

From: *The Lernerian Society*
Tom Elster, 1989

2. Bargaining

Introduction

Bargaining occurs when there are several cooperative arrangements and the parties have conflicting preferences over them. By a cooperative arrangement I mean any outcome (a) that is better for everybody than the state of anarchy, (b) in which there are no exploiters, defined as noncooperators whose cooperation would cost them less than it would benefit them and others, and (c) in which nobody ends up being exploited, that is, as a cooperator whose cooperation costs him more than it benefits himself and others.¹ By anarchy I mean, following the typology of cooperation set out in the Introduction, the absence of any of the following: actions with positive externalities,² helping, a convention equilibrium, a feasible joint venture or a private ordering.

The parties are assumed, that is, to have a common interest in arriving at some agreement but a conflict of interest over which agreement that is to be. The central problem of bargaining, in theory and practice, is that *the very plurality of cooperative arrangements may prevent any of them from coming about*. Bargaining differs from the narrowly defined collective action problem, in which there is typically a unique cooperative arrangement: that in which *everybody* participates *equally*. It also differs from cases in which there are several cooperative arrangements, but no conflict of interest, either because all are indifferent among the arrangements or because one of them is better for everybody than the others.

Bargaining must be distinguished from attempts to reach agreement by rational discussion.³ One way of characterizing the latter is as bargaining in which strategic misrepresentation and other forms of jockeying for position are not allowed. Although this may capture part of the idea of ra-

tional discussion, it gives too much weight to the bargaining power of the parties. In rational discussion, the only thing supposed to count is the 'power of the better argument', including arguments that are radically dissociated from the bargaining power of the parties. Arguments from behind the 'veil of ignorance' stipulate, for example, that certain actual features of the parties are irrelevant, be it their wealth (meritocratic theories), wealth and skills (Ronald Dworkin)⁴ or wealth, skill and preferences (Rawls). Bargaining, by contrast, takes account of all actual features of the parties. This is why, for instance, no bargaining process, however untainted by strategic elements, would leave the severely handicapped with anything.⁵ Since they make no contribution to the net social product, they have no bargaining power. Similarly, the interests of future generations cannot be represented in a process of bargaining.⁶ It follows that justice cannot be based on bargaining if one believes, as I do, that any theory of justice is constrained by an intuition that the handicapped and the future generations should not be left to their own devices.

In this chapter I discuss mainly a narrow range of bargaining problems, those, namely, that arise in joint ventures and private orderings. Bargaining problems arising out of externalities are postponed until Chapter 4. As a paradigm case of a joint venture I use cooperation between labour and capital in production, giving rise to negotiations over the division of gains from cooperation. As a paradigm case of private ordering I use bargaining over financial custody and financial settlement in the aftermath of divorce.⁷

These examples are convenient in that they involve two parties only, since theories of bargaining have been developed mainly with the two-person case in mind. Although they can be extended to the general n -person case, this extension is often artificial. n -Person bargaining theory rests on the assumption that cooperation is either total or totally absent. More precisely, the only coalition that can form is the grand coalition involving all agents, agreeing to coordinate their actions for mutual benefit. If that coalition does not form, no cooperation occurs. Sometimes this conception is empirically adequate and the general conclusions of two-person theory

⁴ Dworkin (1981).

⁵ It is no counterargument to say that their welfare might enter as an argument into the utility functions of other bargainers. The welfare of the handicapped should not rest on this fragile and contingent basis.

⁶ For the same reason, it is no counterargument to say that the welfare of future generations may enter into the utility functions of the currently living, by concern for one's children, for example.

⁷ This example is discussed at some length in Elster (1989a), ch. 3.

¹ Or, more briefly, Pareto optimality without exploitation.

² A notion that can be extended to include the presence of negative externalities.

³ For the latter, see, e.g., Midgard (1980) and Habermas (1982).

apply. Cartel formation, for instance, is often pointless unless all firms participate, since a single outsider might corner the market.⁸ In joint ventures that require the participation of all partners, smaller coalitions are pointless. Usually no single member of a firm has the power to bring production to a complete halt, but the members may always be partitioned into groups in such a way that each group is indispensable. If the workers are organized in different unions that correspond to one such partition, *n*-party bargaining will be the rule.⁹

If, however, the unions that organize the firm's workers cut across functional divisions, so that two workers doing similar work could belong to different unions, bargaining that allows for coalitions of any size will tend to occur. This is also true if unions are organized along functional lines, but not in such a way that each corresponds to an indispensable group of workers. In these cases, management will try to ally itself with one union against the other. Each union has the choice between forming a united front and allying itself with the management. In such cases, *n*-person bargaining theory is of little help in predicting the outcome. Although there are many theories of coalition formation, I think it is fair to say that none of them is very satisfactory. Indeed, the very fact that there are so many suggests that none of them is very useful.

I shall consider three ways of approaching bargaining phenomena. First, we can try to *predict* the outcome of bargaining from the assumption that people's behaviour is guided by specific principles. In this chapter, I assume throughout that bargainers are rational. In Chapter 6 the idea that they are guided by social norms is introduced as an alternative hypothesis. Next, we can try to *describe* the pattern of outcomes that are realized in actual (experimental and real-life) bargaining. Finally, we may lay down normative principles to *evaluate* the outcomes of bargaining, by comparing them with the outcome that ought to be reached. These principles might, for instance, guide an arbitrator. I refer to these as the analytical, behavioural and normative aspects of bargaining. The main problems in the literature can then be formulated as follows. Will rational bargaining lead to a normatively acceptable outcome? Are the predictions of rational bargaining theory confirmed by behavioural evidence? If not, is there an alternative theory that performs better in this respect? If so, are the outcomes predicted by this theory normatively acceptable?

From the pioneering work of John Nash,¹⁰ bargaining has been considered mainly as a cooperative game. In this approach, the problem is not to predict whether a Pareto-optimal agreement will be reached, but to determine which settlement the parties will agree on. Failures to realize gains from cooperation are excluded by definition. (Somewhat paradoxically, the possibility of disagreement nevertheless plays a role in determining what the agreement will be.) The next section states the basic assumptions and results of this approach, which remains a fundamental tool for understanding bargaining processes.

Beginning with Nash himself, many writers have felt, however, that the cooperative theory of bargaining is an unsatisfactory description of behaviour. Pareto optimality should be derived as a theorem from individualistic premises, not stipulated as an axiom. If bargaining is understood in a normative sense, as (costless) arbitration rather than as a process of proposals and counterproposals, the stipulation that the gains from cooperation be fully realized is more acceptable, but in analytical and behavioural context the possibility of bargaining failure cannot be excluded *a priori*. There are two ways to handle this issue. One is to search for microfoundations for collective rationality, to argue that individually rational players will avoid bargaining failure. Another is to offer a positive theory of disagreement in bargaining, distinguishing between the conditions under which agreement will be reached and those in which failure may be expected. In particular, if the parties have less than full information about each other's preferences and information, bargaining may break down as each party forms unrealistic expectations about the concessions the other is willing to make.

The present chapter surveys a variety of theories and models, without displaying great faith in any of them. The reader might well wonder about the point of the exercise. The justification for my procedure is that by working through and reflecting upon these models, we enhance our understanding of the underlying issues. By seeing why and where a particular model goes wrong, we become aware of features of bargaining that otherwise would have gone unnoticed or been taken for granted. Also, each of the models probably has substantial explanatory power in special cases. In rational or 'norm-free' bargaining with full information, for instance, the standard economic models probably perform quite well. If the assumptions of rationality and full information are violated, the process of bargaining

⁸ Olson (1965), pp. 40-1. ⁹ Horn and Wolinsky (1988).

¹⁰ Nash (1950, 1953).

becomes more opaque, yet nonstandard models may at least suggest in which direction the outcome will differ from that predicted by the standard model.

Cooperative models of bargaining

Bargaining can occur over divisible or indivisible objects. Let us first assume that the objects are continuously divisible and that the point of contention is how to divide them among the parties. The objects may be one-dimensional or many-dimensional. Labour and management negotiate simultaneously over working conditions, salary, employment, the contract period and the like. Any given proposal or counterproposal is a multidimensional package. It is often convenient, however, to represent the package in terms of the utilities which the bargainers assign to it. The set of feasible bargains is represented by, or reduced to, the set of feasible utility pairs. As a result, much information about the physical features of the bargaining situation is discarded. Within a given bargaining situation, two physically different proposals may be indistinguishable in terms of the utilities assigned to them by the bargainers. Two bargaining situations which represent totally different physical problems may be represented by the same set of feasible utility pairs.

The source of utility may be one's own consumption of the object of bargaining, or someone else's consumption, possibly even that of the other bargainer, which may have positive or negative weight in one's own utility function. There is no need to assume that bargainers are selfish. Even a society of altruists would have to bargain over the allocation of goods among them. If I derive utility from your consumption of a good and you from mine, each of us will want to shift consumption towards the other. The conflict will be resolved by a sequence of offers and counteroffers that is formally indistinguishable from, say, the process of labour-management bargaining described later. The basic source of bargaining problems is scarcity of resources, not selfish motivations. Note also that because the parties are bargaining over utilities, it makes no sense to suggest that malevolent bargainers, who derive pleasure from each other's nonconsumption, have an incentive to let bargaining break down. The suggestion would involve double counting, since any externalities in the utility function would already be incorporated into the representation of the feasible utility pairs.

For analytical purposes, the representation of the situation as bargaining over utility is often quite acceptable. If the task is to predict the outcome of bargaining among rational individuals, it is not implausible to assume that they are interested only in the utilities they derive from the outcomes, and not in the physical carriers of these utilities. If, however, one suspects that bargainers are not always fully rational, the simplification may be less defensible. If they are subject to cognitive bias and distortion, they may be distracted by irrelevant physical aspects of the situation.¹¹ Attempts to predict the outcome that do not take account of such psychological tendencies are then likely to fail. Also, in the absence of full information about preferences, salient features of the physical situation may be important determinants of the outcome. Equal or proportional physical diversion are obvious focal points when utilities are unknown.¹² Moreover, for normative purposes the physical aspects of the situation may be directly relevant. It has been shown experimentally that the problem of distributing a given number of avocados and grapefruit between two people yields very different ethical intuitions when we are told that these fruits are valued for their taste and when we are told that they are valued for their content of vitamin C, even if the utility functions are identical in both cases.¹³

Assuming, then, a situation in which two parties are bargaining over divisible objects represented by their utilities, we can state the cooperative approach in terms of a diagram (Fig. 2.1). The *bargaining situation* is fully described by a set of feasible utility pairs (derived from the set of feasible physical bargains) and a disagreement point which specifies the utility of the outcome that will be produced if the parties fail to reach agreement. (The role of the disagreement point is controversial, and discussed later.) We assume that the set of feasible utility pairs includes all points in the area circumscribed by *OPABTO* and that *d* is the disagreement point. The feasible set is assumed to be convex, meaning that all points on a line between two feasible points are also feasible. If, for instance, the bargaining is over the division of a sum of money, the decreasing utility of money will ensure the convexity of the feasible set. The question is which if any of the feasible points will be chosen as the outcome or the 'solution' to the

¹¹ See notably Bazerman and Carroll (1987). ¹² Schelling (1963).

¹³ Yaari and Bar-Hillel (1984). They also show, more disturbingly perhaps, that intuitions differ in situations that have identical representations in utility space and rest on similar evaluations.

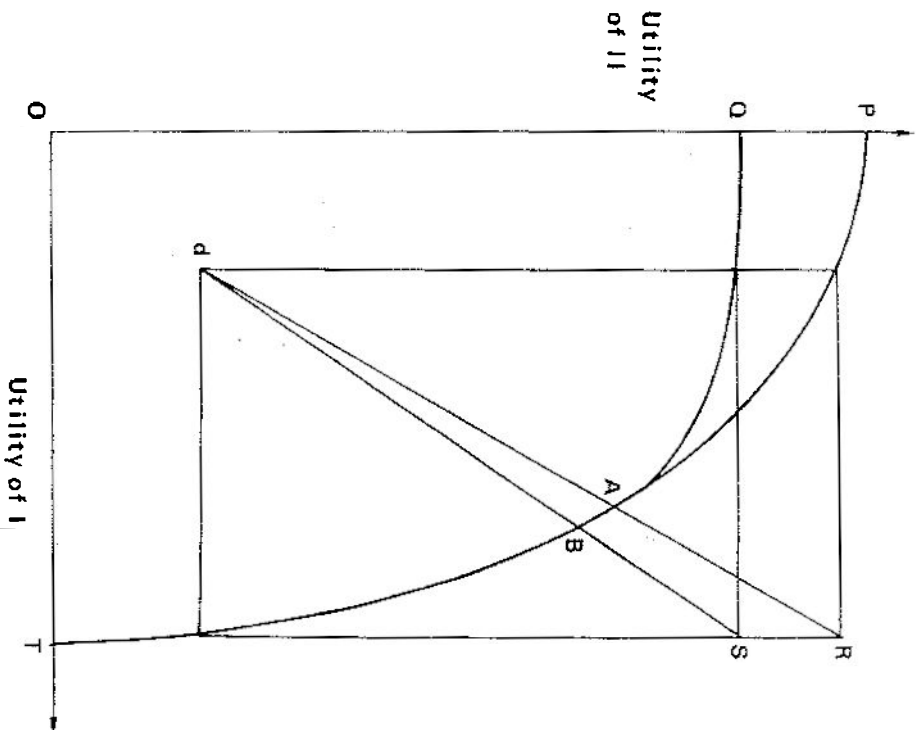


Fig. 2.1

bargaining problem. A *theory of bargaining* – analytical or normative – can be summarized by a function f which for any bargaining problem – any convex set S and any disagreement point d – picks an element in S as the outcome that will be reached or ought to be reached.

The construction of such a function is a nontrivial task. Nash did it by laying down conditions which one would expect any pair of rational bargainers (or any fair arbitrator) to respect and showing that there was only one function satisfying all of them. Specifically, he stipulated four conditions:

1. *Pareto optimality.* It should not be possible to improve the outcome for one party without loss for the other. The solution should be on the Pareto frontier, $PABT$ in the diagram.
2. *Invariance.* The solution should be invariant with respect to positive linear utility transformations. This condition is explained later.
3. *Symmetry.* If the feasible set is symmetrical around the 45° line, with the disagreement point on that line, the solution should also be on that line.
4. *Independence of irrelevant alternatives.* If we have two bargaining problems (S, d) and (T, d) with S included in T and if the solution to (T, d) is a member of S , it should also be the solution to (S, d) . Or more simply, if the solution in a larger game remains feasible in a smaller game, it should also be the solution in the latter.

Nash proved that these conditions, together with the assumption that the set of feasible utility pairs is convex, uniquely define a solution concept: the outcome of bargaining is constrained to be the point in the feasible set which *maximizes the product of the utility gains* of the parties, compared with the disagreement point. In the game $(OPABTO, d)$ this is the point A . From a normative point of view, this solution concept has no special appeal, apart from the axioms which jointly imply it. Indeed, it might appear positively unattractive, because of the following property. If we assume that a poor man and a rich man are bargaining over the way to divide some amount of money large enough to be very important to the poor man, the Nash solution will assign most of it to the rich man, because he can more credibly make a proposal favourable to himself and say, 'Take it or leave it'.¹⁴ This is the 'Matthew effect' in bargaining.

The Nash solution often corresponds well to institutions about the way people actually behave in bargaining situations, but one might argue that it ought not to be chosen by an arbitrator who tries to reach a fair decision. Other solutions might appear more attractive, such as the point which maximizes the sum of the utility gains of the parties, the point on the Pareto frontier point that equalizes their utility gains or, more strongly, some point that implies a larger utility gain to the poor man than to the rich. These proposals, however, violate the invariance condition (about which more later). And in any case one might argue that even from an arbitration

¹⁴ Luce and Raiffa (1957), pp. 129–32.

point of view, the task is not necessarily to find the abstractly just outcome, but one that is *appropriate* given the bargaining power of the parties.¹⁵

Before I discuss the axioms, it is time to meet an obvious objection: How does bargaining theory handle situations in which there is only a small number of feasible outcomes? Does it single out one of them as the solution? In that case, how does it handle symmetrical cases in which each of two nondisagreement outcomes gives everything to one party and nothing to the other? We can use Solomon's judgement to illustrate the problem. Before he knew anything about the preferences of the parties, he had to treat them as identical. There was no reason for preferring one woman over the other. The set of feasible outcomes not being convex, the symmetrical solution prescribed by the Nash axioms is not available. Solomon's first proposal, cutting the child in half, effectively amounted to an arbitration impasse. The reactions of the women to that proposal enabled him to form a better impression of their preferences and, indeed, to make a better decision. But what if both women had reacted as the true mother did, the false one hoping that her willingness to give the child to the other would stop Solomon from doing so? In that case, what should he have done?

Let us consider a similar problem that arises between spouses in child custody bargaining, conceived as private ordering and not as arbitration. I shall assume that there are two children, a boy and a girl, with four possible custody arrangements (with respective utility assignments) corresponding to the vertices of Fig. 2.2.

In most divorces, bargaining over custody takes place simultaneously with bargaining over the financial settlement. The two negotiations tend to be coupled, so that the parents try to extract financial advantages by offering custody and vice versa. Let us suppose that the vertices represent the parties' utility for the various custody arrangements together with one particular financial settlement, namely that which the court will make if the parties cannot reach agreement by themselves.¹⁶ Intermediate outcomes can then be generated by allowing the parties to transfer money to each other, if necessary by borrowing. These outcomes will *not* correspond to the straight lines between the vertices. Because the parties have decreasing marginal utility of money and because the utility of money interacts with

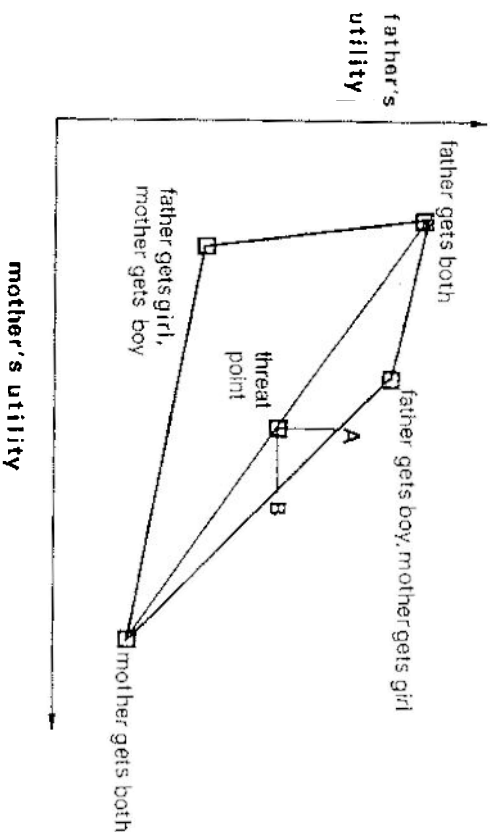


Fig. 2.2

that of custody, the relation will be more complex. But the set of intermediate outcomes will be convex, as required by bargaining theory, so that a solution can be derived.

Side payments are not always feasible. Perhaps one of the parents has no money and is unable to get a loan. Or perhaps the situation is such that side payments are thought to be ethically unacceptable. The parents might agree, for instance, to make the financial settlement before the custody decision, because it will hurt the children to think that they are convertible into money. Another, more general way of generating intermediate outcomes is then available, at least in principle. Solomon could have flipped a coin to decide between the women. The parents can use a lottery in which the various allocations are assigned definite probabilities, adding up to 1. To each such lottery corresponds a point within the quadrangle defined by the four allocations. The utility of a lottery is simply the sum of the utilities of the allocations weighted by their probability. (This is explained later.) Thus the points on a line between two vertices correspond to lotteries in which the allocations underlying these vertices are assigned probabilities p and $1-p$, with p taking on all values between 0 and 1. In this way, the set of feasible outcomes is rendered convex, so that bargaining theory can be applied.

To predict the outcome of the bargaining, we must first make an assumption about the threat point. We can stipulate, as in Fig. 2.2, that the

¹⁵ Selten (1987), pp. 46-7.

¹⁶ Courts, unlike bargaining spouses, dissociate financial settlements from custody settlements.

parents believe that in a legal dispute each of them has a 50 per cent chance of getting custody of both children. This belief could be brought about in several ways. If the legal rule is to give custody according to what is in the best interests of the child, the parents might well believe that there is no detectable difference in fitness between them, so that, for all they know, each has as good a chance as the other of getting custody. Or the legal rule might actually be to use a fair coin to settle the issue.¹⁷ Whatever the grounds for the fifty-fifty expectation, it follows that the expected utility of the parents in the case of a legal dispute is midway on the line between the vertices corresponding to maternal custody of both children and paternal custody of both. Both parents know that by going to court they can achieve at least this level of expected utility; hence they will reject any proposed solution which offers them less. (For simplicity, I ignore the costs of litigation.) If, moreover, we assume that they will not accept any solution which is worse for both than some other feasible outcome, we see that if they reach an agreement, it will be somewhere on the line *AB*. Each point on *AB* assigns a probability *p* that the mother gets custody of both children and a probability $1 - p$ that the father gets the boy and the mother gets the girl. By inspecting Fig. 2.2 we can see that *p* ranges between .22 and .45 (approximate values).

When parents bargain over custody, they will rarely if ever choose a point on *AB*. Indeed, I think lotteries are virtually never used to settle private, nontrivial disputes. I have no systematic empirical evidence to back this claim, only casual observation, together with some general arguments. First, of course, the conditions under which lotteries or probabilistic compromises are superior to physical compromises may not often be realized. In particular, when side payments are available and acceptable, they provide a much more robust form of compromise. Second, even when a lottery seems to be called for, lack of enforceability might prevent it from being used. Each party might agree to a lottery in the hope that the outcome will be his or her preferred alternative, and then renege if it turns out differently. The knowledge that this may happen could easily prevent a lottery from being attempted in the first place.¹⁸ And as far as I know, no country has a public official or public institution with the power to carry

out and enforce lotteries privately agreed upon by the parties. Hence I shall not consider lotteries as a serious way of resolving bargaining impasses. Let us, then, limit our attention to bargaining over divisible objects and to bargaining over indivisible objects in which side payments are possible. With convexity thus ensured, the plausibility of the Nash solution depends on the plausibility of the four axioms. I have already discussed the condition of Pareto optimality, and I shall have more to say about it later. First, however, I discuss the other Nash conditions one by one.

Invariance has the effect of imposing a special kind of utility function on the bargaining situation. At one extreme, one can show that with purely ordinal utility functions the bargaining problem cannot be defined.¹⁹ At the other extreme, one may argue that well-known problems associated with interpersonal comparison of utilities prevent us from defining such solution concepts as 'maximize the sum of the utility gains' or 'equalize the utility gains'. An intermediate category, represented by the invariance condition, is that of a von Neumann-Morgenstern utility function, in which utility is 'uniquely given up to an arbitrary positive multiplicative constant and an arbitrary additive constant.'²⁰ The relation between any two utility functions which represent the preferences of a given individual is like the relation of Celsius to Fahrenheit temperature scales. Statements such as 'The sum of temperatures in New York and Chicago is larger than the sum of temperatures in London and Paris' are not meaningful, since they do not always retain their truth value when we go from Celsius to Fahrenheit. By contrast, the statement 'The difference in temperature between New York and Chicago is larger than that between London and Paris' is meaningful, since its truth value does not depend on the choice of temperature scale. Analogously, interpersonal comparisons of utility levels are meaningless with von Neumann-Morgenstern utility functions, but some interpersonal comparisons of intrapersonal differences are feasible. In particular, statements comparing the rate of change of the marginal utility of money and commodities are meaningful. In the bargaining problem between the rich man and the poor man, the latter is at a disadvantage because for him the marginal utility of money decreases rapidly while for the rich man it is approximately constant.

The invariance condition is, however, implausible, on analytical, be-

¹⁷ I discuss this proposal in Elster (1989a), sec. 3.5.

¹⁸ Lotteries in private bargaining without a third-party enforcer will be used only if the parties are moved by 'self-interest without guilt', as explained in the concluding chapter.

¹⁹ Shubik (1982), pp. 92-8.

²⁰ For a lucid exposition, see Luce and Raiffa (1957), ch. 2.

havioural and normative grounds. To see why, consider two bargaining problems A and B, both involving the same bargainers I and II.²¹ In both problems the bargainers are assumed to be indistinguishable, in the sense that any utility function that represents I's preferences can also represent those of II. In both games, the object of the bargaining is to agree on two numbers p and q between 0 and 1 and summing to 1 or less. If they agree, I has probability p of winning the prize in a certain lottery and II probability q of winning the prize in another lottery. If they cannot agree, neither gets anything. In bargaining problem A both lotteries involve similar prizes, namely two bicycles. By symmetry and Pareto optimality, the outcome of the bargaining must be $p = q = .5$. In bargaining problem B everything is unchanged, except that the prize in I's lottery now is a Rolls Royce. Assuming invariance, it is easy to see that the solution must be the same, that is, $p = q = .5$. Consider, namely, the utility to I of any outcome (p, q) in II. Writing u for the utility to him of the Rolls Royce and v for the utility to him of the bicycle, the utility of (p, q) to him is $p \cdot u + p(u/v)v = [u/v]pv$. In other words, the utility he derives from any given probability in the second bargaining problem equals the utility he derives from the same probability in the first problem, multiplied by a positive constant u/v . But this means that the second bargaining problem can be derived from the first by a positive linear transformation, so that the solution must be the same.

This result is analytically implausible. In problem B, II would certainly be able to demand $q > .5$. He could say, credibly, that since I's desire for a Rolls Royce was much stronger than his own desire for a bicycle, I must accept $p < .5$: 'Take it or leave it'. To be sure, this involves interpersonal comparison of utilities, but not a very difficult one since the prizes are so different, assuming that I and II are reasonably similar persons.²² From a normative point of view it is also arbitrary that I should get a 50 per cent chance of a very valuable object and II a 50 per cent chance of a much less

²¹ The following draws upon Kalai (1985) and Roth (1987).

²² For an argument that interpersonal comparisons of utility are not only possible but inevitable see Davidson (1986). A method for constructing interpersonal comparisons from interpersonal comparisons is proposed by Ormrod-Orlin and Roemer (1987). It is clear that sometimes we have no difficulty carrying out such comparisons. It would be tempting to conclude that with more progress in psychology an increasing number of cases should lend themselves to comparison. This presupposes, however, that utility (or happiness, welfare or well-being) is a one-dimensional concept. If, however, utility is many-dimensional (as suggested in Sen 1980-1), we may never be able to do better than a partial ordering of welfare levels.

valuable one. Surely, in the absence of further information about them²³ the reasonable solution would be to choose p and q so as to equalize expected monetary reward — not to equalize the chances of getting two very different rewards. When we throw away information not only about the physical nature of the problem, but also about interpersonal utility information, we lose an essential aspect of the bargaining process. In fact, our intuitions about the bargaining problem between the rich man and the poor man probably derive as much from interpersonal comparisons of utility as from the fact that the poor man's marginal utility decreases more rapidly. To a poor man, an extra dollar simply means much more than it does to the rich, who has, therefore, much less to lose if no bargain is struck.²⁴

Alvin Roth and his collaborators conducted extensive behavioural studies of the invariance condition.²⁵ They found that with different prizes the outcome of bargaining was not a fifty-fifty allocation of the chances. On average, the outcomes favoured the party with a smaller prize. They did not find, however, that the outcomes clustered around the allocation that would give equal expected monetary value. Rather the distribution was bimodal and tended to 'cluster around two "focal points": the equal probability agreement and the "equal expected value" agreement that gives each bargainer the same expected value'.²⁶ This finding suggests that to have an equal division of something is more important than the nature of the dividendum. This idea, which is also supported by studies from non-bargaining contexts,²⁷ is explored in Chapter 6.

Symmetry is intended to capture the idea that when the parties have the same bargaining power, the outcome should in some sense reflect that equality. Given invariance, 'bargaining power' cannot here refer to absolute levels of utility. It has to be understood in terms of features which remain invariant under positive linear transformations. If we think again in terms of an underlying monetary bargain, one such feature could be the rate of decrease of the marginal utility of money. The notion of bargaining power is discussed separately later.

²³ It could be, of course, that A is so inept at transforming goods into utility that he requires a Rolls Royce to achieve the same utility level for which B needs only a bicycle. But surely the burden of proof will then be on A to produce evidence about this unlikely state of affairs.

²⁴ I am grateful to Luc Bovens for helping me see more clearly the relation between interpersonal and interpersonal comparisons of utility in such cases.

²⁵ Summarized in Roth (1987).

²⁶ *Ibid.*, p. 21. ²⁷ Harris and Joyce (1980).

Independence of irrelevant alternatives implies, in terms of Fig. 2.1, that when the feasible set is restricted from $OPABTO$ to $OQABTO$, the solution should remain at A since this point remains feasible. One might want to object to this condition. It would appear that the bargaining strength of II is weakened when the part of the feasible set which is most favourable to him is eliminated. Hence if outcomes reflect bargaining strength, II should fare worse in the smaller bargaining problem than in the larger. At the very least, he should not be able to improve his position when the odds change against him. To capture this intuition Ehud Kalai and Meir Smorodinsky have proposed an alternative condition, to be substituted for the Nash independence condition.²⁸ This *axiom of monotonicity*, like the independence condition, is stated as a comparison between two bargaining problems. It says that if, for every utility level that player I may demand, the maximum feasible utility level for player II is at least as large in the second game as in the first, then the utility level assigned to player II according to the solution should not be less in the second game than in the first. Or, more briefly, no one should suffer from the feasible set expanding in his favour.

Kalai and Smorodinsky proved that this condition, together with Pareto optimality, invariance and symmetry, implies the following solution concept. The utility gains should be proportional to the maximum feasible gains which the parties could achieve. In Fig. 2.1, the (nonfeasible) combination of the maximal feasible gains for the problem $OPABTO$ is represented by the point R .²⁹ The solution occurs at the intersection between the Pareto frontier $PABT$ and the line from R to the disagreement point d . In this case, the Nash solution and the Kalai-Smorodinsky solution happen to coincide. If, however, we restrict the feasible set to $OQABTO$ the Nash solution remains at A while the Kalai-Smorodinsky solution is moved to B .

This solution concept is arguably more plausible than that of Nash, both on behavioural and on normative grounds. 'Intuitions about "bargaining power" and "fairness" might include the notion that if A could win a lot in a bargaining situation, he or she is "entitled" to more than if he or she could only, in the best of circumstances, win a little'.³⁰ Although the ex-

²⁸ Kalai and Smorodinsky (1975).

²⁹ The maximum feasible gains are constrained by the disagreement point. What I can get in the event that II gets less than his disagreement payoff cannot be relevant to the outcome.

³⁰ McDonald and Solow (1981), pp. 905-6.

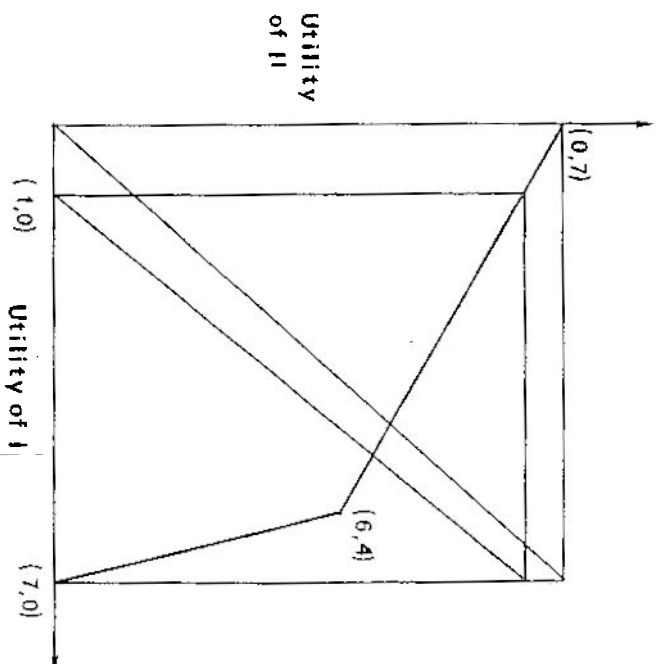


Fig. 2.3

perimental evidence is ambiguous,³¹ one would expect wage bargaining, for instance, to be sensitive to the maximal feasible gain. Higher unemployment benefits, which increase the workers' reservation wage and hence reduce the maximal feasible profit, should strengthen the bargaining power of the workers. Under the Kalai-Smorodinsky solution this will always happen. In the cooperative Nash model, it may or may not happen. (We shall later see that in noncooperative models it essentially never happens.) Thus in Fig. 2.3, define S as the set spanned by $(0, 0)$, $(0, 7)$, $(6, 4)$ and $(7, 0)$ and compare the two bargaining games with disagreement points $d = (0, 0)$ and $d' = (1, 0)$. We may think of the second game as defined by an increase in the reservation wage for the workers (player I). We observe that under the Kalai-Smorodinsky solution, the outcome is shifted

³¹ In their experiments, Nydegger and Owen (1975) found that the Nash solution was a better predictor than the Kalai-Smorodinsky solution. So did Roth and Malouf (1979). In support of their view, however, Kalai and Smorodinsky cite Croit (1971). The results reported by Nydegger and Owen are dominated by their subjects' massive preference for equality. A better test would involve a problem with an asymmetrical Nash solution.

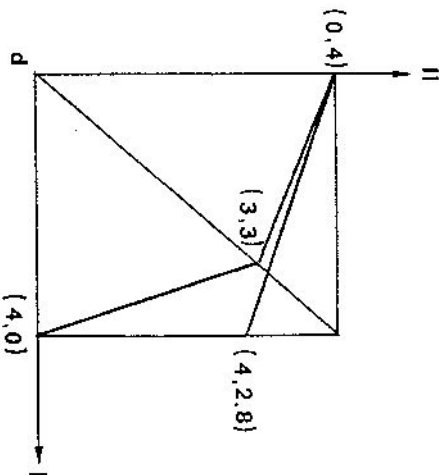


Fig. 2.4

in favour of the workers, whereas under Nash it can easily be shown to be the same in both cases, namely (6, 4).

Normatively, the monotonicity axiom seems more plausible than the independence axiom. It would seem perverse to accept a solution concept which entailed that one person could get less as a result of the feasible set expanding in his favour. Figure 2.4 (from Kalai-Smorodinsky) shows how this might happen. There are two games, both with disagreement point (0, 0) and spanned by [(0, 0), (0, 4), (3, 3) and (4, 0)], [(0, 0), (0, 4), (4, 2.8) and (4, 0)], respectively. The Kalai-Smorodinsky solutions are found at the intersections between the diagonal and the bargaining frontiers: (3, 3) in the smaller game and (40/13, 40/13) in the larger. The Nash solutions are at the corners in both games. In the second game II is more highly favoured than in the first game,³² and yet under Nash he ends up with less in the second.

If we try to assess the two conditions from an analytical point of view, to determine which, if any, would be respected by rational players, intuition can easily lead us astray. Both conditions have analogues in the theory of individual choice, where they are plausible and perfectly consistent with each other. Here the independence condition says that choices should not change when the feasible set contracts to exclude an item that was not chosen in the first place. If the menu offers beef, chicken and salmon and

³² Actually, both parties are more highly favoured. However, I gains under both solution concepts when the feasible set is expanded in his favour.

the customer has chosen chicken, he should not, barring special circumstances,³³ switch to beef upon hearing that the restaurant is out of salmon. The monotonicity condition for individual choice simply says that less is never preferred to more. Barring special cases,³⁴ this condition also seems innocent and compelling.

When we go from one-person choice to two-person bargaining, however, the conditions are far from compelling. This is partially reflected in the fact that they cannot be satisfied simultaneously, given the other Nash axioms, but since those axioms themselves are far from unquestionable we should not place too much weight on the inconsistency. One should just be wary in general of thinking that the constraints on the outcome of individual choice carry over to the outcome of bargaining.³⁵ The conditions must be justified directly as conditions on bargaining, not by analogy from individual choice.³⁶

Both the independence condition and the monotonicity conditions are stated in terms of a comparison between two bargaining situations. There is no reason for these to involve the same persons, or the same bargaining objects. All that matters is that the feasible utility sets and the disagreement points are related in certain ways. Yet the conditions should also apply to the special case when the same individuals are bargaining under different circumstances. When interpreted in this way, Nash's independence condition appears implausible. Consider, for instance, wage bargaining under full employment, before and after protectionist measures have been passed for the industry in question. Michael Wallerstein has shown³⁷ that (a close relative of) the Kalai-Smorodinsky solution ensures that both workers and management benefit when protectionism is introduced in a situation of full employment. The workers get higher wages, the management higher prof-

³³ He might rationally do so, however, if he believes that restaurants which make good chicken always make a point of being well stocked with salmon. Levi (1986) shows that similar behaviour might also occur as a by-product of a rational way of coping with value conflicts. Less rational ways in which the feasible set can affect preferences are discussed in Elser (1983a), ch. 3.

³⁴ A restaurant customer might prefer less to more if he believes that the quality of each item is inversely proportional to the number of items from which he can choose. A rational individual who knows his own propensity to overeat might want to have less food in the house rather than more. Elser (1984), ch. 2, offers a survey of such cases.

³⁵ Crawford (1984), p. 378.

³⁶ An attempt to provide strategic foundations for the independence condition is that of Binmore (1987b). The argument is less compelling, however, than the corresponding argument for the irrelevance of outside options, further discussed later.

³⁷ Wallerstein (1987).

is. Under the Nash solution, however, the workers do not benefit at all from the larger earning power of their firm.³⁸ By contrast, in a situation of unemployment workers do benefit from protectionism under both solution concepts. Hence Wallerstein concludes, 'Unions, according to the Nash solution, are foul-weather allies in protectionist coalitions'. While the Nash solution, like any bargaining solution, specifies that both parties benefit from cooperating, it does not always imply that both parties gain from an increase in the gains from cooperation. Intuitively, this makes no sense. Surely, the workers would insist on a share in the increased earnings made possible by protectionism.³⁹ If the Nash solution predicts otherwise, this goes only to show how implausible it is.

This observation can be generalized. All varieties of formal bargaining theory, when applied to wage negotiations, imply that the primary thing to be explained is the wage level. Wage increases, by contrast, are secondary — to be derived by subtracting one wage level from another. There is no analytical difference between wage differences and wage increases. It makes no difference whether (S, d) and (S', d') are two bargaining games that take place on two different planets or two games that take place between the same management and the same union in two successive years. In actual wage bargaining, of course, these two cases differ vastly, since the workers remember what they got last time. The baseline for bargaining is given not only by what they would get in the absence of any cooperation, but also by what they got in the previous round of negotiations. If (S, d) and (S', d') took place on two different planets, the outcome might conceivably be the same even though the first occurred in a nonprotectionist and the second in a protectionist context. It is *not* conceivable, however, that a union would accept an unchanged wage when the firm suddenly benefited from protectionist measures. One cannot assume that wage increases can simply be derived from wage levels. Sometimes a separate analytical apparatus may be needed to explain wage increases. I return to these matters in Chapter 6.

Noncooperative bargaining theory

Usually, we think of competition and bargaining as intensely interactive and conflictual processes, involving winners and losers, failures as well as

successes. Mainstream economic theory has taken the bite out of both, representing them by the equilibrium features of their outcomes rather than by their internal dynamics. Reactions to the aseptic and sanitized notion of competition have come mainly from the Austrian school of economics, including contemporary followers of Joseph Schumpeter.⁴⁰ By and large, the mainstream remains unaffected by the criticism. Although the objections are often telling, their target is well and alive, protected by the fact that you cannot beat something with nothing. Where the alternative writers pride themselves on the realism of their models, the mainstream economist sees only ad hoc assumptions in stark contrast to the simplicity and power of equilibrium theory.⁴¹

Attempts to desanitize bargaining theory have proved more successful, but only after several false starts. The features of bargaining that must be incorporated into a dynamic model with rational players include the following. First, bargaining is a process that can be broken down into successive offers and counteroffers. A bargaining solution must be defined as the outcome of a process of bargaining or, more subtly, as driven by the anticipation of this process, which need not actually take place. Second, bargaining is *costly*. For one thing, players who care more about the present than the future always suffer from a delay. A fifty-fifty split of a dollar between you and me tomorrow is worth less to each of us than the same division today. 'If it did not matter when people agreed, it would not matter whether or not they agreed at all'.⁴² Other costs of bargaining are discussed later. Third, threats made in the course of bargaining must be *credible*. A rational bargainer with no access to precommitment devices will not be taken seriously if he makes a threat which it will not be in his interest to carry out when the time to do so arrives. A father might say, for instance, that if the mother gets custody he will not exercise his visitation rights, thus harming the child and, through the child, the mother. But if the mother knows that the father is too rational to cut off his nose to spite his face, she will not take the threat seriously. Management cannot credibly threaten with lock-out if the workers know that the ensuing loss of customers would cripple the firm.

An early attempt to provide noncooperative foundations for cooperative bargaining theory was made by John Harsanyi.⁴³ He showed in 1956 that a model of stepwise bargaining, governed by a concession rule proposed

³⁸ The result presupposes a production function of the Cobb-Douglas kind. With other

functions, the counterintuitive conclusion does not follow.

³⁹ I am indebted to Fredrik Engelstad for forcing this point on me.

⁴⁰ Nelson and Winter (1982) offer the most fully developed version of this view.

⁴¹ For elaborations of this argument, see Elster (1983c, 1986).

⁴² Cross (1965), p. 196. ⁴³ Harsanyi (1955), more fully set out in Harsanyi (1977a).

by Fredric Zeuthen some twenty-five years earlier, converges to Nash's solution of bargaining as a cooperative game. One problem with Harsanyi's argument is that there is far from full agreement that the concession rule is rational. A deeper problem stems from the fact that although bargaining in his model takes place in real time, so that the sequence of proposals and counterproposals actually has to be gone through, it is assumed to be costless. He neglects the fact that in a temporally extended sequence of proposals and counterproposals the parties are in effect bargaining over a shrinking pie. An agreement may finally be reached, but in the meantime much of the gain from cooperation has been squandered.

Harsanyi's model satisfies only the first requirement, that the model of bargaining be process-oriented. Other models, which also satisfy the second requirement, rest on artificial assumptions about motivation and expectations.⁴⁴ Ariel Rubinstein's path-breaking 1982 article provided the first model in which all three requirements are satisfied.⁴⁵ To explain the workings of the model, I shall proceed in three steps. First, I shall explain the idea of a *perfect equilibrium*, which is central to these noncooperative models of bargaining. Next, I shall illustrate the idea with respect to a particular bargaining problem. Finally, I shall use an ingenious technique invented by A. Shaked and J. Sutton to derive the solution to a simple, although representative, bargaining game.

The traditional equilibrium concept in noncooperative game theory is, like the best-known solution concept in cooperative bargaining theory, associated with John Nash.⁴⁶ As explained in the Introduction, an equilibrium (or Nash equilibrium, as it is usually called) is a set of strategies that are best replies to each other. In equilibrium, nobody can improve his outcome by unilateral deviation. In games with several equilibria, game theory often has no way of determining which will in fact be chosen.⁴⁷ In such cases, it was usually assumed that one equilibrium is as likely to be realized as any other, until Reinhard Selten demonstrated that only perfect

⁴⁴ Cross (1965); Coddington (1968).

⁴⁵ Rubinstein (1982). An early forerunner of Rubinstein's model is found in Stahl (1972).

⁴⁶ Nash (1951). The relation between the two models is explained in Stahl (1988).

⁴⁷ Nash (1951).

⁴⁸ Harsanyi and Selten (1988), who offer a 'general theory for equilibrium selection', point out (p. 366) that 'Rubinstein's approach provides an interesting alternative in many cases to our own theory for selecting a unique solution to sequential games. But in its present form it seems that it cannot be extended to games involving simultaneous moves by the players'. My concern here is with sequential games, for which Rubinstein provides a simple and tractable analysis.

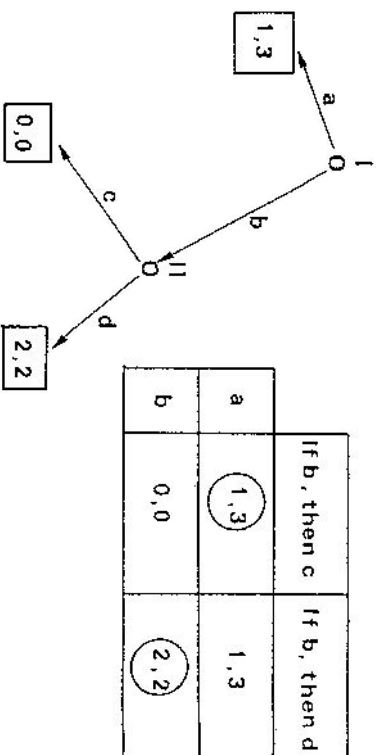


Fig. 2.5

equilibria – to be defined in a moment – will in fact be chosen.⁴⁸ Nash equilibria that are not perfect, such as equilibria that rest on noncredible threats, will never be realized.

Figure 2.5 offers two representations of the same game.⁴⁹ The left-hand representation shows the game in *extensive form*, as a sequence of moves and countermoves. Player I moves first. If he chooses *a*, the game is over. If he chooses *b*, player II chooses between *c* and *d*. Numbers at the end nodes represent payoffs to the players, the first number being the payoff to I. The right-hand representation has reduced the game to the *normal form*, which states the relations between strategies and outcomes in a compact way. In fact, the normal-form representation is too compact, since vital information is lost. In the normal form there are two equilibria, (1, 3) and (2, 2). For all we know, either might be realized. The extensive form makes it clear, however, that the outcome (1, 3) will never be reached, unless II can precommit himself to use *c* in case I plays *b*. The threat to use *c* is not credible, since it will not be in II's interest to execute the threat if the second node should be reached. If we assume that precommitment is unfeasible, I will play *b*.⁵⁰

I shall now extend this reasoning to sequential bargaining games, in which the parties take turns making proposals and counterproposals. The game comes to an end when one party makes an offer that is accepted by the other. For simplicity, assume that the parties are bargaining over the

⁴⁸ Selten (1975). ⁴⁹ The example is taken from Harsanyi (1977b).

⁵⁰ In the next chapter and in the concluding chapter I discuss how *social norms* could lend credibility to I's threat.

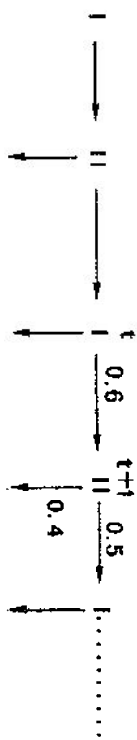


Fig. 2.6

division of a dollar. A *strategy* in sequential bargaining is a response function that for any sequence of offers and counteroffers up to a given point defines a unique behaviour: either acceptance of the previous offer or, if that is rejected, a new offer. An *insistent offer* is a response function that always makes the same demand and accepts only if the same proposal is made by the other party. Clearly, any Pareto-optimal pair of insistent offers is an equilibrium. If one party insists on getting $1 - x$ and the other insists on x , we are in equilibrium, for any x between 0 and 1. But are all these equilibria perfect? Is the threat to hold out credible?

Consider the game depicted in Fig. 2.6.⁵¹ Players I and II are to divide a dollar. We assume that the parties incur costs in each bargaining period — for instance, because of the need to pay lawyers. (Time discounting is *not* assumed here.) The utility to I of getting s at t is $s - 0.1t$. The utility to II of getting s at t is $s - 0.2t$. Clearly, II is at a disadvantage, since his lawyer is twice as expensive as I's.⁵² Suppose that II has decided to hold out for 0.5 and that at time t I makes an offer of $(0.6, 0.4)$. Can II credibly hold out? If II accepts, she gets $0.4 - 0.2t$. If she holds out for 0.5, she can *at most* get $0.5 - (t + 1)0.2 = 0.3 - 0.2t$. But that is less than she could get by accepting the offer; hence the threat to hold out is not credible. Note that the credibility of threats is intimately linked to the costs of bargaining.

Consider next a worker and a firm bargaining over a dollar, with the cost of bargaining represented by the fact that future payoffs are discounted to present value by a factor d , the same for both parties. This is the only cost of bargaining. Offers are made and accepted or rejected in the same time period. But a new offer has to be made in a new period. In other words, to refuse an offer always involves a costly delay, which may or may not be offset by the prospect of getting a better deal. The firm makes the first offer. Then consider the subsequence shown in Fig. 2.7.

⁵¹ From Rubinstein (1982).

⁵² This difference might reflect superior bargaining abilities of II's lawyer. But there is no way in which this element can be incorporated into this model.



Fig. 2.7

Considered from $t = 0$, the discounted value of what the firm and the worker will have to divide at $t = 2$ is d^2 . Let us look (still from the point of $t = 0$) at the subgame that begins at $t = 2$. Suppose that M cents is the maximum the firm can get in any perfect equilibrium of this subgame. We do not know what this maximum is, except that it is determined by the feasible set and by the *bargaining power* (explained later) of the parties. In this case, the only source of asymmetrical bargaining power derives from the fact that the firm moves first. Discounted to $t = 0$, the value of this maximum is $d^2 \cdot M$. Consider now, still from the point of view of $t = 0$, what the worker should do at $t = 1$. He does not have to offer the firm more than $d^2 \cdot M$, because the firm cannot credibly hold out for more. If the firm rejects the offer, it has to go into a new period, in which it can get at most $d^2 \cdot M$. At $t = 1$, the total value of the dividend is d . Thus the worker gets at least $d - d^2 \cdot M$. Consider now the offer made by the firm in the first period. Any offer it makes has to leave the worker with at least $d - d^2 \cdot M$. Hence the maximum of what the firm can get is $1 - (d - d^2 \cdot M)$. *But the game at $t = 0$ is identical to the game at $t = 2$* . In both cases, the two players look down the same infinite path of offers and counteroffers. Hence the maximal amounts the firm can get in these two games must be the same.

From $M = 1 - (d - d^2 \cdot M)$ we derive $M = 1/(1 + d)$. This also turns out to be the minimum of what the firm can get, since the game argument can be repeated, minima and maxima being interchanged throughout. Since M is both the maximum and the minimum of what the firm will get, it defines the outcome of the bargaining game. The worker's share is $d/(1 + d)$. The agreement will be reached in the first bargaining round, since neither party can gain from holding out. Assume that $d = 0.9$, so that both are quite patient. Then the firm gets 0.53 and the worker gets 0.47. Assume that $d = 0.5$, so that both are quite impatient. Then the firm gets two-thirds and the firm gets one-third. In other words, the heavier the parties discount the future, the larger the advantage of being the first player. This argument relies heavily on backward induction and hence is vul-

nerable to the objections stated in the Introduction. On the one hand, the argument is supposed to show that the firm and the worker will reach agreement instantaneously, agreeing to share the dollar as just stated. On the other hand, what forces agreement is the prospect of offers and counteroffers in later bargaining rounds. *But the assumption that there will be later bargaining rounds is inconsistent with the conclusion derived from that assumption.* When the firm contemplates the idea of being at $t = 2$, it should know that this can occur only if something goes wrong, since if both are rational they will never get that far down the path. Perhaps the reason they are at $t = 2$ is that the worker is irrational. But in that case, he might be so stubborn as to hold out for something he should know he cannot get. Faced with a potentially irrational opponent, the firm might have to concede more than it would otherwise have done. The worker, of course, has to go through similar reasoning. The outcome is essentially indeterminate. As I said in the Introduction, I suspect that the last word on the matter has not been said.

By now, a variety of noncooperative bargaining models have been proposed, differing mainly in the determinants of the costs of bargaining.⁵³ First, there are time-preference models similar to the one just discussed, but allowing for the possibility that the parties may have different rates of time discounting. Second, there are fixed-cost models of the kind mentioned earlier, with, for instance, the need to pay lawyers being the main cost of bargaining. Third, there are models that stipulate an exogenously given probability that bargaining might break down — for instance, because the opportunity for a joint venture ceases to be present. A firm and an inventor bargaining over a contract run the risk, for instance, that if negotiations drag out another firm might preempt the idea. Some models relax the assumption of alternate offers and counteroffers by stipulating that after the proposal and rejection of an offer there is a positive probability that the same party will make the next offer.⁵⁴

Bargaining power

Both cooperative and noncooperative models of bargaining try to capture the notion of bargaining power. Focusing on the simple case of dividing a sum of money, deviation from equal division can be explained only by

unequal bargaining power. Here I consider mainly bargaining power derived from the material preferences and resources of the parties as well as from the temporal structure of the bargaining process. In Chapter 6 I also discuss *social norms* as a determinant of the outcome of bargaining.

In some models, bargaining power is simply taken as a parameter to be estimated. The *generalized Nash solution* is often used for this purpose. Whereas the simple Nash solution states that the outcome of bargaining will be the utility pair (u_1, u_2) which maximizes the product $u_1 \cdot u_2$, the generalized solution defines it as the pair which maximizes the product $u_1^\alpha \cdot u_2^{1-\alpha}$, with $0 < \alpha < 1$. [I assume that the disagreement point is $(0, 0)$.] Here α is a parameter that is supposed to capture all determinants of relative bargaining power: bargaining ability, resources to hold out during a conflict, support in public opinion or anything else that might be relevant, including normative considerations. Econometric work can then be carried out to estimate the parameter. In addition, by stipulating that bargaining power is a linear function of variables like unemployment and cost of living, one can estimate the importance of each determinant of bargaining power in shaping the outcome.⁵⁵ I will not comment on this approach, except to say that it is vitiated by its lack of microfoundations and the mechanical character of the assumptions.

There have been attempts, however, to provide noncooperative foundations for the generalized Nash solution.⁵⁶ In models with an exogenously given probability of break-down, the outcome favours the party whose estimate of this event is lower. In models with different rates of time preference, the outcome favours the less impatient party who can say, credibly, that he does not mind waiting. In both models, it can be shown that as the length of the bargaining intervals goes to zero, the outcome of the bargaining game converges to a generalized Nash solution. In the first model, the bargaining parameter is a function of the rates of time preference and in the second a function of the subjective probabilities assigned to a breakdown. With positive intervals between the offers, the party who makes the first proposal has an advantage, but in the limit it does not matter who moves first. In the special case where the parties have the same time preferences or the same beliefs, the noncooperative outcome converges to the simple Nash solution.

⁵³ For surveys, see Sutton (1986) and Birmore, Rubinstein and Wolinsky (1986).

⁵⁴ See, e.g., Moene (1988b).

⁵⁵ Sveinjar (1986) is an example of this procedure.

⁵⁶ The following draws on Birmore, Rubinstein and Wolinsky (1986).

This demonstration does not automatically provide microfoundations for the Nash solution as traditionally conceived. For one thing, when the bargaining interval is incompressible — perhaps because the union leader has to go back to his constituency to discuss the offer — the Nash solution will not obtain. More important, the disagreement point in the noncooperative version of the Nash solution differs from the traditional conception of the status quo in bargaining. Because the point is fundamental, it should be discussed at some length.

Using wage bargaining as an example, we can ask what happens if management and the union fail to reach agreement. There are two ways of looking at the matter. One is to assume that the parties fall back on their *outside options*, that is, on the state which would obtain if the joint venture were definitely dismantled. The workers may find a job in another firm or live on unemployment benefits. Assuming that managers act only as representatives of the owners, they may sell off the physical assets of the firm and redeploy the capital elsewhere. In child custody bargaining, the outside option of the parties is represented by the expected legal decision. An alternative is to assume that the joint venture is only temporarily dismantled until agreement is reached. In that case the parties must do with their *inside options*, that is, what they can get during the conflict. The workers might have to rely on their strike funds. The firm might get support from the employers' association. One of the parents usually has temporary custody until they agree on a final settlement or refer the matter to the court.

It seems clear that both inside and outside options are relevant to the outcome of bargaining.⁵⁷ Noncooperative theorists argue, however, that they matter in fundamentally different ways.⁵⁸ Outside options *constrain* the outcome but do not influence it in any other way. Inside options affect

⁵⁷ A good example is provided by inside subcontracting in Hungarian enterprises (Szirmai 1989). In this system, skilled workers in the firm are allowed to set up work partnerships that 'have authority and legal status as semi-autonomous economic units to enter into contracts [usually with the mother enterprise] to produce goods and services during free hours, using factory equipment'. The outside options in firm-partnership bargaining are, for the firm, the price of outside labour (often Polish workers or workers hired from cooperatives) and, for the workers, normal overtime pay. Inside options also affect the outcome, since workers frequently use their tacit knowledge of the economic difficulties of firms to bargain for better rates for their partnerships'. Sabel and Stark (1982), p. 458, make the opposite argument: 'The existence of such a secondary economy clearly augments the bargaining power of workers in the primary plants'. To the extent that these secondary jobs are in inside subcontracting, the argument seems incorrect. Since only workers who already hold a regular job in the firm are allowed to enter work partnerships, they cannot credibly threaten to leave the firm and work full time in the second economy.

⁵⁸ See especially Sutton, Shaked and Binmore (1986).

the outcome via the bargaining power the parties can derive from them. Specifically, *inside options determine whether threats are credible*. Hence the noncooperative version of the generalized Nash solution says that the agreement will maximize the weighted product of the utility gains of the parties compared with their inside options. In cooperative models, by contrast, it has been tacitly assumed, albeit with some confusion,⁵⁹ that what is maximized is the weighted product of the utility gains compared with the outside options.

Here is a numerical illustration. Suppose that workers in the firm are currently earning \$8 per hour. They know that they could get a job in another firm at \$6. Their strike fund will ensure them an income equivalent to \$4 per hour. Under these conditions, assume that the outcome of bargaining is \$9. Assume now that the reservation wage increases from \$6 to \$7, while everything else remains constant. The noncooperative theory then predicts that the change will not affect the bargaining outcome. If, however, the strike fund swells to ensure the workers \$5 per hour, the workers might well be able to get \$10 instead of \$9. Outside options serve as floors on what the workers will get but have no role beyond that. The workers can credibly threaten to leave the firm if they are offered less than what they could get elsewhere, but they cannot credibly threaten to leave the firm if the alternative wage is below the management's offer. And it makes no difference to the credibility whether the alternative wage is well below that offer or only a little below. By contrast, the credibility of a strike threat is affected by *any* change in the value of the inside option.

The point can be brought home by a comparison of two varieties of noncooperative models.⁶⁰ In both, there are costs of bargaining from discounting. One case is defined by the feature that if a random event occurs (with known probability), the party whose turn it is to make an offer can either decide to quit the game, in which case both players receive fixed payoffs, or decide to stay in with a new proposal. The other case is defined by the feature that if the random event occurs, the game is over and the players receive their fixed payoffs. In the former case, the fixed payoffs

⁵⁹ McDonald and Solow (1981) seem to confuse inside and outside options. Using the cooperative framework, they write that the disagreement outcome for the workers is an outside option, determined by such elements as unemployment benefits, the value of leisure, the value of working around the house, net gains from illegal activities and the expected value of alternative employment opportunities (p. 899). For the firm, the disagreement is an inside option: zero profits or even negative profits if there are fixed costs that have to be paid (p. 905). The same confusion is found in Svejnar (1986), p. 1057.

⁶⁰ Sutton (1986).

serve as outside options. They constrain the outcome of the game but do not otherwise affect it. In the second case, however, it can be shown that the fixed payoffs do affect the outcome over and above the constraining effect. The intuition behind this result is that 'small options, if chosen voluntarily, have no effect; the "exogenous intervention" mechanism serves to make (even small) threats credible'.⁶¹

Let me summarize. When agreement is forced by the risk of bargaining breaking down, the classical Nash model gives the right result. The outcome maximizes the product of the gains relative to 'fall-back' outside options, but *only because these options cannot be freely chosen*. The outside options shape the outcome because they are, as it were, part of the inside options. By contrast, when agreement is forced by time discounting, outside options have no effect beyond that of constraining the outcome. Within the constraints, the solution is determined by the inside options.

The argument about 'the irrelevance of outside options' is related in spirit to the condition that the outcome be independent of irrelevant alternatives. Changes in options that would not be realized anyway should not affect the outcome. Whether these options are inside the bargaining range or fall-back options in case bargaining fails, they can have no effect on the outcome beyond constraining it. The alternative view — that changes in options far from the outcome might nevertheless affect it — would involve a social analogue to action at a distance. Within the paradigm of 'norm-free' bargaining, these arguments are compelling. Rational players would not take account of changes that have no impact on the bargaining power of the parties.

Behaviourally, however, there is no doubt that irrelevant alternatives and outside options do make a difference. Figure 2.8 represents a symmetrical bargaining game *OCD* with outside options at the origin and the solution at *B*. Consider now a truncation of the feasible set that excludes all alternatives above *AB*. If we believe in the independence of irrelevant alternatives, the solution in the modified game *OABC* should remain at *B*. Common sense suggests that it will not: player I will not accept that II will get his best possible payoff whereas I will have to be content with something well below his maximum. A trade union leader, for instance, could never make his constituents accept an outcome whereby he had to make all the concessions and management none. The Kalai-Smorodinsky solu-

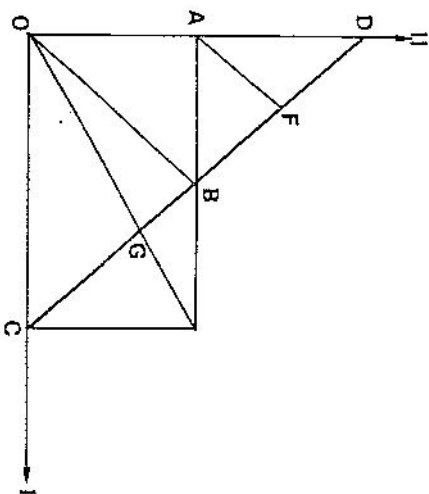


Fig. 2.8

tion, which suggests that the outcome in the modified game will be at *G*, is more plausible.

Consider now a different truncation of the game, from *OCD* to *ABD*, resulting from a change in the outside option of II.⁶² Within the noncooperative framework, this should not make any difference: the solution should remain at *B*. Once again, however, this is behaviourally implausible. We may imagine that I and II are bargaining over the division of a sum of money, for example, \$200. In the original game, they get nothing if they fail to agree. In the modified game, II is sure to get \$100 if they fail to reach agreement. This game can be plausibly described as if I and II were bargaining over \$100, in which case we would expect the solution to yield an equal division at *F* rather than remain at *B*. The noncooperative theory of rational bargaining tells us that all three games should have the same solution at *B*. Common sense tells us that they will have three different solutions. The discrepancy between theory and common sense may be due to a faulty conception of rationality, as suggested by the comments on backward induction. Or, assuming that conception to be correct, common sense may induce deviations from rational behaviour. The latter position is unstable, however, since ultimately the raw material for any theory of rational behaviour is our intuitions about what it makes sense to do in particular cases.

These deviations from what the theory tells rational players to do are

⁶¹ Ibid., p. 715.

⁶² I am indebted to Michael Wallerstein for this example.

frequently found in real-life bargaining. I have already argued that 'relevant' alternatives are relevant to capital-labour bargaining. Similarly, outside options affected the outcome, over and above their constraining effect. Workers look at the wage rate in other firms, to preserve existing wage differentials. The norms of fairness that govern capital-labour and labour-labour relations form the topic of Chapter 6.

The main determinants of bargaining power are time preferences, risk aversion and inside options. The more impatient, the more risk averse and the lower one's disagreement utility, the weaker is one's bargaining position. These subjective elements are often highly correlated with one another and with the objective wealth of the parties. Under standard conditions the less wealthy are more impatient, are more risk averse, and have a lower level of disagreement utility.

The relation between the objective and the subjective elements in bargaining is complicated. In principle, only subjective elements matter, yet in practice only the objective ones are observable. If psychology were essentially similar across people, all subjective differences would be caused by objective differences. Just as one person has different time preferences and risk attitudes at different levels of wealth, subjective differences across persons can be induced by objective differences. But subjective differences cannot be fully reduced to objective ones. The parties' external circumstances may be identical and yet their personalities or temperaments may differ.⁶³

Some people are content because they have much, others because they have learned to be content with little. In either case, their bargaining power is enhanced: by caring less about what they get, they get more.⁶⁴ Attitudes towards risk can similarly arise in two ways. First, there is the wealth-induced effect: a rich man will be more willing to take risks than a poor man. Second, people differ intrinsically in their subjective attitudes towards

⁶³ For a striking example, consider a play once planned by Jean-Paul Sartre. 'Colette Audri, with whom Sartre once discussed this play, tells us that the play was to be called *The Wager* (after Pascal's wager), and would concern a child who is not wanted by his father. The mother, however, does not let herself be pressed into abortion, although a horrible life has been prophesied for the child: severe trials and reverses, poverty, and finally death at the stake. The child is born, grows up, and everything takes place as prophesied. "In fact he changes nothing material in his existence", Sartre says, "and his life ends, as foretold, at the stake. But thanks to his personal contribution, his choice and his understanding of freedom, he transforms this horrible life into a magnificent life" (Føllesdal 1981, pp. 403-4).

⁶⁴ Could this provide an incentive for strategic character planning? Gandhian techniques of nonviolence suggest that the question is not wholly absurd.

risk taking, independently of their external circumstances. Some people are naturally cautious, others are daring or even reckless. In von Neumann-Morgenstern utility functions these two mechanisms are inextricably intertwined, but in principle they can and should be distinguished.⁶⁵ Whatever the source of risk aversion, it usually is a handicap in bargaining.⁶⁶ Time preferences, finally, also arise in two ways. First, again, there is a wealth-induced effect: a rich man can better afford to wait. Second, people may differ intrinsically in their subjective rates of time preference, independently of their external circumstances.

The upshot of these remarks is the following. Bargainers can enhance their bargaining power to the extent that they can credibly communicate low utility gains, low risk aversion and low time discounting. To the extent that these follow naturally from their external situation, no special evidence is needed for these claims to be credible. To the extent that they go against what one might expect from the external situation, special evidence is necessary. The evidence must not be related to the bargaining situation, because if it is, the adversary will assume that it has been produced for the special purpose of gaining an advantage in bargaining. A union that incurs the costs of a strike in order to prove that it is less impatient and risk averse than would otherwise be assumed might end up having the worst of both worlds. It loses in the current round without gaining the credibility that would get it more the next time around.

Temporal asymmetries in the bargaining process can also influence the bargaining power of the parties. I have already mentioned that the party who moves first has an advantage. In addition, a party that needs more time to respond to a proposal by the other party has an edge in the bargaining process. The more the bargaining pie shrinks during the period when the union considers an offer by the employers, the more the latter have to lose by not giving in to the union's claim and hence the more likely they are to do so.⁶⁷ The internal weakness and lack of integration of many

⁶⁵ For attempts to construct cardinal utility functions without any element of intrinsic attitudes to risk see Shubik (1982), pp. 421-4 (reporting a result of L. Shapley) and Sen (1977), p. 339. Unfortunately, both proposals rest on shaky psychological foundations, in that they require subjects to make comparisons whose subjective meaningfulness is highly doubtful.

⁶⁶ In special cases, risk aversion may enhance one's bargaining power. Thus for bargaining games in which potential agreements involve lotteries which have a positive probability of leaving one of the players worse off than if a disagreement had occurred, the more risk averse a player, the better the terms of the agreement which he must be offered in order to induce him to reach an agreement and to compensate him for the risk involved (Roth and Rothbaum 1982).

⁶⁷ Barth (1988). See also de Geer (1986), p. 353.

unions may, paradoxically, enhance their bargaining power because they can claim, *credibly*, that it will take them a long time to respond to the management's offer. Conversely, the management of a subsidiary of a multinational firm may gain bargaining power by pointing out that any counterproposal by the union will have to be sent back to headquarters.

Uncertainty, manipulation, inefficiency.

The noncooperative models of bargaining discussed earlier capture part of what goes on in real-life bargaining, but far from all. While emphasizing the role of threats, they ignore the haggling, bluffing, posturing and jockeying for position which are part and parcel of any negotiation in the real world. No actual haggling takes place in these models, only virtual haggling, as a result of which agreement is reached in the first moves of the game.⁶⁸ There are no elements of uncertainty, nor any possibility for strategic prebargaining moves. The remainder of this chapter is devoted to these issues.

I shall discuss three closely related questions. First, what is the role of information and uncertainty in bargaining? Second, what is the scope for strategic manipulation in bargaining? Third, how many of the benefits from cooperation are realized in actual bargaining? The first and the second questions are related because uncertainty creates an incentive for strategic misrepresentation of preferences and other factual matters (such as wage statistics). The first and the third are related because uncertainty, without misrepresentation, may lead to the break-down of bargaining. The second and third are related because attempts to influence the feasible set or the disagreement point can lead to waste of resources, through failure to reach agreement or for other reasons. The central argument, linking all three questions, is easily summarized: *to increase their distributive shares, bargainers engage in tactics that either decrease the probability of reaching agreement or decrease the size of the total to be shared.* In both cases, social losses result.

Uncertainty and the role of information

Uncertainty is massively important in bargaining. Because the buyer and seller of a house do not know each other's reservation prices, they often

go through an intricate dance of soliciting and misrepresenting information.⁶⁹ Also, the buyer knows that the seller knows more about the house than he can find out from brief or even from extensive inspection. If he buys in the summer, he has largely to take the seller's word for what it will be like in the winter. The example suggests that the distribution of information is doubly asymmetrical. First, on each side there are subjective items that only that side knows, such as that party's preferences and information. Second, there can be information about objective items which in principle is available to both parties but in practice to only one of them.

Uncertainty about preferences has several aspects. In all bargaining, risk attitudes and time preferences are central. In bargaining over multidimensional packages, the parties' subjective trade-offs among the components can be very important. Interpersonal comparisons of utility can have a massive impact. Bargainers who meet each other over and over again usually end up learning a great deal about each other's preferences, but one-off bargains are obviously very different. Even bargainers who know each other intimately may nevertheless be able to exploit third-party uncertainty, if unresolved conflicts are to be arbitrated.⁷⁰ In court a husband may be able to present himself as a caring and competent father, although it is common knowledge between his wife and himself that he would neglect the child were he to get custody. In private ordering between the parties, the wife may therefore have to accept a bad financial settlement in order to get custody.⁷¹

Uncertainty about objective items, although a less fundamental problem, is also widespread. The seller of an oriental carpet can be assumed to know more about its quality than do most prospective buyers. Under conditions of asymmetrical information buyers may be subject to the 'winner's curse': if their offer is accepted, they will suspect that they could have gotten a better deal. There may not even be *any* offer such that, if it is accepted, it should have been made.⁷² There may be *no* club willing to accept them as members which they would want to join. To see this, consider two parties bargaining over a piece of land on which there may or may not be oil. Both parties know that the owner of the land knows the exact value to him of

⁶⁹ Raiffa (1982), ch. 3, uses bargaining over a house as the introductory example in his splendid account. See also Scheppele (1988) for the question of legal redress for misrepresentation.

⁷⁰ For this distinction between 'common knowledge' and 'public knowledge' see the editorial introduction to Binmore and Dasgupta, eds. (1987), p. 19.

⁷¹ I am assuming that it is also common knowledge that the father desires custody, since otherwise his threat to go to court would not be credible.

⁷² Samuelson (1985); Samuelson and Bazerman (1985); Thaler (1988).

⁶⁸ Binmore (1987a), p. 179, has a quote from Hobbes that is also appropriate here: 'For the Schooles find . . . no actual Motion at all; but because some Motion they must acknowledge, they call it Metaphorical Motion, which is but an absurd speech'.

the oil, but that the prospective buyer knows only that the value to the seller is somewhere between 0 and \$100 million, with any value in this range being equally likely. They also know that whatever the value is to the seller, it is half again as large for the prospective buyer, perhaps because he owns adjacent land that will make it easier to refine the oil. Assume now that the buyer makes an offer of \$X, which the seller accepts. From this the buyer can infer that the land is worth at most \$X to the buyer. More precisely, he knows that the value to the buyer is somewhere between 0 and X, with any value in this range being equally likely. From the buyer's point of view, the expected value of the land to the seller is, therefore, $X/2$, and its expected value to him is half again as large, that is, $3X/4$. But this is less than what he offered to pay. Realizing that *any* offer he could make that would be accepted by the buyer would be likely to buy him a piece of land worth less to him than he paid for it, he will decide not to make any offer at all. Although there is room for a mutually beneficial deal, none will be struck.

Management usually knows more than unions about the firm's ability to pay. The firm's duty to disclose this information is often severely limited. In Britain, the Employment Protection Act of 1975 obliges the employer to disclose to trade unions 'information without which the trade union representatives would to a material extent be impeded in carrying out . . . collective bargaining', subject, however, to numerous qualifications and exceptions.⁷³ The employer is not obliged to disclose information unless the amount of work and expenditure it would require is 'proportionate to the value of the information'. Since 'it is extremely difficult to define *ex ante* what the value of the information in collective bargaining will be',⁷⁴ employers can do more or less as they want to. In deciding how much to disclose, they will be guided both by efficiency and by distributive shares, knowing that disclosure may facilitate agreement, but also skew agreement in the union's favour.⁷⁵

There are theories of bargaining under incomplete information,⁷⁶ but I do not think they are very useful. They rest on the assumption that although the bargainers are not certain about each other's preferences or about the

quality of the object of bargaining, they have well-defined subjective probability distributions over these variables. One model, for instance, assumes that the values of the object to the parties 'are independent random variables and that each is uniformly distributed over the interval from 0 to 1 (in some monetary scale)'.⁷⁷ The obvious question is: why the uniform distribution? Perhaps the assumption is supposed to be justified by the principle of insufficient reason. That principle, however, is highly dubious⁷⁸ and in any case rarely appropriate in bargaining situations. A bargainer usually has enough information to entertain a subjective distribution of the *ordinal* probability of the values. He may, for instance, have a notion of the most probable value of the object to the other party, but no idea of how fast the probability of other values tapers off as we go to the extremes. In that case, should he assume that the distribution is normal? Lognormal? Or simply that it is a member of a family of distributions with known properties?⁷⁹ In my opinion, Bayesian theories of bargaining suffer from a fundamental lack of realism, as do the closely related theories of incentive compatibility. Their results are achieved at the cost of assumptions that are not merely heroic, but close to the supernatural.⁸⁰

This being said, I have no alternative theory to offer about the behaviour of rational bargainers in situations of uncertainty. Most likely, no such theory will ever be forthcoming. I do not mind: rational-choice theory cannot explain everything. In fact, the first task of rational-choice theory must be to circumscribe its own limits.⁸¹ This is not to say that the outcome of bargaining is indeterminate, only that a particular theory of bargaining fails to yield determinate results.⁸² To achieve or approach predictive determinacy, we then have to consider other theories. Herbert Simon's theory of bounded rationality suggests that bargainers set themselves a target and give in as soon as it is reached. Thomas Schelling's theory of focal points suggests that psychological salience and prominence are important. The theory of social norms — set out in Chapter 3 and applied to bargaining in Chapter 6 — suggests that notions of fairness matter. All of these theories

⁷³ Myerson (1985), p. 116. ⁷⁴ See, e.g., Luce and Raiffa (1957), pp. 284–5.

⁷⁵ See Hey (1981) for a devastating critique of optimal-search rules that are similarly based on the assumption that people have well-defined subjective probability distributions.

⁷⁶ Elster (1989a), ch. 2, argues, in fact, that Bayesian decision theory is to modern decision making as astrology was to decision making in earlier times.

⁷⁷ This is the central argument in Elster (1989a).

⁷⁸ See Pen (1959), p. 91, for comments on the 'pathetic fallacy' of projecting our ignorance of the outcome of bilateral monopoly onto the situation itself.

⁷³ Here I draw upon Foley and Mauders (1977). ⁷⁴ *Ibid.*, p. 18. ⁷⁵ *Ibid.*, p. 106.

⁷⁶ See, e.g., Myerson (1984, 1985) and Rubinstein (1985a). I do not claim to have fully mastered these highly technical papers. Their results, however, cannot be more robust than their premises, and it is the latter which I criticize in the text. I am not implying that these authors are unaware of the frailty of their premises.

are closer, perhaps, to description than to explanation. Again, this is fine in my book, since I believe that at the present time the social sciences cannot aspire to be much more than a phenomenological study of mechanisms.⁸³

Strategic manipulation of bargaining parameters

The standard bargaining models ignore the possibility of strategic misrepresentation. To see how this problem arises, we may note that a bargaining process can be broken down into the following parts. (a) There is an underlying physical bargaining environment consisting of the physically feasible outcomes and the disagreement outcome. (b) There are the preferences of the players over the physical environment.⁸⁴ (c) Combining (a) and (b), we can define the bargaining problems in utility terms (S and d). (d) Assuming a given theory of rational bargaining, we can determine the solution to the bargaining problem (S , d). (e) The parties choose agents to implement the solution. Of these (c) is merely a mathematical transformation; (d) is assumed to be exogenously given and not subject to strategic manipulation. The remaining elements, however, do lend themselves to such manipulation.

Consider first strategic distortion of preferences. For a given physical environment, a given solution concept and a given implementation, the outcome is a function of the preferences. If the real preferences are unknown, the function might instead take reported preferences as its arguments. In that case, the parties face a noncooperative game in which they have to choose which preferences to report. It has been shown that when the parties bargain over a single good, the dominant strategy is to report linear (risk-neutral) utilities, leading to equal division of the good.⁸⁵ Here strategic distortion of utility at most affects distribution, not efficiency. In bargaining over many goods, misrepresentation can also generate inefficient outcomes.⁸⁶

Consider next strategic action related to the implementation of the solution. If rational bargainers can be expected to reach a conclusion that favours one party, that party may try to turn the tables on the other by

sending an irrational substitute to the bargaining table.⁸⁷ A person who is too stupid to understand the weakness of his bargaining position may refuse to yield where a rational bargainer would back down. Or one might delegate the bargaining to a person who can be counted on to carry out a threat even if it will not be in his interest to do so when the time to execute it arrives. Some societies foster codes of honour that add credibility to threats that otherwise would not be believable. It might make sense to hire a mafioso to represent one at the bargaining table (unless, of course, the other party does the same). I have more to say about this in the next chapter.

Consider, finally, strategic moves that take place before the parties sit down at the bargaining table. Let us assume that the parties know that the actual bargaining will take place according to a specific deterministic model.⁸⁸ The outcome, then, depends wholly on the feasible set S and the disagreement point d . With known preferences over the physical environment, the solution is a function of the latter. *The bargainers will then try to manipulate the physical environment in a direction that skews the outcome in their favour.*⁸⁹ In two-party bargaining, each party has an interest in manipulating the parameters to its advantage. If both parties engage in such manoeuvring, both may lose. I discuss such failures of collective rationality in Chapter 4.

In addition to the parties directly involved, third parties may have an interest in shaping the parameters. The distinction is not sharp, since the parties might try to achieve their goal by influencing a third party with the power to shape these parameters. Unions and employers lobby for laws that, if passed, would enhance their bargaining power. Unions want laws requiring formal training and licensing for certain types of work. This affects d , by preventing employers from threatening to use unskilled labour.

⁸⁷ Schelling (1963) remains the best study of such bargaining ploys.

⁸⁸ Actually, what follows also applies, if more loosely, to the case in which uncertainty prevents the bargaining from being fully deterministic.

⁸⁹ This principle can be used to determine whether a given attempt to reach agreement is a case of bargaining or of, say, rational discussion. Was the Constitutional Convention of 1787 a case of bargaining or of rational discussion? Many aspects of the Constitution, such as the rule whereby a slave was to be counted as three-fifths of a free person for the purpose of representation in Congress, certainly seem like the kind of compromise typically found in bargaining. If the process was purely one of bargaining we would also expect, however, the states to have made strategic moves before the convention for the purpose of strengthening their bargaining position. Minimally, we would expect them to have drummed up public opinion in a way that would make it more difficult for them to make concessions. Without evidence of such strategic behavior, we should be wary of applying the bargaining model.

⁸³ For a defence of this view, see Elster (1989b), ch. 1 and passim.

⁸⁴ For simplicity, I assume that both parties have full knowledge about the physical aspects of bargaining, so that there is no room for strategic misrepresentation on factual matters.

⁸⁵ Crawford and Varian (1979). ⁸⁶ Sobel (1981).

Employers lobby against minimum wage legislation. If they are successful, this affects *S*, by expanding the set of feasible outcomes.

Consider first how the government can use its legislative powers to shape the disagreement point. If the law determines the outcome when private bargaining fails, it serves as a disagreement point for the latter. The decision that would be made in a court or by an arbitrator, as well as the cost of legal fees, will have to be taken into account by the parties in their private bargaining. In divorce bargaining, the financial settlement will be influenced by the law regulating custody disputes. A maternal presumption rule, for instance, enhances the financial bargaining power of women. The shift to the principle that custody should follow the best interests of the child reduces their power correspondingly.⁹⁰ In labour-management bargaining, the government has an incentive to facilitate speedy agreement and to avoid costly strikes. To achieve this goal it may act on the disagreement point, by preventing strikes or lock-outs of indefinite duration. Although the intention may simply be to avoid loss of production, there will be distributional side effects. If management can afford to hold out longer than the union, an upper limit on the duration of strikes and lock-outs will favour the latter.

Third parties can also act on the feasible set – for example, by outlawing certain outcomes. Often, some possible contracts are forbidden because legislators believe (a) that they are inherently undesirable and (b) that they would otherwise be potential outcomes of private bargaining. 'It is well known, for instance, that courts will invalidate contracts by which people would sell themselves into slavery or bind themselves to perform immoral acts. . . . It is less well known that courts will refuse to enforce agreements by which people would waive the right to marry, to divorce, to sue for relief under the bankruptcy laws, to alienate labour freely, or to require a landlord to provide an apartment that meets minimum standards of habitability.'⁹¹ In all cases, the reason is that in the absence of regulation there would be a real danger that such contracts would be made.⁹² In particular, there may be a collective action problem that is overcome by banning certain contracts. If workers bargain individually with their employer, they

⁹⁰ Weizman (1985); Elster (1989a), ch. 3. ⁹¹ Coleman and Silver (1986), p. 109.

⁹² An alternative procedure would be to act on the disagreement point, e.g., to alter the initial endowments of the parties so that such bargains would not be expected to be struck (except under circumstances that would in themselves be grounds for invalidation). Yet if the point is to ensure that, say, contracts to sell oneself into slavery are never made, outright bans are necessary.

may all agree to work long hours even though all would be better off if all worked short hours.⁹³ Each worker may have to accept long hours since the employer may, credibly, point to the existence of other workers who are willing to do so. Unionization is one way out of this collective action problem. Legislation is another.⁹⁴

Minimum-wage legislation and legislation on working conditions, including the length of the working day, affect the bargaining power of the parties. The extent of the impact depends on one's view of the bargaining process. Suppose that if the union and management had bargained over wage and length of the working day without any legal limitation on either, they would have agreed on \$6 per hour and a forty-hour week. If legislation sets the minimum wage to \$5 and the maximum length of the week to forty-two hours, will the outcome be different?⁹⁵ Under the Nash solution, the condition of independence of irrelevant alternatives says that the outcome should not differ. The Kalai-Smorodinsky solution, by contrast, implies that legislation matters even if the outlawed outcome would never have been the outcome of bargaining unconstrained by legislation. Suppose that legislation had set the minimum wage at exactly \$6 or the maximum working week at exactly forty hours. In modern societies, it is surely implausible that the union would not achieve any gains over and above what the law ensured them. In societies less extensively permeated by norms of fair division it is perhaps more plausible that the outcome would be unaffected by irrelevant alternatives.⁹⁶

From third-party manipulation I now turn to manipulation by the parties themselves. Consider first strategic action on the feasible set. In addition to the lobbying efforts discussed earlier, the parties have a powerful incentive to manipulate public opinion and the mass media. If a union leader,

⁹³ Hardin (1988), pp. 92–4.

⁹⁴ For an argument in favour of the latter solution see Fried (1984).

⁹⁵ As earlier, the question can be understood in two ways: as a question in comparative statics or as a question about what will happen following a transition from one regime to the other. Under the second, intuition strongly suggests that the laws will make a difference. The first supports the same intuition, albeit less strongly.

⁹⁶ Intuitively, the following account seems to make sense. Each party possesses a certain amount of bargaining power. If it does not have to spend bargaining power on one issue, because the law ensures that it will get what it wants without bargaining, it has more power to spend on other issues. Similarly, if the law ensures that it can get most of what it wants without bargaining, less expenditure of bargaining power on that issue is necessary than it would be were there no legal constraints. Anyone who has engaged in bargaining will, I believe, recognize that there is something to this intuition, often expressed in phrases like 'I do not want to spend my bargaining chips on this issue'. I do not know, however, how to transform the intuition into a formal theory of bargaining.

management utility

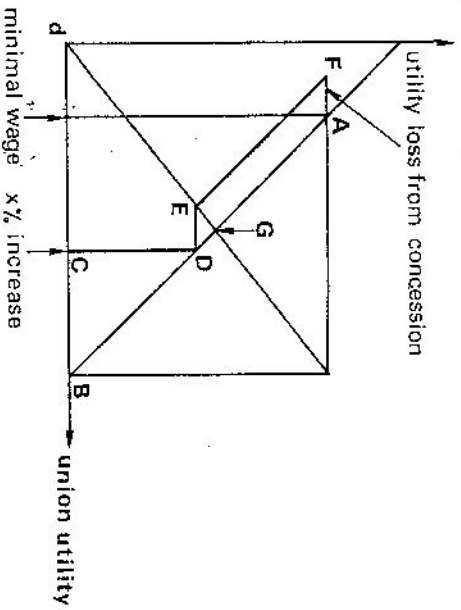


Fig. 2.9

for instance, publicly states that he will resign unless his members get a wage increase of at least x per cent, the announcement amounts to a change of the feasible set of outcomes. To see this, consider Fig. 2.9.

Here, AB is the Pareto frontier in the absence of any public announcements. If the union makes an announcement, this amounts to imposing a cost $ED = FA$ on the union if it accepts a wage increase short of x per cent. The frontier shifts to $CDEF$.⁹⁷ The Nash solution will then shift in the union's favour, as suggested by intuition. The Kalai-Smorodinsky solution shifts from G to E , to the detriment of both parties, as we might expect from the monotonicity argument underlying that solution. Here the Kalai-Smorodinsky solution concept is less adequate intuitively than the Nash solution. An arbitrator might, however, announce that he will impose the Kalai-Smorodinsky solution, to prevent wasteful jockeying for position.

Consider next manipulation of the disagreement point and, more specifically, of the inside options.⁹⁸ This is, I believe, by far the most important

⁹⁷ After the announcement, the feasible set is no longer convex. Strictly speaking, this does not allow us to apply the Nash and Kalai-Smorodinsky solution concepts. It is easy to verify, however, that the reasoning in the text also applies to the convex hull of $CEDF$, obtained by substituting a straight line from D to F for DEF .

⁹⁸ Outside options do not lend themselves to strategic manipulation. The workers in one firm have little influence over what workers earn in other firms to which they could credibly threaten to move. The management may, however, use blacklisting to prevent workers from exiting.

target of strategic action by the parties. Workers may try to build up strike funds to support themselves in case of a disagreement. These funds represent unproductive expenditures. Employers' associations may build up the equivalent of strike funds to support their member firms during strikes.⁹⁹ Management may build up large inventories, 'both to reduce the costs of strikes when they occur and to enhance their bargaining position by reducing their vulnerability to strike threats'.¹⁰⁰ Stockpiling also, however, involves obvious dead-weight losses. Resources that would otherwise be used for productive purposes may be spent on warehouses. Management may deliberately refrain from hiring young productive workers, preferring instead older, married workers with high mortgage payments that make it difficult for them to hold out during a strike. Although older workers produce less, their presence ensures that there will be more time in which to produce.

Moreover, the management may deliberately refrain from investing in capital-intensive technology which would make the firm more vulnerable to pressure by workers.¹⁰¹ The decline of the U.S. steel industry since 1959 has, in particular, been explained as the result of fear of investing in 'hedge capital'.¹⁰² Suboptimal investment induced by fear of worker militancy may doubly harm the workers, not only by the loss of an important bargaining chip, but also by the reduction of the total to be shared, since the use of suboptimal technology reduces the income of the firm.

It would be in the interest of management and workers if the latter could promise not to engage in costly strikes, because this would induce management to invest in more productive technology.¹⁰³ Making credible promises involves problems, however, which mirror those of credible threats. In theory, both problems can be solved by the method of side bets:¹⁰⁴ one lodges a sum with a third party, which is forfeited if the threat is not carried out or the promise not kept. In practice, no societies to my knowledge

⁹⁹ Conflict funds involve a waste of resources because they have to be kept in fairly liquid form, earning a lower interest than in their most productive use. De Geer (1986), pp. 53-4, shows how the Swedish Employers' Association was able to overcome this problem through an agreement with the banks that allowed them to borrow against their nonliquid assets.

¹⁰⁰ Crawford (1983), p. 376. Empirical evidence on the importance of inventories for bargaining strength is found in Holden (1987a).

¹⁰¹ For a discussion of this issue, see Baldwin (1983), Grou (1984), van der Ploeg (1987) and Moene (1988a).

¹⁰² Baldwin (1983).

¹⁰³ For extensive discussion of these issues, see also Williamson (1985), chs. 7 and 8.

¹⁰⁴ Schelling (1963).

have evolved institutions that make and enforce side bets of this kind. Because of the losses that would occur if both sides of a conflict used them to make binding threats, the absence of these institutions might, on balance, be a good thing. An alternative solution might emerge if the union and firm knew that they would have to bargain again on later occasions. If the game between the union and the firm is played over and over again, it is possible to sustain the inconsistent (Pareto-superior) outcome as long as the discount rate is small enough and/or the length of the punishment interval is long enough, even though explicit binding contracts are unavailable.¹⁰⁵ I return to this issue in the concluding chapter.

I have cited cases in which inside options are affected by the directly involved in a conflict. They can also be shaped, however, by the associations to which the parties belong or by legislative action. From the bargainers' point of view, in other words, inside options can be either strategic weapons or institutionally given constraints. The latter case has been studied by Karl Ove Moene in a noncooperative model of labour-management bargaining.¹⁰⁶ By varying the threats at the disposal of labour and management, he shows that bargaining environments differ systematically in their impact on wages, profit and employment.

Moene assumes that the union's objective function is set by majority voting among the workers and that layoffs occur in inverse order of seniority.¹⁰⁷ If the workers know that layoffs will never concern as much as half of the work force, the union's only interest will be to maximize wages. It will not care about employment. (From the social point of view, needless to say, employment matters.) Hence the firm can set employment unilaterally, taking account of the wage effects of its decision. Wages are set by bargaining with the union. The inside options shaping the outcome could be any of the following: go-slow, work-to-rule, official strikes or illegal wildcat strikes. The first two can be reduced to a common formula: the workers reduce their work effort somewhat and receive some fraction of the going wage. In go-slow actions the fraction is strictly smaller than 1; in work-to-rule it equals unity. The last two can also be reduced to a common formula: no work is done, workers receive some income during the strike, whereas the firm receives some support from the employers' association to which it belongs. The difference between the two forms of in-

dustrial action is that wildcat strikes have a smaller strike fund (and that unions may have to pay a fine).

Moene shows that from the point of view of maximizing employment, wages and profits, these bargaining environments can be ranked in the following order:

Employment	Wages	Profits
1. Wild-cat strikes	1. Work-to-rule	1. Wild-cat strikes
2. Official strikes	2. Go-slow	2. Official strikes
3. Go-slow	3. Official strikes	3. Go-slow
4. Work-to-rule	4. Wild-cat strikes	4. Work-to-rule

Workers want high wages, firms want high profits and the government wants high employment. We observe that the interests of government and employers coincide fully, both being opposed to the interests of the workers. These conclusions rest on two assumptions. First, the workers' income during a go-slow action is at least as high as their strike support during a legal strike. Second, the support to firms during a strike is at least as high as the net profit during a go-slow action. The second assumption is empirically vulnerable, since in most countries firms receive no strike support from central funds. In these countries, the assumption holds only when the firm is totally crippled by a go-slow action. But then the first assumption is not very plausible.

Assume instead, therefore, that firms receive no central support and that go-slow income is the same percentage of the going wage as go-slow output is of normal output. If we also assume that support during a legal strike equals go-slow income, the above conditions are violated and the conclusions do not hold. Under these new and more realistic assumptions, Moene shows that a movement from a go-slow regime to one with legal strikes leads to higher profits, higher employment and higher wages. If the firm faces a downward-sloping demand curve, this implies lower output prices as well, benefiting consumers. All is for the best in the best of all possible worlds. Needless to say, this is a special case with few implications about actual bargaining. The argument nevertheless is important, because it shows that the effects of the bargaining environment can be subtle and not immediately detected by intuition. It is clear enough that workers do better

¹⁰⁵ van der Ploeg (1987), p. 1488.

¹⁰⁶ Moene (1988b).

¹⁰⁷ Oswald (1985, 1986).

for themselves when there are some legal forms of industrial action than when all actions are illegal. It is less obvious which form of legal action they should prefer and what the consequences are for other parties.

The inefficiency of bargaining

One of the main points of bargaining is to make joint ventures possible, by enabling the parties to agree on the division of the gains to be made from cooperation. If they cannot agree on how to share the gains, there may be no gains to share. Bargaining, however, has costs of its own. In the words of the late Leif Johansen, '*Bargaining has an inherent tendency to eliminate the potential gain which is the object of the bargaining*'.¹⁰⁸ The reasons he cites for this tendency can be paraphrased as follows. (a) Because of uncertainty about the range of realistic proposals, the parties may begin with excessive claims and never be able to meet half-way. (b) There is a pervasive tendency to bias the presentation of information in one's favour, so that even unbiased information is not believed, leading to inefficiency. (c) To make information credible, mere words are not enough: one must put one's money where one's mouth is and actually expend resources on credibility. (d) Similarly, threats may not be credible unless one carries them out, with a socially undesirable waste of resources. (e) In particular, parties may carry out a threat to establish a reputation for being tough negotiators. (f) The strategy of precommitting oneself to a particular claim can be disastrous if both parties follow it. (g) In particular, if the parties are organizations that try to mobilize their members, they may end up playing the sorcerer's apprentice.¹⁰⁹

Most of these problems, as well as some not included in Johansen's list, were discussed earlier in this chapter. To supplement the list, I shall propose another typology of bargaining costs and bargaining failures, drawing on the typology of cooperative problems set out in the Introduction.

The cost of bargaining failures. If people fail to reach agreement in a joint venture, the production forgone can be a substantial loss. To quote one example at random, 'There began in the UK during 1979 some 1080 stop-pages of work due to industrial disputes, involving 4.548 million workers and resulting in 29.474 million working days lost'.¹¹⁰ Failure to agree on

¹⁰⁸ Johansen (1979), p. 520. Italics in original.

¹¹⁰ Sapsford (1982), p. 3.

¹⁰⁹ *Ibid.*, pp. 518-19.

a private ordering can block Pareto-efficient outcomes. The 'winner's curse', for instance, may prevent mutually beneficial deals from being struck. Failure to capture positive externalities can be serious, as when OPEC countries fail to agree on an allocation of quotas. Failure to agree on the terms of mutual helping can occur between neighbours, if there is disagreement over what constitutes fair reciprocity. Failure to agree on a convention equilibrium can yield large inefficiencies, as in the coexistence of VHS and Beta videocassette recorders or, more importantly, of different systems of weights and measures. The causes of these failures include uncertainty and the various forms of strategic manipulation mentioned earlier and further discussed in Chapter 4. In Chapter 6 I argue that the appeal to norms of equity and equality can also lead to bargaining failures. Finally, theory¹¹¹ and experiments¹¹² indicate that the probability of disagreement increases when the potential gains from agreement increase, contrary to what intuition might suggest.

The costs of preparing for bargaining. These costs derive from strategic manipulation of the bargaining environment. Investments in improving one's bargaining position are a dead-weight loss for society, although they may increase the share and the final outcome of the investor.¹¹³ When all parties deploy such strategies, everybody may end up being worse off than if nobody had prepared for bargaining. I do not know of any empirical studies of the magnitude of these effects, but I suspect they might be non-negligible.

The costs of conducting bargaining. The main task of unions is to prepare for and conduct wage bargaining. Workers pay substantial membership dues, which have to be counted among the costs of bargaining. The magnitude of these costs is indicated by the fact that if workers instead used these funds to buy shares in their firm, most firms would be worker-owned after a few decades.¹¹⁴ There are similar costs on the employer's side, and in maintaining the arbitration system.

The costs of decentralized bargaining. As I further discuss in Chapter 4, local and sectoral wage bargaining give rise to collective action problems.

¹¹¹ Crawford (1982). ¹¹² Malouf and Roth (1981).

¹¹³ When they increase the share without improving the final outcome, such strategies are individually irrational.

¹¹⁴ Moene and Ognedal (1987).

Separate bargaining creates externalities that can make everybody worse off than if a single, encompassing union had negotiated on behalf of all.¹¹⁵ There are exceptions to this statement. Sometimes all are better off by virtue of not being able to present a united front. By and large, however, unity makes for moderation and collective gains.

¹¹⁵ This is a major theme in Olson (1982).

3. Social norms

Introduction

One of the most persisting cleavages in the social sciences is the opposition between two lines of thought conveniently associated with Adam Smith and Emile Durkheim, between *homo economicus* and *homo sociologicus*. Of these, the former is supposed to be guided by instrumental rationality, while the behaviour of the latter is dictated by social norms. The former is 'pulled' by the prospect of future rewards, whereas the latter is 'pushed' from behind by quasi-inertial forces.¹ The former adapts to changing circumstances, always on the lookout for improvements. The latter is insensitive to circumstances, sticking to the prescribed behaviour even if new and apparently better options become available.² The former is easily captured as a self-contained, asocial atom, and the latter as the mindless plaything of social forces or the passive executor of inherited standards. In this chapter I attempt to characterize this contrast more fully. I also argue that while social norms are extremely important for solving the first problem of social order, their contribution to the second problem is more ambiguous. Social norms do coordinate expectations. They may or may not help people to achieve cooperation.

Generally speaking, three solutions have been proposed to resolve the opposition between the two paradigms. First, there is the eclectic argument that some forms of behaviour are best explained on the assumption that people act rationally, whereas others can be explained by something like the theory of social norms. Alternatively, the eclectic view could be that both rationality and social norms are among the determinants of most actions. By and large, I shall adopt one or the other of these eclectic views.

¹ For a useful exploration of this contrast, see Gambetta (1987).

² The theory of social norms must be supplemented by a theory of what happens if the prescribed behaviour ceases to be feasible. Durkheim's theory of anomie was in part intended to answer this question (Besaard 1987).