1
Spontaneous Order

In Britain, drivers almost always keep to the left-hand side of the road. Why? It is tempting to answer: 'Because that is the law in Britain'. Certainly someone who drove on the right would be in danger of prosecution for dangerous driving. But British drivers don't keep slavishly to *all* the laws governing the use of the roads. It is a criminal offence for a driver not to wear a seat belt, to drive a vehicle whose windscreen wipers are not in working order, or to sound a horn at night in a built-up area; but these laws are often broken. Even people who cheerfully break the law against drunken driving – a very serious offence, carrying heavy penalties – usually keep left.

The answer to the original question, surely, is: 'Because everyone else drives on the left'. To drive on the right in a country in which people normally drive on the left is to choose a quick route to the hospital or the cemetery. The rule that we should drive on the left is self-enforcing.

So we do not always need the machinery of the law to maintain order in social affairs; such order as we observe is not always the creation of governments and police forces. Anarchy in the literal sense ('absence of government') cannot be equated with anarchy in the pejorative sense ('disorder; political or social confusion'). The notion of spontaneous order – to use Friedrich Hayek's phrase[1] – or orderly anarchy – to use James Buchanan's (1975, pp. 4–6) – is not a contradiction in terms. Perhaps driving on the left is a rare example of spontaneous order, and in most cases the absence of government does lead to disorder and confusion; but this is not a self-evident truth. The possibilities of spontaneous order deserve to be looked into.

In this book I shall investigate the extent to which people can co-ordinate their behaviour – can maintain some sort of social order – without relying on the formal machinery of law and government. In short, I shall study human behaviour in a state of anarchy. I shall begin by explaining why I believe such a study to be worthwhile, and in particular, since I am an economist, why I believe it has something to contribute to economics.

1.1 ECONOMICS AND SPONTANEOUS ORDER

It might be objected that economics is, and always has been, primarily a study of spontaneous order. The idea that the market is a self-regulating system has been one of the main themes of the discipline ever since Adam Smith. As Smith put it in one of the most famous sentences in economics, the individual who intends only his own gain is led by the institutions of the market to promote the public interest: he is 'led by an invisible hand to promote an end which was no part of his intention' (Smith, 1776, Book 4, Ch. 2). But in modern economics, the market is seen as a complex and imperfect system which has to be carefully looked after by the government: property rights have to be defined and protected, contracts have to be enforced, 'market failures' have to be corrected and income has to be redistributed to ensure social justice. The markets that are represented in economic theory are not states of anarchy.

It is certainly true that economists take a great deal of professional pride in the theory of the workings of an ideal competitive economy. Although this theory has been progressively refined by mathematical economists from Walras in the nineteenth century to Arrow and Debreu (1954) in the twentieth, it is in a direct line of descent from Smith's analysis of the invisible hand. Arrow is speaking for many fellow economists when he writes that the price system 'is certainly one of the most remarkable of social institutions and the analysis of its working is, in my judgement, one of the more significant intellectual achievements of mankind' (Arrow, 1967, p. 221). Nevertheless, the Walrasian (or Arrow-Debreu) theory of competitive equilibrium is not usually taken seriously at the practical as opposed to the theoretical level. Few economists are brave enough to claim that a modern industrial economy organized on *laissez-faire* principles would work much like the Walrasian model. It is more usual to say that the Walrasian model serves as a benchmark or limiting case; it provides a framework for thought. Thus economists tend to view the realities of economic life as divergences from an ideal Walrasian model. (That is why we use such expressions as 'market failure' and 'government intervention': the theoretical norm is a perfect market system and a *laissez-faire* government.) But the prevailing view is that these divergences are many and significant: market failure occurs in most areas of economic life, and so government intervention is required. When Arrow writes about the economics of medical care, for example, he presents a long list of ways in which reality diverges from the ideal model for which, as a theoretician, he has such admiration; he concludes by remarking that a study of these divergences 'force[s] us to recognize the incomplete description of reality supplied by the impersonal price system' and by endorsing the 'general social consensus . . . that the *laissez-faire* solution for medicine is intolerable' (Arrow, 1963, p. 967). The medical-care industry is perhaps an extreme case, but there can be few real world markets in which economists have not diagnosed some kind of market failure and prescribed some kind of government intervention.

Most modern economic theory describes a world presided over by *a government* (not, significantly, by governments), and sees this world through the government's eyes. The government is supposed to have the responsibility, the will and the power to restructure society in whatever way maximizes social welfare; like the US Cavalry in a good Western, the government stands ready to rush to the rescue whenever the market 'fails', and the economist's job is to advise it on when and how to do so. Private individuals, in contrast, are credited with little or no ability to solve collective problems among themselves. This makes for a distorted view of some important economic and political issues.

1.2 THE PROBLEM OF PUBLIC GOODS

In one important respect economists are very pessimistic about the chances of promoting the public interest through the self-interested actions of individuals. The market, it is conventional to say, cannot solve the problem of supplying public goods. A public good is one that is consumed jointly by a number of individuals, so that the benefit each person derives from the good depends not on how much *he* buys or supplies, but on how much *everyone* does. (For example, suppose that a community of ten households is connected to the outside world by a single road. If one household spends time or money repairing the road, all the others benefit; so the activity of repairing the road is a public good to the community.) Conventional economic theory predicts that public goods will be seriously under-supplied unless the government intervenes (the US Cavalry again). If the supply of public goods is left to private individuals, it is said, everyone will try to take a free ride at the expense of everyone else; the result will be one that *no one* wants.

Everyone ends up trapped in a situation in which everyone would prefer it if everyone contributed towards the supply of the public good; but each person finds it in his interest not to contribute. This problem is known variously as the 'problem of public goods', the 'problem of collective action', the 'prisoner's dilemma problem' and the 'tragedy of the commons'.

In reality, however, some public goods *are* supplied through the voluntary contributions of private individuals, without any pressure from the government. In Britain, for example, the lifeboat service is paid for in this way. If you are in danger at sea, the boats of the Royal National Lifeboat Institution will come to your rescue, even if you have never contributed a penny towards their costs; and you will be charged nothing. So the existence of the lifeboat service is a public good to everyone who might have to call on its services. The same applies to blood banks. If you need a blood transfusion in Britain, the National Blood Transfusion Service will supply the blood without charge; so the existence of a blood bank is a public good to everyone who might some day need a transfusion. This public good is supplied by unpaid donors. Countless more examples could be given. Works of art, great houses and stretches of countryside are bought for the nation through fund-raising appeals; the work of the churches in many countries is almost entirely financed by private gifts; and so on, and so on.

Economics has found it extremely difficult to explain this sort of activity, which runs counter to the theoretical prediction of free-riding behaviour.[2] It seems that economics underestimates the ability of individuals to coordinate their behaviour to solve common problems: it is unduly pessimistic about the possibility of spontaneous order.

1.3 THE LIMITS OF GOVERNMENT

The power of governments is not unlimited: some laws have proved almost impossible to enforce. The most famous example is probably the American experience of Prohibition. Prostitution, too, is notoriously resistant to the laws of puritanical governments. In centrally planned economies, black markets are similarly resistant to the forces of law. The attempts of successive British governments to regulate trade union activity have had, at best, mixed success.

Wise governments do not risk losing credibility by passing laws that cannot be enforced; and when such laws are passed, wise police forces turn a blind eye to violations of them. British policy towards speed limits on roads provides an interesting example. The speed at which most people actually drive on a particular road is used to help determine the level at which the speed limit is set. If the vast majority of drivers are observed to break a speed limit on a particular stretch of road, this is taken to be evidence in favour of raising the limit.

One implication of this is that governments must, if only as a matter of prudence, take some account of the possibility that the laws they might wish to pass may be unenforceable. The willingness or unwillingness of individuals to obey the law is a constraint on the government's freedom of action. Obviously, for any law, there will always be some people who will not obey it except under threat of punishment; but the system of policing and punishment is liable to break down if everyone is in this position. In other words, if a law is to work it must not go too much against the grain of the forces of spontaneous order. Adam Smith put this point well in *The Theory of Moral Sentiments*:

> The man of system . . . is often so enamoured with the supposed beauty of his own ideal plan of government, that he cannot suffer the smallest deviation from any part of it. . . . He seems to imagine that he can arrange the different members of a great society with as much ease as the hand arranges the different pieces upon a chess-board. He does not consider that the pieces upon the chess-board have no other principle of motion besides that which the hand impresses upon them; but that, in the great chess-board of human society, every single piece has a principle of motion of its own, altogether different from that which the legislature might chuse to impress upon it. If those two principles coincide and act in the same direction, the game of human society will go on easily and harmoniously, and is very likely to be happy and successful. If they are opposite or different, the game will go on miserably, and the society must be at all times in the highest degree of disorder. (1759, Part 6, Section 2, Ch. 2)

A more fundamental implication is that it may sometimes be misleading to think of the law as the creation of the government, imposed on its citizens. This characteristically utilitarian view is the one economists usually take; the law, for most economists, is a 'policy instrument' to be controlled by a benevolent social-welfare maximizing government. (Economists often recommend that the government 'corrects' a market failure by means of some change in the law – for example that monopoly power should be limited by anti-trust law, or that the law of property should be changed so as to 'internalize' external effects.) But it may be that some important aspects of the law merely formalize and codify conventions of behaviour that have evolved out of essentially anarchic situations; as in the case of the speed limits, the law may reflect codes of behaviour that most individuals impose on themselves.

The British rule of driving on the left provides another example. If you were caught driving on the right-hand side of the road you would normally be charged, not under any law specifically requiring you to drive on the left, but with the catch-all offence of 'dangerous driving'. Clearly it *is* dangerous to drive on the right, but only because everyone else drives on the left. In other words, driving on the right is illegal *because* it is contrary to convention: the law follows the regularity in behaviour, and not the other way round. To admit this possibility is to say that if we are to understand why the law is as it is, and how it works, we must study anarchy as well as government.

There is another way in which the power of governments is limited: every government inhabits a world that contains other governments. The difficulties that this creates are often swept under the carpet in theoretical economics, where the typical model is of a self-contained society presided over by a single government. As I mentioned earlier, economists tend to talk about 'the government' rather than 'governments'.

Writing in the seventeenth century, Thomas Hobbes noted that international affairs provided one of the best examples of pure anarchy (Hobbes, 1651, Ch. 13). Three hundred years later, that insight remains true; we are no nearer to the prospect of a world government with the power to enforce its rulings on recalcitrant states. In the meanwhile, unfortunately, the dangers of international anarchy have increased immeasurably. Quite apart from the continuous increase in the destructive power of military weapons, there is a growing tendency for the peacetime activities of one nation to impinge on the citizens of others. Consider the problems of acid rain, pollution of the sea, over-fishing and deforestation. In all these cases – and there are many others – conservation is a public good on an international scale. Each nation has an incentive to take a free ride on the conservation efforts of other nations.

In cases like these, the economist's traditional recommendation of 'government intervention' is useless; there is no government to intervene in the affairs of nations. The institutions and conventions of anarchy are the only ones we have within which to find solutions for some of the most pressing problems of our time. This alone would be sufficient reason for studying spontaneous order.

1.4 MORAL VIEWPOINTS

Practical economists are expected to draw 'policy conclusions' from their studies of human behaviour. A policy conclusion, as every economist knows, is a recommendation about what the government ought to do. The economist's job, it seems, is to observe and explain the behaviour of private individuals – workers, consumers and entrepreneurs – and then to advise the government. It is a curious fact that academic economists rarely think of doing things the other way round – observing the behaviour of governments and then using their findings to advise private individuals. Such work *is* done, but only in response to a demand, and for pay: it is 'consultancy'. In contrast, advice to governments – unasked for and unheeded – can be found in any economics journal.

In a sense, the usual stance of the economist is to pretend he *is* the government, and free to implement any policy he wishes. Applied economics is largely about predicting the consequences of alternative policies that the government might adopt. In order to make these predictions, the economist has to try to model the behaviour of private individuals as accurately as possible; he has to understand how they actually behave. But there is no need to do the same for the government; the concern is not with how governments actually behave, but with what would happen if particular policies were adopted. In the jargon of economics, the behaviour of governments is exogenous to the theory. This is part of what I have called the 'US Cavalry' model in which the government is an unexplained institution, always on call to implement whatever solutions to social problems the economist can devise.

This government's-eye view of the world has led economists to take a rather one-sided view of moral questions. The normative branch of economics – called, significantly, 'welfare economics' or 'social choice theory' – is concerned with questions of the kind: 'What is best for society?' or 'What would generate most welfare for society?' or 'What ought society to choose?' Notice that these are the sort of moral questions that would be faced by a benevolent and all-powerful government of the kind that the economist imagines himself to be advising.

This, however, is only one aspect of morality, and one that is rather remote from the concerns of the ordinary individual. For most of us, 'What ought *I* to choose?' is a much more pressing moral question than 'What ought *society* to choose?' Economists have had very little to say about the morality of individual behaviour. The prevailing view, I think, is that we should take individual morality as we find it, and treat it as a kind of preference. There is even a tendency to restrict the words 'moral' and 'ethical' to judgements about the good of society as a whole. Many economists (including myself, I must confess) have made use of Harsanyi's (1955) distinction between 'subjective preferences' and 'ethical preferences'. A person's subjective preferences, which by implication are non-ethical, are those that govern his private choices, while his ethical preferences are his disinterested judgements about the welfare of society as a whole. Harsanyi argues that in order to arrive at ethical preferences, a person must try to imagine himself in a position where he doesn't know his own identity, and has an equal chance of becoming anyone in society. The logic of this approach is that the proper viewpoint from which to make moral judgements is that of an impartially benevolent observer, looking on society from above; in a literal sense we can never take this viewpoint, but when we think morally we must try our best to imagine how things would look to us if we were impartial observers. This conception of moral thinking is not, of course, peculiar to modern economics; it is the view characteristically taken by writers in the utilitarian tradition and can be traced back at least as far as Smith (1759) and Hume (1740).

One of the main themes of this book is that there is another viewpoint from which an individual can make moral judgements: his own viewpoint. Individuals living together in a state of anarchy, I shall argue, tend to evolve conventions or codes of conduct that reduce the extent of interpersonal conflict: this is spontaneous order. The origin of these conventions is in the interest that each individual has in living his own life without coming into conflict with others. But such conventions can become a basic component of our sense of morality. We come to believe that we are entitled to expect that other people respect these conventions in their dealings with us; when we suffer from other people's breaches of conventions we complain of injustice.

So, I shall argue, some of our ideas of rights, entitlements and justice may be rooted in conventions that have never been consciously designed by anyone. They have merely evolved. A society that conducts its affairs in accordance with such standards of justice may not maximize its welfare in any sense that would be recognized by an impartial observer. To put this the other way round, a benevolent government may find that it cannot maximize social welfare, evaluated from some impartial viewpoint, without violating conventions that its citizens regard as principles of justice.

It is, of course, open to the utilitarian or the welfare economist to say, 'Too bad about the citizens' ideas of justice: our duty is to maximize social welfare.' But to say this is to take a viewpoint something like that of a colonial administrator benevolently trying to advance the welfare of a native population.[3] Or, as Buchanan (1975, p. 1) puts it, there is a suggestion of playing God. In a democratic and open society, public morality cannot be something separate from the morality that guides private individuals in the conduct of their own affairs. A good deal of our private morality, I shall be arguing, has nothing to do with the rational reflections of an impartial observer. To understand it we must understand the forces of spontaneous order.

2 Games

2.1 THE IDEA OF A GAME

The notion of spontaneous order can, I shall argue, best be understood by using the theory of games. In this chapter I shall explain how I intend to use this theory, illustrating my argument with a very simple game which provides a useful model of how social conventions might evolve.

A game is a situation in which a number of individuals or players interact, and in which the outcome for each of them depends not only on what he or she chooses to do, but also on what the others choose to do. Here is a simple example, which I shall call the 'banknote game'. Two people, A and B, are taken to different rooms, and are not allowed to communicate with one another. The organizer of the game then tells each player: 'I have donated a £5 note and a £10 note to enable this game to be played. You must say which of the two notes you want to claim. If you claim the same note as the other player, neither of you will get anything; but if you claim different notes, you will each get the note you claim.' Notice that both players have an interest in the existence of some convention about who takes which note, even though they would not agree about which convention was best.

I have deliberately chosen a game that could be played in a controlled experiment, so as to exclude the complications that would be bound to arise in any discussion of real social relationships. For the present, my purpose is only to set out the logic of game theory. However, there are many real problems whose structure is similar to that of the banknote game; some of these will be discussed in chapter 3.

In the banknote game, each player has to choose one of two strategies. A can claim the £5 note (which may be called strategy A_1) or the stable equilibrium strategy. This, of course, is conditional on the value of π being sufficiently high, as defined by (7A.1)–(7A.3). Since $c_2 \geqslant c_q$, these conditions may be compressed into the single condition

$$\pi > \max\left[\frac{c_q}{v}, \frac{c_2}{c_1 - c_q}, \frac{c_1 - v}{c_1 - c_q}\right] \tag{7A.4}$$

It is a special case of this proof that the strategy of licensed free riding is a stable equilibrium for the (extended) snowdrift game. It is also a special case of this proof that the tit-for-tat strategy T_1 is a stable equilibrium for the (extended) snowdrift game and for the (extended) prisoner's dilemma game. For the snowdrift game, $q = 2$ and $v > c_1$, so the restriction on π that makes T_1 a stable equilibrium strategy is

$$\pi > \max\left[\frac{c_2}{v}, \frac{c_2}{c_1 - c_2}\right]$$

This is the result I stated in Section 7.3. For the prisoner's dilemma game, $q = 2$ and $c_1 > v$, so the restriction is

$$\pi > \max\left[\frac{c_2}{v}, \frac{c_2}{c_1 - c_2}, \frac{c_1 - v}{c_1 - c_2}\right]$$

We may reproduce the exchange-visit version of the prisoner's dilemma game – the version set out in Figure 6.1 and discussed in chapter 6 – by substituting $v = b$, $c_1 = b + c$ and $c_2 = c$. Then the restriction on π reduces to $\pi > c/b$, the result proved in Sections 6.2–6.3.

8 Natural Law

8.1 CONVENTIONS AS NATURAL LAW

In the preceding chapters I have shown how social life can be regulated by rules that evolve spontaneously and that, once established, are self-enforcing. These rules are conventions.

The conventions I have analysed fall into three broad categories. The first of these is made up of conventions of coordination – the kinds of convention I examined in chapter 3. These conventions evolve out of repeated play of games of pure coordination, like Schelling's rendezvous game, or out of games of the crossroads or 'leader' kind, in which the degree of conflict of interest between the players is relatively minor. Typical examples of these conventions in social life are: 'keep left' (or 'keep right') and 'give way' rules on the roads; the use of money; weights and measures; market-places and market days; and languages.

The second class of conventions is made up of what I shall call conventions of property – the kinds I examined in chapters 4 and 5. These conventions evolve out of the repeated play of games of the hawk–dove or chicken kind, or related games such as the war of attrition and the division game. In all of these games there is a real conflict of interest between the players: they are in dispute over something that they all want, but all cannot have. This something may be a physical object, like the $20 bill of Friedman's example (Section 5.1), or an opportunity such as the use of a public telephone or a seat on a train, or the privilege of taking a free ride on other people's contributions towards the supply of a public good. Typical examples of these conventions in social life are: the 'finders keepers' rule; the principle of 'prescriptive rights' (that is, the principle that a right can be established by long occupation or usage); the importance of 'custom and practice' in labour disputes; queues; and the principle that everyone is responsible for the tidiness of his own front garden.

The final class of conventions is made up of conventions of reciprocity – the kinds I examined in chapters 6 and 7. These conventions evolve out of the repeated play of games of the exchange-visit or prisoner's dilemma kind, or related games like the mutual-aid, snowdrift and public-good games. In these games individuals choose between strategies of 'co-operation' and 'defection'; it is contrary to the immediate interest of an individual to choose 'co-operate' but by doing so he confers benefits on others. Conventions of reciprocity prescribe that individuals should co-operate with those people who co-operate with them – but not with others. Conventions of this kind can be found in practices of mutual restraint (I respect your interests if you respect mine), mutual aid (I help you when you need my help if you help me when I need yours), trade and exchange (I keep my promises if you keep yours), and contributions towards the supply of public goods (I contribute towards goods that benefit both of us if you contribute too).

These conventions regulate interactions between individuals in situations in which their interests are in conflict. (The conflict of interest is most obvious in the case of conventions of property and of reciprocity, but there is some conflict of interest in many of the games from which conventions of coordination evolve. Only in the special case of a pure coordination game do individuals have completely common interests.) Situations of conflict of interest are ones in which we typically invoke ideas of *justice*; in cases of serious conflict we may be able to appeal to the courts. Thus conventions fulfil some of the same functions as positive laws (that is, laws promulgated by some authority, such as Parliament or Congress or the King); but whereas positive laws are the product of conscious human design, these conventions have evolved spontaneously, out of the repeated interactions of individuals with conflicting interests. In this sense, conventions of coordination, property and reciprocity are natural laws.

In saying this I am – as in so much else – following Hurne. The account I have given of the evolution of conventions is, I believe, essentially the same as Hume's account of the origin of justice – fleshed out with more details and formulated in game-theoretic terms. Hume argues that justice is a virtue 'that produce(s) pleasure and approbation by means of an artifice or contrivance, which arises from the circumstances and necessities of mankind' (1740, Book 3, Part 2, Section 1), by which he seems to mean that principles of justice are social conventions: our sense of justice is not innate in the way that our 'natural affections' (such as our feelings towards our own children) are. Hume's way of putting this is to say that justice is an 'artificial' rather than 'natural' virtue. But:

> when I deny justice to be a natural virtue, I make use of the word, *natural*, only as oppos'd to *artificial*. In another sense of the word; as no principle of the human mind is more natural than a sense of virtue; so no virtue is more natural than justice. Mankind is an inventive species; and where an invention is obvious and absolutely necessary, it may as properly be said to be natural as any thing that proceeds immediately from original principles, without the intervention of thought or reflexion. Tho' the rules of justice be *artificial*, they are not *arbitrary*. Nor is the expression improper to call them *Laws of Nature* . . .. (1740, Book 3, Part 2, Section 1)

Notice that for Hume justice is a *virtue*. Our sense of justice has evolved out of repeated interactions between individuals pursuing their own interests; but it is a *moral* sense: we believe we *ought* to keep to the 'Laws of Nature'. In Hume's words, we 'annex the idea of virtue to justice' (1740, Book 3, Part 2, Section 2).

In this respect Hume's conception of natural law should be distinguished from another conception that has been much discussed by political theorists – that of Thomas Hobbes's *Leviathan* (1651). Hobbes starts unashamedly from each individual's pursuit of his own interests. Natural law, for Hobbes, is a system of rules that it is in each individual's interest to follow – and nothing more. So far, I must concede, my approach has been essentially Hobbesian. Conventions, I have argued, are stable because once they have become established, it is in everyone's interest to keep to them. I have been more optimistic than Hobbes about the possibilities for co-operation in a state of nature, but my starting point has been the same as his. (The similarities between Hobbes's theory and those presented in this book are explored in an Appendix.) However, I now wish to follow Hume in suggesting that natural laws can come to have moral force for us. Let me make it clear that I am not presenting a moral argument: I am not going to argue that we ought to behave according to natural law. What I am going to argue is that we tend to believe that we ought to.

8.2 BREACHES OF CONVENTIONS

On a strict application of my definition of 'convention' (Section 2.8), it can never be in a person's interest to behave contrary to a convention, provided he can be sure that other people will abide by it. Nevertheless, people sometimes *do* behave contrary to the sort of practical rules that I have claimed are conventions.

One reason for this is that people can make mistakes. Conventions have to be *learned*, and a person may fail to grasp the principles of a convention, or interpret it incorrectly (that is, unconventionally) in a particular case. For example, there may be a convention that disputes over resources are resolved in favour of the possessor; but it is not always clear which disputant is the possessor. (Compare the discussion of mistakes in Section 4.7.) This problem may be compounded by wishful thinking.[1] If, according to established convention, I am the challenger, it is not in my interest to act as though I were the possessor; but I still wish the convention made me the possessor. It is a human weakness to allow our judgements about what really *is* the case to be confused by our thoughts about what we should *like* to be the case. We may also be absent-minded and behave in ways that we would not have done, had we thought carefully first. (We all occasionally break the conventions of the roads through absent-mindedness.)

A second reason is that people can suffer from weakness of will, and yield to temptations to act in ways they know to be contrary to their long-term interests. For example, in the extended prisoner's dilemma game it is in each player's short-term interest to defect: this gives the best outcome in the round in which a player first defects. If a tit-for-tat convention is established, the defector will be punished in the next round, provided the game does not come to an end first; taking a long-term view, defection does not pay. Nevertheless, there is a temptation to go for the immediate benefit that comes from defection.

A third reason is that it sometimes *is* in a person's interest to break the kind of rules that I have described as conventions. Take, for example, a 'give way' convention at a crossroads. If I am driving a large car and you are riding a bicycle, and if it is clear that you have seen me and have time to stop, it may be in my interests to pull out in front of you, even if there is an established rule that you have the right of way. Or take the case of a convention of mutual restraint between neighbours. Normally it pays me not to annoy my neighbours too much because if I do, they may retaliate; but if I am about to move house, it may no longer be in my interest to show restraint.

That such cases can arise reflects the fact that the games I have been analysing are no more than models of real life; equally, the theoretical definition of a convention as a certain kind of equilibrium strategy in a game is only a model of the practical rules that in ordinary speech would be called conventions. For the purposes of theory, it is convenient to use a single matrix of utilities to describe a whole class of interactions, such as all cases in which two vehicles meet at a crossroads, or all cases of disputes between neighbours. Nevertheless, such a matrix can represent only some kind of average of many matrices, each slightly different from every other. A convention, then, is a rule that, in the typical case, is a stable equilibrium – that it is in each player's interest to follow, provided that his opponents follow it too. But there can be atypical cases in which it is in a player's interest to break the convention, even if his opponents do not.

If we take a Hobbesian approach we shall say that the first two kinds of breaches of conventions – those arising out of mistakes and weakness of will – generate their own punishment. Thus although natural law will sometimes be broken, there will be a strong *tendency* for it to be kept. But the Hobbesian response to the third kind of breach must, I think, be that in these atypical cases conventions *will* be broken. Or, more accurately, it must be that maxims like 'Give way to the vehicle on the right' and 'Show restraint towards neighbours who show restraint towards you' are only rules of thumb. If natural law is no more than a system of rules for promoting an individual's own interest, then a fuller specification of the appropriate rules would be something more like 'Give way to the vehicle on the right – unless you are sure you can get away with not doing so' and 'Show restraint to neighbours who show restraint towards you – unless you are sure they cannot retaliate if you annoy them.'

I wish to suggest, however, that this is *not* how we typically think of the conventions that regulate social life. When we meet people who break conventions through carelessness or stupidity or weakness of will, we are not content to let them go to the devil in their own way; if their behaviour harms us we feel resentment and anger; we believe that we have been wronged. (Suppose you are driving along a British road and narrowly avoid a collision with a car that is being driven on the right-hand side of the road. Perhaps the driver is drunk, or perhaps he is an absent-minded French tourist. How would *you* react?)

Further, we recognize that conventions are rules that apply to the atypical as well as the typical cases. It is usually in our interest to follow these rules; but even if it is not, we still believe that we ought to follow them. And if we meet with opponents who break these rules we believe we have been wronged – even if we know that those opponents were acting in their own interests. (Suppose you are travelling on a crowded train, and you leave your seat to go to the lavatory. Following the usual practice, you leave a coat on the seat to mark it as yours. When you return you find the coat has been moved on to the luggage rack and a hefty young man has taken your seat. Don't you feel some sense of resentment against the man who has taken your seat, some sense that he has not merely *harmed* you but *wronged* you?)

The point of these examples is that conventions of the kind I have analysed in this book are maintained by something more than the interest that each individual has in keeping to them – most of the time. We expect that our dealings with other people will be regulated by convention, but this expectation is more than a judgement of fact: we feel *entitled* to expect others to follow conventions when they deal with us, and we recognize that they are entitled to expect the same of us. In other words, conventions are often also norms, or, to use Hume's expression, principles of natural law.

8.3 WHY OTHER PEOPLE'S EXPECTATIONS MATTER TO US

Suppose you want me to perform some action X. You also have a confident expectation, based on your experience of other people's behaviour in similar circumstances, that I will do X. In the event I do something else, leaving you worse off than you had expected to be. Then you will probably feel some resentment against me.

In order to explain this sense of resentment, I suggest, it is not necessary to call on any sophisticated moral theory. You had expected me to do X; other people, in my situation, would have done X; my not doing X has hurt you. In these circumstances resentment is a primitive human response.

It is another natural human response to feel uneasy about being the focus of another person's resentment. Because of this, our actions – and our evaluations of our actions – are influenced by other people's expectations of us. We can probably all remember foolish actions – actions that we knew to be foolish at the time – that we did merely because other people wanted us to do them and expected us to do them. Curiously, we can be motivated by what we take to be other people's expectations about us even when those other people are total strangers, and when there seems to be no solid reason for us to care about their opinions of us.

Suppose you are driving a car and waiting to pull out into a main road. It is difficult to find a gap in the traffic and you have waited some time. A queue of vehicles has built up behind you. Doesn't the mere presence of these other vehicles, with drivers who are waiting for you to pull out, put psychological pressure on you? I can only record my own response to this sort of situation. I know that the drivers behind will never remember me, even if we do happen to meet again, so there is no way they can reward me for acting in their interests or punish me for not doing so. (Thus the relationship between them and me is not like those of the games that generate conventions of mutual assistance.) But at the time it does seem to matter what the driver behind thinks of me. I know he wants me to pull out as quickly as possible; I know he has expectations about normal driving behaviour; and because of this I feel under some kind of pressure not to behave in a way that he might judge over-cautious.

Here is another example. Suppose you take a taxi ride. You know it is normal to give the driver a tip, but you have reached your destination safely and you can be as good as sure that you will have no more dealings with this particular driver. (Perhaps you are a tourist in a city to which you do not expect to return.) In any case, the driver is unlikely to remember your face. So, in the ordinary sense of the words, it is not in your interest to give a tip. Nevertheless, many people do tip in circumstances like these; others (and here, I regret to confess, I speak also from personal experience) keep their hands in their pockets – but with sensations of unease and guilt. It is one thing to adopt a policy of not tipping, and another thing to carry it off with panache. Why isn't it *easy* for us not to tip? A large part of the answer, I suggest, is that it matters to us what the taxi-driver thinks of us. We know he wants a tip. We know he expects a tip. We know that he knows that we know he expects one. If we don't tip we shall be the focus of his ill-will, if only for a few minutes. Admittedly, there is very little he can do to us; the worst we can expect is a sarcastic remark. But isn't the mere knowledge of his ill-will towards us a source of unease?

In each of these examples it is important that the other person not only *wants* us to do something but also *expects* us to do it; and his expectation is based on his experience of what other people normally do. If we were motivated simply by a desire that other people's wants should be satisfied – that is, by altruism – their expectations would not matter to us. But in these kinds of cases expectations do matter. We feel under pressure not to slow down other road users by driving in unusual ways, but we do not feel under the same kind of pressure to speed them along by showing them unexpected degrees of courtesy. We don't feel under pressure to give the taxi-driver a bigger tip than we think he expects. No doubt bus-drivers are as much in need of extra income as taxi-drivers, but we don't feel under pressure to tip them.

In the examples I have given, the people whose opinions of us matter to us are, in the game-theoretic sense, our opponents: they are people whose interests are directly affected by our actions. But theirs are not the only opinions that matter to us. When we play a game we also care about the opinions of third parties – people with no direct interest in the game, but who happen to observe it, or who are told about it afterwards. This is evident from the impulse we seem to feel, when engaged in any quarrel, to appeal to others to take our part. When we have been – as we see it – wronged, we want our interpretation of events to be confirmed by others. Even though other people may be unable to give us material help, we want them to share our resentments.

One consequence of this was noted by Adam Smith with his usual realism:

> we are not half so anxious that our friends should adopt our friendships, as that they should enter into our resentments. We can forgive them though they seem to be little affected with the favours which we may have received, but lose all patience if they seem to be indifferent about the injuries which may have been done to us: nor are we half so angry with them for not entering into our gratitude, as for not sympathizing with our resentment. (1759, Part 1, Section 1, Ch. 2)

We can take more satisfaction in our resentments when other people share them. (This is one instance of what Smith called the 'pleasure of mutual sympathy'.) Because of this, we are more prone to express – and indeed to cultivate – feelings of resentment, the more confident we are that other people regard us as in the right. Conversely, our unease at being the focus of one person's ill-will is compounded if that person has the sympathy of others. (Suppose you go back to a shop to complain about the quality of some goods you have bought. The manager is called and he refuses your demand for a refund. Aren't you emboldened if you sense that other customers in the shop are taking your side? And don't you feel under more pressure to back down if instead they seem to be on the manager's side?)

So other people's expectations of us do matter. They matter because we care what other people think of us. Our desire to keep the good will of others – not merely of our friends, but even of strangers – is more than a means to some other end. It seems to be a basic human desire. That we have such a desire is presumably the product of biological evolution. We are social animals, biologically fitted to live in communities. Some in-built tendency to accommodate oneself to others – some natural inclination to what Hobbes called 'complaisance'[2] – must surely be an aid to survival for a social animal.

It might be objected that all this has nothing to do with morality. That the taxi-driver expects me to tip him, that he wants me to tip him, that he would resent my not tipping him, that other people's sympathies would be on his side if I didn't tip him, that I should be uneasy at being the focus of this resentment and ill-will – these propositions may all be true, but can they entail that I *ought* to give a tip? If this is a question about the logic of moral propositions, then the answer must be 'No'. 'Ought' statements cannot be derived from 'is' statements by any logically valid chain of reasoning: this is Hume's Law[3], which I have no intention of questioning. There would be nothing self-contradictory in saying: 'I know the taxi-driver expects to be tipped, I know he wants to be tipped, etc., but I have no moral obligation to tip him'.

But my argument is not about the logic of moral propositions; it is about the psychology of morals. It is a matter of common experience, I suggest, that we are strongly inclined to believe that we ought to do what other people want and expect us to do; when we go against other people's wants and expectations, we are inclined to feel guilt. Any plausible theory of moral learning – of how we judge some things right and others wrong – would surely have to assign a good deal of importance to praise and blame. We learn to think wrong those actions that other people censure. That other people censure them does not *make* them wrong; but it is a powerful force influencing us to *judge* them wrong.

Some readers, I suspect, will still object that I am misusing the terms 'right' and 'wrong', 'praise' and 'blame'. The psychological urge we feel to meet other people's wants and expectations, the objection would run, may be real enough; but it is an improper use of words to describe this urge as a sense of moral obligation. Equally, it might be said, it is improper to describe the resentment we feel when other people frustrate our expectations as moral censure, or to describe our pain at being the subject of such resentment as a sense of guilt. If I am to answer this objection I must make clear what I mean when I use words like 'right' and 'wrong'.

My position is a Humean one. The famous passage in which Hume presents what has come to be called his 'law' is part of a section headed 'Moral Distinctions not deriv'd from Reason'. He presents the following argument in support of his 'law':

> But can there be any difficulty in proving, that vice and virtue are not matters of fact, whose existence we can infer by reason? Take any action allow'd to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. Here is a matter of fact; but 'tis the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it. (1740, Book 3, Part 1, Section 1)

One implication of this position – which is fairly conventional among economists, if not among philosophers – is that moral propositions can be derived only from other moral propositions. A person's moral beliefs must therefore include some beliefs that cannot be justified by any appeal to reason or evidence. Following Sen (1970, pp. 59–64), economists often call these non-justifiable beliefs 'basic value judgements'. If we want to explain why a person subscribes to one basic value judgement rather than another, we cannot use *moral* reasoning. (We cannot say that the reason why we tend to believe that killing people is wrong is because killing people *is* wrong.) The explanation must be a psychological one: as Hume put it, we must look to the constitution of human nature.

Does this mean that a moral judgement is no more than a personal preference? No, because a moral judgement is a statement of *approval* or *disapproval*, and this is not the same thing as a like or dislike. And moral judgements are more than bald reports of sensations of approval or disapproval. Hume put it like this:

> The approbation of moral qualities most certainly is not deriv'd from reason, or any comparison of ideas; but proceeds entirely from a moral taste, and from certain sentiments of pleasure or disgust, which arise upon the contemplation and view of particular qualities or characters. Now 'tis evident, that those sentiments, whence-ever they are deriv'd, must vary according to the distance or contiguity of the objects; nor can I feel the same lively pleasure from the virtues of a person, who liv'd in *Greece* two thousand years ago, that I feel from the virtues of a familiar friend and acquaintance. Yet I do not say, that I esteem the one more than the other . . .. Our situation, with regard both to persons and things, is in continual fluctuation . . . and 'tis impossible we cou'd ever converse together on any reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. In order, therefore, to prevent those continual *contradictions*, and arrive at a more *stable* judgement of things, we fix on some *steady* and *general* points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation. (1740, Book 3, Part 3, Section 1)

In other words, it is a *convention of language* that moral judgements are invariant with respect to changes in viewpoint. I stress 'convention of language' because this is not a hypothesis about human psychology. We are naturally inclined to disapprove more strongly of actions, the more their ill-effects impinge on us; but we use moral language to express our general disapproval of *classes* of actions. Moral judgements are in this sense universalizable.[4]

My position is that any universalizable statement of approval or disapproval is a moral judgement. Suppose there is some established convention that anyone in a particular situation should behave in a particular way, for example the convention that queues should be respected. Suppose that I should feel a strong sense of disapproval if you pushed past me in a queue. Suppose also that if I, as a bystander, saw you push past someone else in a queue, I should still disapprove, even if my resentment against you would be less lively than in the first case. Finally suppose that if *I* pushed past *you* in a queue, I should feel some degree of guilt, that is, I should not altogether approve of my own action. Then my disapproval of queue-jumping is universalizable: it is, on my account, a moral judgement.

8.4 CONVENTIONS, INTERESTS AND EXPECTATIONS

It is in the nature of an established convention that everyone *expects* everyone else to keep it. Is it also in everyone's *interest* that everyone else keeps to it? If it is, then we should expect breaches of conventions to provoke general resentment and censure; and this will predispose people to the moral judgement that conventions ought to be kept.

I shall now examine the extent to which it is in one individual's interest for other individuals to follow conventions. I shall consider these conventions in their pure forms, that is, as equilibrium strategies in 'typical case' games. Strictly speaking, then, my arguments will apply only to those breaches of conventions that arise out of mistakes and weakness of will. I am inclined to think that my conclusions will carry over to most of the atypical cases in which it is in an individual's interest to behave contrary to convention, but I cannot prove this. (A proof would require a catalogue of all the atypical forms that the various games could take, and an analysis of each; this would be a huge task.) I shall argue that conventions are normally maintained by *both* interest *and* morality: it pays us to keep to conventions and we believe we ought to keep to them. If this belt-and-braces conclusion is correct, then it is at least plausible to suppose that morality may sometimes motivate us to behave according to natural law even when interest does not.

The question, 'Is it in one person's interest that other people follow a convention?' can be posed in various ways; but if we are concerned with the possibility that breaches of conventions will provoke resentment, the most obvious question to ask seems to be this: 'If one individual follows a convention, is it in his interests that his opponents follow it?'.

The connection between this question and the possibility that people will resent breaches of convention *by their opponents* is obvious enough. If a convention is established, each individual will expect his opponents to follow it. Because he has this expectation, it is in his interests to follow it himself. Then if the answer to my question is 'Yes', each individual will not only *expect* his opponents to follow the convention; he will also *want* them to. Resentment in the face of breaches of the convention would be a natural response.

What about the responses of bystanders? In what sense can anyone have an interest in whether or not other people keep to a convention in a game in which he is not involved? The most obvious answer seems to be this. If the bystander is a member of the community within which the game is played, then the individuals whom he observes may be his opponents in future games. Thus the bystander's interest in the behaviour of other people is the interest of a potential opponent. Take an example. Suppose you are driving along a British road, and see two drivers, A and B, narrowly miss an accident. The cause of the problem was that A was driving on the left while B was driving on the right. It will surely occur to you that B and drivers like him are a threat, not just to A, but *to you.* So you have some cause for resentment against B, which will predispose you to take A's side. This brings us back to the question, 'If one individual follows a convention, is it in his interests that his opponents follow it?' If, for a given convention, the answer is 'Yes', then breaches of the convention are likely to provoke *general* resentment – from bystanders as well as from opponents.

However, *is* the answer to this question 'Yes'? There is nothing in my definition of 'convention' that entails that everyone wants his opponent to follow an established convention (although this *is* entailed by David Lewis's definition: cf. Section 2.8). I shall therefore consider in turn the three kinds of convention I have investigated in this book.

Conventions of Coordination

The crossroads game (or leader game) is typical of the games from which conventions of coordination evolve. As a typical convention, consider the rule 'Give way to vehicles approaching from the right' in the crossroads game. It is easy to see that an individual who follows this convention has an interest in his opponents doing the same. This is true even if the convention favours the opponent. If you are approaching from my right, then the convention of 'priority to the right' favours you; in this instance, at least, I should prefer the convention to be 'priority to the left'. Nevertheless, it is in my interest to slow down, because I expect you to follow the established convention; and because I shall slow down, it is in my interest that you maintain speed. In other

words, I want you to follow the convention that has actually become established, even though I may wish we all followed some other convention.

This example may seem slightly artificial, because it is in the nature of a convention that it applies to a wide class of cases, and *in the long run* the rule of 'priority to the right' does not seem to favour anyone relative to anyone else. But consider the convention 'Cars give way to buses'. (Clearly this rule would resolve only some instances of the crossroads game, but for games involving a car and a bus it is a perfectly workable convention. Compare the old nautical rule 'Steam gives way to sail'.) If I am a car driver who never travels by bus, this convention favours my opponent in every game to which it applies; taking a long-run view, then, I should certainly prefer the opposite convention to be established. But even so, if 'Cars give way to buses' is the established convention, I do not want *individual* bus drivers to try to give way to cars.

There is, however, one significant difference between 'Priority to the right' and 'Cars give way to buses'. Notice that they are both asymmetrical conventions – that is, they both exploit some asymmetry between the roles of the two players in a game. (In contrast, 'Keep left' as a convention for dealing with cases where two vehicles approach one another head-on is a symmetrical convention.) But 'Priority to the right' exploits what I shall call a cross-cutting asymmetry. By this I mean that in the community of players of the crossroads game, each individual will sometimes find himself on one side of the asymmetry and sometimes on the other. (If, as is surely true in the case of 'Priority to the right', each individual has *the same* probability of being assigned each role as every other individual has, I shall say that the asymmetry is perfectly cross-cutting.) In contrast, the asymmetry exploited by 'Cars give way to buses' is *not* cross-cutting: there are many car drivers who never drive or ride in buses and perhaps a few bus drivers who never have anything to do with cars.

This distinction is significant because of its implications for third parties. If a convention of coordination exploits a cross-cutting asymmetry, *everyone* stands to be harmed by the existence of mavericks who break the convention. (Suppose the established convention is 'Priority to the right'. My friend tells me that *he* never bothers about this convention, but instead always tries to force the other driver to slow down. It is bound to occur to me that some day I might meet a fool like him approaching from my left.) In contrast, if a convention exploits an asymmetry that is *not* cross-cutting, there are at least some people who can never meet one another as opponents in games to which the convention applies. Such people may be more inclined to make light of one another's breaches of conventions.

To sum up, then, provided an established convention exploits a cross-cutting asymmetry, everyone has an interest in everyone else's keeping to it. This is true even if some people (or even if everyone) would prefer a different convention to have become established. This suggests that conventions of coordination are likely to acquire moral force: they are likely to become norms, or principles of natural law.

Conventions of Property

The hawk–dove or chicken game is typical of the games from which conventions of property evolve. As a typical convention, consider the rule: 'If possessor, play "hawk"; if challenger, play "dove"'. This, like most conventions of property,[5] is asymmetrical. In most practical cases, the asymmetry between possessor and challenger is cross-cutting: everyone is the possessor in *some* conflicts. In some cases this cross-cutting is close to perfect – consider the convention of queueing, or the convention that a passenger on a train may reserve his seat by leaving his coat on it. In other cases it is not – consider the principle of prescriptive rights, which systematically favours those who have been most successful in holding on to valuable resources in the past, or who are lucky enough to have the right kind of ancestors. But for the argument that follows, all that matters is that conventions of property exploit asymmetries that are *to some degree* cross-cutting.

Suppose some individual follows the convention: 'If possessor, play "hawk"; if challenger, play "dove"'. Is it in his interests that his opponents follow it too? This clearly depends on the individual's role in a game. When he is the possessor it is in his interests that his opponent follows the convention: if someone commits himself to fighting, he wants his opponent to back down. Equally obviously, when an individual is the challenger, it is *not* in his interests that his opponent follows the convention: he is ready to back down in the face of aggression, but doesn't *want* to meet aggression. Similar conclusions apply to the division game and the war of attrition.

In all these games of property, strategies vary in their degrees of aggression – whether this is represented by the difference between 'dove' and 'hawk', by the amount of a disputed resource that is claimed, or by the length of time a player is prepared to fight before surrendering. Conventions prescribe appropriate degrees of aggression for each player. It is characteristic of all these games that each player prefers that his opponent should be less aggressive rather than more. (There are some cases in which a player may be indifferent about his opponent's degree of aggression, at least within some range; but there are no cases in which a player would positively prefer greater aggression by his opponent.) Thus if an individual follows a convention, he wants his opponent's behaviour to be no more aggressive than the convention prescribes.

This suggests that conventions of property are likely to be associated with norms forbidding excessive aggression: people will tend to believe that it is wrong for anyone to press a claim that is not supported by convention. In contrast, unconventional meekness – as prescribed, for example, by the Christian ethic of turning the other cheek – seems less likely to provoke moral censure.[6] Of course, this is not to say that we should expect to find much in the way of Christian meekness, since it is highly *imprudent* to be less aggressive than an established convention allows one to be.

Notice that a convention of property may become a generally accepted norm even though it cannot be justified in terms of any external standard of fairness. Having become a norm, a convention *becomes* a standard of fairness; but, on my account, it does not become a norm *because* it is seen to be fair. Equally, a person may believe that everyone ought to follow an established convention even though that convention systematically favours others at his expense. For example, suppose there is an established convention that each person retains possession of those things he has possessed in the past. The corresponding norm against over-aggression is the Old Testament one: 'Thou shalt not steal'. Clearly, this convention favours some people much more than others. Those who start out in life possessing relatively little would much prefer many other conventions – for example, a convention of equal division – to the one that has become established. Nevertheless, it is in each individual's interest to follow the established convention, given that almost everyone else does. And once a person has resolved to follow the convention, his interests are threatened by the existence of mavericks who are aggressive when the convention prescribes submission. Or in plainer English: provided I own *something*, thieves are a threat to me. So even if the conventions of property tend to favour others relative to me, I am not inclined to applaud theft.

Conventions of Reciprocity

The extended exchange-visit or prisoner's dilemma game is typical of the games from which conventions of reciprocity evolve. As a typical convention, consider the tit-for-tat convention I called T_1 (Sections 6.2–6.3). Recall that this convention is defined for an extended game in which players occasionally make mistakes. T_1 is a symmetrical convention prescribing that each player should co-operate as long as his opponent co-operates, and should punish to a prescribed degree an opponent who defects in breach of the convention. It also prescribes that a player who defects by mistake should accept his punishment and not retaliate. Suppose that some individual A follows this convention. Is it in his interests that his opponents do the same?

Suppose that A follows the convention T_1 in a game against B, and suppose that A never makes a mistake. How will he want B to behave? One feature of T_1 is that, although the punishment prescribed for breaches of the convention is enough to deter deliberate defections, it is not enough to make full reparation to the injured party. This reflects the fact that T_1 is the most 'forgiving' tit-for-tat strategy that is capable of constituting a stable equilibrium (Section 6.3). Thus if A follows T_1 he will not want B to defect in breach of the convention; even if B were then to accept his punishment without retaliating, A would still be left worse off than if there had been no breach. But if B *does* defect, it is in A's interest that B should not retaliate when A punishes him. So if A follows T_1 without making mistakes, it is in his interest that B should do the same; and if B does make a mistake, it is in A's interest that B should continue the game in the way prescribed by T_1.

What if A tries to follow T_1 but defects by mistake? Suppose that in some round i, T_1 prescribes co-operation for A, but he defects; B co-operates. Then, following T_1, A will co-operate in round $i+1$; and, irrespective of what B does in round $i+1$, A will co-operate again in round $i+2$. (In other words, if B defects in round $i+1$, A will accept this as a legitimate punishment and not retaliate.) If B follows T_1 he *will* defect in round $i+1$; he will punish A for his defection. But A does not want to be punished; he would prefer it if B co-operated. So in this case, but only in this case, A would prefer B not to keep to the convention.

To sum up: if an individual follows a convention of reciprocity, he wants his opponents to be no less co-operative than the convention prescribes. This suggests that conventions of reciprocity are likely to be associated with norms forbidding non-co-operative behaviour in situations in which the convention calls for co-operation. In other words, people are likely to subscribe to an ethic of reciprocity, according to which each person ought to co-operate with those who are prepared to co-operate with him.

Earlier in this section I suggested that our sense of morality may sometimes motivate us to follow the dictates of natural law even when it is contrary to our interests to do so. Clashes between interest and morality seem particulary likely to occur in cases where public goods can be supplied only through the voluntary sacrifices of many individuals. From our experience of playing the public-good game in small groups – within the family, among neighbours and friends and workmates – we learn that there are established conventions of reciprocity and that it is generally in our interest to follow them. Around these conventions a system of morality grows up; we come to recognize a moral obligation to play our part in co-operative arrangements and learn to condemn those who try to take free rides on other people's efforts. When we play the public good game in large groups, however, we find that conventions of reciprocity are fragile. It is often *not* in our interests to be co-operative: free riding often pays. But we may still feel the force of the ethic of reciprocity. We may still believe that we *ought* to shoulder our share of the costs of co-operative arrangements – that if others are doing their part, we ought to do ours. It is because we subscribe to some such ethic, I suggest, that even within large groups public goods are sometimes supplied through voluntary contributions.[7]

8.A APPENDIX: RECIPROCITY IN HOBBES'S THEORY OF NATURAL LAW

Hobbes's theory, as set out in *Leviathan* (1651), begins with the following definition:

> A LAW OF NATURE, *lex naturalis*, is a precept or general rule, found out by reason, by which a man is forbidden to do that, which is destructive of his life, or taketh away the means of preserving the same; and to omit that, by which he thinketh it may be best preserved. (Ch. 14)

For Hobbes, then, natural law is sheer prudence. Men are basically self-interested; natural laws are answers to the question, 'How best can I promote my self-interest?' or – since Hobbes always emphasizes the dangers men face from one another – 'How best can I ensure my survival in a dangerous world?'. My starting-point – which I think is broadly in line with Hume's – is not dissimilar to this. I have not gone so far as Hobbes in assuming individuals to be selfish; I have assumed only that they have *conflicting* interests (Section 2.3). But I have assumed that each individual is concerned only with promoting *his own* interests (whether selfish or not: my interest in my son's welfare is not selfish, but it is *my* interest). The conventions of coordination, property and reciprocity that

I have called natural laws are grounded in each individual's pursuit of his own interests; they are answers to the question, 'How best can I promote my interests in a world in which other people are promoting theirs?'.

Notice, however, that Hobbes's natural laws are found out by reason. The idea seems to be that natural laws can be *deduced* by a chain of logic from a few self-evident first principles. This is in marked contrast to Hume's idea that natural laws *evolve* and are *learned by experience.* If natural laws can be found out by reason, then presumably there is a unique code of natural law that can be discovered by any rational person in any society. This leaves no room for the possibility that some natural laws might be conventions – rules that have evolved in particular forms in particular societies, but that might have evolved otherwise.

So much for what Hobbes *means* by natural law. What about its *content*? Hobbes formulates no fewer than nineteen laws of nature, but the core of his system seems to be contained in his first three laws. The first law of nature is that every man should 'seek peace, and follow it'; this is presented as part of a more comprehensive rule:

> it is a precept, or general rule of reason, *that every man, ought to endeavour peace, as far as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of war.* (Hobbes, 1651, Ch. 14)

In the state of nature – that is, in a society without government – no man has any hope of obtaining peace; and so, in accordance with Hobbes's 'general rule of reason', there is a state of war of all against all: every man, Hobbes says, has a right to every thing. The second law of nature expands on the precept that every man ought to endeavour towards peace. This law is:

> that a man be willing, when others are so too, as far-forth, as for peace, and defence of himself he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himself. (Hobbes, 1651, Ch. 14)

This requires that men be willing to make covenants with one another; if these covenants are to be more than empty words, there must be a third law of nature: *'that men perform their covenants made'.* This law is 'the fountain and original of JUSTICE' (Ch. 15). But in the state of nature this law has little force, because:

> If a covenant be made, wherein neither of the parties perform presently, but trust one another; in the condition of mere nature, which is a condition of war of every man against every man, upon any reasonable suspicion, it is void: but if there be a common power set over them both, with right and force sufficient to compel performance, it is not void. For he that performeth first, has no assurance the other will perform after . . .. And therefore he which performeth first, does but betray himself to his enemy; contrary to the right, he can never abandon, of defending his life, and means of living. (Hobbes, 1651, Ch. 14)

Hobbes is describing a problem whose structure seems rather like that of the prisoner's dilemma game – and particularly like the version of that game that I called the trading game (Section 6.1; see also Taylor, 1976, pp. 101–11). In the state of nature, two or more people may be able to benefit from some agreement – provided that all the parties to the agreement keep it. But each party is tempted to make the agreement and then break it, in the hope that the others will still keep to it. Since everyone knows that everyone else is tempted in this way, no one can trust anyone else to keep an agreement; and so agreements can never be made. This problem stands in the way of any escape from the war of all against all, since the only way to achieve peace – which everyone prefers to war – is by an agreement to cease fighting.

Contrary to my arguments in this book, Hobbes seems to be claiming that this problem has no solution *within the state of nature*; agreements can be made only if there is a 'common power' set over all individuals with sufficient force to compel them to keep agreements. This is why, according to Hobbes, everyone will agree to subject himself to some sovereign power, provided that everyone else does the same; once this agreement has been made, the state of nature is at an end.

For my point of view, however, what is most interesting about Hobbes's laws of nature is that their central principle seems to be reciprocity. Each person must 'endeavour peace, as far as he has hope of obtaining it'; as the second law of nature makes clear, this means that each must be prepared to make peace provided that everyone else makes peace too. Similarly, each person must keep his part of any agreement he has made, provided that the other parties to the agreement keep theirs. To aim for more than this – to refuse to make peace even though everyone else is willing, to break your side of an agreement even though you know the other side will be kept – is contrary to natural law. In other words, it is contrary to rational self-interest. As Hobbes puts it, 'justice [is] not contrary to reason' (1651, Ch. 15).

Hobbes takes the case of an agreement between two individuals, according to which one performs his part of the agreement before the other. (Compare Hume's case of the two farmers – see Section 6.1.) Suppose 'one of the parties has performed already'. Then:

> there is the question whether it be against reason, that is, against the benefit of the other to perform, or not. And I say it is not against reason. For the manifestation whereof, we are to consider; first, that when a man doth a thing, which notwithstanding any thing can be foreseen, and reckoned on, tendeth to his own destruction, howsoever some accident which he could not expect, arriving may turn it to his benefit; yet such events do not make it reasonably or wisely done. Secondly, that in a condition of war, wherein every man to every man, for want of a common power to keep them all in awe, is an enemy, there is no man who can hope by his own strength, or wit, to defend himself from destruction, without the help of confederates; where every one expects the same defence by the confederation, that any one else does: and therefore he which declares he thinks it reason to deceive those that help him, can in reason expect no other means of safety, than what can be had from his own single power. He therefore that breaketh his covenant, and consequently declareth that he thinks he may with reason do so, cannot be received into any society, that unite themselves for peace and defence, but by the error of them that receive him; nor when he is received, be retained in it, without seeing the danger of their error; which errors a man cannot reasonably reckon upon as the means of his security . . .. (Hobbes, 1651, Ch. 15)

Hobbes's argument here seems rather like the game-theoretic arguments I presented in chapters 6 and 7, showing that strategies of reciprocity can be stable equilibria. Hobbes is saying that in the state of nature self-interest will lead each individual to follow the strategy: 'Keep agreements only with those who keep agreements with others'. (This seems to amount to what I have called 'multilateral reciprocity' – compare the mutual-aid game of Section 7.2.) It is in each individual's self-interest to follow this strategy, provided everyone else does.

It seems that, after all, Hobbes recognizes that *some* agreements will be made and kept in the state of nature. How else could men combine into 'confederations' for self-defence? And Hobbes specifically mentions the case of a 'covenant to pay a ransom . . . to an enemy' which 'in the condition of mere nature [is] obligatory' (1651, Ch. 14). These agreements work because self-interest leads everyone to follow a strategy of reciprocity. However, Hobbes is extremely pessimistic about the *extent* of co-operation that can be expected in the state of nature. Here I have to confess that I cannot understand Hobbes's argument. He seems to be saying that, although it is rational (i.e. in your self-interest) to perform your part of an agreement if the other party has already performed, it is not rational to perform first because you have no assurance that the other party will perform after you. This seems inconsistent: if you know that it is in the other party's interest to perform, why doesn't this give you the assurance you need? Hobbes's picture of the state of nature seems something like an extended prisoner's dilemma game in which everyone is following a strategy of cautious reciprocity; since no one is willing to make the first move, no one ever co-operates. But if the analysis of Section 6.4 is correct, this is not necessarily a stable equilibrium. In a world of cautious reciprocators it may pay to be brave: it may be in each individual's interest to make the first move. In other words, co-operation *can* evolve in a Hobbesian state of nature.

9
Rights, Co-operation and Welfare

9.1 SYMPATHY AND SOCIAL WELFARE

If my argument so far is right, a rule is likely to acquire moral force if it satisfies two conditions:

1 Everyone (or almost everyone) in the relevant community follows the rule.
2 If any individual follows the rule, it is in his interest that his opponents – that is, the people with whom he deals – follow it too.

Any rule that is a convention necessarily satisfies a third condition:

3 Provided that his opponents follow the rule, it is in each individual's interest to follow it.

Notice that none of these conditions requires any comparison to be made between a world in which the rule is generally followed and one in which it is not. This leads to an implication that many will find surprising: a convention can acquire moral force without contributing to social welfare in any way.

Take, for example, those conventions of property that resolve disputes in favour of possessors. Such conventions, I have argued, are likely to become norms. Yet in many cases they maintain inequalities that seem arbitrary from any moral point of view (except, of course, a viewpoint that makes morality a matter of convention). The asymmetry between possessor and challenger tends to be prominent, unambiguous and cheat-proof, so we can easily understand how such a convention might have

evolved; but there seems no good reason to expect that the distribution of property that it generates will be one that can be justified in terms of any coherent conception of social welfare.

It might be objected that although a convention favouring possessors is morally arbitrary, it is in everyone's interest that everyone follows *some convention or other*. Perhaps the established convention is not quite the best there could possibly be; but things would be much worse for everyone if there were no established convention at all. On this argument, conventions of property *do* contribute to social welfare, however arbitrary they may be. But is it true that conventions of property necessarily work to everyone's benefit?

The hawk–dove game provides a theoretical counter-example. Suppose there is an established convention that possessors play 'hawk' and that challengers play 'dove'. Then, using the utility indices I presented in Table 4.1, possessors derive a utility of 2 from each game while challengers derive a utility of 0. So if a particular individual has a probability p of being the possessor in any game, the expected utility of each game for him is $2p$. What if instead there were no established convention at all? One way of representing such a state of affairs is as an equilibrium state of a *symmetrical* game. (Recall that this is how I modelled the Hobbesian state of nature.) The symmetrical hawk–dove game has a unique and stable equilibrium in which the probability that any player will play 'dove' is 0.67; this gives each individual an expected utility of 0.67 (Section 4.2). More pessimistically, we might represent the absence of any established convention by assuming that everyone chooses strategies entirely at random. If the probability that any player will play 'dove' is 0.5, the expected utility for each player is 0.25. In either event, a player with a sufficiently low – but still non-zero – value of p would be better off without the convention. In other words, a person who possesses relatively little might be better off taking his chances in a state of nature. Nevertheless, the convention that favours possessors may acquire moral force for *everyone* in the community, including those who would be better off without it.

This conclusion may be surprising, but the form of my argument is not new. Lewis argues that conventions of coordination[1] often become norms. He writes that if other people

> see me fail to conform [to a convention of coordination], not only have I gone against their expectations; they will probably be in a position to infer that I have knowingly acted contrary to my own preferences, and contrary to their preferences and their reasonable expectations. They will be surprised, and they will tend to explain my conduct discreditably. (1969, p. 99)

This is essentially the same kind of argument as I presented in chapter 8; notice how it appeals to the fact that a convention of coordination satisfies the three conditions I have listed. However, the kinds of rules that Lewis considers can be said to serve everyone's interests. Lewis defines conventions as solutions (of a certain kind) to 'coordination problems'. These are games whose structure approximates to that of a pure coordination game. The players' main concern is to coordinate their strategies in some way or other; in *which* way they coordinate their strategies is a matter of relatively little importance to them (Lewis, 1969, pp. 5–24). Thus although individuals may have preferences between alternative conventions, each has a stronger interest that some convention becomes established rather than none. In this sense, at least, an established convention serves everyone's interests.

A similar argument can be found in Ullman-Margalit's *The Emergence of Norms* (1977, p. 88). Ullman-Margalit follows Lewis's argument very closely – as far as conventions of coordination are concerned. But, significantly, she is unwilling to apply to other kinds of games the same logic as she applies to coordination problems. She recognizes that rules of property can correspond with stable equilibrium strategies in games of conflict,[2] and that these rules can become norms; but her explanation of *how* these rules become 'norms of partiality' seems entirely separate from her earlier account of the evolution of 'coordination norms'. If I read her correctly, norms of partiality are supposed to come about because they fortify a discriminatory status quo and in doing so promote the interests of those who are favoured in that status quo (Ullman-Margalit, 1977, Ch. 4, especially pp. 173–97). The implication is that when a convention of property becomes a norm, those who are not favoured by that convention are somehow being hoodwinked into approving it.

There seems no good reason, however, to confine Lewis's argument to conventions of coordination. Lewis's explanation of how these conventions become norms does not depend on his assumption that they work to everyone's advantage. It may be true that we all benefit from the existence of certain established conventions, but this is not *why* we believe we ought to follow them. So we should not be surprised if we find that other conventions, which do not work to everyone's advantage, also become norms.

In arguing that conventions can acquire moral force without working in the interests of society as a whole I am making something of a break with the arguments of Hume. Hume claims that conventions of property ultimately work to everyone's benefit:

> 'Tis impossible to separate the good from the ill. Property must be stable, and must be fix'd by general rules. Tho' in one instance the public be a sufferer, this momentary ill is amply compensated by the steady prosecution of the rule, and by the peace and order, which it establishes in society. And even every individual person must find himself a gainer, on ballancing the account; since, without justice, society must immediately dissolve, and every one must fall into that savage and solitary condition, which is infinitely worse than the worst situation that can possibly be suppos'd in society. (1740, Book 3, Part 2, Section 2)

I have already explained why I cannot accept that this optimistic conclusion is necessarily true: some people may be better off in a state of nature than in a society with rules of property that discriminate against them.

The claim that conventions of property work to everyone's benefit plays an important part in Hume's account of how these conventions become norms – or, in Hume's words, of why we annex the idea of virtue to justice. For Hume the rules of justice are what I have called conventions of property. These conventions evolve spontaneously in a state of nature in which each individual pursues his own interests. The original motive for individuals to follow these conventions is simple prudence:

> To the imposition then, and observance of these rules, both in general, and in every particular instance, they are at first mov'd only by a regard to interest; and this motive, on the first formation of society, is sufficiently strong and forcible.

But these conventions come to have moral force:

> But when society has become numerous, and has encreas'd to a tribe or nation, this interest is more remote; nor do men so readily perceive, that disorder and confusion follow upon every breach of these rules, as in a more narrow and contracted society. But tho' in our own actions we may frequently lose sight of that interest, which we have in maintaining order, and may follow a lesser and more present interest, we never fail to observe the prejudice we receive, either mediately or immediately, from the injustice of others; as not being in that case either blinded by passion, or byass'd by any contrary temptation. Nay when the injustice is so distant from us, as no way to affect our interest, it still displeases us; because we consider it as prejudicial to human society, and pernicious to every one that approaches the person guilty of it. We partake of their uneasiness by *sympathy*; and as every thing, which gives uneasiness in human actions, upon the general survey, is call'd Vice, and whatever produces satisfaction, in the same manner, is denominated Virtue; this is the reason why the sense of moral good and evil follows upon justice and injustice. And tho' this sense, in the present case, be deriv'd only from contemplating the

> actions of others, yet we fail not to extend it to our own actions. The *general rule* reaches beyond those instances, from which it arose; while at the same time we naturally *sympathize* with others in the sentiments they entertain of us. (Hume, 1740, Book 3, Part 2, Section 2)

I have quoted this passage at length because it so closely parallels my own argument about how conventions become norms. When other people breach conventions in their dealings with us, we are harmed; and we resent this. This, I take it, is the prejudice we receive 'immediately' from the injustice of others. When other people breach conventions in dealings in which we are not involved, our interests are still endangered, because we may have to deal with these people in the future. This is perhaps what Hume means when he speaks of the prejudice we receive 'mediately'.[3] And we feel uneasy about being the focus of other people's resentment: we 'naturally sympathize with them in the sentiments they entertain of us'. Our disapproval of other people's breaches of conventions, and our uneasiness about our own breaches, are universalized in our acceptance of the general rule that *all* breaches are to be disapproved of: conventions ought to be kept.

Hume's argument diverges from mine, however, in stressing the role of *sympathy*. According to Hume, we are displeased by breaches of conventions even in cases in which our interests are completely unaffected; and our displeasure stems from sympathy. Whether this claim is compatible with my own argument depends on how we suppose sympathy works.

One conception of sympathy is endorsed by Hume when he says that the happiness or misery of any human being, or indeed of any animal capable of these feelings, can affect us *'when brought near to us, and represented in lively colours'* (1740, Book 3, Part 2, Section 1; my italic). The idea here is that the extent of our sympathy with another person is usually a product of the relationship between him and us: his happiness or misery has to be *brought near to us* if it is to engage our sympathies. Thus, other things being equal, we tend to sympathize most strongly with those people whose situations are most like our own: these are the people with whom we can most easily identify. Now suppose that I always follow a particular convention. For example, suppose that I never pick pockets. This respect for established rules of property may be a matter of simple prudence: I am not particularly dextrous, and afraid of being caught. But whatever the reason for my keeping to the convention, I have no *experience* of the pickpocket's satisfaction at successfully completing a job of work. In contrast, I do have experience of the fear of having my pocket picked; I may also have experience of feeling anger and resentment after being robbed in this way. Thus, I suggest, I shall be inclined to sympathize less with pickpockets than with their victims. More generally, if I follow a convention I shall be inclined to sympathize less with those who breach it than with those who are harmed by these breaches. Thus, to use Hume's words, we can be displeased by injustice because of a natural tendency to sympathize with the uneasiness of those who are the victims of injustice.

So far there is no contradiction with the argument of this book. But, according to Hume, we sympathize with the victims of injustice because we consider injustice 'prejudicial to human society'. His position is made more clear in his summing-up of the passage I have just quoted:

> *Thus self-interest is the original motive to the* establishment *of justice: but a* sympathy *with public interest is the source of the* moral approbation, *which attends that virtue* (1740, Book 3, Part 2, Section 2).

Notice that Hume is talking about sympathy *with public interest*. The idea here seems to be that we sympathize impartially with everyone's pleasures and pains; because principles of justice work in the public interest, the balance of our sympathies come down on the side of justice. On this view of sympathy, conventions can acquire moral force only if they contribute to the overall welfare of society.

It is at this point that I part company with Hume. The idea that we sympathize on the basis of some kind of cost–benefit analysis seems psychologically implausible. No doubt we are *capable* of imagining ourselves into the position of what Adam Smith (1759, Part 3, Ch. 3) called an 'impartial spectator', sympathizing equally with everyone in society; but, as Smith himself recognized,[4] this is not how our sympathies *naturally* work.

To this it might be objected that the concept of a moral judgement requires a degree of impartiality; however partial our sympathies may be, our moral judgements must be universalizable. Or, as Hume put it, it is a convention of language that we 'fix on some *steady* and *general* points of view' when we make moral judgements (cf. Section 8.3). I accept this; but we can be impartial without being impartially *sympathetic*. A judge who impartially upholds the law of property need not (and, I think, will not) always decide cases in the way that would be dictated by an equal sympathy for every member of society; but he is nevertheless impartial. Similarly, if I am prepared to condemn all breaches of a convention, including my own, my condemnation is sufficiently steady and general to be recognizable as moral; I need not believe that the convention works for the overall welfare of society.

Some readers may still think that conventions are too arbitrary to form the basis for a system of morality; moral judgements, it might be argued, should follow from the impartial application of a few simple and general moral principles. To answer this objection, I shall try to show that the morality that grows up around conventions – the morality of natural law – *does* have a unifying principle.

9.2 THE PRINCIPLE OF CO-OPERATION

In this book I have argued that certain kinds of conventions tend to evolve spontaneously in human society, and that these conventions come to have the moral status of principles of justice, of natural law. It is tempting to suppose that if the members of a society subscribe to a common moral code, then that code must serve some social purpose. There must be *some* sense, we are tempted to say, in which this code is good for society. But this is a mistake. As I have argued in Section 9.1, conventions can acquire moral force without contributing to the overall welfare of society. So if there is a unifying principle behind natural law, it is not a principle of social welfare.

Nevertheless it *is* possible to extract a general principle from my argument about how conventions acquire moral force. Recall that, according to my argument, a convention is likely to acquire moral force if it satisfies two conditions: first, that almost everyone in the relevant community follows it; and second, that it is in each individual's interest that the people with whom he deals follow the rule, provided that he follows it too (Section 9.1). (The first condition ensures that everyone *expects* everyone else to follow the convention; the second condition ensures that everyone *wants* everyone else to follow it.) So the moral rules that grow up around conventions are likely to be instances of the following principle:

> *The principle of co-operation.*[5] Let *R* be any strategy that could be chosen in a game that is played repeatedly in some community. Let this strategy be such that if any individual follows *R*, it is in his interest that his opponents should do so too. Then each individual has a moral obligation to follow *R*, provided that everyone else[6] does the same.

Let me make it clear that I am not claiming that this principle constitutes the *whole* of our morality. I am claiming only that there is a strong tendency for us to subscribe to moral rules that are instances of this principle; to put this another way, we are inclined to give this principle *some* moral weight.

It does, I believe, appeal to some common moral intuitions. Suppose that almost everyone in the community follows *R*. Then if I play a game against you, it is reasonable for me to expect that you will play *R*. For me to play *R* is for me to act in a way that I can reasonably expect to be in your interests (since if you do play *R*, my playing *R* will be in your interests). Equally, for you to play *R* is for you to act in a way that you can reasonably expect to be in my interests. Then if I play *R* but you do not, I have acted in the way best calculated to accommodate you, but you have failed to reciprocate. The moral intuition behind the principle of co-operation is that in such a case I have a legitimate complaint against you. Take, for example, the crossroads game. Suppose it is the general practice at crossroads to give priority to the vehicle approaching from the right. Suppose you always follow this practice. Then, given the way other drivers can be expected to behave, you are behaving in the way that is best calculated to benefit them. I eccentrically choose to adopt the strategy of never giving way to anyone, and in doing so put your life at risk. Then you have grounds for complaint against me.

The morality that grows up around conventions, then, is a morality of *co-operation*. It is also a morality of *rights*. If everyone else in my community is following *R*, I am obliged to do the same. Notice that this obligation arises out of my relationship with other individuals: I am obliged to benefit them because they are benefiting me. My obligation, then, is *to particular other people*. Corresponding with my obligation to follow *R* is everyone else's *right* that I should do so; each other person is entitled to demand that I meet my obligation to him. This is quite different from the sort of obligation imposed by a morality of maximizing social welfare, or of maximizing the sum of happiness in the world – which are obligations to no one in particular.

Any system of morality that rests on an idea of co-operation must incorporate some reference point from which benefit or disbenefit is to be measured. The idea is that if I benefit you, I am entitled to demand that you benefit me in return; but benefit is a comparative concept.[7] When I say that I have benefited you, I am saying that I have made you better off than you would have been in some other state of affairs: this state of affairs is the reference point. What, then, is the reference point for the principle of co-operation that underlies natural law? The reference point is the status quo.

Why do I say this? Notice that the principle of co-operation obliges an individual to follow a strategy *R* only if everyone else is doing so. Thus there can be a general obligation for everyone to follow *R* only in a state of affairs in which *R* is being generally followed. In other words, to suppose that there is such a general obligation is to suppose that the status quo is one in which everyone follows *R*. Now the idea behind the principle of co-operation is that if any individual *unilaterally* defects from the established practice of following *R* he will disadvantage his opponents; an individual who chooses *not* to disadvantage his opponents in this way is entitled to expect in return that his opponents will not disadvantage him. Here disadvantage is being measured relative to the state of affairs in which everyone follows *R*; and this is the status quo.

To recognize that natural law is based on a principle of mutual advantage, and that the reference point for measuring advantage is the status quo, is to recognize that rights and obligations are matters of convention. In a community in which everyone follows some rule *R*, everyone may have a moral right to expect everyone else to follow this rule; and yet it may be equally true that if everyone followed a different rule *S*, everyone would have a moral right to expect everyone else to follow *S*.

Many readers, I imagine, will object strongly to this idea. Moral theories are usually constructed so that the question, 'What rights and obligations do individuals have?' has a unique answer. Take, for example, the theory that morality is concerned with the maximization of some conception of social welfare. Following Sen (1979), I shall call this kind of theory 'welfarist'; the classical utilitarian position, that the sum of happiness should be maximized, is a special case of welfarism. For a welfarist, rights and obligations can be justified only as means for achieving the end of maximum social welfare. Welfare, unlike benefit or advantage, is *not* a comparative concept; so once the concept of social welfare is defined, the problem of how to maximize it will normally have a unique solution.

Or take Rawls's (1972) theory of justice. For Rawls, justice is a matter of mutual advantage, but advantage is defined in relation to a fixed 'initial arrangement' in which what Rawls calls 'social primary goods' – including income and wealth – are distributed equally (Rawls, 1972, p. 62). Principles of justice are those principles that would secure the unanimous agreement of individuals in this position of initial equality. Now, of course, this is not enough to ensure a unique set of principles. (There may be a choice between alternative sets of principles, all of which make everyone better off than in the initial position; some individuals may benefit more from one set of principles, others from another.) Rawls recognizes this problem but, significantly, responds by designing the initial position so that a unique set of principles of justice *will* be generated (Rawls, 1972, pp. 139–40). The device that ensures this is the 'veil of ignorance': no one is allowed to know his or her identity, and so no one can know which principles would work out most to his or her advantage.

As a final example, take Nozick's (1974) entitlement theory of justice. Nozick simply *assumes* the existence of a unique code of natural law. He provides no real justification for this assumption beyond an appeal to the authority of Locke – while remarking that Locke 'does not provide anything like a satisfactory explanation of the status and basis of the law of nature' either (Nozick, 1974, p. 9). For Locke, natural law is a system of moral rights and duties that embodies the will of God, as interpreted by natural reason – that is, without the aid of divine revelation (Locke, 1690, Book 2, Ch. 2). It is not clear whether Nozick shares Locke's theism; certainly God plays no explicit part in Nozick's theory. It seems that for Nozick it is a matter of simple moral intuition that individuals have certain clearly defined rights.

Against this background, the claim that rights and obligations are matters of convention will be controversial. Let me therefore repeat what I have said several times before. It is no part of my argument that the morality that evolves in human society is the morality that we *ought* to follow. I am not presenting a moral argument; I am trying to explain how we come to have some of the moral beliefs we do. My claim is that human beings tend to use the status quo as a moral reference point. Whether they ought to do so is another matter, and (at least for those of us who accept Hume's Law) one that can never be resolved by an appeal to reason.

It is, therefore, open to anyone to say that he or she refuses to accept the moral legitimacy of claims based on what I have called natural law. He or she may say: 'The only right thing to do is . . .; if this means violating what some people regard as their moral rights, then that's just too bad. I don't accept that these so-called rights have any claim on my respect'. A consistent welfarist, for example, would say that the right thing for a government to do is whatever maximizes social welfare. In his rejection of claims based on natural law, the welfarist would (I imagine) be emboldened by the reflection that, after all, these 'laws' are only conventions. He may even be tempted to think about moral engineering. If people's moral beliefs are based on conventions, they can be changed. Rather than accept the convention that happens to have become established, we can work out which of the many possible conventions would maximize social welfare, and then require everyone to follow it. At first, no doubt, some people will resist the change, claiming that they have been cheated out of their entitlements; but with time these ideas of rights and obligations will die out and be replaced by new ones that are more compatible with the maximization of welfare.

The analysis of this book provides no moral counter-argument to such a committed welfarist. (Nor does it provide a counter-argument to someone who is committed to a Rawlsian or a Nozickian theory of justice.) However, it does provide some warnings. Established conventions *can* be overturned – consider the success of the metric system – but history is littered with failed attempts to reform conventions. The same generation of rationalistic Frenchmen and Frenchwomen who introduced the metric system attempted to reform the calendar; but we all still use the old irrational one. Successive Irish governments have made great efforts to re-establish the Gaelic language, but they have been unable to prevent the overwhelming majority of the Irish people from speaking and writing in English. The convention that gold is money has lasted for thousands of years and is still resistant to attempts to reform the international monetary system.

Consider the convention that conflicts are resolved in favour of possessors. This may have no rational foundation; from a welfarist point of view it may be morally perverse. But is it in the power of any government to eradicate it? Think of all the situations in which this convention is used to resolve disputes, and of how it can spread by analogy from one situation to another (Section 5.3). Think how many disputes are settled outside the influence of the legal system – between neighbours and workmates, in clubs and associations, on the roads, in school playgrounds, in youth gangs, ... How is any government going to change the way people behave in *these* situations? Think also of the way the convention favouring possessors is used to settle international disputes. (Think how the political map of Europe has been determined by the positions reached by the Soviet and American armies in 1945.) How can any one government renounce this convention in its own dealings with other governments?

If, as I have argued, the conventions by which people resolve disputes come to have the status of moral rights and obligations, any government that tries to overturn these conventions must expect its actions to be viewed as morally wrong – as illegitimate invasions of individuals' rights. The consistent welfarist must be prepared to accept the blame his policies will attract from the members of the society whose welfare he is seeking to promote. Recalling Bernard Williams's analogy between utilitarian theory and colonial administration (see Section 1.4), one is reminded of Rudyard Kipling's account of the white man's burden.[8]

However, the viewpoint of the colonial administrator, seeking to promote the welfare of a society without being constrained by its prevailing morality, is a peculiar one: there is no reason why we must adopt it. There is no reason why we must justify our moral beliefs by showing that they are beliefs that would be shared by an impartially benevolent observer, looking on society from some distant vantage point.

We are entitled to make moral judgements from *where we are* – as members of a society in which certain ideas of obligations and rights have become established and accepted by us as part of our morality. In Sen's language, these ideas may be basic value judgements for us (cf. Section 8.3), just as the welfarist's judgements are basic for him. We cannot give any ultimate justification for our adherence to a morality of co-operation; but neither can the welfarist justify his welfarism.

For most of us, I suggest, morality cannot be reduced to welfarism. Whatever political principles we may profess, most of us believe that we each have rights that cannot be legitimately overridden merely to increase the overall welfare of society. When what we take to be our rights or legitimate expectations are under threat, we feel entitled to ask: 'But *why* must I sacrifice my expectations for the good of society?' This question is not a stupid one. It does not reveal a failure to grasp the logic of moral reasoning. It is a real moral question. When the chips are down – when it is *our* expectations that are at stake – this is a question that we *do* ask. As Rawls (1972, pp. 175–9) puts it, it is a general fact of moral psychology that we are inclined to conceive of society as a system of co-operation through which *our* interests, as well as everyone else's, are to be advanced; we are not psychologically equipped to take the welfarist's view of society in which our interests are subsumed in some social whole.

If the argument of this book is correct, at least some of what we take to be our rights are grounded in nothing more than convention; in accepting them as morally significant we are using the status quo as a moral reference point. This is an unmistakably conservative idea. There is no claim, however, that the status quo is *better*, all things considered, than any other possible state of society. To suppose that some claim of this kind is being made is to interpret a conservative theory as a species of welfarism; the theory has nothing to say about the overall welfare of society, or about how one possible state of society should be compared with another. The significance of the status quo is simply that, as James Buchanan (1975, p. 78) has put it, we start from here, and not from some place else.

Many readers, I suspect, will still balk at such conservatism, and will look for some other way of rationalizing their moral convictions. I wish them luck, but I suspect that the task is hopeless. Certain moral beliefs, I have argued, have a natural tendency to evolve. We cannot easily shake these beliefs off. Nor are we inclined to try, since they are *our* beliefs: they are part of our moral view of the world. However much we might wish to deny it, our morality is in important respects the morality of spontaneous order; and the morality of spontaneous order is conservative.

Notes

1 SPONTANEOUS ORDER

1. Spontaneous order is the central theme of Hayek's three-volume work *Law, Legislation and Liberty* (1979). In *The Constitution of Liberty* (1960, p. 160), Hayek attributes the phrase 'spontaneous order' to Polanyi.
2. Some economists (e.g. Becker, 1974) have tried to explain voluntary contributions to the supply of public goods in terms of conventional economic theory: each individual is assumed to choose the contribution that maximizes his own utility, taking other people's contributions as given. It is claimed that such a theory is consistent with the observation that *some* voluntary contributions are made – even though the theory also predicts that public goods will be supplied in less-than-efficient amounts. I have argued that the evidence of voluntary contributions to public goods cannot plausibly be explained in this way (Sugden, 1982, 1985; see also Margolis, 1982, pp. 19–21).
3. This shrewd analogy is due to Williams (1973, p. 138).

2 GAMES

1. This piece of terminology is due to Gibbard (1973).
2. A somewhat similar conception of utility has been developed by Camacho (1982).
3. This work is summarized by Maynard Smith (1982). I shall be saying more about the biological literature in chapter 4.
4. Axelrod uses only condition (2.1) and not (2.2) to define his concept of a 'collectively stable strategy'. In the language I shall be using, what Axelrod calls a collectively stable strategy is an equilibrium strategy, but not necessarily a stable one.
5. This assumption, which makes the position of each individual symmetrical with that of every other, is convenient but not strictly necessary. All that is

required for the argument that follows is that in any game each individual has *some* (i.e. some non-zero) probability of being A and *some* probability of being B.
6. Lewis (1969, p. 22) uses the concept of a 'proper equilibrium', which he defines as a situation in which each player's strategy is his *unique* best reply to his opponent's strategy (or opponents' strategies). This is sufficient but not necessary for stability on my definition.

3 COORDINATION

1. Nozick (1974, p. 18) offers an alternative reading of these passages from Locke. Nozick argues, as I do, that the use of money is a convention that could have evolved spontaneously, but he claims that Locke believed that the 'invention of money' was the result of 'express agreement'. I think Nozick is taking Locke's use of the concept of agreement too literally; Locke does, after all, say that this agreement is 'tacit' and 'without compact'.
2. Schotter (1981, Ch. 2) presents a model of the evolution of market-day conventions.

4 PROPERTY

1. It would be anachronistic not to follow Hobbes in speaking of 'men' rather than 'persons'.
2. Most versions of the teenagers' game are probably better modelled by the 'war of attrition' game presented in Section 4.3.
3. The assumption that utility is lost at a constant rate through time is not really necessary, although it allows the analysis to be simplified slightly. All that is necessary is that utility is lost continuously until one player surrenders. To generalize the argument in the text, a pure strategy should be interpreted as an 'acceptable penalty' (i.e. the maximum loss of utility a player will sustain before surrendering) rather than as a persistence time (cf. Norman *et al.*, 1977).
4. More formally, let $f(t)$ be the probability density function for persistence times in a given strategy. Then $S(t)=f(t)/[1-\int_0^t f(t)\,dt]$.
5. I have avoided using differential calculus so as to make the argument comprehensible to as many readers as possible. This result could have been derived more elegantly by taking limits as $\delta t \to 0$.
6. This result corresponds with one in the theory of 'rent-seeking': if individuals can compete for access to economic rents, free entry to the competition will ensure that in the long run the value of the rents is entirely dissipated (cf. Tullock, 1967 and Krueger, 1974).
7. A slightly different game can be constructed by assuming instead that where the claims sum to less than 1, the unclaimed portion of the resource is divided equally between the players. The analysis of this game is very similar to that of the division game.
8. In Schelling's game, when players' claims are incompatible, each simply receives nothing.

2. A strategy corresponding to T_1 (i.e. 'Play "dove" if your opponent is in good standing, or if you are not; otherwise play "hawk"') can be defined for the hawk-dove game. But with the particular utility values I used in presenting the hawk-dove game, this strategy turns out *not* to be a stable equilibrium, however close the value of π is to 1. My version of the hawk-dove game is equivalent to the snowdrift game with $v=4$, $c_1=2$, $c_2=1$; this gives max $[c_2/v, c_2/(c_1-c_2)]=1$.

3. Here I am assuming the probability of mistakes to be insignificantly small.

4. Mackie (1980, pp. 88-9) offers two different interpretations of Hume's rowing example. On one interpretation the rowers' problem is merely to coordinate their strokes (perhaps each rower has one oar): this is a pure coordination game. On Mackie's other interpretation it is a prisoner's dilemma game in which the players adopt tit-for-tat strategies.

5. The idea that the voluntary supply of public goods can be modelled by the chicken game has been developed by Taylor and Ward (1982); as I pointed out in Section 7.3, the snowdrift game is a form of the chicken game.

6. To simplify the presentation, I shall ignore the possibility that v might be *exactly* equal to one of $c_1, \ldots, c_n$.

7. Models in which public goods are supplied in this kind of way have been presented by, among others, Taylor (1976, Ch. 3) and Guttman (1978).

8. This idea is developed by Nozick, (1974, Ch. 2).

9. Lifeboat services could no doubt be organized on the club principle, provided lifeboatmen were prepared to be sufficiently ruthless and rescue only those who had paid subscriptions. But given the principles on which the lifeboat service actually operates, its life-saving activities are to the benefit of anyone who goes near the sea.

8 NATURAL LAW

1. The problem of wishful thinking is explored by Elster (1983).

2. Hobbes's fifth law of nature 'is COMPLAISANCE; that is to say, *that every man strive to accommodate himself to the rest*' (1651, Ch. 15). Hobbes's argument is that a man who wishes to survive and prosper among other men ought, as a matter of prudence, to cultivate a complaisant or sociable character. If this argument is correct, we should expect some tendency for biological natural selection to favour the same character traits.

3. See Hume (1740, Book 3, Part 1, Section 1) for a statement of this 'law'.

4. The idea that moral judgements are universalizable is developed by Hare (1952). It will be clear from what follows that I do not accept Hare's (1982) more recent argument that universalizability entails some kind of utilitarianism.

5. In the division game there is one symmetrical convention – that of equal division. But all the other conventions in this game, and all the conventions in the hawk-dove and war of attrition games, are asymmetrical.

6. An individual who benefits from the fact that a particular convention has become established might perhaps resent such meekness because of its tendency to undermine the convention; but as long as the convention is secure, the existence of a meek minority works to everyone else's benefit.

7. I have developed this idea more fully in a recent paper (Sugden, 1984). As I emphasize in that paper, the ethic of reciprocity is not to be confused with the principle – often called Kantian – that in games with a prisoner's dilemma structure each individual has an unconditional moral obligation to play the co-operative strategy (cf. Laffont, 1975; Collard, 1978; Harsanyi, 1980). A principle of reciprocity obliges a person to co-operate only if others co-operate too.

9 RIGHTS, CO-OPERATION AND WELFARE

1. Lewis calls these simply 'conventions'; his definition of convention excludes the rules I have called conventions of property and conventions of reciprocity (see Section 2.8).

2. The game that she takes as her paradigm has the same structure as the banknote game I presented in chapter 2: there are two players and two alternative conventions, one favouring one player and one the other.

3. An alternative reading is that every breach of a convention tends to weaken it. On *this* reading, we receive 'mediate' prejudice from acts of injustice only if the rules of justice work to our advantage in the long run.

4. In the *Theory of Moral Sentiments* Smith argues against the view that our approval of principles of justice is grounded in sympathy with public interest: 'The concern which we take in the fortune and happiness of individuals does not, in common cases, arise from that which we take in the fortune and happiness of society' (1759, Book 2, Section 2, Ch. 3).

5. This principle is a close relative of the 'principle of fairness' formulated by Hart (1955). It also has some similarities with the principle of reciprocity that I have presented in a recent paper (Sugden, 1984).

6. Or, if we are concerned with practical morality, *almost* everyone else. Such a qualification is theoretically awkward, but unavoidable.

7. This point is made by Buchanan (1985, p. 63).

8. 'Take up the white man's burden / And reap his old reward / The blame of those ye better / The hate of those ye guard.'

References

Allais, M. (1953) 'Le comportement de l'homme rationnel devant le risque; critique des postulats et axiomes de l'ecole Americaine'. *Econometrica*, **21**, 503–46.

Arrow, K. J. (1963) 'Uncertainty and the welfare economics of medical care'. *American Economic Review*, **53**, 941–73.

Arrow, K. J. (1967) 'Values and collective decision-making'. In Laslett, P. and Runciman, W. G. (eds), *Philosophy, Politics and Society*. London: Blackwell.

Arrow, K. J. and Debreu, G. (1954) 'Existence of an equilibrium for a competitive economy'. *Econometrica*, **22**, 265–90.

Axelrod, R. (1981) 'The emergence of cooperation among egoists'. *American Political Science Review*, **75**, 306–18.

Bacharach, M. (1976) *Economics and the Theory of Games*. London: Macmillan.

Becker, G. S. (1974) 'A theory of social interactions'. *Journal of Political Economy*, **82**, 1063–93.

Bell, F. (1907) *At the Works: A Study of a Manufacturing Town*. London: Edward Arnold.

Bishop, D. T. and Cannings, C. (1978) 'A generalized war of attrition'. *Journal of Theoretical Biology*, **70**, 85–124.

Bishop, D. T., Cannings, C. and Maynard Smith, J. (1978) 'The war of attrition with random rewards'. *Journal of Theoretical Biology*, **74**, 377–88.

Buchanan, A. (1985) *Ethics, Efficiency, and the Market*. Oxford: Clarendon Press.

Buchanan, J. M. (1965) 'An economic theory of clubs'. *Economica*, **32**, 1–14.

Buchanan, J. M. (1975) *The Limits of Liberty*. Chicago: University of Chicago Press.

Camacho, A. (1982) *Societies and Social Decision Functions*. Dordrecht: Reidel.

Collard, D. A. (1978) *Altruism and Economy*. Oxford: Martin Robertson.

Dawkins, R. (1980) 'Good strategy or evolutionarily stable strategy?'. In Barlow, G. W. and Silverberg, J. (eds), *Sociobiology: Beyond Nature/Nurture*. Boulder: Westview Press.

Elster, J. (1983) *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Friedman, M. (1962) *Capitalism and Freedom*. Chicago: University of Chicago Press.

Gibbard, A. (1973) 'Manipulation of voting schemes: a general result'. *Econometrica*, **41**, 587–601.

Guttman, J. M. (1978) 'Understanding collective action: matching behavior'. *American Economic Review*, **68**, Papers and Proceedings, 251–5.

Hammerstein, P. and Parker, G. A. (1982) 'The asymmetric war of attrition'. *Journal of Theoretical Biology*, **96**, 647–82.

Hardin, G. (1968) 'The tragedy of the commons'. *Science*, **162**, 1243–8.

Hare, R. M. (1952) *The Language of Morals*. Oxford: Oxford University Press.

Hare, R. M. (1982) 'Ethical theory and utilitarianism'. In Sen, A. K. and Williams, B. (eds), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

Harsanyi, J. C. (1955) 'Cardinal welfare, individualistic ethics and interpersonal comparisons of utility'. *Journal of Political Economy*, **63**, 309–21.

Harsanyi, J. C. (1980) 'Rule utilitarianism, rights, obligations and the theory of rational behavior'. *Theory and Decision*, **12**, 115–33.

Hart, H. L. A. (1955) 'Are there any natural rights?' *Philosophical Review*, **64**, 175–91.

Hayek, F. A. (1960) *The Constitution of Liberty*. London: Routledge and Kegan Paul.

Hayek, F. A. (1979) *Law, Legislation and Liberty*. London: Routledge and Kegan Paul. (In three volumes, Vol. 1 published 1973, Vol. 2 published 1976, Vol. 3 published 1979).

Hobbes, T. (1651) *Leviathan*, edited by M. Oakeshott. London: Macmillan, 1962.

Hume, D. (1740) *A Treatise of Human Nature*, edited by L. A. Selby-Bigge (2nd edition). Oxford: Clarendon Press, 1978.

Johnson, N. (1981) *Voluntary Social Services*. Oxford: Basil Blackwell.

Kahneman, D. and Tversky, A. (1979) 'Prospect theory: an analysis of decision under risk'. *Econometrica*, **47**, 263–91.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan.

Kramer, R. M. (1981) *Voluntary Agencies in the Welfare State*. Berkeley: University of California Press.

Krueger, A. O. (1974) 'The political economy of the rent-seeking society'. *American Economic Review*, **64**, 291–303.

Laffont, J.-J. (1975) 'Macroeconomic constraints, economic efficiency and ethics: an introduction to Kantian economics'. *Economica*, **42**, 430–7.

Lewis, D. K. (1969) *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.

Locke, J. (1690) *Two Treatises of Government*, edited by P. Laslett. Cambridge: Cambridge University Press, 1960.

Loomes, G. and Sugden, R. (1982) 'Regret theory: an alternative theory of rational choice under uncertainty'. *Economic Journal*, **92**, 805–24.

Mackie, J. L. (1980) *Hume's Moral Theory*. London: Routledge and Kegan Paul.

Margolis, H. (1982) *Selfishness, Altruism and Rationality*. Cambridge: Cambridge University Press.

Maynard Smith, J. (1974) 'The theory of games and the evolution of animal conflicts'. *Journal of Theoretical Biology*, **47**, 209–21.

Maynard Smith, J. (1982) *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Maynard Smith, J. and Parker, G. A. (1976) 'The logic of asymmetric contests'. *Animal Behaviour*, **24**, 159–75.

Maynard Smith, J. and Price, G. R. (1973) 'The logic of animal conflicts'. *Nature*, **246**, 15–18.

Milgrom, P. and Roberts, J. (1982) 'Predation, reputation and entry deterrence'. *Journal of Economic Theory*, **27**, 280–312.

Morgenstern, O. (1979) 'Some reflections on utility'. In Allais, M. and Hagen, O. (eds), *Expected Utility Hypotheses and the Allais Paradox*. Dordrecht: Reidel.

Neumann, J. von and Morgenstern, (1947) *Theory of Games and Economic Behavior* (2nd edition). Princeton: Princeton University Press.

Norman, R., Taylor, P. and Robertson, R. (1977) 'Stable equilibrium strategies and penalty functions in a game of attrition'. *Journal of Theoretical Biology*, **65**, 571–8.

Nozick, R. (1974) *Anarchy, State and Utopia*. New York: Basic Books.

Parker, G. A. and Rubinstein, D. I. (1981) 'Role assessment, reserve strategy, and acquisition of information in asymmetrical animal conflicts'. *Animal Behaviour*, **29**, 135–62.

Rapoport, A. (1967) 'Exploiter, Leader, Hero and Martyr: the four archetypes of the 2×2 game'. *Behavioral Science*, **12**, 81–4.

Rapoport, A. and Chammah, A. M. (1965) *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.

Rawls, J. (1972) *A Theory of Justice*. Oxford: Oxford University Press.

Samuelson, P. (1954) 'The pure theory of public expenditure'. *Review of Economics and Statistics*, **36**, 387–9.

Schelling, T. (1960) *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

Schoemaker, P. (1982) 'The expected utility model: its variants, purposes, evidence and limitations'. *Journal of Economic Literature*, **20**, 529–63.

Schotter, A. (1981) *The Economic Theory of Social Institutions*. Cambridge: Cambridge University Press.

Selten, R. (1975) 'Reexamination of the perfectness concept for equilibrium points in extensive games'. *International Journal of Game Theory*, **4**, 25–55.

Selten, R. (1978) 'The chain store paradox'. *Theory and Decision*, **9**, 127–59.

Selten, R. (1980) 'A note on evolutionarily stable strategies in asymmetric animal conflicts'. *Journal of Theoretical Biology*, **84**, 93–101.

Sen, A. K. (1970) *Collective Choice and Social Welfare*. Edinburgh: Oliver and Boyd.

Sen, A. K. (1977) 'Rational fools: a critique of the behavioural foundations of economic theory'. *Philosophy and Public Affairs*, **6**, 317–44.

Sen, A. K. (1979) 'Personal utilities and public judgements: or what's wrong with welfare economics?' *Economic Journal*, **89**, 537–58.

Shubik, M. (1971) 'The Dollar Auction game: A paradox in noncooperative behavior and escalation'. *Journal of Conflict Resolution*, **15**, 109–11.

Slovic, P. and Tversky, A. (1974) 'Who accepts Savage's axiom?' *Behavioral Science*, **19**, 368–73.

Smith, A. (1759) *The Theory of Moral Sentiments*, edited by D. D. Raphael and A. L. Macfie. Oxford: Clarendon Press, 1976.

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, edited by R. H. Campbell, A. S. Skinner and W. B. Todd. Oxford: Clarendon Press, 1976.

Sugden, R. (1982) 'On the economics of philanthropy'. *Economic Journal*, **92**, 341–50.

Sugden, R. (1984) 'Reciprocity: the supply of public goods through voluntary contributions'. *Economic Journal*, **94**, 772–87.

Sugden, R. (1985) 'Consistent conjectures and voluntary contributions to public goods: why the conventional theory does not work'. *Journal of Public Economics*, **27**, 117–24.

Taylor, M. (1976) *Anarchy and Cooperation*. London: John Wiley and Sons.

Taylor, M. and Ward, H. (1982) 'Chickens, whales and lumpy goods: alternative models of public-goods provision'. *Political Studies*, **30**, 350–70.

Tullock, G. (1967) 'The welfare costs of tariffs, monopolies and theft'. *Western Economic Journal*, **5**, 224–32.

Ullman-Margalit, E. (1977) *The Emergence of Norms*. Oxford: Clarendon Press.

Walmsley, L. (1932) *Three Fevers*. London: Collins.

Williams, B. (1973) 'A critique of utilitarianism'. In Smart, J. J. C. and Williams, B. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Index