In Beliefs, Interactions, and Preferences in Decision Making, M. Machina and B. Munier (eds.), Kluwer Academic Publishers, 1999

ARBITRAGE, INCOMPLETE MODELS, AND OTHER PEOPLE'S BRAINS

ROBERT NAU
FUQUA SCHOOL OF BUSINESS
DUKE UNIVERSITY

1. Introduction

Over the last 50 years, the theory of rational choice has emerged as the dominant paradigm of quantitative research in the social and economic sciences. The idea that individuals make choices by rationally weighing values and uncertainties (or that they ought to, or at least act as if they do) is central to Bayesian methods of statistical inference and decision analysis; the theory of games of strategy; theories of competitive markets, industrial organization, and asset pricing; the theory of social choice; and a host of rational actor models in political science, sociology, law, philosophy, and management science.

Rational choice theory is founded on the principles of *methodological individualism* and *purposive action*. Methodological individualism means that social and economic phenomena are explained in terms of "a particular configuration of individuals, their dispositions, situations, beliefs, and physical resources and environment" rather than in terms of holistic or emergent properties of groups. (Watkins 1957; *c.f.* Ordeshook 1986, Coleman 1990, Arrow 1994) Purposive action means that those individuals have clear and consistent objectives and they employ reason to find the best means of achieving those objectives. They intend their behavior to cause effects that they desire, and their behavior does cause those effects for the reasons they intend. (Ordeshook 1986, Elster 1986) In most versions of the theory, the objectives of the individuals are expressed in terms of *preferences*: they choose what they most prefer from among the alternatives available, and the (only) role of social and economic institutions is to enable them to satisfy their preferences through exchange or strategic contests with other individuals

This paper sketches the outline of *arbitrage choice theory* (ACT), a new synthesis of rational choice that weakens the emphasis on methodological individualism and purposive action and departs from the traditional use of preference as a behavioral primitive, building on earlier work by Nau and McCardle

(1990, 1991) and Nau (1992ab, 1995abc); a more complete treatment appears in Nau (1998). The main axiom of rationality in this framework is the requirement of *no arbitrage*, and it leads to a strong unification of the theories of personal decisions, games of strategy, and competitive markets. It also yields a very different perspective on the purpose of social and economic interactions between agents, suggesting that group behavior is to some extent emergent and suprarational, not merely an aggregation or collision of latent individual interests.

2. The Standard Theory Of Rational Choice

Standard rational choice theory begins with the assumption that the infinitely detailed "grand world" in which real choices are made can be adequately approximated by a "small world" model with a manageable number of parameters. The formal theory is then expressed in terms of assumptions about the structure of the small-world environment, the modes of behavior that take place in that environment, and the conditions that behavior must satisfy in order to qualify as rational.

Environment: The environment is inhabited by one or more human *agents* (also called *actors* or *players*), and in the environment *events* may happen, under the control of the agents and/or nature. An event that is under an agent's control is an *alternative*, and an event that is under nature's control is a *state of nature*. (States of nature just *happen*, while alternatives are *chosen*.) A realization of events is called an *outcome*. For each agent there is a set of material or immaterial *consequences* (wealth, health, pleasure, pain, etc.) that she may enjoy or suffer, and there is a known mapping from outcomes to consequences. Every alternative for an agent is a feasible mapping of events not under that agent's control to consequences. An *act* for an agent is an *arbitrary* mapping of events not under that agent's control to consequences. Thus, an act is a (usually) hypothetical alternative, while an alternative is a feasible act. The set of acts is typically much richer than the set of alternatives for each agent.

Behavior: Within the environment, several kinds of behavior occur. First and most importantly, there is *physical* behavior, which affects the outcomes of events. Physical behavior by agents[1] consists of choices among feasible alternatives. However, it does not suffice to model only physical behavior, because the set of feasible alternatives usually is not rich enough to support a tight mathematical representation and because it is of interest to predict choices from other kinds of antecedent behavior. The other kind of behavior most often modeled in rational choice theory is *preference* behavior. Preferences are hypothetical choices between hypothetical alternatives (acts). An agent prefers act x to act y if she imagines that she would choose x rather than y if given the choice between them. (Actually, we are getting ahead of ourselves: at this point preference is merely an undefined

Table 1. Elements of standard rational choice theory

1. Environment: – Agents – Events (states of nature and alternatives for agents) – Consequences – Mappings of events to consequences for all agents – Acts (hypothetical mappings of events not under an agent's control to consequences)
2. Behavior: – Physical behavior: choices among alternatives – Mental behavior: preferences among acts
3. Rationality: (a) Individual rationality – Axioms of preference (completeness, transitivity, independence, separability of belief and value) – Axiom of rational choice (choices agree with preferences) (b) Strategic rationality – Common knowledge of individual rationality – Common knowledge of utilities – Common prior probabilities – Probabilistic independence (c) Competitive rationality – Price-taking – Market clearing (d) Social rationality – Pareto efficiency (e) Rational asset pricing – No arbitrage (f) Rational learning – Bayes' theorem (g) Rational expectations – Self-fulfilling beliefs

primitive, but it is eventually linked to physical behavior through the "axiom of rational choice," which asserts that choices follow preferences.) Preference behavior is thus a kind of *mental* behavior rather than physical behavior. The agent is assumed to have preferences with respect to all possible acts, even those that involve counterfactual assumptions, such as: "If I had a choice between riding in a cab (which is not running today) or walking to work with an umbrella (which I don't have), I would take the cab." The domain of preferences is therefore very rich, providing the quantitative detail needed to support a fine-grained mathematical representation of behavior.

Rationality: Assumptions about what it means to behave rationally are typically imposed at several levels: the level of the agent (individual rationality), the level of the small group (strategic rationality), and the level of the large group (competitive or social rationality). The lowest (agent) level assumptions apply in all cases, while the higher (group) level assumptions are applied in different combinations depending on the situation.

The principal assumptions of individual rationality are axioms imposed on mental behavior—that is, on preferences (von Neumann and Morgenstern 1947, Savage 1954). Preferences are assumed to establish an *ordering* of acts from least-preferred to most-preferred, and they are usually taken to satisfy the axioms most often imposed on ordering relations so as to ensure the existence of a convenient numerical representation. For example, preferences are usually assumed to be *complete*, so that for any acts *x* and *y* an agent always knows whether she prefers *x* to *y* or *y* to *x* or is perfectly indifferent between the two. Preferences are also usually assumed to be *transitive*, so that if *x* is preferred to *y* and *y* is preferred to *z*, then *x* is also preferred to *z*. Where uncertainty is involved, preferences are assumed to satisfy additional assumptions which ensure that preferences are *independent* of common consequences and that the effects of beliefs and values are *separable*—i.e., that "belief can be discovered from preference" and "value can be purged of belief" (Shafer 1986).

The assumptions imposed on individual behavior yield a representation theorem stating that to every event not under her control the agent can assign a numerical degree of belief called a *probability*, and to every consequence she can assign a numerical degree of value called a *utility*, according to which the more preferred of two acts is the one that yields the higher *expected utility*. (See Wakker 1989 for an elegant recent derivation.) In models where decisions occur sequentially over time, the rational learning assumption requires that agents update their probabilities in response to new information through a mechanical application of Bayes' theorem. In models of games with two or more players, the assumptions of strategic rationality require that their utilities should be commonly known and their probabilities should be mutually consistent (e.g., derived from a "common prior" distribution, as proposed by Harsanyi 1967), so that the choices of different individuals form a *Nash equilibrium* or some refinement thereof. In models of markets and other large-group phenomena, assumptions of competitive or social rationality give rise to other varieties of equilibria. The group-rationality assumptions are usually made strong enough to ensure that the final equilibrium is uniquely determined by the initial conditions, enabling a "comparative statics" analysis to be performed.

Puzzles, paradoxes, and pathologies: Despite its ubiquitous application to the modeling of economic and social institutions, the standard theory of rational choice is troubled by empirical violations of its key axioms, unobservability of its primitive elements, disharmony among its assumptions about individual and group rationality, and assorted methodological pathologies. In particular:

- Real decision makers satisfice rather than optimize (Simon 1997), they exhibit biases and heuristics (Kahneman et al. 1982), their decision processes are contingent on how their attention is focused (Payne et al. 1993), and they reason inductively rather than deductively (Arthur 1994).
- Subjects routinely and systematically violate the preference axioms in laboratory experiments and even in simple thought experiments (Allais 1953, Ellsberg 1961, Machina 1987).
- It is often impractical or even impossible to carry out Savage's construction of a "small world"—especially a suitable set of consequences—which agrees with a given set of concrete acts and in which it is reasonable to obey the preference axioms (Shafer 1986).
- It is not clear how the consequences or preferences experienced by one individual can be observed or measured by another with much precision.
- Probability and utility cannot be separated by observing choice behavior (Kadane and Winkler 1988, Schervish et al. 1990, Nau 1995b, Karni and Safra 1995); hence it is dubious that true probabilities can be mutually consistent or that true utilities can be common knowledge
- The subjective theory of probability is incompatible with the assumptions of game theory (Kadane and Larkey 1982, 1983; Sugden 1991; Hammond 1997).
- It is hard to aggregate the beliefs and values of perfectly rational agents—impossibility theorems (Arrow 1951, Genest & Zidek 1986) and incentive compatibility problems (Groves and Ledyard 1987) arise.
- Rational choice models do not satisfactorily explain basic economic and social phenomena such as price formation (Arrow 1987), the structure of firms (Simon 1997), or voting (Green and Shapiro 1994).
- Rational choice models are not often helpful to individual decision makers, and even in the exceptional cases when they are successfully applied to personal or corporate decisions, their main benefit is often to help frame the issues, foster communication, and rewrite the rules of the game, activities which take place outside the bounds of formal theory.

3. Arbitrage Choice Theory (ACT)

This section presents the outline of *arbitrage choice theory*, a reformulation of rational choice that circumvents the difficulties of the standard theory mentioned above. In some respects, it avoids trouble merely by making weaker assumptions and not trying to predict exactly what every agent will do in every situation. But in other respects it makes fundamentally different assumptions about the nature of interactions among agents: what is knowable, how it comes to be known, and what it means for agents to act rationally. The result is a theory that is much simpler and more strongly unified than the standard one. It also leaves the door open for some unorthodox speculation on the purpose of interactions between agents, which will be offered up in a later section. The main elements are as follows:

Table 2. Elements of arbitrage choice theory

1. Environment: – A set of agents – A set of events (states of nature and alternatives for agents) – Money (and perhaps other divisible, transferable commodities) – A contingent claims market – An outside observer
2. Behavior: – Physical behavior: choices among alternatives – Verbal behavior: small gambles or trades that agents are willing to accept
3. Rationality: – Additivity of acceptable gambles or trades – No ex post arbitrage

Environment: As before, the small-world environment includes one or more human agents and a set of events that are under the control of nature and/or the agents: an event controlled by nature is a state of nature and an event controlled by a human agent is an alternative.[2] But there are no "acts" in ACT—that is, there are no arbitrary, imaginary mappings of "consequences" to events—there are only feasible alternatives. Indeed, the troublesome abstract consequences of the standard theory do not appear here at all. Rather, we assume that the environment includes *money* (a divisible, transferable, and universally desirable commodity) and perhaps other commodities, and that there is a *market* in which contracts can be written to exchange money and other commodities contingent on the outcomes of events. (The set of events is now essentially *defined* by the possibility of writing contingent contracts on them.) Of course, agents' choices presumably do have consequences for them other than money and commodities, as in the standard theory, but such consequences are not assumed to be directly observable or exactly quantifiable: they are not explicitly modeled as part of the small world. Instead, they are indirectly modeled through the effects that they exert on the agents' willingness to trade contingent claims to money and commodities. Finally, the environment includes an outside observer, who merely embodies what the agents know about each other (and what we theorists know about them). The observer's role is one that can be played by any of the agents at any time.

Behavior: Two kinds of behavior are of interest, one of which is physical behavior—choices among alternatives—as in the standard theory. The second kind of behavior will be loosely called "verbal" behavior, and it determines the gambling or trading contracts that are written in advance of the outcomes of events. In particular, verbal behavior consists of public offers by agents to accept low-stakes contracts with specified terms, as well as counter-offers by other agents to enforce (i.e., take the other side of) those contracts. Verbal behavior is never counterfactual nor is it merely cheap talk: it has direct, though usually only incremental, economic consequences for the agents. If one agent says "I'm willing to accept contract *x*," where *x* is a vector of amounts of money (and perhaps other commodities) to be exchanged as a function of the outcomes of events, then the observer is free to choose a small non-negative multiplier α, and a contract will be enforced in which the vector of money and/or commodities actually exchanged is αx. (Both the offer *x* and the response α are considered as instances of verbal behavior: communication is a two-way street.) For example, if the agent says "I would give a dollar for a glass of lemonade if it's hot today," this means that she will pay $1 per glass for any small (possibly fractional) quantity of lemonade at the discretion of anyone else, in the event that the day is hot.[3]

The gambles or trades that agents are willing to accept play the same role that preferences play in the standard theory, in the sense that they provide the rich set of ancillary measurements needed for a quantitative theory of choice. However, the acceptance of a gamble or trade is not merely synonymous with the existence of a preference. An agent who says "I would give a dollar for an glass of lemonade" is not only saying that she *prefers* a glass of lemonade to a dollar in the usual sense of the term. She is also saying *she doesn't mind who knows it*, and furthermore she will still trade the dollar for the lemonade *even if someone else wants to take her up on it*. Thus, verbal behavior has an intersubjective quality that is not inherent in the concept of preference: it is a process of reciprocal influence, not merely a process of thought.[4] It has the potential to alter the agents' stakes in the outcomes of events, as small gambles may accrue into finite transfers of wealth during preplay communication, and in this way it partly endogenizes the "rules of the game" that the agents end up playing. (In the standard theory, by comparison, the rules of the game are always fixed in advance and are not changed by the measurement of preferences.) By construction, all verbal behavior is common knowledge in the sense of the specular common knowledge of a public market, and no additional assumptions about common knowledge will be needed later on.

Rationality: The most striking difference between the present theory and the standard one is the radical streamlining of the assumptions about rationality. There are only two such assumptions: *additivity* and *no ex post arbitrage*. The additivity assumption is more structural than substantive: it means that an active agent's willingness to accept a contract is unaffected by an observer's simultaneous enforcement (or not) of other contracts with the same agent or other agents. This assumption is justified in part by the smallness of the stakes and in part by the fact that the preplay communication process is assumed to have converged. An observer's enforcement of any one contract at this point does not appreciably alter the distribution of wealth or information, and hence it should not dull anyone's appetite for other contracts they are simultaneously offering to accept. But it is also partly justified as being merely a rule of the game: the agents are aware that the contracts they offer to accept may be bundled together by an observer, so they should only accept contracts that are desirable under those conditions. No assumption is made concerning the completeness of verbal behavior—i.e., the number or scope of the contracts that agents must accept. They need not offer prices or exchange rates for all commodities or contingent claims, nor are they required to name prices at which they would indifferently buy or sell any given claim. Since they are not required to accept any contracts at all, they are not unduly burdened by the requirement that acceptable contracts should be additive.

The substantive rationality postulate is the no-arbitrage axiom, which embodies rationality as judged from the perspective of an observer. Recall that the observer's role is one that can be played by any of the active agents: he knows only what is common knowledge and he trades only what is commonly traded. We assume the observer to be completely naïve concerning the likelihoods of events, and consequently he is interested only in opportunities for riskless profit—that is, arbitrage. An *arbitrage opportunity* is a collection of acceptable contracts that wins money for the observer (and loses money for the agents in the aggregate) in at least one possible outcome of events, with no risk of loss for the observer in any outcome. There is *ex post arbitrage* if the outcome that occurs is one in which the observer could have won a positive amount of money without placing any money at risk. The principal axiom of rationality, then, is that there should be *no* ex post arbitrage. In this way, rationality depends on both verbal and physical behavior—whether what the players *do* is sufficiently consistent with what they *say* in order to avoid material exploitation by the naïve observer—and it is evaluated posterior to the outcomes of events. This is the only consistency condition imposed by the theory, and it refers only to "primal" variables—i.e., to observable behavior. By comparison, the standard theory imposes many different kinds of consistency conditions, some of which refer to primal variables such as choices and preferences and others of which refer to dual variables such as probabilities and utilities.

The ex post evaluation of rationality is essential for dealing with situations in which the agents assign chances of 0 or 1 to events for betting purposes. If the players jointly accept contracts that are equivalent to assigning a 100% chance to an event E that (they believe) is controlled by nature, then they are deemed irrational if event E fails to occur, so they should accept such contracts only if they *are sure* that E will occur. If they jointly accept contracts that are equivalent to assigning a 100% chance to an event E that (they believe) is under their own control, then they had better *make sure* that E occurs. Whether the events are really under their control is irrelevant to the observer. The players are deemed irrational after the fact if they should have *known better* or *acted otherwise*.

Results: Arbitrage choice theory leads to conclusions about rational behavior that resemble those of the standard theory in many respects. (Examples and proofs are given in Nau and McCardle 1990, 1991; Nau 1992b, 1995abc, 1998.) First, every agent individually must behave in a manner that maximizes expected utility with respect to some probability distribution and utility function. Second, the agents jointly must behave as if implementing a competitive equilibrium in a market or a strategic equilibrium in a game. Third, where uncertainty is involved, there must exist a "common prior" probability distribution. And fourth, where agents anticipate the receipt of information, they must *expect* their beliefs to be updated according to Bayes Theorem. However, these results also differ from those of the standard theory in some important respects, namely:

- Probability distributions and utility functions need not be uniquely determined (i.e., belief and preference orderings of individuals need not be complete).
- Allais' paradox does not arise: consistency is not forced between counterfactual pairs of choices.
- "True" probabilities and utilities, to the extent that they exist, need not be publicly observed or commonly known.
- Utility functions need not be state-independent.
- Bayes' Theorem need not describe the *actual* evolution of beliefs over time (as has been argued by Hacking 1967 and Goldstein 1983 & 1985).
- Equilibria need not be uniquely determined by initial conditions.
- The uncertain moves of different agents need not be regarded as probabilistically independent.
- Common prior probabilities are *risk neutral probabilities* (products of probabilities and relative marginal utilities for money) rather than true probabilities (Nau and McCardle 1991, Nau 1995ab).
- The natural strategic equilibria are varieties of correlated equilibria (Aumann 1974 & 1987, Nau and McCardle 1990, Nau 1992b & 1995c) rather than Nash equilibria.

Among these departures from the standard theory, the reinterpretation of the Common Prior Assumption is the most radical: it is literally at odds with the most popular solution concepts in game theory and information economics. It calls attention to a potentially serious flaw in the foundations of game theory and justifies much of the skepticism that has been directed at the CPA over the years.

The other departures are all in the direction of weakening the standard theory: personal probabilities and utilities are typically set-valued rather than point-valued, beliefs and values are confounded in the eye of the beholder, equilibria are coarsened rather than refined, and beliefs are stochastic rather than deterministic.

Much of this ground has been traveled before in models of partially ordered preferences, state-dependent utility, correlated strategies, and temporally coherent beliefs. Assumptions of the standard theory are weakened and, not surprisingly, weaker conclusions follow. The arbitrage formulation scores points for parsimony and internal cohesiveness—especially in its tight unification of decision analysis, game theory, and market theory—but otherwise it appears to settle for rather low predictive power.

But there is one other very important respect in which arbitrage choice theory departs from the standard theory, which was alluded to in the introduction: the arbitrage theory does not rest on assumptions about the rationality of *individuals*. Its axioms of rational behavior (additivity and no ex post arbitrage) apply to agents as a *group*: they carry no subscripts referring to specific individuals. It is the group which ultimately behaves rationally or irrationally, and the agents as individuals merely suffer guilt-by-association. The fact that the group is the unit of analysis for which rationality is defined admits the possibility that the group is more than just an uneasy alliance of individuals who are certain of and wholly absorbed in their own interests, as the next section will explore.

4. Incomplete Models And Other People's Brains

Standard choice theory assumes that a decision model is *complete* in every respect: the available alternatives of agents and the possible states of nature are completely enumerated, the consequences for every agent in every event are completely specified in terms of their relevant attributes, and the preferences of the agents with respect to those consequences are (among other things) completely ordered. For example, consider a generic decision problem under uncertainty with two alternatives to choose from, two states of nature that might obtain, and two attributes of consequences that are experienced. Because the agent is also assumed to have completely ordered preferences over a much larger set of acts, we can infer the existence of unique numerical probabilities of states (denoted by p and $1-p$) and unique (up to affine transformations) utilities of consequences. Under suitable additional axioms on preferences (Keeney and Raiffa 1976), the utilities of consequences can be further decomposed into functions of the utilities of their attributes. For example, it might be the case here that the utility of a consequence is an additive function of the utilities of its attributes, so that, after suitable scaling and weighting, the utility of the consequence received when alternative a_i is chosen and state s_j occurs is $u_{ij} + v_{ij}$, where u_{ij} is the utility of the level of attribute 1 and v_{ij} is the utility of the level of attribute 2. Thus, we can compute the agent's expected utilities for alternatives a_1 and a_2:

$$EU(a_1) = p(u_{11} + v_{11}) + (1-p)(u_{12} + v_{12})$$

$$EU(a_2) = p(u_{21} + v_{21}) + (1–p)(u_{22} + v_{22})$$

The optimal alternative is the one with the higher expected utility, and it is uniquely determined unless a tie occurs, in which case there is a set of optimal alternatives among which the agent is perfectly indifferent. This is a satisfactory state of affairs for the agent, who knows exactly what she ought to do and why she ought to do it, and it is also satisfactory for the theorist, who can predict exactly what the agent will do and why she will do it. And if there are many agents, the omniscience of the theorist is merely scaled up: she can predict what everyone will do and why they will do it. A complicated economic or social problem has been reduced to a numerical model with a small number of parameters that easily fits inside one person's head, perhaps aided by a computer.

Of course, a "small world" model such as the one described above is merely an idealization of a more complicated "grand world" situation. More realistically, the agent's beliefs and values may be only *incompletely ordered*, in which case probabilities and utilities should be represented by intervals rather than point values. For example, in our 2×2×2 decision problem, let p and P denote lower and upper bounds, respectively, on the probability of the first state of nature, and u_{ij} and U_{ij} denote lower and upper bounds, respectively, on the utility of attribute 1 of the consequence received when alternative a_i is chosen and state s_j occurs, and so on. Then we can no longer compute an exact expected utility for each alternative, but we can compute lower and upper expected utilities (denoted *eu* and *EU*, respectively):

$$eu(a_1) = \min \{x(y_{11} + z_{11}) + (1–x)(y_{12} + z_{12})\},$$

$$EU(a_1) = \max \{x(y_{11} + z_{11}) + (1–x)(y_{12} + z_{12})\},$$

where the minimum and maximum are taken over all x, y_{11}, z_{11}, etc., satisfying $x \in [p, P]$, $y_{11} \in [u_{11}, U_{11}]$, $z_{11} \in [v_{11}, V_{11}]$, etc. If it turns out that $EU(a_i) < eu(a_j)$—i.e., if the intervals of expected utilities are disjoint—then we may still conclude that there is a unique optimal solution. But, in general, it is possible to have multiple "potentially optimal" solutions when there is overlap among the expected utility intervals for different alternatives. The optimal solution is now set-valued, as are the parameters of the model. (Models of this kind have been studied by Smith 1961, Aumann 1962, Giron and Rios 1979, Rios Insua 1990, Walley 1991, Nau 1989 & 1992a, Seidenfeld et al. 1988 & 1998, among others.) In the worst case, we may not be able to pin the optimal solution down very precisely, but we can often at least narrow the range of alternatives that ought to be considered.

A model with incomplete preferences is undoubtedly more realistic than a model with complete preferences, but it is somewhat unsatisfying for the agent and perhaps even more unsatisfying for the theorist. The agent merely fails to get a unique recommended solution from the model, whereas the theorist's entire house of cards begins to wobble: the infinite regress, the common knowledge assumptions, the refinements of equilibria, all become hard to sustain in the face of indeterminacy at the agent level.

But the true situation is even worse: realistically, most models are not only preferentially incomplete, they are also *dynamically and/or structurally incomplete*: the attributes of consequences, states of nature, and available alternatives may not be fully enumerated, and the timing of choices—even the necessity of making choices—may be indeterminate, as bounded-rationality theorists have long emphasized.[5] This is the situation normally encountered at the first stage of decision analysis: the tree has not yet been drawn and the decision maker has not yet thought of everything she might do, everything that might happen, and all the consequences that might befall her—let alone quantify her beliefs and values. In such a situation, the set of alternatives that are even *potentially* optimal may be indeterminate. This is not to say that analysis might not be helpful—indeed, the very lack of clarity that a decision maker typically experiences upon facing an unfamiliar problem is the raison d'être for professional decision analysts. Structured techniques of "value-focused thinking" (Keeney 1992) and "strategy generation" (Howard 1988) can be used to articulate values and create strategic alternatives, the views of other stakeholders can be solicited to provide different perspectives and new ideas, prototype models can be tested to identify the most important tradeoffs or uncertainties, and finally the relevant probabilities and utilities can be assessed. Eventually a complete decision tree model may be constructed, and it is hoped to be a "requisite" model of the real-world situation.

What is significant about such prescriptive applications of rational choice theory to real decisions is that most of the effort typically goes into *completing the model* rather than solving a model that is already complete, and the process of completing the model usually involves *other people's brains.*[6] The same is true of "rational" solutions to most everyday problems, particularly where novelty or complexity is an issue: an individual begins by asking the advice of colleagues and then proceeds to get help from experts in the subject and/or from the marketplace. For example, a person requiring a new home appliance does not undertake an elaborate decision analysis to determine how much she ought to pay for it, wracking her own brain to assess the expected utility of the uncertain stream of benefits produced by owning the appliance and then inverting it to obtain the corresponding certainty equivalent in dollars. She just goes to a local store and pays the competitive market price, knowing that it embodies the uncertainty assessments and value tradeoffs made by millions of other consumers interacting with scores of producers and retailers.[7] In the spectrum from cheap to expensive models, she focuses on a price range and brand image that is appropriate for someone in her demographic cohort, an identity that has been impressed on her by countless TV advertisements. If she undertakes any analysis at all, it may be to ask her friends for recommendations, read consumer publications for product ratings, and/or compare models and prices at a few stores, which further enlarges the pool of brains she draws upon. In the process, she will discover a few things about the product category that she was unaware she didn't know. Her own private knowledge and idiosyncratic values will play some role in her eventual "choice" between model X and model Y—which is her own small contribution to the collective wisdom of the market—but by then most of the important work of creating and pricing the alternatives and focusing her attention will have been done by others.

If the decision problem arises in an organizational setting, standard administrative procedures will be followed to assign personnel, gather information, construct and evaluate alternatives, build consensus, and generate action. For example, a strategy team may be formed, focus groups and decision conferences may be held, consultants may be brought in, and dialogues among different levels of management may be conducted. Higher levels of management are usually occupied by individuals who scaled the corporate ladder precisely by virtue of their knack for harnessing the abilities of subordinates and providing leadership in ambiguous situations (Isenberg 1984, March and Simon 1993, March 1994). Eventually a decision is made and an individual may assume responsibility, but by then many individuals have contributed to the process, and it is possible that no one has a clear view of the entire problem—or even the same view as anyone else. As on a professional sports team, there are different views from the playing field, the coaching box, and the front office, and all must contribute in order for the team to succeed.[8]

Another aspect of model incompleteness is that the mere passage of time often lends detail to a decision model because unforeseen things happen, as the Austrian economists (Kirzner 1992) and complexity theorists (Arthur 1994, Lane et al. 1996) emphasize. Nature or a human competitor makes an unexpected move, objectives wax or wane in perceived importance, opportunities that were not previously envisioned suddenly present themselves, while alternatives that were believed to be available sometimes turn out to be illusory. Indeed, the subjective experience of time is measured by the accumulation of small surprises: if nothing that you did not precisely foresee has happened lately, time has stood still for you. For this reason, obtaining the most "perfect" information that can be envisioned today may not be equivalent to obtaining an option to put off a decision until tomorrow, contrary to textbook decision analysis.

The preceding description of the decision making process suggests that all decisions are to some extent group decisions, and they are acts of creation and discovery, not merely acts of selection. This view is compatible with arbitrage choice theory, but it is largely incompatible with the standard theory. The standard theory assumes that every agent begins with a complete mental model of the problem, and such models are often required to be mutually consistent or even objectively correct. Every agent is an island of rationality with no reason to visit the islands of other agents except to get more of what she already knows she wants through trade or contests of strategy.[9] In the arbitrage theory, by comparison, the rational actor is naturally a group. An individual typically perceives only part of the problem: different agents know the relative values of different objects, the relative likelihoods of different events, the ranges of alternatives available in different situations. Like the proverbial blind man holding one limb of an elephant, an individual has opinions about some things but not necessarily everything, and interactions with others may provide a more complete picture. In the end, “rationality” gets enforced at the point where inconsistencies within or between agents become exploitable, and exploitation usually occurs against groups rather than individuals.

5. The Limits Of Theory

Of course, to be applied quantitatively, arbitrage choice theory requires a certain degree of completeness: events must be well-defined and contingent trades must be possible, or at least imaginable, with respect to those events. As such, the theory recognizes its own limits. Where events are not well-defined and where contingent trades are hard even to imagine, we should not expect to be able to predict or prescribe choice behavior in exact, quantitative terms. Even there, the theory is not altogether silent: it suggests that *completing the model* may be the most important task at hand and that interactive decision processes (other people’s brains) may facilitate that task. But a sobering implication emerges, namely that if the agents typically do not see the whole picture, then neither does the theorist. If the purpose of interactions among agents is to construct, between them, a more complete model of a situation that is too complex for any individual to fully comprehend, then the theorist cannot hope to do that job for them single-handedly. No single mode of analysis or theoretical paradigm can expect to yield the last word in explaining a social or economic phenomenon. This is not an argument for relativism in science, but merely a reflection on the reason why universalist theories and quantitative models have thus far proved less useful in the social sciences than in the natural sciences. Models yield accurate predictions only if their complexity is commensurate with that of the phenomena they are intended to represent, which is easier to achieve with physical systems (where we look into the eyepiece of the microscope) than with social and economic systems (where we look up from the other end). If the complexity of the system is very much greater than that of the observer and her model, she cannot expect to understand every detail by herself—but the scientific establishment to which she contributes will gradually assemble a more complete picture through its network of institutions, its competition among paradigms, and its accumulation of technology and literature.

6. Implications For Modeling

The theory of choice outlined here is broadly consistent with the normative ideals of optimization and equilibrium that are central to standard rational choice theory, notwithstanding some indeterminacies and unobservables. But it also reinforces many of the principal arguments of rational choice critics, and its implications for modeling are distinctly different from those of the standard theory in a number of important respects.

First, decision analysis should embrace the idea that decisions are inherently interactive rather than celebrate the heroically rational individual. The aim of decision analysis should be to use multiple brains to find creative solutions to complex problems, not merely satisfy the latent preferences of a designated decision maker. Some of the brains may be multiple selves of the decision maker, evoked by exercises in reframing, role-playing, and value-focused thinking. But in most cases the individual's colleagues, expert advisers, superiors, subordinates, role models, fellow consumers or investors—and decision analyst, if any—will also play a role as participants in the decision, not mere catalysts or passive data sources. Waiting may sometimes help: it may be advantageous to purchase an option to put off a decision until after some of its complexities are resolved by the passage of time, and the value of the option may reside partly in the opportunity to let unforeseen things happen. Of course, decision analysis practitioners already know these things, but they consider them part of their art rather than their science. Indeed, paradoxically, there is no role for a decision analyst in the standard theory of rational choice—the decision maker already implicitly knows what she ought to do. The view of choice developed here suggests, to the contrary, that a second brain ought to be helpful, particularly one that has broad experience with similar kinds of decisions or substantive knowledge of the problem domain. This may explain why software packages for decision-tree-analysis and expected-utility maximization have not proved particularly helpful to individuals acting alone—even decision theorists do not often use them in their private decisions.

Second, a novel approach to the modeling of games of strategy is suggested: let the common knowledge assumptions be translated into gambles or trades that the players are willing to publicly accept. Then impose a no-arbitrage condition on the resulting (observable) market, rather than imposing an equilibrium concept on the players' (unobservable) probabilities and utilities. The two approaches are in fact dual to each other, in the sense of the duality theory of linear programming, but the arbitrage-free equilibria turn out to be subjective correlated equilibria rather than Nash equilibria and the common prior distribution is a risk-neutral distribution rather than anyone's true distribution. Perhaps more importantly, by operationalizing the common knowledge assumptions in terms of material gambles or trades, the players are given the ability to tweak the rules of the game. The question then becomes not only "how should this game be played?" but also "should another game be played instead?" For example, two risk-averse players would never implement a mixed strategy Nash equilibrium: if side bets were permitted, they would bet with each other (or with an entrepreneurial observer) in such a way as to equalize their rankings of the outcomes, thereby eliminating the need for randomization and perhaps even dissolving the game (Nau 1995c). The theory of games that emerges here is thus a theory of how to make the rules as well as how to play once the rules have been fixed.

Concerning the study of economic and social institutions in which many individuals interact with each other—firms, markets, and governments—the implication is even more fundamental. To assume that it is "as if" every individual has a complete model of her environment and is behaving optimally in it is to deny what is perhaps the most important reason for them to interact, namely that their models are incomplete, and usually profoundly so. No wonder this leads to paradoxes and puzzles. As Arrow (1987) observes: "If every agent has a complete model of the economy, the hand running the economy is very visible indeed." This is not to suggest that mathematical models are unhelpful, but merely that plural analysis and high-bandwidth data are likely to be needed to fully illuminate complex social or economic phenomena. Painstaking field studies of institutions, rich data from real markets and polities, *in vivo* social experimentation, lessons drawn from history and literature—all of which harness the power of many people's brains—may shed more light on human affairs than the exploration of a system of equations with a small number of parameters. Certainly there is an important unifying mathematical principle that underlies rational behavior in decisions, games, and markets: it turns out to be the principle of no-arbitrage, once the knowledge assumptions are operationalized and the dual side of the model is examined. But that principle has little predictive ability by itself: the devil is still in the details.

Acknowledgements

I am grateful to Bob Clemen, Bart Lipman, Mark Machina, Peter Wakker, and Andy Yates for helpful comments, but they are not responsible for the consequences. This research was partially supported by the National Science Foundation under grant 98-09225.

Notes

[1] Nature also "behaves" by choosing among the states, but in standard choice theory nature's moves are modeled differently than those of the human agents. In the alternative theory introduced below, states of nature and alternatives of agents are treated more symmetrically.

---

[2] In arbitrage choice theory, it is not necessary to distinguish *a priori* between events controlled by one player and events controlled by another, nor, indeed, between alternatives and states of nature. The question of who controls what need not be answered in order for the behavior of the players to be judged rational. What matters is the way in which agents are willing to *bet on* the outcomes of events, which may reveal what they *believe* about who is in control. The usual terminology nevertheless will be retained for convenience.

[3] It is also possible to allow statements about large and specific exchanges, such as " I would give exactly ten dollars for exactly one gallon of lemonade." Statements of this kind are needed for modeling cooperative behavior, enabling the agents to escape from prisoner's dilemma games and other troublesome situations.

[4] The type of verbal behavior described here was first introduced by de Finetti (1937, 1974) as a method of eliciting subjective probabilities. De Finetti viewed its reliance on money and its game-theoretic overtones as potential "shortcomings" vis-à-vis an axiomatization based on a qualitative binary relation. (footnote (a) in the 1964 translation). Here, these very qualities are viewed as its strengths vis-à-vis a binary preference relation.

[5] Simon (1947, chapter IV) comments: "The task of decision involves three steps: (1) the listing of all the alternative strategies; (2) the determination of all the consequences that follow upon each of these strategies; (3) the comparative evaluation of these sets of consequences. The word 'all' is used advisedly. It is obviously impossible for the individual to know *all* his alternatives or *all* their consequences, and this impossibility is a very important departure of actual behavior from the model of objective rationality."

[6] At a commencement address in 1992, J.B. Fuqua remarked: "Perhaps the most important thing I learned early in my business career was that I would not attain my personal or business goals unless I used OPM and OPB—other people's money and other people's brains." Much of economic theory is concerned with the use of OPM. Rather less attention has been paid to the use of OPB, although recent work in complexity theory (e.g., Lane et al. 1996) emphasizes the role of "networks of generative relationships" in business decision making.

[7] Significantly, the market also strategically decouples the consumer and the store owner, so they need not haggle over the price. Market-based communication of beliefs and values often has this effect of eliminating the need to play games.

[8] Of course, through experience, individuals can acquire the capability to solve various kinds complex problems single-handedly, and we then call them "experts." But true experts usually find it hard to externalize the source of their expertise, which resides mainly in the ability to recognize patterns and instinctively generate appropriate responses. Hence the behavior of the expert cannot be easily explained or replicated by an analytic model.

[9] Even when the agent seeks to acquire information from other agents, she is assumed to "know what she doesn't know." That is, she is assumed to be able to foresee all her possible future states of information, and the process of information-gathering merely instantiates one of these already-foreseen states. Similarly, when she hires specialists to act on her behalf, she is assumed to be able to foresee the possible consequences of their actions and to quantify the probabilities with which those consequences will obtain—she takes nothing "on faith."

References

Allais, M. (1953) Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Américaine. *Econometrica* **21**, 503—546.

Arrow, K.J. (1951) *Social Choice and Individual Values.* Wiley, New York.

Arrow, K.J. (1987) Economic Theory and the Hypothesis of Rationality. In J. Eatwell, M. Milgrave, and P. Newman (eds.) *The New Palgrave: A Dictionary of Economics*, MacMillan Press Limited, New York

Arrow, K.J. (1994) Methodological Individualism and Social Knowledge. *American Econ. Rev.* **84**:2, 1—9

Arthur, W. B. (1994) Bounded Rationality and Inductive Reasoning. *Amer. Econ. Rev.* **84**:2, 406—411

Aumann, R.J. (1974) Subjectivity and Correlation in Randomized Games. *J. Math. Econ*. **1** 67—96.

Aumann, R.J. (1987) Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica* **55**, 1—18.

Coleman, J.S. (1990) *Foundations of Social Theory,* Harvard University Press, Cambridge

de Finetti, B. (1937) La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Ann. Inst. Henri Poincaré* **7,** 1—68. Translation reprinted in H.E. Kyburg and H.E. Smokler, eds. (1980) *Studies in Subjective Probability*, 2nd ed., Robert Krieger, New York, 53—118

de Finetti, B. (1974) *Theory of Probability, Vol. 1*. Wiley, New York

Ellsberg, D. (1961) Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics* **75**, 643-669.

Genest, C. and J.V. Zidek (1986) Combining Probability Distributions: A Critique and Annotated Bibliography. *Statistical Science* **1,** 114—148

Giron, F.J., and S. Rios (1980) Quasi-Bayesian Behavior: A More Realistic Approach to Decision Making? In J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds.), *Bayesian Statistics*. University Press, Valencia, Spain

Goldstein, M. (1983) The Prevision of a Prevision. *J. Amer. Statist. Assoc.* **78**, 817—819

Goldstein, M. (1985) Temporal Coherence (with discussion). In J. Bernardo, M. DeGroot, D. Lindley, and A. Smith (Eds.), *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting*, (North-Holland, Amsterdam, 231—248.

Green, D.P. and I. Shapiro (1994) *Pathologies of Rational Choice Theory*. Yale University Press, New Haven

Groves, T. and J. Ledyard (1987) Incentive Compatibility Since 1972. In T. Groves, R. Radner, and S. Reiter (eds.), *Information, incentives, and economic mechanisms: Essays in honor of Leonid Hurwicz.* University of Minnesota Press, Minneapolis. 48—111

Hacking, I. (1967) Slightly More Realistic Personal Probability. *Philosophy of Science* **34**, 311—325.

Hammond, P. (1997) Subjective Expected Utility and Non-cooperative Games. In S. Barberà, P. Hammond and C. Seidl (eds.), *Handbook of Utility Theory*. Kluwer Academic Publishers, Dordrecht

Harsanyi, J.C. (1967,1968) Games With Incomplete Information Played By `Bayesian' Players: I-III. *Management Science* **14**, 159—182, 320—334, 486—502.

Howard, R.A. (1988) Decision Analysis: Practice and Promise. *Management Science* **34**:6, 679—695

Isenberg, D. (1984) How Senior Managers Think. *Harvard Business Review*, Nov-Dec, 80—90

Kadane, J.B. and R.L. Winkler (1988) Separating Probability Elicitation from Utilities. *Journal of the American Statistical Association* **83**, 357—363.

Kadane, J.B. and P.D. Larkey (1982), Subjective Probability and the Theory of Games, and Reply, *Management Science* **28**, 113—120, 124.
Kadane, J.B. and P.D. Larkey (1983), The Confusion of Is and Ought In Game Theoretic Contexts, *Management Science* **29,** 1365—1379.
Kahneman, P. Slovic, & A. Tversky (1982, Eds.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.
Karni, E. and Z. Safra (1995) The Impossibility of the Experimental Elicitation of Probabilities. *Theory and Decision* **38,** 313—320
Keeney, R. and H. Raiffa (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley, New York
Keeney, R. (1992) *Value-Focused Thinking: A Path to Creative Decisionmaking*. Harvard University Press:, Cambridge, Massachusetts
Kirzner, I. (1992) *The Meaning of Market Process: Essays in the Development of Modern Austrian Economics*. Routledge, London
Lane, D, F. Malerba, R. Maxfield, and L. Orsenigo (1996) Choice and Action. *J. Evolutionary Economics* **6**:1, 43—76
Machina, M.J. (1987) Choice under Uncertainty: Problems Solved and Unsolved. *Economic Perspectives* **1**, 121—154.
March, J. (1994) *A Primer on Decision Making: How Decisions Happen*. Free Press, New York
March, J. and H. Simon (1993) *Organizations, 2nd Ed.* Blackwell, New York (1st ed. 1958, J. Wiley, New York)
Nau, R.F. (1989) Decision Analysis with Indeterminate or Incoherent Probabilities. *Annals of Operations Research* **19,** 375—403
Nau, R.F. (1992a) Indeterminate Probabilities on Finite Sets. *The Annals of Statistics* **20**:4, 1737—1767
Nau, R.F. (1992b) Joint Coherence in Games of Incomplete Information. *Management Science* **38**:3, 374—387
Nau, R.F. (1995a) The Incoherence of Agreeing to Disagree. *Theory and Decision* **39,** 219—239
Nau, R.F. (1995b) Coherent Decision Analysis with Inseparable Probabilities and Utilities. *Journal of Risk and Uncertainty* **10,** 71-91
Nau, R.F. (1995c) Arbitrage-Free Correlated Equilibria. Working Paper, Fuqua School of Business
Nau, R.F. (1998) Arbitrage, Incomplete Models, and Interactive Rationality. Working Paper, Fuqua School of Business
Nau, R.F. and K. F. McCardle (1990) Coherent Behavior in Noncooperative Games. *Journal of Economic Theory* **50**:2, 424-444
Nau, R.F. and K.F. McCardle (1991) Arbitrage, Rationality, and Equilibrium. *Theory and Decision* **31,** 199-240
Ordeshook, P. (1986) *Game Theory and Political Theory: An Introduction.* Cambridge University Press, New York
Payne, J., J. Bettman and E. Johnson (1993) *The Adaptive Decision Maker*. Cambridge University Press, New York
Rios Insua, D. (1990) *Sensitivity Analysis in Multi-objective Decision Making*. Springer-Verlag, Berlin
Savage, L.J. (1954) *The Foundations of Statistics*. Wiley, New York. (Second edition 1972, Dover, New York)
Schervish, M.J., T. Seidenfeld, & J.B. Kadane (1990) State-Dependent Utilities. *Journal of the American Statistical Association* **85**, 840—847.
Seidenfeld, T. (1988) Decision Theory Without "Independence" or Without "Ordering:" What's the Difference? *Economics and Philosophy* **4,** 267—315
Seidenfeld, T., M.J. Schervish and J.B. Kadane (1998) A Representation of Partially Ordered Preferences. *The Annals of Statistics,* forthcoming

Shafer, G. (1986) Savage Revisited (including comments), *Statistical Science* **1,** 463-501

Simon, H. A. (1947) *Administrative Behavior; A Study Of Decision-Making Processes In Administrative Organizations*. Macmillan Co., New York

Simon, H.A. (1997) *Models of Bounded Rationality, Vol. 3*. MIT Press, Cambridge Mass.

Smith, C.A.B. (1961) Consistency in Statistical Inference and Decision. *J. Roy. Stat. Soc. B* **23,** 1—25

Sugden, R. (1991) Rational Choice: A Survey of Contributions from Economics and Philosophy. *Economic Journal* 101, 751-785

von Neumann, J. & O. Morgenstern (1947) *Theory of Games and Economic Behavior*. Princeton University Press, Princeton NJ.

Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London

Wakker, P.P. (1989) *Additive Representations of Preferences: A New Foundation of Decision Analysis*. Kluwer Academic Publishers, Dordrecht.

Watkins, J.W.N. (1957) Historical Explanation in the Social Sciences. *British J. for the Philosophy of Science* **8** 104—117.