Draft of an article to appear in *The MIT Encyclopedia of the Cognitive Sciences* (Rob Wilson and Frank Kiel, editors), Cambridge, Massachusetts: MIT Press, 1997. 

# Rational Decision Making

Jon Doyle
Massachusetts Institute of Technology
Laboratory for Computer Science
545 Technology Square
Cambridge, Massachusetts 02139
`http://www.medg.lcs.mit.edu/doyle`
`doyle@mit.edu`

January 20, 1998

RATIONAL DECISION MAKING: Choosing among alternatives in a way that "properly" accords with the preferences and beliefs of an individual decision maker or those of a group making a joint decision; in particular, the subject as developed in decision theory [16] (see RATIONAL CHOICE THEORY), decision analysis [24], GAME THEORY [33], political theory [20], psychology [14] (see DECISION MAKING), and economics [10, 11] (see ECONOMICS AND COGNITIVE SCIENCE), in which it is the primary activity of homo oeconomicus, "rational economic man." The term refers to a variety of notions, with each conception of alternatives and proper accord with preferences and beliefs yielding a "rationality" criterion. At its most abstract, the subject concerns unanalyzed alternatives (choices, decisions) and preferences reflecting the desirability of the alternatives and rationality criteria such as maximal desirability of chosen alternatives with respect to the preference ranking. More concretely, one views the alternatives as actions in the world, and determines preferences among alternative actions from preference rankings of possible states of the world and beliefs or probability judgments about what states obtain as outcomes of different actions, as in the maximal expected utility criterion of decision theory and economics. UTILITY THEORY and the FOUNDATIONS OF PROBABILITY theory provide a base for its developments. Somewhat unrelated, but common, senses of the term refer to making decisions through reasoning [3] (see DECISION MAKING), especially reasoning satisfying conditions of logical consistency and deductive completeness (see DEDUCTIVE REASONING, LOGIC) or probabilistic soundness (see PROBABILISTIC REASONING, FOUNDATIONS OF PROBABILITY). The basic elements of the theory were set in place by Bentham [5], Bernoulli [6], Pareto [22], Ramsey [25], de Finetti [8], VON NEUMANN and Morgenstern [33], and Savage [28]. Texts by Raiffa [24], Keeney and Raiffa [15], and Jeffrey [13] offer good introductions.

The theory of rational choice begins by considering a set of alternatives facing the decision maker(s). Analysts of particular decision situations normally consider only a restricted set of abstract alternatives that capture the important or interesting differences among the alternatives. This often proves necessary because, particularly in problems of what to do, the full range of possible actions exceeds comprehension. The field of decision analysis [24] addresses how to make such modeling choices and provides useful techniques and guidelines. Recent work on BAYESIAN NETWORKS [23] provides additional modeling techniques. These models and their associated inference mechanisms form the basis for a wide variety of successful KNOWLEDGE-BASED SYSTEMS [35].

The theory next considers a binary relation of preference among these alternatives. The notation $x \precsim y$ means that alternative $y$ is at least as desirable as alternative $x$, read as $y$ is weakly preferred to $x$; "weakly" since $x \precsim y$ permits $x$ and $y$ to be equally desirable. Decision analysis also provides a number of techniques for assessing or identifying the preferences of decision makers. Preference assessment may lead to reconsideration of the model of alternatives when the alternatives aggregate together things differing along some dimension on which preference depends.

Decision theory requires the weak preference relation $\precsim$ to be a complete preorder, that is, reflexive ($x \precsim x$), transitive ($x \precsim y$ and $y \precsim z$ imply $x \precsim z$), and relating every pair of alternatives (either $x \precsim y$ or $y \precsim x$). These requirements provide a formalization in accord with ordinary intuitions about simple decision situations in which one can readily distinguish different amounts, more is better, and one can always tell which is more. Various theoretical arguments have also been made in support of these requirements;

for example, if someone's preferences lack these properties, one may construct a wager against him he is sure to lose.

Given a complete preordering of alternatives, decision theory requires choosing maximally desirable alternatives, that is, alternatives $x$ such that $y \precsim x$ for all alternatives $y$. There may be one, many, or no such maxima. Maximally preferred alternatives always exist within finite sets of alternatives. Preferences that linearly order the alternatives ensure that maxima are unique when they exist.

The rationality requirements of decision theory on preferences and choices constitute an ideal rarely observed but useful nonetheless (see [14], DECISION MAKING, JUDGMENT HEURISTICS, and ECONOMICS AND COGNITIVE SCIENCE). In practice, people apparently violate reflexivity (to the extent that they distinguish alternative statements of the same alternative), transitivity (comparisons based on aggregating sub-comparisons may conflict), and completeness (having to adopt preferences among things never before considered). Indeed, human preferences change over time and through reasoning and action, which renders somewhat moot the usual requirements on instantaneous preferences. People also seem to not optimize their choice in the required way, more often seeming to choose alternatives that are not optimal but are nevertheless good enough. These "satisficing" [30], rather than optimizing, decisions constitute a principal focus in the study of BOUNDED RATIONALITY, the rationality exhibited by agents of limited abilities [12, 27, 31]. Satisficing forms the basis of much of the study of PROBLEM SOLVING in artificial intelligence; indeed, NEWELL [21, p. 102] identifies the method of problem solving via goals as the foundational (but weak) rationality criterion of the field ("If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action."). Such "heuristic" rationality lacks the coherence of the decision-theoretic notion since it downplays or ignores issues of comparison among alternative actions that all lead to a desired goal, as well as comparisons among independent goals. In spite of the failure of humans to live up to the requirements of ideal rationality, the ideal serves as a useful approximation, one that supports predictions, in economics and other fields, of surprisingly wide applicability [4, 32].

Though the notions of preference and optimal choice have qualitative foundations, most practical treatments of decision theory represent preference orders by means of numerical utility functions. We say that a function

$U$ that assigns numbers to alternatives represents the relation $\precsim$ just in case $U(x) \leq U(y)$ whenever $x \precsim y$. Note that if a utility function represents a preference relation, then any monotone-increasing transform of the function represents the relation as well, and that such representation does not change the set of maximally-preferred alternatives. Such functions are called ordinal utility functions, as the numerical values only indicate order, not magnitude (so that $U(x) = 2U(y)$ does not mean that $x$ is twice as desirable as $y$).

To formalize choosing among actions that may yield different outcomes with differing likelihoods, the theory moves beyond maximization of preferability of abstract alternatives to the criterion of maximizing expected utility, which derives preferences among alternatives from preference orderings of the possible outcomes together with beliefs or expectations that indicate the probability of different consequences. Let $\Omega$ denote the set of possible outcomes or consequences of choices. The theory supposes that the beliefs of the agent determine a probability measure $Pr$, where $Pr(\omega|x)$ is the probability that outcome $\omega$ obtains as a result of taking action $x$. The theory further supposes a preference relation over outcomes. If we choose a numerical function $U$ over outcomes to represent this preference relation, then the expected utility $\hat{U}(x)$ of alternative $x$ denotes the total utility of the consequences of $x$, weighting the utility of each outcome by its probability, that is $\hat{U}(x) = \sum_{\omega \in \Omega} U(\omega) Pr(\omega|x)$. Because the utilities of outcomes are added together in this definition, this utility function is called a cardinal utility function, indicating magnitude as well as order. We then define $x \precsim y$ to hold just in case $\hat{U}(x) \leq \hat{U}(y)$. Constructing preferences over actions to represent comparisons of expected utility in this way transforms the abstract rational choice criterion into one of maximizing the expected utility of actions.

The identification of rational choice under uncertainty with maximization of expected utility also admits criticism [17]. Milnor [19] examined a number of reasonable properties one might require of rational decisions, and proved no decision method satisfied all of them. In practice, the reasonability of the expected utility criterion depends critically on whether the modeler has incorporated all aspects of the decision into the utility function, for example, attitudes toward risk.

The theory of rational choice may be developed in axiomatic fashion from the axioms above, in which philosophical justifications are given for each of the axioms. The complementary "revealed preference" approach uses the axioms instead as an analytical tool for interpreting actions. This approach, pioneered by Ramsey [25] and de Finetti [8] and developed into a useful mathematical and practical method by VON NEUMANN [33] and Savage [28], uses real or hypothesized sets of actions (or only observed actions in the case of [9]) to construct probability and utility functions that would give rise to the set of actions.

When decisions are to be made by a group rather than an individual, the above model is applied to describing both the group members and the group decision. The focus in group decision making is the process by which the beliefs and preferences of the group determine the beliefs and preferences of the group as a whole. Traditional methods for making these determinations, such as voting, suffer various problems, notably yielding intransitive group preferences. Arrow [1] proved that there is no way, in general, to achieve group preferences satisfying the rationality criteria except by designating some group member as a "dictator," and using that member's preferences as those of the group. May [18], Black [7] and others proved good methods exist in a number of special cases [29]. When all preferences are well-behaved and concern exchanges of economic goods in markets, the theory of general equilibrium [2, 10] proves the existence of optimal group decisions about allocations of these goods. GAME THEORY considers more refined rationality criteria appropriate to multiagent settings in which decision makers interact. Artificial markets [34] and negotiation techniques based on game theory [26] now form the basis for a number of techniques in MULTIAGENT SYSTEMS.

# References


[1] [1] K. J. Arrow. Social Choice and Individual Values. Yale University Press, second edition, 1963.

[2] [2] K. J. Arrow and F. H. Hahn. General Competitive Analysis. North Holland, Amsterdam, 1971.

[3] [3] J. Baron. Rationality and Intelligence. Cambridge University Press, Cambridge, 1985.

[4] [4] G. S. Becker. The Economic Approach to Human Behavior. University of Chicago Press, Chicago, 1976

[5] [5] J. Bentham. Principles of Morals and Legislation. Oxford University Press, Oxford, 1823. Originally published 1789.

[6] [6] D. Bernoulli. Specimen theoriae novae de mensura sortis. Comentarii academiae scientarium imperialis Petropolitanae (for 1730 and 1731), 5:175-192, 1738.

[7] [7] D. Black. The Theory of Committees and Elections. Cambridge University Press, Cambridge, 1963.

[8] [8] B. de Finetti. La prévision: see lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré, 7, 1937.

[9] [9] D. Davidson, P. Suppes, and S. Siegel. Decision making; an experimental approach. Stanford University Press, Stanford, CA, 1957.

[10] [10] G. Debreu. Theory of Value: an axiomatic analysis of economic equlibrium. Wiley, New York, 1959.

[11] [11] J. M. Henderson and R. E. Quandt. Microeconomic Theory: A Mathematical Approach. McGraw-Hill, New York, third edition, 1980.

[12] [12] E. J. Horvitz. Reasoning about beliefs and actions under computational resource constraints. In *Proceedings of the Third AAAI Workshop on Uncertainty in Artificial Intelligence*. AAAI, 1987.

[13] [13] R. C. Jeffrey. The Logic of Decision. University of Chicago Press, Chicago, second edition, 1983.

[14] [14] D. Kahneman, P. Slovic, and A. Tversky, editors. Judgement under Uncertainty: Hueristics and Biases. Cambridge University Press, 1982.

[15] [15] R. L. Keeney and H. Raiffa. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley and Sons, New York, 1976.

[16] [16] R. D. Luce and H. Raiffa. Games and Decisions. Wiley, New York, 1957.

[17] [17] M. J. Machina. Choice under uncertainty: Problems solved and unsolved. Journal of Economic Perspectives, 1(1):121-154, Summer 1987.

[18] [18] K.O. May. Intransitivity, utility, and the aggregation of preference patterns. Econometrica, 22:1-13, 1954.

[19] [19] J. Milnor. Games against nature. In R. M. Thrall, C. H. Coombs, and R. L. Davis, editors, Decision Processes, pages 49-59. Wiley, New York, 1954.

[20] [20] D. C. Mueller. Public Choice II. Cambridge University Press, Cambridge, second edition, 1989.

[21] [21] A. Newell. The knowledge level. Artificial Intelligence, 18(1):87-127, 1982.

[22] [22] V. Pareto. Manual of Political Economy. Kelley, New York, 1971. Originally published 1927. Translated by A. S. Schwier, edited by A. S. Schwier and A. N. Page.

[23] [23] J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA, 1988.

[24] [24] H. Raiffa. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Addison-Wesley, Reading, MA, 1968.

[25] [25] F. P. Ramsey. Truth and probability. In H. E. Kyburg, Jr. and H. E. Smokler, editors, Studies in Subjective Probability. John Wiley and Sons, 1964. originally published in 1926.

[26] [26] Rosenschein, J. S. and G. Zlotkin 1994, Rules of Encounter: Designing Conventions for Automated Negotiation among Computers. MIT Press.

[27] [27] S. J. Russell. Do the Right Thing: Studies in Limited Rationality. MIT Press Cambridge, MA 1991.

[28] [28] L. J. Savage. The Foundations of Statistics. Dover Publications, New York, second edition, 1972.

[29] [29] A. Sen. Social choice theory: A re-examination. Econometrica, 45:53-89, 1977

[30] [30] H. A. Simon. A behavioral model of rational choice. Quarterly Journal of Economics, 69:99-118, 1955.

[31] [31] H. A Simon. Models of Bounded Rationality: Behavioral Economics and Business Organization, volume 2. MIT Press, Cambridge, MA, 1982.

[32] [32] G. J. Stigler and G. S. Becker. De gustibus non est disputandum. American Economic Review, 67:76-90, 1977.

[33] [33] J. von Neumann and O. Morgenstern. Theory of Games and Economic Behavior. Princeton Universtiy Press, Princeton, third edition, 1953.

[34] [34] M. P. Wellman. A market-oriented programming environment and its application to distributed multicommodity flow problems. Journal of Artificial Intelligence Research, 1:1-23, 1993.

[35] [35] M. P. Wellman, J. S. Breese, and R. P. Goldman. From knowledge bases to decision models. The Knowledge Engineering Review, 7(1):35-53, 1992.