How and Why Did Fairness Norms Evolve?

KEN BINMORE

What is just . . . is what is proportional.
Aristotle, *Nicomachean Ethics*

WHAT IS FAIR?

WHEN A DISH IN SHORT SUPPLY is shared at a polite dinner party, there is seldom any verbal dispute. If things go well, the dish is divided without any discussion or intervention by the host. When questioned, everybody will agree that each person should take his or her fair share. But how do we know what is fair?

This is not a simple question. What is judged to be fair according to our current standards of morality depends on a complex combination of contingent circumstances — like who is fat and who dislikes cheese. Moreover, if we observe what actually happens, rather than what people say should happen, we will find that it also depends on how each person at the table fits into the social pecking order.Woe betide the poor relative sitting at the table on sufferance in the nineteenthcentury who helped himself to an over-generous portion of his favourite dish!

Numerous scholars have tried to make sense of the calculations that people must implicitly have made when they co-ordinate on an outcome that they afterwards describe as fair. It surely can be no accident that the consensus is firmly in favour of some type of do-as-you-would-be-done-by principle. Moralists down the ages have offered numerous arguments that seek to explain why it is morally imperative that each person should follow such a golden rule. But none of these traditional arguments is founded on anything solid. I think we become suckered into taking them seriously because we are too ready to confuse a fairly accurate description of *what* we do in certain circumstances with an explanation of *why* we do it.

Rather than resorting to metaphysical speculation, I think that the first step on the road to understanding the human thirst for justice lies in the recognition that variants of the do-as-you-would-be-done-by principle are *already* firmly entrenched among the instincts and customs that regulate our lives. The relevant norms do not survive because we consciously cherish them. On the

Proceedings of the British Academy, **110**, 149–70, © The British Academy 2001.

contrary, I think that most of our habituated behaviour is acquired using processes that operate below the level to which our conscious minds have easy access. Like monkeys, we are programmed to imitate the behaviour of our more successful neighbours. If those in thrall to a particular habit or custom are perceived as being winners, then their habituated behaviour will be copied, without any need for anyone to understand *why* the habituated behaviour works well in the current social environment.

A fairness norm may be a do-as-you-would-be-done-by principle, but many such principles can be formulated. Which of these should we study? To my knowledge, only one principle has been proposed that adequately responds to objections such as: Don't do unto others as you would have them do unto you — they may have different tastes from yours. This chapter will need to refer to both Rawls (1972) and Harsanyi (1977) in studying this fairness principle, but the terminology will be that of Rawls' *Theory of Justice*. Rawls proposes the *original position* as a hypothetical standpoint to be used in making judgements about how a just society should be organized. Each citizen is asked to envisage the social contract to which he or she would agree *if* his or her current role in society were concealed behind a *veil of ignorance*. In considering the social contract on which to agree under such hypothetical circumstances, each person will pay close attention to the plight of those who end up at the bottom of the social heap. Devil take the hindmost is not such an attractive principle when you yourself may be at the back of the pack.

I think that the reason most people find the device of the original position intuitively attractive as a fairness criterion has nothing to do with the Kantian arguments offered by Harsanyi and Rawls. I believe that its appeal lies in the fact that we recognize it as a stylized version of a principle that we already unconsciously apply every day when interacting with our peers. From such a perspective, fairness is interpreted entirely in naturalistic terms. The original position is merely a device that has been washed up on the beach along with the human race by the forces of biological and social evolution. If we can figure out precisely how we use it at present to avoid inefficient disputes over small matters, perhaps we will also be able to use it to achieve stable political compromises over large-scale issues. The defence for such a proposal is entirely pragmatic. Here is a tool supplied by Nature. Let us use it to improve our lives, just as we use whatever tools we find in our toolbox when making repairs around the house. But we shall get nowhere in this enterprise if we refuse to be realistic about how the device of the original position functions in our daily life at present.

Psychological equity theory

Our capacity for objective introspection is notoriously limited. What we say about our beliefs and motivations is often absurdly at variance with our

behaviour. Experimental work is therefore necessary to discover how we actually split a surplus when we believe ourselves to be acting fairly.

Social psychologists who have conducted experiments on fairness have been led to an empirically based law that resolves problems of social exchange by equalizing the ratio of each person's gain to his or her worth (Furby 1986; Mellers 1982; Mellers and Baron 1993; Walster *et al*. 1978). People who are deemed worthy therefore get more of the gravy than others. As in Wilson (1993), this theory is usually referred to as 'modern equity theory', although it originates with Aristotle's *Nicomachean Ethics* and has been little developed since it was introduced to social psychologists by Homans (1961) and Adams (1963, 1965) more than thirty years ago. Selten (1978) provides an account of the theory which is easily accessible to economists.

The psychological theory of equity requires that a surplus be shared in proportion to each person's worthiness. Written as an equation:

$$\frac{g_A}{w_A} = \frac{g_E}{w_E} \tag{1}$$

where g_A and g_E are the respective gains to Adam and Eve, and w_A and w_E quantify how worthy they are. But how are gains to be measured? Where is the zero to be located on whatever scale is chosen? How is worthiness to be construed? Is it to be measured in terms of social status, merit, effort, need or what? My understanding of the pyschological literature is that the answers to these questions depend on the context. But what is the rule that maps a context onto the relevant scales for measuring gain and worthiness?

To answer such questions, one needs a background theory to suggest critical experiments. I believe that such a theory can be constructed by asking how the apparatus of the original position proposed by Harsanyi and Rawls may have evolved from prehistoric food-sharing agreements between members of the same family. In seeking to construct such a theory in the following pages, one needs to make hypotheses about the social contracts that held sway among human foraging bands in prehistory. The second half of this chapter takes up the main purpose of the paper, which is to comment on the claim that modern hunter-gatherer societies provide a suitable model for their prehistoric counterparts. The rest of the chapter briefly relates this commentary to my adaptation of the theories of Harsanyi and Rawls.

Natural duty?

Rawls (1972) invented the device of the original position to provide a properly argued alternative to utilitarianism. Harsanyi (1977) appealed to precisely the same device when defending utilitarianism. I support Harsanyi in this dispute,

since Rawls succeeds in evading a utilitarian conclusion only by throwing orthodox decision theory overboard. However, I think that Rawls' intuitive grasp of the type of outcome to which one is led by applying the original position under realistic conditions is much sounder. Rawls advocates redistributing worldly goods according to the maximin criterion, which demands that we give priority to ensuring that the worst-off members of society gets as much as possible.

Figure 1(a) compares Rawls' maximin outcome R with Harsanyi's utilitarian outcome H. In this diagram, Adam and Eve are the two members of a society inhabiting the Garden of Eden. A social contract is modelled as a pair $x = (x_A, x_E)$ of utilities. The set X contains the social contracts that are feasible. Figure 1(b) shows three bargaining solutions from co-operative game theory: the Nash bargaining solution n; a weighted utilitarian solution h; and the proportional or egalitarian bargaining solution r. The point ξ represents Adam and Eve's current *status quo*. Our focus for the moment is on the third of these.

Two important features of the egalitarian bargaining solution should be noted. The first is that r can be identified with the result of applying the psychological equity law if Adam and Eve's respective gains are taken to be $g_A = r_A - \xi_A$ and $g_E = r_E - \xi_E$, and their worthiness coefficients are chosen so that w_E/w_A is the slope of the line joining ξ and r. The second point is that r is also the result of applying the maximin criterion after correcting x_A and x_E to $(x_A - \xi_A)/w_A$ and $(x_E - \xi_E)/w_E$. Such a correction corresponds to relocating the zeros and units on Adam and Eve's utility scales in order to ensure that our standard of measurement matches the manner in which interpersonal comparisons of welfare are made in the society under study.

I think Rawls' attempt to derive the maximin criterion from an analysis of how Adam and Eve will bargain behind the veil of ignorance goes awry at two

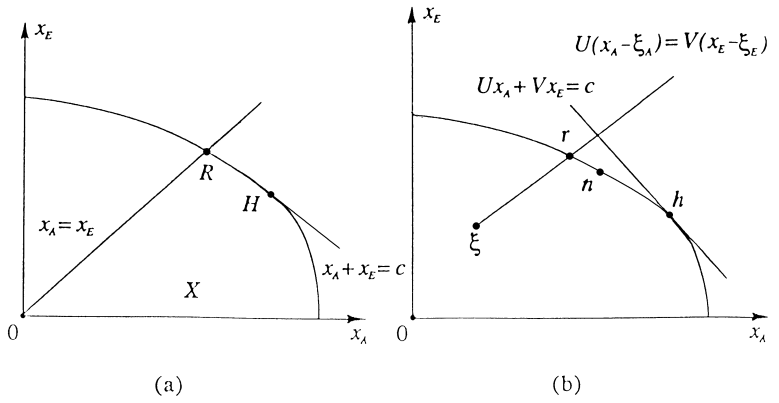

Figure 1. Rawls and equity theory.

points. He should not have adopted the iconoclastic expedient of denying orthodox decision theory, and he should not have joined with Harsanyi in assuming that Adam and Eve are *committed* to the hypothetical deal reached in the original position. Rawls (1972: 115) says that we have a 'fundamental natural duty . . . to comply with just institutions', but I think that he and Harsanyi are really just indulging in some wishful thinking. It would certainly make life more pleasant if we instinctively rated the call of justice above our own selfish concerns, but the evidence for such a claim is not very favourable.

The commitment problem arises in its starkest form in the study of the Prisoners' Dilemma of Figure 2(a). If Adam and Eve discuss how they should play this game, whether behind a veil of ignorance or not, they are likely to agree that both should play *dove*. Each will then receive a payoff of 2. If they are committed to the agreement, this is the end of the story. But if they are not committed, then they have the opportunity to cheat on the agreement when the time comes to play. Since cheating on the deal by playing *hawk* is optimal for each player whatever strategy the other chooses, the result will be that both play *hawk*. Each then receives a payoff of 0.

When Adam and Eve both choose *hawk* in the Prisoners' Dilemma, each is using a strategy that is an optimal reply to the strategy choice of the other. Game theorists register that a pair of strategies has this property by calling it a Nash equilibrium. If an authoritative book on game theory records the rational solution to a game, it must be a Nash equilibrium — otherwise it would be rational for at least one player to deviate from the book's recommendation.

Various attempts to escape the conclusion that rational play calls for both players to cheat in the one-shot Prisoners' Dilemma have been proposed which

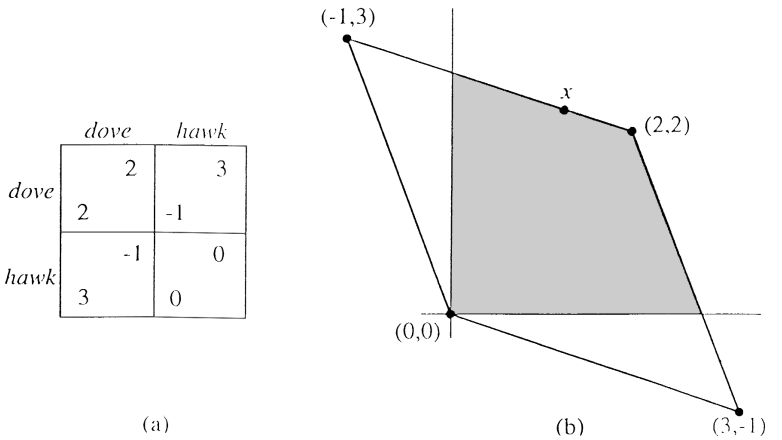

Figure 2. Prisoners' Dilemma.

postulate that Nature has equipped us with *internal* commitment mechanisms whose engagement can be convincingly transmitted to an opponent (Binmore 1994). But where is the evidence that such internal mechanisms exist? Why would they be evolutionarily stable? Since no adequate answers are on offer, game theorists restrict their attention to *external* commitment mechanisms. For example, if Adam and Eve sign a legal contract under modern circumstances to play *dove* in the Prisoners' Dilemma, then each will regard themselves as committed to the agreement, since any breach of the contract will be punished by our judicial system. Other external enforcement agencies have operated in other places and at other times. Fear of ostracism by one's peer group is a particularly effective form of disciplining agreements.

In postulating an evolutionary history for the device of the original position, it is therefore important that we take a view on the extent to which an external source of authority for policing agreements was available in the relevant period of prehistory. If an external enforcement agency were available in the form of a dominant leader or strong peer pressure, then Harsanyi's (1977) analysis suggests that our fairness norms would be utilitarian in character. However, when a similar analysis is applied to the case when no external enforcement agency at all exists, Binmore (1994, 1998) is led to fairness norms that implement the egalitarian bargaining solution. The latter conclusion is more in line with modern experiments on fairness, but one would have to dismiss this as a coincidence if a case could not be made for the claim that prehistoric hunter-gatherer bands operated so anarchic a social contract that agreements between pairs of individuals were viable only if the pair themselves were able to police the agreement without help from other members of the band.

Reciprocity

The one-shot Prisoners' Dilemma is very misleading if used as a model of the human game of life. Since its only Nash equilibrium requires each player to cheat on any co-operative agreement, we would not have evolved as social animals if it were our game of life. As explained first by Hume in his *Treatise of Human Nature* in 1739, the mechanism that sustains human co-operation is *reciprocity*. But Adam cannot threaten not to scratch Eve's back if she won't scratch his, without presupposing that they have an ongoing relationship to nourish. To model such self-policing, long-term relationships, we need to study the Nash equilibria of games that are to be *repeated* an indefinite number of times. If the players are sufficiently forward-looking that future payoffs seem nearly as good as current payoffs, they will be reluctant to cheat on their partners today for fear of losing the fruits of co-operation tomorrow.

Trivers (1971) introduced this idea into biology under the name of *reciprocal altruism*. Axelrod (1984) popularized the notion further by explaining why

it is a Nash equilibrium in the indefinitely repeated Prisoners' Dilemma for each player to use the strategy TIT-FOR-TAT. Since the resulting outcome is that each player receives a payoff of 2 each time the Prisoners' Dilemma is repeated, one learns that rational co-operation is possible without any need to call upon the services of an external enforcement agency. However, the fuss about TIT-FOR-TAT obscures the fact that the problem in studying an indefinitely repeated game is not *whether* co-operative equilibria exists, but *which* of the many co-operative equilibria should be selected.

In the early 1950s, before Trivers or Axelrod, several game theorists independently discovered the *folk theorem* that characterizes the whole set of Nash equilibrium outcomes of an indefinitely repeated game (Aumann & Maschler 1995). For example, the shaded region of Figure 2(b) is the set of all per-game payoff pairs that can be supported as equilibria by sufficiently forward-looking players in the indefinitely repeated Prisoners' Dilemma. The *equilibrium selection problem* in such a game consists of predicting which of these outcomes will actually be observed. The symmetry of the Prisoners' Dilemma makes the symmetric outcome (2,2) focal but real-life games are seldom symmetrical. To create a co-operative species, Nature therefore had to find a way of allowing equilibrium selection devices to evolve. I believe that fairness is one of Nature's solutions to this problem.

If one accepts that fairness norms evolved to co-ordinate behaviour on an equilibrium in a repeated game of life in the absence of any external enforcement agency, then one must also accept that the procedure required to implement the fairness norm must be as self-policing as the equilibrium it is designed to select. Far from postulating a natural duty to be just, I therefore assume that people will cheat on the judicial procedure whenever they can. The only procedures that are viable are therefore those that provide nobody with a motive to cheat. As observed in the previous section, adopting this principle requires that the approaches of both Harsanyi and Rawls be very substantially modified. Rather than being led to the utilitarian outcome that results if one applies orthodox decision theory with external enforcement, one is led instead to an egalitarian outcome r as illustrated in Figure 1(b).

However, a major problem remains. The worthiness coefficients w_A and w_E are undetermined in our specification of the egalitarian bargaining solution. But we need to know what they are if we are to apply the egalitarian solution to the problem of selecting an equilibrium in a game like the indefinitely repeated Prisoners' Dilemma.

Interpersonal comparison of utility

The laboratory experiments that led psychologists to formulate their equity law suggest that modern fairness norms are egalitarian rather than utilitarian, but

further experimentation has been hindered by lack of a background theory able to make predictions about how the worthiness coefficients w_A and w_E should be anticipated to vary with the context. So what does my theory have to say on this subject?

I argue that the food-sharing agreements with which human co-operation presumably began originated within the family. Since we share genes with our kin, it would be surprising if we were not biologically programmed to write their welfare into our utility functions according to their degree of relationship to us. For example, according to Hamilton's (1963, 1964) rule, if Eve is Adam's full cousin, then he should care for her one-eighth as much as he cares for himself. The reason is that the probability that her body is playing host to any specific gene in his body is 1/8. My guess is therefore that we are biologically hardwired to assess the probable degree of relationship to those we encounter within the family circle, and to use this as a standard for making interpersonal comparisons when comparing their lot with our own.

But the interesting case consists of our fairness transactions with strangers. I believe that the fairness algorithm itself is biologically hardwired, but that its adaptation for use with strangers must have been contrived by *cultural* evolution. We learned to adopt strangers into our clans by treating them as relatives. But the degree of relationship attributed to such adopted strangers must have been socially determined. However, if the worthiness of someone outside the family circle is a social convention, then it need not be constant as the context varies. Nor need it be invulnerable to change over time.

The latter consideration is particularly important, since it allows predictions to be made about how worthiness coefficients will adjust over time in a fixed context. In Binmore (1998), I argue that one must expect social evolution to change the way in which people perceive the worthiness of others until the egalitarian bargaining solution r of Figure 1(b) coincides with the Nash bargaining solution n. In principle, one can then predict the relative size of w_A and w_E under ideal conditions. First locate the Nash bargaining solution for the feasible set X with *status quo* ξ. The ratio w_A/w_E is then the slope of the line joining ξ and n.

ANTHROPOLOGICAL EVIDENCE

In this section, I turn to the main business of this chapter, which is to assess the anthropological evidence for and against the assumptions of the theory briefly outlined in the foregoing pages. It first needs to be noted that the consensus is strong among anthropologists that uncontaminated hunter-gatherers, from Greenland eskimos to Kalahari bushmen, operated sharing-caring societies without bosses or social distinctions. The sharing of food, especially meat, is reported to be universal. (Bailey 1991; Damas 1972; Erdal & Whiten 1996;

Evans-Pritchard 1940; Gardner 1972; Hawkes *et al*. 1993; Helm 1972; Isaac 1978; Kaplan & Hill 1985; Knauft 1991; Lee 1979; Megarry 1995; Meggitt 1962; Riches 1982; Rogers 1972; Sahlins 1974; Tanaka 1980; Turnbull 1965. Usually, modern foraging societies are said to be egalitarian, but I prefer not to use this word in a sense that would include a utilitarian society.)

Why share food?

I follow the traditional line that attributes the evolutionary origins of the food-sharing phenomenon to the need for individuals to insure each other against privation. As Evans-Pritchard (1940: 85) explains:

> The habit of share and share alike is easily understandable in a community where everyone is likely to find himself in difficulties from time to time, for it is scarcity and not sufficiency that makes people generous, since everybody is then insured against hunger. He who is in need today receives help from him who may be in need tomorrow.

At least three criticisms of this explanation of food sharing need to be mentioned. The first is that prehistoric hominids are unlikely to have been provided with the 'Machiavellian intelligence' necessary to sustain such insurance contracts (Byrne & Whiten 1988). This piece of jargon expresses the familiar claim that to explain a piece of behaviour in terms of rational self-interest is to assert that it was carefully planned in advance by a coldly calculating intellect. Economists commonly disclaim such straw men by pointing out that a person riding a motorbike is implicitly solving a very difficult mathematical control problem, but nobody would think to deduce that Hells Angels must therefore be master mathematicians. As Evans-Pritchard explains, people in hunter-gatherer societies acquire the *habit* of sharing — and this habit survives because it co-ordinates behaviour on an equilibrium of the game of life without anyone even needing to be aware that a game is being played.

The second and third criticisms arise from differences between the situation envisaged by Evans-Pritchard and more recent reports of modern hunter-gatherer societies. The second criticism disputes the suggestion that hunter-gatherer societies commonly live on the edge of extinction. Sahlins (1974) observes to the contrary that modern hunter-gatherer societies have a relatively affluent lifestyle if one compares the amount of leisure they enjoy after meeting their needs with that of an agricultural labourer or a university professor. But we should not assume that the natural methods of birth control with which modern hunter-gatherers help to regulate their populations preceded the evolution of the food-sharing phenomenon. My guess is that population control is a relatively recent adaptation to the marginal territories currently occupied by hunter-gatherers. But, without controls of some kind, the iron law of Malthus would soon turn plenty into scarcity. How else does one explain the

spread of hunter-gatherer societies over the whole world, even to the most inhospitable of environments? Even for modern hunter-gatherers, every year cannot be a fat year, and it is in the lean years that the invisible hand of evolution strikes down unfit groups. Nor can hunters rely on bringing home the bacon even in the fattest years, so that there will always be good reasons for sharing meat on a reciprocal basis.

The third criticism challenges Evans-Pritchard's appeal to reciprocity as an explanatory factor in food sharing. For example, Erdal and Whiten (1996) conclude that their survey of more than a hundred studies demonstrates that the sharing of food observed 'goes beyond the explanatory power of either kinship or reciprocation. Individuals do sometimes attempt to obtain a disproportionate share of resources or influence for themselves, but this is contained through vigilance and counter-dominant behaviour by their group members.' But what is the second sentence about if not a social contract in which everybody looks after everybody else because those who don't are punished by their fellows? It is true that the mechanism that supports the reciprocal arrangement is not one of the simple models of bilateral exchange that people usually have in mind when they refer to TIT-FOR-TAT. But the punishment strategies that support efficient equilibria in repeated games do not necessarily require that the player injured by a deviant is also the person who punishes the deviation. In the case of modern hunter-gatherer societies, the whole band combines to act as an external enforcement agency in punishing anyone who fails to co-operate in operating the scheme of mutual insurance by means of which it succeeds in surviving when times are bad.

However, although I think that Erdal and Whiten go astray in thinking that the data they survey cast doubt on theories that model hunter-gather societies in terms of rational self-interest, their survey makes it necessary to think twice about the important issue of how authority operates among modern hunter-gatherer societies as compared with the prehistoric societies in which our capacity for making fairness judgements presumably evolved.

Enforcement in foraging societies

Knauft (1991) argues that the evolution of authority in human societies can be seen in terms of a U-shaped curve, in which dominance-structured pre-human societies gave way to anarchic bands of human hunter-gatherers that were then replaced by the authoritarian herding and agricultural societies with which recorded history begins. As Erdal and Whiten (1996) document, the evidence is strong that leadership in modern hunter-gatherer societies lies only in influencing the consensus: 'But when a consensus has been reached, no-one has to follow it against their will — there is no enforcement mechanism.'

At first sight, the apparently anarchic structure of modern hunter-gatherer societies would seem to support my claim that external enforcement structures were indeed absent when the fairness algorithms I believe to be biologically determined were evolving. However, one has to be careful not to put the cart before the horse. One cannot argue that food sharing is the key to human sociality, and simultaneously proceed as though humans were already living in organized communities in the style of modern hunter-gatherers when the fairness norms governing the sharing of food evolved. Nor does the fact that modern hunter-gatherers operate social mechanisms that prevent potentially authoritarian leaders from becoming established imply that their societies do not enforce norms. On the contrary, the evidence is that the social contract operated by a hunter-gatherer community is enforced with a rod of iron. No individual occupies the role of a policeman, but the relatively small size of a hunter-gatherer band makes it possible for *public opinion* to fulfil the same function. When Adam asks himself whether he should offer some of his meat to Eve, he knows very well that he will be relentlessly mocked and ridiculed by the band as a whole should he fail to share in the customary fashion. Full-scale ostracism would follow if he nevertheless persisted in behaving unfairly.

Reports that modern hunter-gatherer communities share on a quasi-utilitarian basis are consistent with the view that public opinion serves as a substitute for an external enforcement agency in such societies. But it is hard to share the enthusiasm expressed by some anthropologists for the oppressive social mechanisms by which discipline is maintained. Envy is endemic. For example, among the !Kung of the Kalahari desert, nobody cares to keep a particularly fine tool for too long. It is passed along to someone else as a gift lest the owner be thought to be getting above him- or herself. But such gifts do not come without strings. In due course, a fair return will be expected. Such close attention to the accountancy of envy in such a social contract makes progress almost impossible. According to Hayek's (1960: 153) definition, the citizens of such a society are free because they are subject to no single individual's will, but it would be a bad mistake for libertarians to idolize such societies. They would do better as a role model for the socialist utopia that Marx envisaged would emerge after the apparatus of the state had withered away.

I therefore diverge from evolutionary psychologists such as Erdal and Whiten who believe that the social contracts of prehistoric hunter-gatherers are preserved in fossilized form by the foraging bands of today. I don't doubt that prehistoric bands were equally free of bosses, but I think it unlikely that they operated a form of social contract that seems to me at least as sophisticated as the authoritarian alternatives operated by ancient tillers of the soil. This claim is of considerable importance to my speculations about the circumstances under which fairness norms evolved, and so it will be necessary for me to defend it at some length.

Farming versus foraging

Cohen (1977) attributes the origins of agriculture to a food crisis in prehistory that arose when human hunter-gatherer bands had expanded until the available habitat was no longer able to support their economies. The response to this over-population problem was twofold. I am particularly interested in the adaptations that allowed foraging to continue in marginal habitats, but anthropologists naturally concentrate on what proved to be the mainstream cultural adaptation — the emergence of agriculture and herding as new modes of production.

The organization necessary both to exploit the increasing returns to scale available in these new modes of production and to prevent the surplus from being appropriated by outsiders made it necessary to abandon the anarchic structure of prehistoric foraging bands. Instead authority was vested in leaders. This readoption of the hierarchical organization typical of ape societies did not require a new set of biological adaptations. We did not lose our capacity to submit to leadership when we acquired the new program that permitted our proto-human ancestors the flexibility necessary to sustain the anarchic lifestyle of hunter-gatherers with a whole world into which to expand. Even in modern foraging societies, our natural urge to dominate our fellows continues to operate in an uneasy relationship with our natural urge to be fair. Otherwise social mechanisms that inhibit dominance behaviour would not be necessary. Anthropologists attribute the social retooling necessary for the transition back to the type of hierarchical social contract needed to maintain a communal farming society to *cultural* evolution. The time available seems too short for a further *biological* adaptation to have been responsible.

It is frequently argued that the human species paid a heavy price for the opportunity to become farmers. When social evolution erected an authoritarian superstructure on a biological foundation that had evolved to permit our ancestors to live a free-wheeling leaderless existence, a war began between part of our biological nature and our social conditioning. Social commentators such as Maryanski and Turner (1992) argue that we are still fighting this war. I express their characterization of a modern industrial society as a *social cage* in the language of game theory by saying that the social conditioning that habituates us to using leadership as an equilibrium selection device conflicts with our natural instinct to employ fairness for this purpose. My guess is that we succeed in tolerating leaders by inventing the social fiction that they are responsible as individuals for the capabilities of the groups they coordinate. The worthiness that would be attributed to the group if it were a person is then conferred on its leader. The leader's claim to more than their fair share is thereby rationalized away. But maintaining such a charade is endemically stressful.

Speculative though it is, such a story about the origins of the farming communities from which our own industrial societies are descended seems relatively uncontroversial, but the same is not true of my belief that the social contracts of proto-human foraging bands were too unlike the complex social contracts of modern foragers to allow the analogy to be useful.

Recall Cohen's (1977) suggestion that a prehistoric food crisis caused by over-population spelled the end of foraging as the normal productive mode among humans. But it did not wipe hunter-gatherers out altogether from those old-world territories where the problem arose. Foragers continued to survive in marginal habitats on the fringes of deserts or the polar ice-cap, where growing crops or herding animals is not feasible. Indeed, the fact that such habitats were colonized is one piece of evidence that favours the over-population theory.

To survive in such marginal habitats without the possibility of emigration, foragers had to develop a culture that was no less at variance with their natural instincts than those who took up the farming option. Three cultural adaptations universally observed in modern hunter-gatherer societies seem especially significant. The first is their use of 'natural' methods of birth control — such as the delayed weaning of children. This adaptation goes some way towards solving their population problem. The harshness of their environment when times are bad probably does the rest. The second cultural adaptation was the development of extremely effective social mechanisms that prevent the emergence of leaders or entrepreneurs — except temporarily in emergencies.

Why should mechanisms that inhibit leadership confer an evolutionary advantage? The reason is presumably that innovators are poison for foraging bands occupying marginal habitats. The survival of the memes that regulate the life of a hunter-gatherer society depends on the equilibrium on which its members co-ordinate in a bad year, when food is scarce. If the crisis is sufficiently severe, some members of the band die, and the rest seek refuge with neighbouring bands. But such lean years are infrequent (Sahlins 1974). In the fat years that intervene, memories of the privations of the last lean year will fade. The band will then be at risk of being seduced by a charismatic entrepreneur into co-ordinating on a new equilibrium that does better at exploiting the surpluses available in fat years. Disaster will then ensue if this new social contract is being operated when a lean year comes along.

In brief, the memes that inhibit the appearance of leaders, who are likely to tamper with a traditional social contract tailored to the conditions that prevail in lean years, serve as a kind of collective unconscious that preserves a folk memory of disasters narrowly avoided in the past. The stubborn conservatism of supposedly stupid peasants occupied in subsistence farming doubtless has a similar explanation.

The third cultural adaptation has already been mentioned. Public opinion can serve as a substitute for an external enforcement agency in small close-knit

communities. Harsanyi's (1977) analysis of the device of the original position under such circumstances then provides us with an explanation of how the use of the fairness algorithm that I believe is written in our genes can result in quasi-utilitarian sharing of the type reported to be universal among modern hunter-gatherers. (I am suppressing the sceptical streak that reminds me of nineteenth-century reports that the infirm and elderly were abandoned by various nomadic tribes of North America when times became hard. Such behaviour is consistent with my theory, because standards of interpersonal comparison are likely to develop that result in the powerless being deemed unworthy when food is shared. However, such behaviour is not compatible with anthropological reports that food is shared strictly according to need.)

Anarchy in prehistory?

Figure 3 illustrates the speculations about the evolutionary history of modern social contracts offered above. Its significant feature is that the sophisticated social contracts of contemporary hunter-gatherer societies appear on a twig on the branch of the tree that leads to societies like our own. To understand the origins of our instinct for justice, we therefore need to go back to the common ancestor of both types of social contract. But this does not imply that nothing is to be learned from the social contracts of modern foraging bands. They show that human biological hardwiring allows us to operate social contracts without ape-like dominance hierarchies. Since modern hunter-gatherer societies manage without bosses, humans do not need bosses to survive as social animals.

Knauft's (1991) U-shaped curve must therefore be correct insofar as it embodies the claim that the imperative for authoritarianism was somehow cleansed from the genes of our pre-human ancestors, only to be revived in relatively recent history as a *cultural* adaptation to the need to domesticate plants and animals in response to population pressures. To understand the circumstances under which human fairness algorithms evolved, we therefore need to look back to a time after the biological imperative to organize in terms of dominance structures had weakened to an extent that made it possible for more flexible social systems to evolve, but before population pressures had led those societies that continued to forage to develop socially sophisticated methods of controlling both their population as a whole and their selfishly inclined entrepreneurs.

Such prehistoric foraging bands must have differed from their modern descendants in several important respects. In particular, barriers to emigration were absent when the whole world was available for colonization. Under such circumstances, their social organization must have been anarchic to an extent that would make modern hunter-gatherer societies look positively paternalistic. How could it have been otherwise when a dissident group always had the

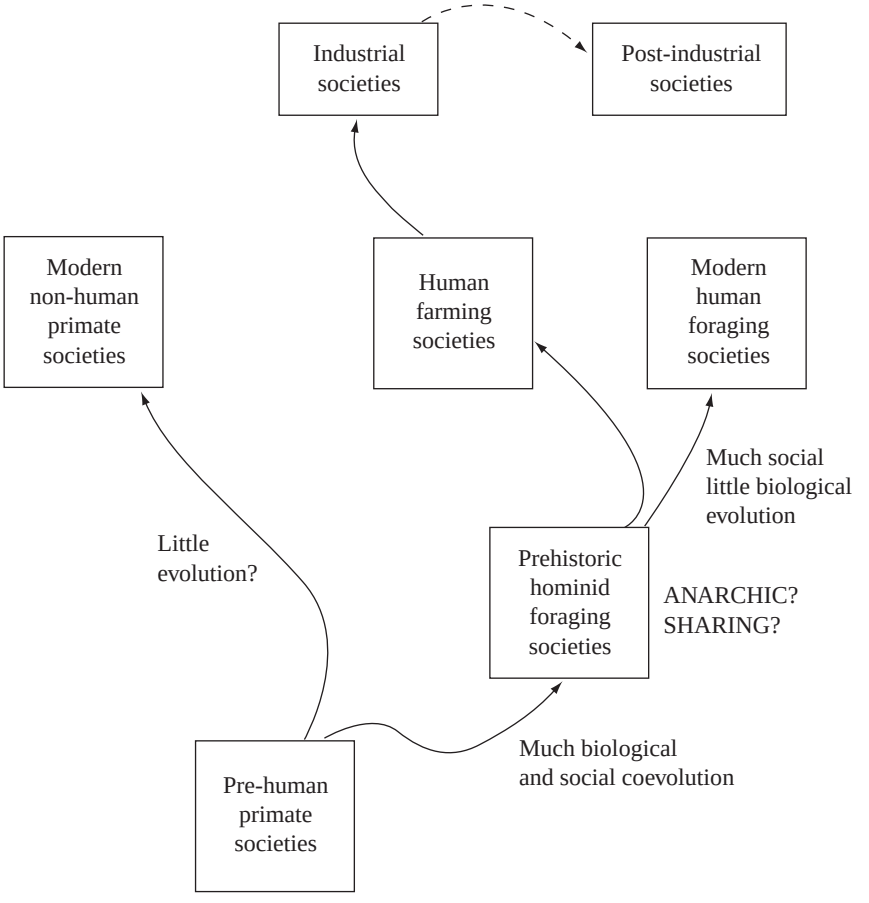

Figure 3. In what environment did fairness evolve?

Lockean option of breaking away at relatively low cost to set up shop in pastures new? Neither public opinion nor personal authority can act as Big Brother when punishment can always be evaded simply by walking off.

With no external enforcement agency available to enforce food-sharing agreements, the Adams and Eves of prehistoric times presumably must have shared food in much the same manner as vampire bats share blood. Vampire bats starve if deprived of blood for more than 60 hours. Wilkinson (1984: 182) observes that close kinship is not necessary for a bat to regurgitate blood for a regular roost-mate. Not only are bats able to recognize each other; the evidence shows that they are more likely to help out a neighbour who has helped them out in the past. Just as each bat in a reciprocating pair has to act as its own policeman in disciplining any tendency by its partner to cheat, so each Adam

and Eve who learned to co-operate would have found it pointless to appeal to the rest of the band about any bad behaviour by their partners.

In short, co-operation must originally have been based on *pairwise* interactions, with each pair responsible for policing its own affairs. One can follow Axelrod (1984) in using TIT-FOR-TAT as a representative of the type of reciprocating strategy necessary to support co-operation in such circumstances. However, it is important to bear in mind that it would be a mistake to do the same when discussing societies whose structure cannot be seen as a collection of overlapping *two-person* subsocieties.

Kinship in small groups

Game theorists never tire of warning against proceeding as though the camaraderie that enlivens small groups who work or play together extends to the world in general. The daily life of such small groups awakens the memories buried in our genes of how to interact with others in the small hunter-gatherer communities of prehistory. The brotherhood of man then manifests itself because we are literally programmed to treat each other like brothers and sisters in such circumstances. But Dunbar (1992) has plausibly argued that the size of a group within which it is possible for people to treat each other like family is limited by the capacity of the human neocortex to sustain a social model in which each person in the group and their relationships with others are modelled individually.

This theory about the social dynamics of small groups would seem to be undermined by Erdal and Whiten's (1996) claim that kinship theories fail to explain how food is shared among modern hunter-gatherers. It is doubtless true that a player's share cannot be calculated from his or her family relationships using some simple formula. But it is hard to see how reports that food is shared according to *need* can be explained without assuming that the players actively sympathize with each other's plight, as predicted by Hamilton's rule.

An example will be useful in illustrating the mechanism envisaged. The same example will then be used to compare the food-sharing norms of modern hunter-gatherers with the different food-sharing norms that my theory attributes to the more anarchic hunter-gatherers of prehistory.

Modern hunter-gatherers

Adam and Eve's individual utility functions are normalized so that $u_A(0) = 0$ and $u_A(1) = u_E(1) = 1$. Their shape is determined by the following considerations. Adam and Eve are assumed to be indifferent between obtaining all of the kill brought home by a hunter and receiving some smaller share σ of the kill. In Adam's case, $\sigma = \frac{2}{3}$. In Eve's case, $\sigma = \frac{5}{6}$. Finally, both Adam and Eve are

assumed to be risk neutral about shares x that lie between 0 and σ. (The graphs of their individual utility functions therefore consist of two line segments, one joining (0, 0) to (σ, 1) and another joining, (σ, 1) to (1,1). A better definition of need in this context would replace the first line segment by a line segment joining (0, 0) to (σ, 0). A player would then reject any share $x < \sigma$ in favour of any lottery with prizes $x = 0$ and $x = \sigma$. The shares in Table 1 would then be replaced by the probabilities that a player's need is met. However, such a model would have to be complicated by providing players with some form of compensation for settling for a lower probability of winning σ.)

Table 1. Families sharing in modern foraging bands.

Adam	Eve	Adam's share	Eve's share	w_E	w_A	w_E^*	w_A^*
stranger	stranger	½	½	4	5	4	5
cousin	cousin	6/11	5/11	3	4	4	5
brother	sister	⅔	⅓	1	2	4	5
son	mother	⅔	⅓	3	10	4	5
father	daughter	⅙	⅚	4	3	4	5

Need is a complex concept, but only one aspect of what it means to say that a person is in need will be relevant here — the extent to which he or she is in want. Adam and Eve's need to eat will be measured by the smallness of the largest share σ for which they are willing to take a risk of ending up with nothing. Their desperation is strengthened by the assumption that they are risk neutral over smaller shares. Since $\frac{2}{3}+\frac{5}{6} > 1$, Adam and Eve's needs cannot be satisfied simultaneously, even though we shall assume that they are the only members of the band who are thought worthy of a share of this particular kill. To what extent will the norm they operate recognize that Adam's need is greater than Eve's?

If I am right about public opinion in a modern foraging band acting as an external enforcement agency, then Adam and Eve's prowess at hunting or gathering will be irrelevant to how the surplus is divided. The social contract will therefore make the lucky hunter's share contingent only on whether that hunter is Adam or Eve. An able hunter may protest and seek to monopolise the kill but will be only one against the combined might of the whole band. The Marxian principle that each should contribute according to his or her ability is therefore realized.

Distribution is not as easy to deal with as production. However, I will proceed as though the band operates a utilitarian norm in which the standard of interpersonal comparison has adjusted until pairs of players divide the surplus according to the Nash bargaining solution. The band as a whole is assumed to enforce its social contract by confiscating the kill should squabbling replace a dignified application of the relevant fairness norm. The relevant *status quo* for

the Nash bargaining solution is therefore (0, 0). It is then easy to show that the kill will be split fifty:fifty between Adam and Eve — whatever values of σ we write into their individual utility functions.

This split seems to be egalitarian, but it appears less so when translated into utility terms. The slope of the relevant part of the Pareto-frontier to their bargaining set X is $-\frac{4}{5}$. To implement the outcome in which Adam's share is $x = \frac{1}{2}$ using a weighted utilitarian solution, we can therefore take $w_E = 4$ and $w_A = 5$. Eve's lesser need is then reflected in her individual utils being counted as worth only $\frac{4}{5}$ of Adam's. But since the moral content of a fairness norm is eroded away as w_A and w_E adjust until $h = n$ in Figure 1(b), this difference in their perceived worthiness does not result in a move in Adam's direction away from the fifty:fifty split.

Now alter the story by making Adam and Eve relatives. Their individual utility functions then have to be replaced by personal utility functions that incorporate Hamilton's inclusive fitness criterion. If Adam's degree of relationship to Eve is r, his individual utility $u_A(x)$ for a split of the kill in which he gets x and Eve gets $1 - x$ must be replaced by his true personal utility:

$$v_A(x) = u_A(x) + ru_E(1 - x).$$

Similarly, if Eve's degree of relationship to Adam is s, then $u_E(x)$ must be replaced by $v_E(x) = u_E(x) + su_A(1 - x)$.

The inbreeding that is inevitable in small isolated groups will be ignored in the first instance. When Adam and Eve are siblings, gene-counting arguments imply that $r = s = \frac{1}{2}$. When they are first cousins, $r = s = \frac{1}{8}$. The possibility that $r \neq s$ is included to take account of parent–child relationships. If a son has survived until puberty and his mother is no longer nubile, then the degree of their relationship has to be altered to take account of their different chances of reproducing their genes. In this example, I consider the extreme cases when $r = \frac{1}{2}$ and $s = 0$, and $r = 0$ and $s = \frac{1}{2}$.

Table 1 shows how the kill is divided when family relationships are taken into account. The constants w_E and w_A are weights whose use ensures that the agreed split maximizes the weighted utilitarian solution, provided that Adam and Eve's personal utilities are properly evaluated to show their sympathy with each other. The constants w_E^* and w_A^* perform the same function in the case when Adam and Eve's personal utilities are mistakenly replaced by their individual utilities.

Except when Adam is Eve's father, the closer the relationship between Adam and Eve, the more his greater need is recognized. When he interacts with his mother or his sister, his needs are met in full. Such recognition of Adam's need is also evident in the standards of interpersonal comparison that operate in the different cases. For example, when Adam and Eve are siblings, one of

Adam's utils is deemed to be worth only half of one of Eve's. When they are cousins, his utils are worth three-quarters of hers. In the exceptional case when Adam is Eve's father, his unreciprocated concern for her welfare results in her needs taking total precedence over his. Her utils are then deemed to be worth three-quarters of his.

Although no simple formula connects who gets what with how they are related, kinship clearly provides a good explanation of why the needy receive special treatment in modern hunter-gatherer societies. The phenomenon is strengthened if we take account of the fact that inbreeding will increase the degrees of relationship. A simple model that assigns probability ρ to the event that any married couple share a particular gene, attributes a degree of relationship $r = \frac{1}{2} + \frac{1}{2}\rho$ to siblings, and $r = \frac{1}{2} + \frac{7}{8}\rho$ to cousins. If $\rho = \frac{1}{7}$, then Adam's needs are met in full, although Adam and Eve may have no obvious family connection at all.

An observer will then see all members of the same generation sharing food as though only need matters. But one cannot deduce that kinship is irrelevant to the way food is shared. On the contrary, the needy are cared for *because* they are kin.

Prehistoric hunter-gatherers

I have argued against assuming that modern foraging bands will serve as a model of the prehistoric foraging bands of our ancestors. If I am right, the social contracts of prehistoric hunter-gatherers were enforced neither by a powerful leader nor by the whole group acting in concert. The parties to a sharing agreement therefore had to police the deal themselves. A form of social organization in which each citizen produces according to his or her ability and consumes according to his or her need would have been beyond their comprehension.

In the simple example we have been studying, the difference between the two forms of social contract emerges in the location of the state of nature. In a prehistoric foraging band, the band as a whole would not have disciplined Adam and Eve by confiscating their product if they fought over its division instead of operating the conventional fairness norm. The analysis therefore needs to be modified so that the *status quo* used when applying the Nash bargaining solution becomes some analogue of Buchanan's (1975) *natural equilibrium*. (My theory predicts that prehistoric foragers used the egalitarian bargaining solution rather than the weighted utilitarian solution, but the result will be the same in both cases after cultural evolution has finished operating on the worthiness coefficients.)

I assume that the probability that the kill is left in Adam's hands after a failure to agree on an insurance contract is $p = \frac{1}{5}$. In the cleanest case, there is no

fighting and a failure to agree simply leaves the carcase in the hands of the player who made the kill. With no fighting, we can identify p with the probability that Adam is the successful hunter, and so p serves as a measure of ability.

The parameter choices in the model imply that we are to study the case in which Eve is more able and Adam is more needy. Table 2 compares the shares each now receives with the case of a modern hunter-gatherer society. Notice that the standards for making interpersonal comparisons have not changed, but the new power structure in their game of life dramatically alters Adam and Eve's share of the surplus. Only Eve's needs are satisfied when Adam and Eve are no more related than cousins. When Eve is Adam's mother, she still reserves more than half the surplus for herself. Even if a high level of inbreeding with $\rho = \frac{1}{3}$ is postulated, Adam still gets less than $\frac{4}{7}$ of the surplus when Eve is as closely related as a sister.

Table 2. Families sharing in prehistoric foraging bands.

Adam	Eve	Adam's share	Eve's share	w_E	w_A	w_E^*	w_A^*
stranger	stranger	$\frac{1}{6}$	$\frac{5}{6}$	4	5	4	5
cousin	cousin	$\frac{1}{6}$	$\frac{5}{6}$	3	4	4	5
brother	sister	$\frac{1}{3}$	$\frac{2}{3}$	1	2	4	5
son	mother	$\frac{7}{15}$	$\frac{8}{15}$	3	10	4	5
father	daughter	$\frac{1}{6}$	$\frac{5}{6}$	4	3	4	5

CONCLUSION

This paper has first offered a summary account of a theory of fairness developed in a much more leisurely style in Binmore (1994, 1998) and then discussed the relevance of the available anthropological data on modern hunter-gatherers to the evolutionary assumptions of the theory. To sustain the theory, it turns out to be necessary to believe that modern hunter-gatherers operate considerably more sophisticated social contracts than did our primitive ancestors.

This paper is a digest of ideas that are explained at greater length in my book *Game Theory and the Social Contract*, *Vol. II*: *Just Playing* (1998).

Note. The support of the ESRC Centre for Economic Learning and Social Evolution and of the Leverhulme Trust is gratefully acknowledged.

REFERENCES

ADAMS, J. 1963: Towards an understanding of inequity. *Journal of Abnormal and Social Psychology* 67, 422–36.

ADAMS, J. 1965: Inequity in social exchange. In Berkowitz, L. (ed.), *Advances in Experimental Social Science* (New York, Academic Press), Vol. II.

ARISTOTLE. 1985: *Nicomachean Ethics*, trans. T. Irwin (Indianapolis, Hackett).

AUMANN, R. & MASCHLER, M. 1995: *Repeated Games with Incomplete Information* (Cambridge, MA, MIT Press).

AXELROD, R. 1984: *The Evolution of Cooperation* (New York, Basic Books).

BAILEY, R. 1991: The behavioral ecology of Efe pygmy men in the Atari Forest, Zaire (Technical Report Anthropological Paper 86, University of Michigan Museum of Anthropology).

BINMORE, K. 1994: *Game Theory and the Social Contract, Vol. I: Playing Fair* (Cambridge, MA, MIT Press).

BINMORE, K. 1998: *Game Theory and the Social Contract, Vol. II: Just Playing* (Cambridge, MA, MIT Press).

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

BYRNE, R. & WHITEN, A. 1988: *Machiavellian Intelligence* (Oxford, Clarendon Press).

COHEN, M. 1977: *The Food Crisis in Prehistory: Overpopulation and the Origins of Agriculture* (New Haven, Yale University Press).

DAMAS, D. 1972: Copper Eskimo. In Bicchieri, M. (ed.), *Hunters and Gatherers Today* (New York, Holt, Rinehart & Winston).

DUNBAR, R. 1992: Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 20, 469–93.

ERDAL, D. & WHITEN, A. 1996: Egalitarianism and Machiavellian intelligence in human evolution. In Mellars, P. & Gibson, K. (eds.), *Modelling the Early Human Mind* (Oxford, Oxbow Books).

EVANS-PRITCHARD, E. 1940: *The Nuer* (Oxford, Clarendon Press).

FURBY, L. 1986: Psychology and justice. In Cohen, R. (ed.), *Justice: Views from the Social Sciences* (Cambridge, MA, Harvard University Press).

GARDNER, P. 1972: The Paliyans. In Bicchieri, M. (ed.), *Hunters and Gatherers Today* (New York, Holt, Rinehart & Winston).

HAMILTON, W. 1963: The evolution of altruistic behavior. *American Naturalist* 97, 354–6.

HAMILTON, W. 1964: The genetic evolution of social behavior, parts I and II. *Journal of Theoretical Biology* 7, 1–52.

HARSANYI, J. 1977: *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (Cambridge, Cambridge University Press).

HAWKES, K., O'CONNELL, J. & BLURTON-JONES, N. 1993: Hunting income patterns among the Hadza. In Whiten, A. & Widdowson, E. (eds.), *Foraging Strategies and Natural Diet of Monkeys, Apes and Humans* (Oxford, Clarendon Press).

HAYEK, F. 1960: *The Constitution of Liberty* (Chicago, University of Chicago Press).

HELM, J. 1972: The Dogrib Indians. In Bicchieri, M. (ed.), *Hunters and Gatherers Today* (New York, Holt, Rinehart & Winston).

HOMANS, G. 1961: *Social Behavior: Its Elementary Forms* (New York, Harcourt, Brace & World).

HUME, D. 1978 (1739): *A Treatise of Human Nature* (*Second Edition*), edited by L.A. Selby-Bigge, revised by P. Nidditch (Oxford, Clarendon Press).

ISAAC, G. 1978: The food-sharing behavior of protohuman hominids. *Scientific American* 238, 90–108.

KAPLAN, H. & HILL, K. 1985: Food sharing among Ache foragers: tests of explanatory hypotheses. *Current Anthropology* 26, 223–45.

KNAUFT, B. 1991: Violence and sociality in human evolution. *Current Anthropology* 32, 223–45.

LEE, R. 1979: *The !Kung San: Men, Women and Work in a Foraging Society* (Cambridge, Cambridge University Press).
MARYANSKI, A. & TURNER, J. 1992: *The Social Cage: Human Nature and the Evolution of Society* (Stanford, Stanford University Press).
MEGARRY, T. 1995: *Society in Prehistory: the Origins of Human Culture* (London, Macmillan).
MEGGITT, M. 1962: *Desert People: a Study of the Walbiri Aborigines of Central Australia* (Chicago, University of Chicago Press).
MELLERS, B. 1982: Equity judgment: a revision of Aristotelian views. *Journal of Experimental Biology* 111, 242–70.
MELLERS, B. & BARON, J. 1993: *Psychological Perspectives on Justice: Theory and Applications* (Cambridge, Cambridge University Press).
RAWLS, J. 1972: *A Theory of Justice* (Oxford, Oxford University Press).
RICHES, D. 1982: Hunting, herding and potlatching: toward a sociological account of prestige. *Man* 19, 234–51.
ROGERS, E. 1972: The Mistassini Cree. In Bicchieri, M. (ed.), *Hunters and Gatherers Today* (New York, Holt, Rinehart & Winston).
SAHLINS, M. 1974: *Stone Age Economics* (London, Tavistock).
SELTEN, R. 1978: The chain-store paradox. *Theory and Decision* 9, 127–59.
TANAKA, J. 1980: *The San Hunter-Gatherers of the Kalahari Desert: a Study of Ecological Anthropology* (Tokyo, Tokyo University Press).
TRIVERS, R. 1971: The evolution of reciprocal altruism. *Quarterly Review of Biology* 46, 35–56.
TURNBULL, C. 1965: *Wayward Servants* (London, Eyre & Spottiswoode).
WALSTER, E., WALSTER, G. & BERSCHEID, E. 1978: *Equity: Theory and Research* (London, Allyn & Bacon).
WILKINSON, G. 1984: Reciprocal food-sharing in the vampire bat. *Nature* 308, 181–4.
WILSON, J. 1993: *The Moral Sense* (New York, Free Press).