical and methodological arguments in its support can be found, for instance, in Churchland (1981). An economics which takes its microfoundations entirely from cognitive science could look extremely different than the economics of today. But even if one is more hopeful than the Churchlands for the utility of the economist's folk-psychology, the goal is far off. As one leading neuroimaging specialist puts it, 'Despite fantastic technical developments, lingering methodological and conceptual limitations hinder progress in understanding how mental processes (wrapped up in folk psychology) reduce to or emerge from neural processes (Schall 2004, p. 44).' Savoy (2001, p. 36) has a bleaker view: 'Do the new discoveries about human brain function

based on neuroimaging experiments really teach us things that are relevant for the study and understanding of behaviour? That is a question which you must answer. My own impression is that, at present, the overwhelming thrust of these data are toward understanding brain organisation, rather than human behaviour. Of course, we assume that when brain organisation is sufficiently well understood, it will lead to increases in our understanding of behaviour. But I do not think, as yet, there is a great deal of progress in that direction.'

Neuroeconomics hopes to replace the belief/desire folk psychology that informs most of modern analytical economics with a more accurate scientific psychology. Alternatively, one could construct a different folk-psychology which, like utility maximization, has no scientific pretensions, but is more descriptively accurate. Models of *intrapersonal* conflict are the most familiar example of this kind of framework. Strotz (1955) demonstrated the possibility of time-inconsistent planning in intertemporal utility maximization problems, and Pollak (1968) subsequently displayed the essentially strategic nature of the planning problem as a problem of competition between the selves choosing at different dates. Schelling (1984) described a variety of decision problems with aspects of intrapersonal conflicts, and discussed them from a game-theoretic perspective. Schelling, for instance, wrote about the competition between that part of him which desires nicotine and that part which wants to give it up. There is a contest for self-control. The literature today contains a number of intertemporal models which, following Pollak (1968) distinguish two kinds of behavior: Sophisticated behavior chooses today with full knowledge that her future selves may try to undo her decision. A choice is a subgame perfect equilibrium of a game played by all her selves. Naive behavior chooses today assuming, perhaps incorrectly, that her future selves will stick with her decisions. These two models are intrinsically no more realistic than GCT, just different.

For the working economist, the ultimate test of a psychologically more accurate theory of individual choice is how they perform in explaining market and other social outcomes rather than how well they predict the behavior of an individual. How could theory A, more informed with insights from psychology and cognitive science, possibly be less useful for economists? Here are three possibilities: (1) Theory A might be extremely complex. Its application to a heterogeneous-agent financial market model, for instance, is simply impossible to work with, and no conclusions can be derived. (2) Theory A might require for its application data that we can observe in a controlled and heavily instrumented setting but simply cannot collect in the field. (3) Theory A may not be posed with concepts which are useful for the economists interpretation of social outcomes. For instance, theory A might be couched in terms of chemical states of the brain, and not speak at all about agents intentions, beliefs or desires. While it may be possible to construct a biochemical model of the invisible hand, it would not be useful for welfare economics.

Evidence on the question of whether these models lead to better market analyses is sparse,

and mixed, and there is no evidence on how these models perform relative to menu choice models, which address the same questions from a rational choice perspective. More generally, more work needs to be done in the evaluating behavioral models with respect to their economic performance. How useful are they for deriving implications about the performance of aggregate economic variables such as prices. This kind of research is already underway. Two examples are Kocherlakota (2001) and Laibson (1997).

An instance of point (3) can be seen in the time-inconsistency literature. Pollak (1968) and models following him (O'Donoghue and Rabin 1999) see choice not as the expression of a single desire, but as the outcome of conflict, perhaps inefficient and destructive, between competing desires. Now the Pareto ranking of alternatives in a social interaction either becomes dependent on which of the many competing preference orders we modelers choose for each individual or it becomes empty if we try to respect them all. The advantage of menu choice models, the rational choice alternative, is that there is a well-defined notion of preference for each agent, from which a Pareto ranking can be constructed. To be fair, rational choice modeling also poses problems for welfare economics. If individuals make consistent errors in a class of choice problems, what can revealed preference say about intentions? In the presence of systematic error, a welfare economics built from revealed choice is at best misleading.

Conclusion

The purpose of decision models in economics is to explain the behavior not of a single individual but of aggregates of individuals. Sometimes economists explain by appeal to 'Laws', such as 'the Law of Supply and Demand'. But this mode of explanation is mostly an intermediate product; useful, perhaps, for generating back-of-the-envelope predictions about the effects of a tax on market price, but not a source of understanding. There are few natural laws in the social sciences, and the domains of the few we can identify are very limited.

More often, economists appeal to 'mechanism'. We try to understand economic phenomena, such as the determination of prices in different kinds of markets, in terms of the mechanisms which generate them. Given our commitment to methodological individualism, this requires an explanation of how individual economic actors interact with one another. This is where rational actor theories are employed, and it is with respect to how these models do in this discussion rather than how they do in other domains, such as explaining individual behavior, that the rationality principle should be evaluated.

Unfortunately, perhaps, at this point there are no serious alternatives to the rationality principle. For all of its buzz, proponents of *bounded rationality*, by which we mean models of behavior that consider beliefs and desires but that do not optimize, have so far failed to deliver decision models which are robust and not tightly tied to a small class of decision problems.

It is perhaps too early in its intellectual history to ask for as much from *cognitive models*. We are skeptical about the value for social and economic systems analysis of unpacking the black box of consumer behavior by deploying a rich and sophisticated model of cognitive process within a general equilibrium or game theoretic model. There is a point to reductionism. On the other hand, we are enthusiastic about the possibility that cognitive science will contribute to sharpening the rationality principle. The focus of much modern decision theory, such as Kahneman and Tversky (1979), Gilboa and Schmeidler (1989) and Gul and Pesendorfer (2004), has been to make the black-box model better by looking for formulations of rational choice models that better conform to the data. A better understanding of decision mechanism will doubtless suggest constraints on black box behavior which can be captured in reduced-form decision models, and perhaps it will uncover constraints that cannot be observed from behavior alone.

Evolutionary models have also been proposed as an alternative framework to rational choice decision-making. Market forces, or a combination of markets and biology, favor some decision rules over others. In the long run, the market will be populated mostly by those decision rules that are 'most fit', rational or not. One can indeed ask if the forces of market selection favor rational decision rules (Blume and Easley 1992, Sandroni 2000), but the study of market population dynamics is complementary to rather than a substitute for rational choice models. Blume and Easley (2006), for instance, demonstrate how market forces select within the class of rational decision rules, favoring some kinds of preferences and beliefs over others.

Although there appear no be no serious alternatives to the rational choice paradigm on the near horizon, there is much to regret in how the rationality principle is discussed. The following statements should be self-evident, but clearly are not, judging by our reading of the literature: (1) Rationality does not mean complete or symmetric information. In fact, much of rational actor social science attempts to understand social outcomes when these conditions do not obtain. (2) Rationality does not require individuals to be entirely selfish. While much effort has been made to understand social norms from the point of view of entirely individualistic preferences, the insistance on relying on self-regarding rather than pro-social preferences is a matter of style, rather than an axiom of rationality. (3) Rationality does not mean expected utility. Expected utility is one small class of decision models for choice under uncertainty. Its dominance in application was understandable 30 years ago when few alternatives were on the table. Since then decision theorists have been creative in developing better-behaved alternatives, and equilibrium and game theorists have been clever in applying

them. (4) Rationality does not mean ‘rational expectations’. For a belief restriction to be a requirement of rationality, it must be clear that all those who are not ‘all fools all the time’ must have correct beliefs. No research into learning in economics suggests this is the case in any kind of complex environment.

There is also much to regret in how the rationality principle has been deployed in economic analysis. Given the explosion of decision theoretic research over the last three decades, it is surprising how little this research has affected market and game theoretic analysis. The norm still seems to be self-interested preferences, expected utility and rational expectations (or Nash equilibrium). At this point the question of whether contemporary decision models such as Choquet expected utility and cumulative prospect theory have anything new to say about, say, asset pricing, is open. The value to economists of new decision theories, rational choice or not, is not in how they perform in a laboratory but how they perform in the analysis of markets and other social systems. Too rarely have modern decision theories been exposed to this test.

Rational actor social science is a broader tent than both its supporters and its critics make it out to be. We expect the rational choice framework to be as dominant when the next edition of the New Palgrave Dictionary goes to press as it is today. But we also expect the set of decision-theoretic models deployed in the analysis of social systems will be quite different, and probably more diverse, than it is now.

References

ALLAIS, M. (1953): "Le comportement de l' homme rationel devant le risque: Critique del postulats et axiomes de l'Ecole Américaine," *Econometrica*, 21(4), 503–46.

AUMANN, R. J. (1962): "Utility Theory without the Completeness Axiom," *Econometrica*, 32, 445–462.

BENTHAM, J. (1789): *An Introduction to the Principle of Morals and Legislations*. Based on information from English Short Title Catalogue. Eighteenth Century Collections Online. Gale Group., Farmington Hills, MI.

BERGSTROM, T. C. (2002): "Evolution of Social Behavior: Individual and Group Selection," *Journal of Economic Perspectives*, 16(2), 67–88.

BLUME, L., AND D. EASLEY (1992): "Evolution and Market Behavior," *Journal of Economic Theory*, 58(1), 9–40.

BLUME, L. E., AND D. EASLEY (1982): "Learning to be Rational," *Journal of Economic Theory*, 26(2).

——— (2006): "If You're so Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets," *Econometrica*, 74(4), 929–966.

BOWLES, S. (1998): "Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions," *Journal of Economic Literature*, 36(1), 75–111.

CHIPMAN, J. S., L. HURWICZ, M. K. RICHTER, AND H. F. SONNENSCHEIN (1971): *Preferences, Utility, and Demand*. Harcourt Brace Jovanovich, New York.

CHURCHLAND, P. M. (1981): "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy*, 78(2), 67–90.

CYERT, R. M., H. A. SIMON, AND D. B. TROW (1956): "Observation of a Business Decision," *Journal of Business*, 29(4), 237–48.

DAVIDSON, D., J. C. C. MCKINSEY, AND P. SUPPES (1955): "Outlines of a Formal Theory of Value, I," *Philosophy of Science*, 22(2), 140–160.

DENNETT, D. C. (1971): "Intentional Systems," *Journal of Philosophy*, 68(4), 87–106.

FISHER, I. (1892): *Mathematical Investigations in the Theory of Value and Prices*. Connecticut Academy of Arts and Sciences, New Haven CN.

FRIEDMAN, M., AND L. J. SAVAGE (1948): "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56(4), 279–304.

——— (1952): "The Expected–Utility Hypothesis and the Measurability of Utility," *Journal of Political Economy*, 60(6), 463–474.

GALE, D., AND A. MAS-COLLEL (1975): "An Equilibrium Existence Theorem without Ordered Preference," *Journal of Mathematical Economics*, 2(1), 9–15.

GILBOA, I., AND D. SCHMEIDLER (1989): "Maximin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18(2), 141–153.

GRANOVETTER, M. (1985): "Economic Action and Social Structure: The Problem of Embeddedness," *American Journal of Sociology*, 91(3), 481–510.

GUL, F., AND W. PESENDORFER (2004): "Self-Control and the Theory of Consumption," *Econometrica*, 72(1), 119–58.

——— (2005): "The Case for Mindless Economics," Princeton University.

GÜTH, W., AND H. KLIEMT (1998): "The Indirect Evolutionary Approach: Bridging the Gap between Rationality and Adaptation," *Rationality and Society*, 10(3), 377–399.

HARSANYI, J. C. (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63(4), 309–21.

HERSTEIN, I., AND J. MILNOR (1953): "An Axiomatic Approach to Measurable Utility," *Econometrica*, 47(291–7).

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47(2), 263–291.

KOCHERLAKOTA, N. (2001): "Looking for Evidence of Time-Inconsistent Preferences in Asset Market Data," *Federal Reserve bank of Minneapolis Quarterly Review*, 25(3), 13–34.

KREPS, D. M. (1979): "A Representation Theorem for "Preference for Flexibility"," *Econometrica*, 47(3), 565–78.

LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 62, 443–77.

MACHLUP, F. (1946): "Marginal Analysis and Empirical Research," *American Economic Review*, 36(4), 519–554.

MANSKI, C. F. (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14(3), 115–36.

MARSHAK, J. (1950): "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*, 18(2), 111–141.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*. Oxford University Press, Oxford.

NISBETT, R. E., AND T. D. WILSON (1977): "Telling More than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, 83(3), 231–259.

O'DONOGHUE, T., AND M. RABIN (1999): "Doing It Now or Later," *American Economic Review*, 89(1), 103–24.

ORDESHOOK, P. C. (1986): *Game Theory and Political Theory: An Introduction*. Cambridge University Press, Cambridge.

PARETO, V. (1895): "Considerazioni sui Principi Fondamentali dell'Economia Politica Pura, part V," *Giornale degli Economisti*, ser. 2(5), 119–57.

PASCAL, B. ([1672] 1958): *Pascal's Pensées*. E. P. Dutton and Co., New York.

POLLAK, R. A. (1968): "Consistent Planning," *The Review of Economic Studies*, 35(2), 201–8.

PUTNAM, H. (2002): "The Entanglement of Fact and Value," in *The Collapse of the Fact/Value Dichottomy and Other Essays*, pp. 28–45. Harvard University Press, Cambridge MA.

RAMSEY, F. P. (1931): "Truth and Probability," in *The Foundations of Mathematics and other Logical Essays*, ed. by R. B. Braithwaite. K. Paul, Trench, Trubner and Co.

SAMUELSON, P. A. (1938a): "A Note on the Pure Theory of Consumer's Behavior," *Economica*, 5(17), 61–71.

——— (1938b): "The Empirical Implications of Utility Analysis," *Econometrica*, 6(4), 344–356.

——— (1947): *Foundations of Economic Analysis*. Harvard University Press, Cambridge MA.

SANDRONI, A. (2000): "Do markets Favor Agents Ablee to Make Accurate Predictions?," *Econometrica*, 68(6), 1303–42.

SAVAGE, L. J. (1951): "The Theory of Statistical Decision," *Journal of the American Statistical Association*, 46(253), 55–67.

——— ([1956] 1972): *The Foundations of Statistics*. Dover Publications, New York, 2nd edn.

SAVOY, R. L. (2001): "History and Future Directions of Human Brain Mapping and Functional Neuroimaging," *Acta Psychologica*, 107(1–3), 9–42.

SCHALL, J. D. (2004): "On Building a Bridge Between Brain and Behavior," *Annual Review of Psychology*, 55, 23–50.

SCHELLING, T. (1984): *Choice and Consequence: Perspectives of an Errant Economist*chap. The Intimate Contest for Self-Command, pp. 00–00. Harvard University Press.

SETHI, R., AND E. SOMANATHAN (2003): "Understanding Reciprocity," *Journal of Economic Behavior and Organization*, 50(1), 1–27.

SMITH, A. ([1789] 1976): *An Inquiry into the Nature and Causes of The Wealth of Nations*. University of Chicago Press, Chicago.

STROTZ, R. H. (1955): "Myopia and Inconsistency in dynamic Utility Maximization," *The Review of Economic Studies*, 23(3), 165–180.

VON NEUMANN, J., AND O. MORGENSTERN (1947): *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 2nd edn.

WALD, A. (1939): "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, 10(4), 299–326.