RATIONAL CHOICE THEORY

RATIONAL CHOICE THEORY

RATIONAL CHOICE THEORY

RATIONAL CHOICE THEORY
RATIONAL CHOICE THEORY

Nearly a century after Hobbes, Mandeville (§3) argued that private vices beget public virtues. In particular, universal avarice produces wide prosperity. This is a grim thought for virtue theory. All of us must wish others were avaricious because then they would serve our interest well. Hence, the widespread, well-founded, helpful arguments are hard results. Smith, after writing his own book of near-virtue theory, The Theory of Moral Sentiments, then went on to write an economic theory in *The Wealth of Nations* that matched Mandeville's and severely undercuts virtue theory (see Smith, A. §3). The moral move, again, was to cut the supposedly conceptual connection between self interest and collective interest. There could be no conceptual, but only a causal, connection, if any.

2 Game theory

Game theory is typically not about the source of the incentives we face but about the force of them, as is also true of rational-choice theory more generally. For most of its use, game theory is applied to actual payoff structures or to idealized structures. The payoffs are simply given. An economic theory of knowledge and preference formation can be brought to bear to explain preferences and, hence, to determine payoffs.

In social thought and in rational choice theory, games may be associated with three broad classes. There are games that involve pure conflict, games that involve pure coordination, and games that mix these two. If the only way for you to do better is for me to do worse, and we play chess, we are in a game of pure conflict. If the way for each of us to do best is for both of us to do the same thing (in some sense, we are in a game of pure coordination, as when we drive our cars in North America and hope that every other driver stays to the right as we do (or in the UK, to the left). If I can sometimes do better only when you do worse and sometimes we can both do better by doing the same thing, we are in a mixed-motive game, which involves a mixture of conflict and coordination.

These three kinds of game can be represented very simply in two-person formats; each intersection of strategy choices yields a pair of payoffs. The first payoff in each pair is to Row, the second is to Column. The payoffs here are ordinal, with 1 as the first (best) choice, 4 the last (worst) choice. In the coordination and mixed-motive games, each player has two strategies from which to choose. In the pure conflict game, Column has no choice but will get whichever payoff Row chooses.

There could be any number of strategies available to both players in all categories of games. Moreover, the payoffs could be in almost anything, such as money, position, goods, or abstract utility. In most of social thought of rational-choice theory, ordinal games are adequate because there is no good sense of quantity that could cover the range of payoffs. Even in games in which one might state what the payoffs are in each cell, one might be unable sensibly to assign them weights in anything like monetary or utility terms, so that ordinal representations still are adequate. However, ordinal representations lose important information that might alter the way people choose in a game, for example if the difference between one's first and second ranked outcomes is trivial with that between the second and third is enormous.

Game theory first grew out of games of pure conflict (see game 1), perhaps especially on the model of poker. In a pure conflict game there is nothing to be gained by cooperation. When there are more than two players, there is typically the possibility of cooperation that makes the game not pure conflict. Then the analysis of the game becomes an analysis of coalitions of players that can form. Poker can be played under rules that block collusion or cooperation between players, and other games can be contrived that have more than two players but that are still games of pure conflict between every pair of players. Von Neumann's saddle-point or minimax theorem for two-person pure conflict games shows that one cannot do better against a symmetrically rational player than to choose a strategy that includes the so-called minimax payoff. That payoff is the best of the set of worst payoffs one could get from each strategy choice, or the maximum minimum (see NEUMANN, JOHN VON).

One might think that pure conflict games are a very large and important class for social interaction. However, most social interactions that seem superficially to be pure conflict actually involve extensive opportunity for mutual benefit. For example, war might be seen as a game for victory or defeat, an analogy of chess. But actual wars, other than wars of attempted annihilation, can be fought in restricted ways that make all potentially better off. Hence, there is room for coordination as well as conflict.

Virtually the opposite of games of pure conflict are games of pure coordination, in which the incentive is to achieve harmony. Indeed, these games have been called common-interest games. Harmony can often be achieved without communication or anything even vaguely approaching agreement. The players need only recognize that their jointly best payoffs are in one or more outcomes and that there is some reason for choosing some particular strategy that includes one of those outcomes. If I am playing you in game 2 without opportunity for communication, I might choose strategy I instead of strategy II simply because the I-I outcome would come first for anyone reading English, which starts at the top left of the page and I presume you might think the same way. Thomas Schelling (1960) has shown that people are remarkably good at coordinating in such ways without conversation.

In the pure coordination game of game 2, note that there are two outcomes with equally good payoffs for each player. Row does not care whether Column chooses strategy I or strategy II, but only cares whether both of them choose strategy I or strategy II. So long as they both choose the same strategy, they do well. Because there are two possible coordination outcomes, it is possible for the players to misjudge and to finish in one of the second-best payoffs. Hence, even pure coordination problems, as sanguine and simple as they may seem, can pose difficult personal and social choice problems.

When such games in practical life are iterated, they are often resolved by convention (Hume 1739–40; Lewis 1969). Once we reach one of the best outcomes in an iterated class of interactions, we thereafter stick with the successful strategy, thereby establishing a convention.

3 Mixed-motive games

The mixed-motive game of game 3 merits extended discussion. The mixed-motive category includes many strategically distinct games. Game 3 is specifically a variant of the prisoner’s dilemma game, which is by far the most studied of all games. In the 1960s it captured the imagination of social psychologists, for whom it became a lasting focus of research. It is central in theoretical and simulation studies of altruism. And it has been associated more generally with the problem of collective action (§6). Despite all the attention it has got, however, it may have been relatively neglected in many circles because it has a misleading name.

As can be seen from the order of payoffs in game 3, prisoner's dilemma has two striking features that suggest a conflict of motivations for the players. Suppose you are Column and I am Row. You can readily see that you rank your payoff from defection higher than your payoff from cooperation no matter what I do. If I cooperate, you get your best outcome in defecting. If I defect, you get your third best rather than your fourth best if you defect. Hence, you have a dominant strategy: defect no matter what. My position is symmetrically equivalent to yours, so that I have the same dominant strategy. If we both follow our dominant strategy, we both defect and we reach an equilibrium in the game. But it is also obvious in game 3 that we both will be better off if we both cooperate than if we both defect. We may have no device outside the payoffs of game 3 to enable us to contract to secure mutual cooperation, we would benefit from doing so.

The prisoner's dilemma is unique among the seventy-eight strategically distinct two-person ordinal games in that it has an equilibrium that is Pareto inferior to some other outcome (Rapoport and Guyer 1966). (One outcome is Pareto inferior to another if moving from it to the other makes at least one person better off while harming no one.) Defection dominates cooperation individually but mutual cooperation is better than mutual defection. The crossed incentives are maddening. But they are nothing new—face them daily in our lives. Moreover, resolving their large-number analogue is the crux of political order (§6).

The discoverers of prisoner's dilemma were Flood and Dresher at the Rand Corporation around 1950. Flood had been trying to represent the sale of a used car as a game. Perhaps from a sense of that game's peculiarity, he and Dresher then designed the original prisoner's dilemma game to test how people involved in an interaction repeatedly would behave. In essence, the first prisoner's dilemma experiment was of an iterated game. The game was sufficiently peculiar in comparison to other games, as noted, that they showed it to many people, including the Princeton mathematician, A.W. Tucker, speculating on what it might mean or
represent. Tucker proposed the morality tale of two prisoners caught up in the American legal system. Knowing that either the accomplices or the prisoners will get a chance at a reduced sentence if they will turn state's evidence against their partner. If both turn state's evidence, both will get lighter than the maximum sentences. If only one does, the other will get the maximum sentence, while the confessor will get a very light sentence. Both refuse to talk, both will get a maximum sentence for a corrobory crime (illegal possession of a firearm). The individual incentive is to turn state's evidence, but the prisoners could jointly do better if they both keep silence. The terms 'cooperate' and 'defect' correspond, for the prisoners, to keeping silence and turning state's evidence.

Unfortunately, this story of the two prisoners seems to have been more captivating than the game. Yet the real structure of that game is so common in life that had it been properly identified in its name, it might have entered into political philosophy and economics a decade or two sooner. In its payoff structure, prisoner's dilemma is fundamentally the model of ordinary exchange. The tale of two prisoners blurs the general applicability of the dilemma is instantly obvious from their identical payoff structures, the morality tale of two prisoners and ordinary exchange are strategically identical. If there are morals to be drawn, they must have been inserted.

A particularly important variant of the prisoner's dilemma is the game in iterated play. When you and I are in an interaction beyond the immediate play of the game, we might both have incentives for action that derive from the further interaction as well as from our immediate game structure. In general, we may expect to see cooperative play in a prisoner's dilemma that is iterated many times, even into the distant future. Because much of our social life involves ongoing relationships, we are more likely to see gains into game theory only through iterated games. If the prisoner's dilemma is played once only between players who do not expect to meet again and without external constraints that would, for example, permissible enforceable contracts, then virtually all game theorists think the players should rationally defect. What we face in real life commonly violates these pristine conditions enough to make it our interest to cooperate in manifold exchange relations (Hardin 1982; Axelrod 1984).

Game theory is a natural ground on which to investigate the plausible meaning of rationality in interactive choice. The standard statement is that it is rational to maximize (or optimize) one's own payoff against the other (or other players) who simultaneously maximizes. This is a patently circular or self-referential definition. But it sometimes works, as in pure conflict games. It also sometimes fails, as in many mixed-motive games. When it fails, we are left free to appropriate solution concepts for games because we cannot stipulate that a particular one is rational in this natural but circular sense - the literature is cluttered with variant solution concepts. We might conclude that rationality is ill-defined and merely wants more analysis. But we might sooner conclude that rationality is inherently indeterminate in many interaction contexts.

Note also that rational choice in the context of strategic interaction does not fit well with much of philosophical action theory. In philosophical action theory, I typically produce outcomes directly from my own action. My action when I choose one strategy rather than another is, of course, motivated by my concern with outcomes. But the choice and the outcome are not directly connected. Rather, the outcome is the joint result of choices of strategy by me and by as many others as there are in the interaction. I might choose a strategy that includes some dreadful outcomes because my choice joined with strategies I expect others to choose would produce a very good outcome. In the terminology of rational choice and standard decision theory, when I do not choose against nature but choose in combination with others with whom I interact, the meanings of my choices can be very complex.

4 The Arrow problem

The basic move from cardinal to ordinal value theory was completed in the 1930s when J.R. Hicks and R.G.D. Allen (1934) worked out the formal structure of an ordinal theory that could then be put to work in standard economic accounts. The next challenge was to move from Benthamite aggregate welfare to ordinal social welfare and hence to aggregate individual social welfare to individual, but now ordinal, preferences. Arrow took on this problem with the apparent expectation of resolving it positively. In a retrospective account, Arrow (1983) implicitly acknowledged that he had expected to demonstrate that it was a fallacy of composition and that Pareto was de facto right. He proved what is now called Arrow's impossibility theorem; that there is no general way to aggregate individual ordinal preferences into collective ordinal preferences of the possible whole states of affairs.

The limited Pareto principle is often indeterminate because of what is called the plausibility clause that would make someone worse off and because, when there is more than one available Pareto improvement (a common state of affairs), the Pareto principle typically cannot say which is better. Similarly, ordinal social choice is indeterminately related to individual preference.

Arrow's theorem is based on several assumptions that sound relatively benign and acceptable. First, he stipulates that only ordinal preference rankings be considered (the ordering condition, O). Then he stipulates that no individual's preference ordering will be an ordering of the society (non dictatorship, D), or restrictions can be placed on any individual's choice of orderings (universal domain, U), if everyone prefers one state of affairs to another, society prefers that state to the other (unanimity or Pareto, P), and if individuals' rankings over the social choices are changed but no individual reverses the order of two possible choices, then the collective ordering over those two choices cannot change (irrelevance of independent alternatives, I). Finally, he implicitly assumes, reasonably enough, that our society has a finite number of individuals.

One might quarrel with each of these assumptions. But there is no point in quarrelling with O, since it is the point of Arrow's exercise. There is also no point in quarrelling with Arrow's unstated but implicitly invoked condition of finitude of the citizenry. And few have ever quarrelled with condition D - except to strengthen it further! Similarly, few argue against condition P, although Amartya Sen (1970) has campaigned hard against it. It is easy, however, to imagine states of affairs that everyone would rank below some other state (thus violating U). And many, perhaps most, commentators do not really know what to think of condition I. Is it naturally compelling? Or does it truly represent the lazy way we think through the process of aggregation in his own proof of the impossibility theorem, Arrow did not explicitly invoke the condition - although he did implicitly assume it.

As will be discussed further in Chapter 5, 99 there is another fundamental issue that seems bothersome for Arrow's theorem: the ranking of whole states of affairs. As Arrow says, a whole state of affairs means everything is determined. We can avoid the burden of such inclusiveness only if ordinal preferences over some things can be stated with a ceteris paribus clause that declares these things to be decoupled from all else. Why? Because ordinal preferences resolve two important and nearly intransigent problems of the older cardinal value theory. Under the older theory, I should apparently value two dinners tonight at twice the value of one - but clearly I do not. Indeed, I might evidently value one dinner by two compared with the other.

Also under the older theory, I should apparently value anything independently of what else there is or what else I have - but this is patently false. Some things are complements, while some others might be enhanced or reduced by my having something else with them, as my dinner and my glass of wine might be better enjoyed jointly than separately.

Now if the value of a bundle of some consumption depends on whether there are substitutes or complements in my bundle of consumables, I cannot rank some part of the whole package. I can only rank whole packages in comparison to other whole packages. That is, as a citizen, I can only rank whole states of affairs fully determined.

Despite this relatively daunting thought, there is a reduced form of Arrow's theorem that is of great practical importance. It was recognized by the Marquis de Condorcet and by C.L. Dodgson. It is called the Condorcet paradox or the problem of cyclic majorities. If three or more people must choose from three candidates by majority rule, candidate A can defeat B, B can defeat C, and C can defeat A. There is then no outright majority winner (see CONDORCET §2). Many voting systems in wide use avoid this problem by various random or constrained devices that sometimes bias outcomes. For example, legislatures commonly consider amendments to a piece of legislation on one at a time. An amendment that wins a majority vote in the first round is then considered, one that loses is dropped from further consideration. In the end, there is a decisive vote for or against the amended legislation, which is pitted only against the status quo and not against other possible variants. The status quo is therefore given privileged status because, if there are cyclic preferences over the status quo, the final piece of legislation and the same legislation with an alternative amendment, the status quo wins by virtue of not having to face the alternative. One might wish to say this is the best of plausible resolutions. But one cannot say any such thing with coherent meaning - that is the point of Arrow's theorem. Our presumptive notions of 'the best' are vacuous in this context.

5 Voting and democracy

Game theory and Arrow's theorem were very slow to enter debates in political philosophy and political science. Downs' economic theory of democracy, however, immediately received the starting point for most subsequent debate. Again, there are two main results of the theory. First, candidates, needing a majority of votes for election in two-party contests, would have to cater to the middle of the range of popular preferences. Hence, in a catch phrase
RATIONAL CHOICE THEORY

of 1950s US politics, the Democratic and Republican parties put up candidates who were close twins, Tweedledee and Tweedledum. This tendency inspired a massive literature on so-called responsible parties, a literature that argued that parties must emphasize their differences and must take strong positions. Downs' theory virtually killed that literature.

Second, the market model's central deduction was that voters have little interest in actually voting if the act of voting has real costs associated with it. If there are millions of votes and victory goes to the majority candidate, then no one has much chance of affecting the outcome. Indeed, if elections were as close as a single vote, they would sometimes be voided and repeated simply because the counting of votes could not be accurate enough to yield a determinant outcome. But, in any case, genuinely close elections are extremely rare above the level of minor local offices. Hence, to weigh the value of my vote to me in affecting policies in my interest, I must discount my vote by the exceedingly low probability that it could matter. Its discounted value to me must be vanishingly small. This is the benefit side of my vote. Now consider the cost side. Suppose it takes a couple of hours out of my hard day for me to vote. This cost is not vanishingly small. On balance, it is therefore unlikely to be in my interest to vote.

Suppose voting costs could be reduced to nearly nothing. Would it then be in my interest to vote? Only if I can be reasonably sure that I know enough to vote in my interest. Karl Marx (1852) thought the French peasants of the mid-nineteenth century did not know enough and that they harmed themselves with their votes. US air-traffic controllers might soon have concluded that their voting for Ronald Reagan in the presidential election of 1980 was stupid (he fired them when they went on strike). Some people know enough to judge the relation between elections and their interests accurately. Most people may have too limited knowledge and theory to judge that relation. If they are to vote rationally, therefore, they must acquire information. Typically it is costly to acquire information, and often the costs of relevant information for voting would far outweigh the discounted value to oneself of one's vote.

It is sometimes suggested that we can overcome the logic of Downs' rational voter by imposing sanctions on those who do not register and vote. But this is not a full solution — indeed, it might have perverse effects on elections because it may typically still not be rational for a voter to learn enough to make intelligent decisions. A requirement that people register and vote could be monitored relatively easily and cheaply. A requirement that they be prepared to vote intelligently could probably not be monitored without becoming massively intrusive. Moreover, the collective benefit from having everyone be well informed might be reckoned less than the collective costs of the required study. Hence, it seems reasonable to conclude from the rational-choice theory of citizen participation in democratic politics that democracy faces severe constraints and that much of philosophically democratic democracy is outside the realm of real-world politics (see Democracy §4).

Suppose the logic of the economic theory of democracy is correct and that the assumptions about the individual voter in this model implies can, it follows that most so-called democracies should not be the democracies of the democratic theorists. Hence, a theory that is about possibilities and that is based on simple assumptions of the weights of costs and benefits to individuals and groups yields fundamentally normative conclusions.

Unfortunately for simple, stable visions of democracy, there is a third major result in the economic theory. One of the obvious simplifications of Downs' theory was his assumption of a one-dimensional policy space along which voters are arrayed from left to right. If the world is more complex than this and there are two or more policy dimensions that do not correlate well with other dimensions, then扩大说 small changes can be important. Slight changes in location of candidates can entail radical changes in election outcomes. If candidates attempt to respond to each other, the winning policy position can be anywhere in the entire policy space. Every outcome is democratic — or none is.

6 The logic of collective action

Suppose we are members of a group who share an interest in having some policy adopted but that every one of us would rather be a free-rider on the efforts of others than contribute toward providing our common benefit. If we all attempt to be free-riders, we will collectively fail to provide any collective provision. It is in the group's interest to have the good provided even at its costs; but it is in no individual member's interest to provide any part of the good. The economist Olson has called this the logic of collective action (Olson 1965: Hardin 1982).

It is commonly said that doing something is in our interest and therefore in the interest of each of us. Indeed, this is the central argument of distastefully many newspaper editorials on such problems as pollution, urban blight, war and peace, and ending poverty. The argument is wrong. Before Olson, the standard theory of groups fallaciously supposed that what was in a group's interest was in the individual group member's interest. In one very influential theory, it was supposed that, in order to predict policy, we need only do a vector sum of the force of all the relevant groups, with each group's vector length given by its number of members and its vector direction given by its policy preference. Hence, if there were only one large group, the vector sum would be policies adopted to satisfy much of the interest. This was a fallacy of composition — perhaps best called a fallacy of decomposition — but it is hard to read the error directly in traditional group theories.

Anyone who stated the matter artifically would plausibly have recognized its illogic. The full panoply of possible groups includes, in Olson's terms, privileged, latent, and intermediate groups. In a privileged group at least one person values provision (of some good that benefits a group) enough to pay fully for its provision — and therefore we might expect provision. In a latent group, every member has the preferences of the free-rider and therefore we expect no provision. In an intermediate group, every member has the preferences of a free-rider but the group is small enough and close enough for interactions that might produce greater cooperativeness and therefore, possibly, provision. In political contexts it is often seemingly true that small, well-organized industrial groups can succeed in acting collectively for political benefit, as in lobbying for privileged legislation or helping the relevant candidate win election. At the same time, the large group of much, even most, of the public is latent and fails to organize well enough to lobby for or against legislation of interest to the group or to support relevant candidates. The rational-choice theory of group behaviour therefore suggests that government outcomes with a bias toward concentrated interests and against wider public interests.

7 Further applications

The rational choice theory of democracy suggests that we should want to understand the relation of individual to the large groups in which they have roles and with which they have to deal in social life. If an institution is comprised of individuals in some structure of roles, then we should be able to give an individual-choice theory of institutions. This has long been a major programme, as in the work of Smith, Bentham, and Mill. It is the reputed next project for bringing Rawls' theory of justice to bear in a real world. It is a frequent worry in jurisprudence and in the resurgence of constitutional thinking after the extraordinary events of 1989 in eastern Europe. And it is a major impetus to the organization theory in economics, psychology, and sociology.

The modern public-choice programme in economics is often attributed to the judgment of Joseph Schumpeter that standard economics applies as well to public as to private goods. It is essentially a theory of the effects of incentives (see Schumpeter, J.A. §3). One could generalize Schumpeter's recommendation to apply standard economic theory to government as well as to business. It is sometimes recommended that we apply it to all of human behaviour. This would be audacious, perhaps, and possibly misguided. But the recommendation has been made by some theorists, who have applied rational choice. game theory, standard price theory, equilibrium analysis, and other devices of economics to family relations, altruism, religion, crime and sex (Becker 1976), to theory of knowledge, norms, and group identification (Hardin 1995), to the rise of states (Levi 1985), to agency, religion, addiction, war, trust, and whatever ails or moves us. Rational choice theorists commonly treat motivation from anything other than interest or personal preference as an anomaly to be explained, not only for actions in the province of public choice but, increasingly, for action across the broad spectrum of life. In return, economics has benefited from feedback from this information economies of Stigler and others, theories of cognitive limits associated with Herbert Simon and others with fairness and distributive justice.

8 Intentional behaviour

In rational-choice explanations there are commonly generalizations that do not turn on asking every relevant person in a decision-making group why they act as they do. Instead of asking, the rational choice theorist simply assumes a pattern of interests or of costs and benefits. Sometimes, it would be wrong to take people's claims for their actions and motives as compelling. Twentieth-century social and psychological thought has been receptive to the supposition that much of our motivation is not recognized by us. From Freudian psychology to mass movement sociology to economic and rational choice, it is commonly assumed that people act intelligibly but with poor understanding of their own actions.

Much of twentieth-century psychology and, to a lesser degree, other social sciences, was a reaction to this assumption. Since it has often been assumed that we cannot know what is going on in other people's minds many psychologists and others have insisted on considering only behavioural data, not putative mental events. In face of the Freudian vision of mental processes, one can imagine that behaviourism was appealing to many psychologists in the early twentieth century. Behaviourism has sometimes been identified with logical positivism in the philosophy of science, but the identification fails for those who are
RATIONAL CHOICE THEORY

10 Preliminary conclusions

To do rational choice theory with interests instead of goods, utility, or satisfaction, and in the context of the social sciences, a striking feature of many of the foundational difficulties is that there is no clear notion of a utility function. Utility is not a property of individual rationality, but of the competition between actions. Rationality, the original concept, is understood in terms of the structure of preferences, and not in terms of the preferences themselves. The social sciences are therefore essentially concerned with the structure of preferences.

Consider the following example. Suppose that I have two choices: I can either go to a movie or stay home and study. If I go to the movie, I will be happy for two hours, but if I stay home, I will be happy for the next two hours. The problem is that I cannot compare the two choices in terms of happiness. I cannot say that going to the movie is better than staying home because I do not know how to compare the two choices.

Similarly, in the context of the social sciences, we cannot compare the choices of different individuals in terms of utility. We cannot say that one choice is better than another because we do not know how to compare the preferences of different individuals.

To overcome this difficulty, we need to develop a new framework for understanding the structure of preferences. This framework should be based on the idea that preferences are not properties of individual rationality, but of the competition between actions. It should also take into account the fact that preferences are not properties of individual choices, but of the structure of preferences.

There are several problems at the foundations of rational choice. In addition, to the problem of the indifference of utilities, there are also problems of the indifference of preferences, of the indifference of goods, and of the indifference of the structure of preferences.

9 Value theory

Value theory is the branch of philosophy that is concerned with the nature of value and the basis of value judgments. Value theory is also concerned with the nature of moral responsibility and the basis of moral judgments.

There are several problems at the foundations of value theory. In addition, to the problem of the exactness of value judgments, there are also problems of the exactness of moral judgments, of the exactness of the basis of value judgments, and of the exactness of the basis of moral judgments.

8 Probability theory

Probability theory is the branch of mathematics that is concerned with the analysis of random events. Probability theory is also concerned with the analysis of the behavior of individuals and the behavior of populations.

There are several problems at the foundations of probability theory. In addition, to the problem of the exactness of probability calculations, there are also problems of the exactness of the behavior of individuals, of the exactness of the behavior of populations, and of the exactness of the analysis of random events.
RATIONAL CHOICE THEORY

ian Press (Rational choice applied to government institutions and to democratic theory).


• Downs, A. (1957) An Economic Theory of Democracy. New York: Harper & Row. (On the problem of collective action that underscores individual incentive to vote and on the incentives of parties and their candidates to adopt policies that produce winning votes.)


— (1988) Morality within the Limits of Reason, Chicago, IL: University of Chicago Press. (Applies game theoretic reasoning to utilitarian moral theory, esp. in chapter 2, and presents an account of institutional utilitarianism in chapters 3 and 4.)


• Lewis, D.K. (1969) Convention, Cambridge, MA: Harvard University Press. (Defines convention as the spontaneous resolution of a repeated coordination interaction.)

• Luce, R.D. and Raiffa, H. (1957) Games and Decisions, New York: Wiley. (Dated but still valuable survey of fundamental issues in game theory, often quite demanding.)


• Sen, A.K. (1970) Collective Choice and Social Welfare, San Francisco, CA: Holden-Day. (Survey of the Arrow problem and related issues with an extensive but dated bibliography; has alternating verbal and technical chapters, the former quite accessible, the latter often difficult. The presentation in §4 uses Sen’s notation.)


RUSSELL HARDIN

RATIONALISM

Rationalism is the view that reason, as opposed to, say, sense, is the source of knowledge, and the source of knowledge is the attempt to gain knowledge. Different forms of rationalism are distinguished by different conceptions of reason and its role as a source of knowledge, by different descriptions of the alternatives to which reason is opposed, by different accounts of the nature of knowledge, and by different choices of the subject matter, for example, ethics, physics, mathematics, metaphysics, relative to which reason is viewed as the major source of knowledge. The common application of the term ‘rationalism’ is very little about what two philosophers have in common.

Suppose we mean by reason our intellectual abilities in general, including sense experience. To employ reason is to use our individual intellectual abilities to seek evidence for and against potential beliefs. To fail to employ reason is to form beliefs on the basis of such non-rational processes as blind faith, guessing or unthinking obedience to institutional authority. Suppose too that we conceive of knowledge as true, warranted belief. Where we lack reason that a belief be beyond a reasonable doubt though not beyond the slightest doubt. Here, then, is a version of rationalism: reason is the major source of knowledge in the sciences. This is a weak version of rationalism which simply asserts that our individual intellectual abilities, as opposed to blind faith and so on, are the major source of knowledge in the natural sciences. It is clearly not very controversial and is widely accepted.

Suppose, however, we take reason to be a distinct faculty of knowledge distinguished from sense experience in particular. To employ reason is to grasp self-evident truths or to deduce additional conclusions from them. Suppose we conceive of knowledge as true, warranted belief, where warrants now requires that a belief be beyond even the slightest doubt. Let us also extend our attention to metaphysics and issues such as the existence of God, human free will and immortality. Here is a much stronger version of rationalism which asserts that the intellectual grasp of self-evident truths and the deduction of ones that are not self-evident is the major source of true beliefs warranted beyond even the slightest doubt in the natural sciences and metaphysics. Clearly it is highly controversial and not very widely accepted.

The term ‘rationalism’ has been used to cover a range of views. Scholars of the Enlightenment generally have in mind something like the first example — a general confidence in the powers of the human intellect, in opposition to faith and blind acceptance of institutional authority, as a source of knowledge — when they refer to the rationalist spirit of the period and the work of such philosophers as Voltaire. More frequently, the term ‘rationalism’ is used to refer to views, like the second one above, which introduce reason as a distinct faculty of knowledge in contrast to sense experience. Rationalism is then opposed to empiricism, the view that sense experience provides the primary basis for knowledge. This entry concentrates on this still very general form of rationalism, reserving the term ‘rationalism’ for it alone.

1. Continental Rationalism and British Empiricism
2. Innate ideas
3. Intuition and demonstration

1 Continental Rationalism and British Empiricism

The rationalist-empiricist division has traditionally played a major role in our understanding of the history of philosophy, particularly that of the modern period of the seventeenth and eighteenth centuries leading up to Kant. The major philosophers of